CODEN: JASMAN

ISSN: 0001-4966

The Journal of the Acoustical Society of America

Vol. 118, No. 2 August 2005

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 118, NO. 2, AUGUST 2005

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 118, NO. 2, AUGUST 2005

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 118, NO. 2, AUGUST 2005

CONTENTS—*Continued from preceding page*

ACOUSTICAL NEWS—USA

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

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

Preliminary Notice: 150th Meeting of the Acoustical Society of America Jointly Held with Noise-Con 2005

The 150th Meeting of the Acoustical Society of America will be held Monday through Friday, 17–21 October 2005 at the Hilton Minneapolis Hotel, Minneapolis, Minnesota, USA. NOISE-CON 2005 will be held in conjunction with the meeting, Monday through Wednesday, 17–19 October 2005. A block of rooms has been reserved at the Hilton Minneapolis Hotel.

Information about the meeting also appears on the ASA Home Page at <http://asa.aip.org/meetings.html>.

Technical Program

The technical program will consist of lecture and poster sessions. Technical sessions will be scheduled Monday through Friday, 17–21 October. The special sessions described below will be organized by the ASA Technical Committees and the Institute of Noise Control Engineering (INCE).

Special Sessions

Acoustical Oceanography (AO)

Ocean ecosystem measurement
(Joint with animal bioacoustics)
Use of acoustics to determine the abundance distribution and behavior of marine organisms in relation to their environment
Inversion using ambient noise sources
Inversion techniques for estimation of ocean environments using natural and manmade noise as sound sources

Animal Bioacoustics (AB)

Cognition in the acoustic behavior of animals
(Joint with Psychological and Physiological Acoustics)
Research on cognitive processes involved in the acoustic behavior of a variety of animals, including dolphins, bats, birds, primates, and others
Frequency weighting for animal species
(Joint with Psychological and Physiological Acoustics and ASA Committee on Standards)
Development of frequency weighting functions for animals
Temporal patterns of sounds by marine mammals
Analysis of the patterns of sound usage in wild and captive pinnipeds and cetaceans

Architectural Acoustics (AA)

Comparison of U.S. and international standards in architectural acoustics
(Joint with ASA Committee on Standards)
Some standards in common use among architectural acousticians in the U.S. are very similar to those commonly used in the U.K., Germany, Japan, etc., and others are very different. Will explore differences and highlight pros and cons
Indoor noise criteria
(Joint with NOISE-CON and Noise)
Investigations into occupant perception of indoor noise, including productivity effects and relationships to indoor noise criteria systems
ISO 3382 and electroacoustics measurement workshop
(Joint with Engineering Acoustics)
Theory and practice of advanced acoustic measurements: Will include discussion of signal acquisition, analysis array test stimulus optimization
Plumbing noise
(Joint with Noise and NOISE-CON)
Experiences regarding plumbing noise, preventive measures/corrective actions in relation to occupant expectations and response
Reflections on reflections
State of the art methods for altering reflections using absorbers, diffusers and room geometry, and also how to measure, predict, and evaluate the effects of these reflections in real and virtual environments
Speech privacy in buildings
Understandings of effects and techniques of speech privacy in public and performance spaces
Special session in honor of Cyril Harris
(Joint with NOISE-CON and Noise)
Presentations honoring Cyril Harris contributions
Safety of acoustical products
(Joint with Noise and NOISE-CON)
Presenting the various hazards (fire, toxic fumes, etc.) of acoustical products commonly in use

Biomedical Ultrasound/Bioresponse to Vibration (BB)

Acoustic radiation force methods for medical imaging and tissue evaluation
(Joint with Physical Acoustics)
Radiation force is used to induce motion in tissue and the response may be measured by a variety of methods including MRI, acoustic, emission and Doppler ultrasound
Medical applications of time-reversal acoustics
(Joint with Physical Acoustics and Signal Processing in Acoustics)
Use of acoustic time reversal for imaging and therapy in the body
Topical meeting on imaging and control of HIFU-induced lesions
Methods for detecting lesions created in tissue by focused ultrasound therapy

Education in Acoustics (ED)

Acoustics demonstrations
A series of acoustics demonstrations
Hands-on workshop for high school students
Hands-on experiments including measurements
"Take 5's"—Sharing ideas for teaching acoustics
Bring short presentations and demonstrations

Engineering Acoustics (EA)

An ANSI standard for measuring *in-situ* directivity of hearing aids in 3-dimensions
(Joint with ASA Committee on Standards)
A review of technical aspects and results using ANSI standard S3.35 revised in 2005 to include directivity index (DI) calculations from manikin directional responses that are produced by sound sources at several azimuths and elevations

Musical Acoustics (MU)

Acoustics of choir singing
(Joint with Architectural Acoustics)
Acoustics of voices in choral ensemble singing and architectural considerations involved in the design of performance spaces
Music information retrieval
(Joint with Signal Processing in Acoustics)
Automatic retrieval or classification of music information such as melodic content, key, tempo, instrumentation, and musical genre

 0001-4966/2005/118(2)/555/10/$22.50

Nonlinear vibrations of strings
Topics can include theory, simulations, and measurements
Patents in musical acoustics
Papers related to patented devices and instruments

Noise (NS)

Advances in noise, vibration, and harshness in automotive design
(Joint with NOISE-CON, Structural Acoustics and Vibration and Physical Acoustics)
How NVH problems in vehicle design are being treated analytically and experimentally
Current status of noise policy
(Joint with NOISE-CON)
The current status of noise policy and the potential for advancements are discussed
Hospital interior noise control
(Joint with NOISE-CON, Engineering Acoustics, and Architectural Acoustics)
Architectural guidelines and materials suitable for noise reduction in hospitals
Laser Doppler vibrometry measurements in underwater and radiation problems
(Joint with NOISE-CON, Structural Acoustics and Vibration and Physical Acoustics)
Will focus on laser Doppler vibrometer measurements of underwater structures and methods of determining radiation from measured velocity response
Special session in honor of William W. Lang
(Joint with NOISE-CON and ASA Committee on Standards)
Presentations on aspects of Bill's life
Specifying uncertainties in acoustic measurements
(Joint with NOISE-CON and ASA Committee on Standards)
Will examine variabilities and motivate understanding error propagation in acoustic measurements
Workshop on methods for community noise and annoyance evaluation II
(Joint with NOISE-CON)
Applied measurements, qualitative, and quantitative methods; synergetic noise

NOISE-CON 2005

Product noise and vibration control—Case studies
(Joint with Noise)
Case studies from practicing engineers working on noise and vibration control problems.
Measurement of information technology product noise emissions
(Joint with Noise)
Measurement techniques and metrics for quantification of noise from information technology equipment
Measurement of product noise emissions
(Joint with Noise)
Measurement techniques and metrics for quantification of noise from consumer products
Array methods for noise source visualization
(Joint with Noise and Signal Processing in Acoustics)
Near-field acoustic holography, acoustic intensity methods, and beamforming are examples of sound field visualization methods. This session will focus on these and other methods to produce visualizations of sound fields with stationary and/or moving sources
Forensic acoustics
(Joint with Noise and Engineering Acoustics)
Many noise control engineers are asked to be expert witnesses in cases. This session is focused on examples of acoustical and noise control engineers' involvement in forensics
Products for noise control
(Joint with Noise)
Session is focused on materials for industrial and architectural noise control. Case histories involving noise control products will be presented
Energy methods in transportation noise
(Joint with Noise)
Will include case studies on the application of energy methods in transportation applications such as aircraft, automobiles, trucks, and spacecraft. Session may also include presentations on new and modifications to existing energy methods
Numerical methods in acoustics
(Joint with Noise)
Focused on developments and applications of numerical modeling techniques for noise and vibration
Active noise control: Centralized versus decentralized control
(Joint with Noise and Engineering Acoustics)
Focused on types of control strategies in active noise control systems
Applications of active noise control
(Joint with Noise and Engineering Acoustics)
Focused on case histories and potential applications of active noise control
From noise control to product design
(Joint with Noise)
Acoustic and nonacoustic constraints and product functionality have to be factored into product design. Product sound can enhance functionality by providing feedback to the user of the machine. This session is focused on how understanding of noise control strategies, sound perception and the desired functionality of the machine can be factored into product design
Sound quality and soundscapes
(Joint with Noise and Psychological and Physiological Acoustics)
Depending on the tasks being performed within an interior space (office, factory, home, car, space station, etc.), which constitutes an ideal acoustic environment, may differ. Objectives for noise control must be set so that the results enhance the overall soundscape. In this session, the relationship between the sound quality of individual sound sources and the quality of the overall acoustic environment will be explored
Environmental sound quality
(Joint with Noise)
Much of noise in outdoor spaces is quantified by using metrics that are functions of the A-weighted sound pressure level. Level, while important, is not the only sound attribute that impacts the quality of the outdoor acoustic environment. Presentations will be focused on the measurement, quantification, and enhancement of environmental sound quality
Methods for predicting and assessing community responses to noise
(Joint with Noise)
Session is focused on tools to assess community response to existing noise sources and to predict community response changes in the local infrastructure that will impact the acoustic landscape
Case studies: The environmental impact analysis process (EIAP)
(Joint with Noise)
Will examine the success of environmental impact analysis process by studying its use in a variety of situations
Power plant noise: Technology that limits power plant noise control
(Joint with Noise)
Will examine how limitations in technology restrict control of power plant noise, with the aim of defining a systematic strategy for technology development to address power plant noise and reduce its impact on the environment
State and local noise policies and noise ordinances
(Joint with Noise)
The session will be focused on local and state, rather than federal, policies
Public policy workshop
(Joint with Noise)
The aim of this workshop is to formulate a national noise policy
Role of vibrations and rattle in annoyance
(Joint with Noise and Psychological and Physiological Acoustics)
Rattle and vibration coupled with sound and knowledge of the source impact people's evaluation of the sound. This session is focused on research studies on multimodal (noise and vibration) responses and research on community response to sounds that are accompanied by vibration and rattle
Rail noise and vibration issues
(Joint with Noise)
Focused on modeling sources, propagation, and the impact in communities of rail transportation noise and vibration
Mittigating the effects of construction noise
(Joint with Noise)
Strategies to mittigate the effect of construction noise will be presented, including scheduling, barriers, and other noise control methods and community engagement
Noise intrusion in the natural landscape

(Joint with Noise)
Quiet sounds that are perceived as "not belonging" to a particular environment often cause people to complain. Similarly, loud sounds that are perceived as being part of the natural landscape are accepted. Clearly the level-based metrics currently in use for environmental noise are not appropriate in these situations. This session focuses on issues related to these intrusive sounds
Noticeability of noise: Time structure
(Joint with Noise and Signal Processing in Acoustics)
Sounds like speech or music attract attention, perhaps because of the information contained within, or perhaps it is because it is difficult to "tune out" sounds where the level is continuously changing. In this session, the characteristics of the temporal structure of sounds and how that impacts noticeability will be examined
Transportation noise criteria
(Joint with Noise)
Noise criteria are used to help engineers set objectives in noise control. However, the following question still remains: was the noise control successful? Did meeting the criteria actually result in an improvement in the quality of the acoustic environment. This session will be a series of case studies where the success of meeting noise criteria will be examined
Issues in aircraft noise analysis
(Joint with Noise)
This session will focus on shortcoming of currently used techniques and proposals for alternative methods to analyze noise due to aircraft
Progress in aircraft noise research
(Joint with Noise)
This is a series of presentations focused around the activities in the FAA/NASA Center for Excellence for Noise and Emissions Mittigation. Results of recent airport noise measurements, a study on noise impact of aircraft landing maneuvering strategies, land usage, and airport encroachment research as well as proposed studies on supersonic aircraft noise will be presented
Aircraft source noise research
(Joint with Noise)
An overview of current understanding of jet noise will start this session that will also include, for example, noise analysis for engine noise cycle design, sound source reduction, and absorption strategies, as well as measurement techniques to help understand jet noise characteristics and to validate jet engine noise reduction strategies
Advances in military jet noise modeling
(Joint with Noise and Physical Acoustics)
Focused on military jet noise modeling including supersonic aircraft
Tire/pavement noise and quiet pavement applications
(Joint with Noise)
Case histories of quiet pavement projects will be presented along with papers on modeling and research of tire/pavement sound and vibration generation and propagation
Vehicle noise measurement
(Joint with Noise)
Session will be focused on techniques to measure interior and exterior vehicular noise. May also include papers on the relationship between the sound analysis results and the perception of vehicle quality, and on the verification of noise-control treatments
Classroom acoustics
(Joint with Noise, Architectural Acoustics, and ASA Committee on Standards)
Session is focused on noise-control and architectural design strategies to meet standards on classroom acoustics. Session will include case histories, success stories, and challenges that remain
Innovative solutions to architectural design and to meeting LEED and HIPAA requirements
(Joint with Architectural Acoustics and Noise)
Energy, environmental, and privacy concerns affect the architectural and acoustic designs of buildings and public spaces. These also impact noise control strategies for existing buildings. While often viewed as a problem, these challenges can sometimes lead to innovation solutions that are improvements on solutions arrived at without these constraints

Physical Acoustics (PA)

Thermoacoustics: What we are doing, and why our customers want it
(Joint with Engineering Acoustics)
Physics and applications of theromacoustics

Signal Processing in Acoustics (SP)

Biomedical acoustic signal processing
(Joint with Biomedical Ultrasound/Bioresponse to Vibration)
Will discuss various signal processing and imaging techniques employed in biomedical acoustic imaging, diagnostics, and therapies
Nonblind and blind deconvolution in acoustics
(Joint with Underwater Acoustics, Animal Bioacoustics, Noise, Acoustical Oceanography, and Engineering Acoustics)
The use of deconvolution to find the input or the impulse response in an experiment when one of the two is known or when neither one is known

Speech Communication (SC)

Phonetic linguistics: Honoring the contributions of Peter Ladefoged
In honor or Peter Ladefoged's 80th birthday and his 50 years of research in linguistic phonetic aspects of speech communication

Structural Acousics and Vibration (SA)

Experimental modal analysis
(Joint with Signal Processing in Acoustics)
Experimental methods for determining resonant frequencies and mode shapes in structures
Vibration in transit systems
(Joint with Noise and NOISE-CON)
Generation of and propagation of vibration in transit systems

Underwater Acoustics (UW)

Head waves and interface waves
(Joint with Acoustical Oceanography)
Propagation of head (lateral) waves, interface waves and other wave phenomena near cutoff and their potential for geoacoustic inversion
Sonar performance and signal processing in uncertain environments
(Joint with Signal Processing in Acoustics and Engineering Acoustics)
Explores and quantifies environmental uncertainty and potentials for robust processing

NOISE-CON 2005

The 21st annual conference of the Institute of Noise Control Engineering, NOISE-CON 2005, will run concurrently with the 150th Meeting of the Acoustical Society on Monday through Wednesday (17–19 October, 2005), culminating with the Closing Ceremony, which will take place with the ASA Plenary Session on Wednesday afternoon (19 October 2005). It is our plan that all of the Noise and some of the Architectural Acoustics Technical Sessions will be part of the joint ASA–NOISE-CON conference, thus forming an exciting and coherent program of noise control related sessions, which reflects the overlap in membership interests between the two organizations, and the spirit of cooperation that led to the decision to have this joint meeting. Note, that there will be one registration fee for both conferences, so NOISE-CON 2005 participants are encouraged to take the opportunity to learn about some of the work being done in other areas of acoustics, not usually part of regular NOISE-CON technical programs, by attending the sessions taking place on Thursday and Friday.

All presentations in NOISE-CON 2005, including those in sessions jointly organized with the ASA Technical Committees, will be accompanied by a 4 to 8 page paper that will be published by the Institute of Noise Control Engineering (INCE) in the NOISE-CON 2005 Proceedings.

Special Sessions that will be part of the NOISE-CON 2005 program are being organized by INCE and also jointly by INCE and the ASA Noise, Architectural Acoustics, Engineering Acoustics, Physical Acoustics, Psychological and Physiological Acoustics, Signal Processing in Acoustics, and Structural Acoustics Technical Committees. The INCE organized Special Sessions are listed under special sessions above in the section titled "Noise-Con 2005."

Other Technical Events

Topical Meeting on Imaging HIFU-Induced Lesions

A one-day colloquium and discussion on the topic "Imaging and Control HIFU-Induced Lesions" will be held. Subtopics will focus on the following areas: Real-time Monitoring (MRI and Ultrasound), Quantitative Imaging for Damage Assessment, Noninvasive Temperature Monitoring and Control, and Contrast-Assisted Imaging, and Lesion Formation. Each subtopic session will consist of invited and contributed papers and be followed by a panel discussion.

Distinguished Lecture

The Technical Committee on Architectural Acoustics will sponsor a distinguished lecture presented by Manfred R. Schroeder titled "How I stumbled onto number theory when in need of more sound diffusion."

NOISE-CON 2005 Plenary Talks

Carl Burleson of the Federal Aviation Administration (FAA) will present a plenary talk on Monday morning, 17 October, titled "Perspectives on noise in the menu of environmental issues, and the role of technical solutions relative to policy approaches."

Paul Donavan of Illingworth-Rodkin will present a plenary talk on Tuesday morning, 18 October, titled "Tire and pavement noise and the potential impact of quiet pavement technology."

James West of the Johns Hopkins University will present a plenary talk on Wednesday morning, 19 October, titled "Hospital noise, its role in patient well-being and the challenges for noise control engineers."

Power Plant Noise Seminar

A seminar on Power Plant Noise will be given by Frank Brittain on Sunday, 16 October 1:00 p.m. to 5:00 p.m. The seminar will review the basics of noise control for combustion turbine power plants—both simple and combined cycle. The major noise sources will be identified, and proven controls will be discussed. A brief description of how power plants and their equipment work will also be included. This seminar is intended for noise control engineers who have either very limited experience with power plants, or are involved with supplying equipment for power plants.

For registration details contact the INCE business office: ibo@inceusa.org

For more information on the seminar content contact: Frank Brittain, fnbritta@bechtel.com

Technical Tours

A tour of the Aero Systems Engineering (ASE) Fluidyne lab is planned for Monday, October 17. A bus will pick up participants in the morning at approximately 9:00 a.m. and return in time for the 1:00 p.m. sessions. We will have a box lunch on the bus ride back. There will be approximately 45 available spots on the bus, and there will be a $5 charge to cover the box lunch (Aero Systems Engineering is subsidizing the lunch and transportation). For more information about ASE, visit <http://www.aerosysengr.com>.

ASA/INCE members are also invited to a rehearsal of the famed Minnesota Orchestra (Osmo Vanska, conductor) on Tuesday afternoon, 18 October. After the rehearsal we will be treated to a guided tour of Orchestra Hall with Cyril Harris as our tour guide. The Minnesota Orchestra is across the street from the Hilton Minneapolis Hotel

Register for the tours by using the registration form in the printed call for papers or online at <http://asa.aip.org>.

Online Meeting Papers

The ASA has replaced its traditional at-meeting "Paper Copying Service" with an online site that can be found at <http://scitation.aip.org/asameetingpapers/>. Authors of papers to be presented at meetings will be able to post their full papers or presentation materials for others who are interested in obtaining detailed information about meeting presentations. The online site will be open for author submissions in September.

Those interested in obtaining copies of submitted papers for this meeting and the immediate past meeting may access the service at anytime. No password is needed.

Exposition

The meeting will be highlighted by a large exposition, jointly sponsored by the ASA and INCE. It will feature 40 displays with instruments, materials, and services for the acoustical and vibration community. The expo is conveniently located near the registration area and meeting rooms. The exposition will open at the Hilton Minneapolis Hotel with a reception on Monday evening, 17 October, and will close Wednesday at noon. Morning and afternoon refreshments will be available in the exposition area.

The exposition will include computer-based instrumentation, sound level meters, sound intensity systems, signal processing systems, devices for noise control, sound prediction software, acoustical materials, passive and active noise control systems, and other exhibits on vibrations and acoustics. For further information, please contact Richard J. Peppin, Exposition Manager, Institute of Noise Control Engineering c/o Scantek, Inc., 7060 Oakland Mills Road, #L, Columbia, MD 21046, Tel: 410-290-7726; Fax: 410-290-9167; E-mail: peppinr@scantekinc.com

Workshop For ASA Journal Authors and Readers

The American Institute of Physics (AIP) will hold a two-hour workshop to describe electronic services provided for ASA journal authors and readers. This will include information on Peer Xpress for JASA and ARLO, multimedia, Scitation searching and other electronic services.

Tutorial Lecture on Diagnostic Imaging in Biomedical Ultrasound

A tutorial presentation on Diagnostic Imaging in Biomedical Ultrasound will be given by E. Carr Everbach on Monday, 17 October at 7:00 p.m.

Lecture notes will be available at the meeting in limited supply. Those who register by 26 September are guaranteed receipt of a set of notes.

To partially defray the cost of the lecture a registration fee is charged. The fee is $15 for registration received by 12 September and $25 thereafter including on-site registration at the meeting. The fee for students with current ID cards is $7.00 for registration received by 12 September and $12.00 thereafter, including on-site registration at the meeting. To register for the tutorial lecture, use the registration form in the call for papers or register online at <http://asa.aip.org>.

Short Course on Statistical Energy Analysis

A short course on Statistical Energy Analysis will be held on Sunday and Monday, 16-17 October.

The objective of this course is to provide an overview of modern predictive SEA methods. The course will discuss the physics of high-frequency noise and vibration transmission from both modal and wave viewpoints. The derivation of the underlying SEA equations will be discussed and the parameters used in a SEA model will be summarized. Particular emphasis will be placed on the theoretical aspects of the wave approach to SEA; the physics of wave propagation in commonly encountered structural and acoustic subsystems will therefore be discussed in detail. The calculation of the parameters that govern vibroacoustic energy input, storage, transmission, and dissipation will be discussed. Typical SEA applications will be reviewed along with areas of current research.

The instructor is Phil Shorter from ESI US R&D. Phil has primary responsibility for the maintenance and development of the theory implemented in the commercial SEA code AutoSEA2.

Further details about the course may be found in the printed call for papers or online at <http://asa.aip.org/minneapolis/minneapolis.html>

The registration fee is $250.00 and covers attendance, instructional materials, and coffee breaks. The number of attendees will be limited so please register early to avoid disappointment. Only those who have registered by 26 September will be guaranteed receipt of instructional materials. There will be a $50.00 discount for registration made prior to 12 September. Full refunds will be made for cancellations prior to 12 September. Any cancellation after 12 September will be charged a $75.00 processing fee. To register for the short course, use the registration form in the call for papers or register online at <http://asa.aip.org>.

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 12 September to: Jolene Ehl, ASA, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502, Tel: 516-576-2359; Fax: 516-576-2377; E-mail: jehl@aip.org. The proposal should include your status as a student; whether you have submitted an abstract; whether you are a member of ASA; method of travel; if traveling by auto; whether you will travel alone or with other students; names of those traveling with you; and the 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 Minneapolis meeting. This award is designed for young professionals who have completed the doctorate in the past five years (not currently enrolled as a student), who plan to present a paper at the Minneapolis meeting. Each award will be of the order of $300. It is anticipated that the Committee will grant a maximum of three awards. Applicants should submit a request for support, a copy of the abstract they have submitted for the meeting, and a current resume/vita that provides information on their involvement in the field of acoustics and to the ASA to: Dr. Peggy Nelson, Department of Speech-Language-Hearing Sciences, University of Minnesota, 164 Pillsbury Drive SE, Minneapolis MN 55455; Fax: 612-624-7586; E-mail: nelso477@umn.edu. Deadline for receipt of applications is 2 September.

Students Meet Members For Lunch

The ASA Education Committee provides a way for a student to meet one on one with a member of the Acoustical Society over lunch. The purpose is to make it easier for students to meet and interact with members at ASA Meetings. Each lunch pairing is arranged separately. Students who wish to participate should contact David Blackstock, University of Texas at Austin, through e-mail dtb@mail.utexas.edu or telephone 512-343-8248 (alternative number 512-835-3374). Please give Dr. Blackstock your name, university, department, degree you are seeking (BS, MS, or Ph.D.), research field, acoustical interests, and days you are free for lunch. The sign-up deadline is ten days before the start of the Meeting, but an earlier sign-up is strongly encouraged. Each participant pays for his/her own meal.

Plenary Sessions, Awards Ceremony, Fellows' Lunch, and Social Events

Buffet socials with cash bar will be held on Tuesday and Thursday evenings at the Hilton Minneapolis Hotel.

The Joint ASA INCE Plenary session will be held on Wednesday afternoon, 19 October, at the Hilton Minneapolis Hotel where ASA Society awards will be presented and the recognition of newly elected Fellows will be announced. INCE will present two awards at this Plenary Session.

The NOISE-CON 2005 plenary sessions will take place on Monday, Tuesday, and Wednesday mornings.

A Fellows' Luncheon will be held on Thursday, 20 October, at 12:00 noon. This luncheon is open to all attendees and their guests. To Purchase tickets use the registration form in the call for papers or register online at <http://asa.aip.org>.

Women in Acoustics Luncheon

The Women in Acoustics luncheon will be held on Wednesday, 19 October. Those who wish to attend this luncheon must register using the form in the printed call for papers or online at <http://asa.aip.org>. The fee is $15 (students $5) for preregistration by 12 September and $20 (students $5) thereafter, including on-site registration at the meeting. To register use the registration form in the call for papers or register online at <http://asa.aip.org>.

Transportation and Hotel Accommodations

Air Transportation

Minneapolis is served by the Minneapolis–St. Paul International Airport (Airport Code MSP). A number of airlines serve Minneapolis including Northwest, Midwest Airlines, SunCountry, and many others. Most flights arrive at the Lindbergh (main) terminal; SunCountry and other charter flights arrive at the smaller Humphrey terminal. Transportation between terminals is free and convenient. For flight information, visit <http://www.mspairport.com/>; for other information of interest, visit <www.minneapolis.org>.

Ground Transportation

Transportation from the Minneapolis–St. Paul International Airport to the Hilton Minneapolis Hotel:

Ground Transportation Information. The Tram Level information booth is staffed seven days a week from 7 a.m. to 11:30 p.m. Staff provides information, directions, and other assistance to travelers. Travelers may also obtain wheelchairs from the information booth.

Light Rail Service at MSP. Light rail service to and from the airport is now available. The light rail stop closest to the hotel is the Nicollet Mall stop. Exit the train at 5th and Marquette. The hotel is 5 blocks south, at 10th and Marquette.

Trains stop at both the Lindbergh and Humphrey terminals and connect travelers to 15 other destinations, including downtown Minneapolis to the north and the Mall of America to the south. Trains run between the terminals every 7 to 8 minutes during peak hours and every 10 to 15 minutes at other times of the day.

There is no charge for light rail travel between Minneapolis–St. Paul International Airport's two terminals. Fares for travel to other locations from the airport are $1.75 during rush hours (Monday through Friday, 6 a.m. to 9 a.m. and 3 p.m. to 6:30 p.m.) and $1.25 at other times. A six-hour pass is available for $3 and an unlimited-ride day pass for $6. Tickets are sold at vending kiosks at the rail stations. For more information, visit the Metropolitan Council's Web site, <http://www.metrotransit.org/rail>.

The Lindbergh Terminal light rail station entrance is located near the Transit Center, between the Blue and Red Parking ramps. From the Lindbergh Terminal tram level (two floors below the Ticketing Lobby), take the automated Hub Tram toward the Transit Center. When you exit the tram, follow the signs to the light rail station, located 70 feet underground.

The Humphrey Light Rail station is located outside of the terminal building. A covered walkway connects the station located at 34th Avenue and 72nd Street, to the Humphrey parking facility, and to the terminal building.

Major car rental companies. Rental car companies have phones and touch screen information kiosks at the Lindbergh Terminal on the Baggage Claim Level opposite baggage carousels 2, 5 and 10. The rental car counters are located in the Hub building located between the Blue and Red parking ramps, on Levels 1–3. Passengers can take the underground tram to go between the Lindbergh Terminal and the Hub building.

SuperShuttle shared-ride, door-to-door service. Travelers wishing to take a taxi or to take a van to the hotel can gain access to those services through the Lindbergh Terminal's Tram Level.

Metro Transit bus service to the Twin Cities metropolitan area and Jefferson Lines scheduled bus service are both accessible at the airport's new Transit Center. You can reach the Transit Center from the Red and Blue parking ramps or by taking the Hub Tram from the Lindbergh Terminal's Tram Level. For a complete list of shuttle companies, visit <http://www.mspairport.com/MSP/Travelers_Guide/>.

Taxicabs and limousines. Many companies provide taxicab service at Minneapolis International Airport. Taxis are available at the Lindbergh and Humphrey Terminals. Taxi service at the Lindbergh Terminal is accessible via the Tram Level. Signs direct passengers one level up to the cab starter booth, where airport staff will assist passengers obtaining a taxi. At the Humphrey Terminal, taxi service is available at the Humphrey Ground Transportation Center that is located in the Humphrey Parking ramp on Level 1.

Downtown Minneapolis is approximately 16 miles from the airport, with fares averaging $25.00. All cab fares are metered and include a $2.25 trip fee that allows drivers to recoup airport permit fees.

Hotel Accommodations

The meeting and all functions will be held at the Hilton Minneapolis Hotel. Please make your reservations directly with the hotel and ask for one of the rooms being held for the Acoustical Society of America (ASA). The reservation cutoff date for the special discounted ASA rates is 16 September 2005; after this date, the conference rate will no longer be available. A block of guest rooms at discounted rates has been reserved for meeting participants at the Hilton Minneapolis Hotel. Early reservations are strongly recommended. Note that the special ASA meeting rates are not guaranteed after 16 September 2005. You must mention the Acoustical Society of America when making your reservations to obtain the special ASA meeting rates.

The Hilton Minneapolis Hotel is located in the heart of downtown. The hotel features a fully equipped health club, indoor heated swimming pool, sauna, jacuzzi, as well as a full service business center. For more details visit: http://www.hilton.com/en/hi/hotels/index.jhtml?ctyhocn=MSPMHHH

Please make your reservation directly with the Hilton Minneapolis Hotel. When making your reservation, you must mention the Acoustical Society of America to obtain the special ASA meeting rates.

Hilton Minneapolis Hotel
1001 Marquette Avenue
Minneapolis, MN 55403
Tel: 612-376-1000
Fax: 612-397-4871
Rates
Single/Double: $134.00
Executive Level: $169.00

Room Sharing

ASA will compile a list of those who wish to share an hotel room and its cost. To be listed, send your name, telephone number, e-mail address, gender, smoker or nonsmoker preference, by 12 September to the Acoustical Society of America, preferably by e-mail: asa@aip.org or by postal mail to Acoustical Society of America, 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

October is one of the most beautiful months in Minneapolis, when the air begins to get crisp. Enjoy fall colors and outdoor activities by packing a windbreaker, umbrella, and hat. A walk along Riverfront District should provide a glimpse of fall color. Even if the worst happens and winter arrives early, most attractions in downtown Minneapolis are accessible through the indoor skyway system. Average high temperatures in mid-October hover a bit below 60°, with average lows around 40°F. For additional information on Twin Cities weather, visit <http://climate.umn.edu/doc/twin_cities/twin_cities.htm>.

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, asa@aip.org.

Accompanying Persons Program

Spouses and other visitors are welcome at the Minneapolis meeting. The registration fee for accompanying persons is $50. A hospitality room for accompanying persons will be open at the Hilton Minneapolis Hotel from 8:00 a.m. to 11:00 a.m. each morning throughout the meeting where information about activities in and around Minneapolis will be provided.

Minneapolis and the Twin Cities area is well known as an important diverse cultural center, offering outstanding arts programs including opera, museums, and theater. A few of these are highlighted here. Other guides to area attractions will be available on-site.

Downtown Minneapolis is home to several areas of interest for visitors. All are within walking distance of the meeting. The Central Business District, downtown's business core, is the heart of the dining and shopping scene. Nicollet Mall—a pedestrian-only thoroughfare—is lined with great restaurants, plus numerous upscale shops and department stores. The Warehouse District is the center of Minneapolis nightlife. This large area of renovated warehouses is home to bars, clubs, alternative theaters, art galleries, and more. The Riverfront District is where Minneapolis got its start. In the Mississippi Riverfront District, attendees can stroll cobblestone streets, dine at a century-old pub, visit historical attractions, and squeeze in some trendy nightlife at the same time. The Theatre District is packed with broadway shows, music venues, and upscale dining. It is also home to some of Minneapolis' most recognizable productions, along with locally produced gems.

The Guthrie Theater is the area's premier theater, and is easily accessible from the Hilton Hotel. Theater offerings for October were not available at press time, but can be found at <http://www.guthrietheater.org>.

The world-famous Walker Art Center is one of the most celebrated art museums in the country, and is known for commissioning and presenting innovative contemporary art. A new 17-acre urban campus will open in spring 2005. The design for the new Walker engages the surrounding neighborhood with a new four-acre park as well as vistas onto the downtown Minneapolis skyline. The expanded facility, nearly double the size of the existing building, will feature new galleries; education areas; a new 385-seat theater; street-level and rooftop terraces; plazas, gardens, and lounges; and increased services and amenities for visitors. The Minneapolis Sculpture Garden, a project of the Walker Art Center and the Minneapolis Park and Recreation Board, is adjacent to the museum. For more information, visit <www.walkerart.org>.

The Mall of America is one of a kind. It is the nation's largest retail and entertainment complex. The 11.6-mile Hiawatha Corridor Light Rail Transit links three of the region's most popular destinations, Downtown Minneapolis, Minneapolis–St. Paul International Airport and Mall of America. Passengers arrive and depart at the newly remodeled Transit Station under the Mall at 24th Avenue. For more information, visit <www.mallofamerica.com>.

Registration Information

The registration desk at the meeting will open on Monday, 17 October at the Hilton Minneapolis Hotel. To register use the form in the call for papers or register online at <http://asa.aip.org>. If your registration is not received at the ASA headquarters by 26 September you must register onsite.

Registration fees are as follows:

Category	Preregistration by 12 September	Registration after 12 September
Acoustical Society/INCE Members	$325	$375
Acoustical Society/INCE Members One-Day	$165	$190
Nonmembers (nonmembers of ASA or INCE)	$375	$425
Nonmembers One-Day	$190	$215
Nonmember Invited Speakers	Fee waived	Fee waived
ASA/INCE Student Members (with current ID cards)	Fee waived	Fee waived
Nonmember Students (nonmembers of ASA or INCE with current ID cards)	$40	$40
Emeritus members of ASA/INCE (Emeritus status pre-approved by ASA or INCE)	$50	$50
Accompanying Persons (Spouses and other registrants who will not participate in the technical sessions)	$50	$50

Nonmembers who simultaneously apply for Associate Membership in the Acoustical Society of America will be given a $50 discount off their dues payment for the first year (2006) of membership. Invited speakers who are members of the Acoustical Society of America are expected to pay the registration fee, but nonmember invited speakers may register without charge.

NOTE: A $25 processing fee will be charged to those who wish to cancel their registration after 12 September.

Online Registration

Online registration is now available at <http://asa.aip.org>.

Members of the Local Committee for the 150th Meeting of the Acoustical Society of America

General Chair—Peggy B. Nelson; Technical Program Chair—Neal F. Viemeister; Food Service/Social Events—Kay Hatlestad; Audio-Visual—Bruce Olson; Accompanying Persons Program/Technical Tours—David Braslau; Signs/Publicity—Derrick Knight, Benjamin Munson

NOISE-CON 2005 Committee

General Chair—Daniel J. Kato; Co-Chair—Robert J. Bernhard; Technical Program Cochair—Patricia Davies; Technical Program Cochair—J. Stuart Bolton; Exhibit Manager—Richard J. Peppin

James L. Flanagan to receive IEEE Medal of Honor

The Institute of Electrical and Electronics Engineers (IEEE) has named James. L. Flanagan, a pioneer in the areas of speech analysis, speech transmission, and acoustics, as recipient of the 2005 IEEE Medal of Honor. The award celebrates Flanagan's sustained leadership and outstanding contributions in speech technology.

The IEEE Medal of Honor, one of engineering's most prestigious awards and the highest award given by the IEEE, is presented to individuals for their exceptional contributions or extraordinary careers in any of the IEEE fields of interest. The award is sponsored by the IEEE Foundation and comprises a gold medal, bronze replica, certificate, and cash honorarium. Flanagan received the Medal at the annual IEEE Honors Ceremony on 18 June in Chantilly, VA.

As former director of the Information Principles Research Laboratory at Bell Laboratories in Murray Hill, NJ, Flanagan led researchers to a greater understanding of how the human ear processes signals and was responsible for the development of advanced hearing aids and improved voice communications systems. His work included the development of an electronic artificial larynx, playback recording systems for the visually impaired and automatic speech recognition to help the motor impaired.

Flanagan was one of the first researchers to see the potential of speech as a means for human-machine communication. He has made seminal contributions to current techniques for automatic speech synthesis and recognition and to signal coding algorithms for telecommunications and voice mail systems, including voicemail storage, voice dialing and call routing. He also created autodirective microphone arrays for high-quality sound capture in teleconferencing and pioneered the use of digital computers for acoustic signal processing.

More recently, as vice president for research and director of the Center for Advanced Information Processing at Rutgers University in Piscataway, NJ, James Flanagan has been a leader in the development of global systems for human computer interfaces that are actuated by speech and that incorporate sight and touch.

James Flanagan is a Fellow of the Acoustical Society America. He was awarded the ASA Gold Medal in 1986 for contributions to and leadership in digital speech communications. He has served ASA in numerous positions including Associate Editor of the J*ournal of the Acoustical Society of America* (1959–1962), Member of the Executive Council (1971–1974), Vice President (1976–1977), and President (1978–79).

An IEEE Fellow, he is a former president of the IEEE Signal Processing Society and received its Achievement Award. He is also a recipient of the IEEE Centennial Medal and the National Medal of Science, and is a member of the U.S. National Academy of Engineering and the U.S. National Academy of Sciences.

He has a bachelor's degree from Mississippi State University and master's and doctoral degrees from Massachusetts Institute of Technology, all in electrical engineering. He has been awarded Doctor Honoris Causa from the University of Paris-Sud, and from the Polytechnic University of Madrid.

Regional Chapter News

Madras Regional Chapter

The Madras Regional Chapter of the Acoustical Society of America (MIRC-ASA) conducted five meetings in 2004, on 9 May, 29 June, 14 August, 25–27 November, and 4 December in 2004, to commemorate its 10th anniversary. Since 1995, two meetings have been held annually, one meeting with invited speakers and another including full technical sessions.

FIG. 1. Miss R. Shruthi (extreme left) receives the first rank award at the Science and Engineering Fair.

FIG. 2. D. Sai Kiran (second left) was awarded the second rank award at the Science and Engineering Fair.

FIG. 3. B. V. A. Rao (r) receives the First Stanley Ehrlich Gold Medal from H.S. Paul (l)

FIG. 4. Baldev Raj (r) receives First Raman-Chandrasekhar interdisciplinary Silver Medal from H.S. Paul (l)

FIG. 5. S. S. Agrawal (r) receives second Speech Communication Silver Medal from H.S. Paul (l)

FIG. 6. S. Narayanan (r) receives First Structural Acoustics Silver Medal from H.S. Paul (l)

On 9 May 2004, the first Stanley Ehrlich Distinguished Lecture was delivered by Professor B. V. A. Rao, President of the Acoustical Society of India, and an invited lecture was delivered by Professor C. Sujata of the Department of Mechanical Engineering, Indian Institute of Technology (IIT), Madras, at the International Research Institute for the Deaf (IRID), Chennai. On 29 June Professor S. Narayanan, Dean of Academic Research, IIT, Madras delivered a lecture on *Science of Sound* to school children at Tamil Nadu Science and Technology Center (TNSTC), Chennai.

On 14 August, a Science and Engineering Fair on Acoustics was held with six students below 17 years of age receiving awards. In Fig. 1, Miss R. Shruthi receives the first rank award and D. Sai Kiran, age 13, received the second rank award. (see Fig. 2).

The award ceremony of MIRC-ASA and the Acoustical Foundation (AFECT) was conducted during the joint meeting of MIRC-ASA and the Acoustical Society of India at S. J. C. E., Mysore on 25 November. Four outstanding acousticians, who actively participated in the first technical sessions of MIRC-ASA in 1995, received Silver Medals and one eminent acoustician received the Stanley Ehrlich Gold medal. As it was the 10th meeting of the chapter, Chapter Representative and President of AFECT & IRID,H. S. Paul presented all awards (see Figs. 3–7). Chapter's officers also presented an honorary membership certificate to B. V. A. Rao (see Fig. 8).

Dr. Baldev Raj, Director of Indira Gandhi Center of Atomic Research, Department of Atomic Energy, Govt. of India, Kalpakkam, delivered the Distinguished lecture to the chapter on 25 November. Six students received best paper awards on 26 November in the joint meeting held at S.J.C.E., Mysore. Three students received best paper awards (see Fig. 9).

FIG. 7. B. Chakraborty (r) receives First Underwater Acoustics Silver Medal from H.S. Paul (l)

FIG. 8. MIRC-ASA Chapter officers present honorary membership certificate to B.V.A. Rao (l to r) M.V. Subramanian (Vice President), H.S. Paul (C.R.), B.V.A. Rao, C.P. Vendhan (Secretary), A. Ramachandraiah (Treasurer), M. Kumaresan (President).

FIG. 9. (l to r) H. S. Paul, N. S. Kale, S. Kaul, A. K. Gupta and A. Ramachandraiah.

On 4 December, two senior students received awards at the fair held at IRID. The above activities of MIRC-ASA clearly indicate that after the 10th Anniversary of its existence, the chapter and its members continue to be very active.

USA Meetings Calendar

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

2005

17-21 Oct.	150th Meeting of the Acoustical Society of America joint with NOISE-CON 2005, Minneapolis, Minnesota, [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
27–29 Oct.	5th International Symposium on Therapeutic Ultrasound, Boston, MA [www.istu2005.org; E-mail: info@istu2005.org].

2006

6–9 June	51st Meeting of the Acoustical Society of America, Providence RI [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].
17–21 Sept.	INTERSPEECH 2006 (ICSLP 2006), Pittsburgh, PA [www.interspeech2006.org <http://www.interspeech2006.org/>].
28 Nov.–2 Dec.	152nd Meeting of the Acoustical Society of America joint with the Acoustical Society of Japan, Honolulu, HI [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: http://asa.aip.org].

2008

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

Cumulative Indexes to the Journal of the Acoustical Society of America

Ordering information: Orders must be paid by check or money order in U.S. funds drawn on a U.S. bank or by Mastercard, Visa, or American Express credit cards. Send orders to Circulation and Fulfillment Division, American Institute of Physics, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2270. Non-U.S. orders add $11 per index.

Some indexes are out of print as noted below

Volumes 1–10, 1929–1938: JASA, and Contemporary Literature, 1937-1939. Classified by subject and indexed by author. Pp. 131. Price: ASA members $5; Nonmembers $10.

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

Volumes 21–30, 1949–1958: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 952. Price: ASA members $20; Nonmembers $75.

Volumes 31–35, 1959–1963: JASA, Contemporary Literature and Patents. Classified by subject and indexed by author and inventor. Pp. 1140. Price: ASA members $20; Nonmembers $90.

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

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

Volumes 45–54, 1969–1973: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 540. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound).

Volumes 55–64, 1974–1978: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 816. Price: $20 (paperbound); ASA members $25 (clothbound); Nonmembers $60 (clothbound).

Volumes 65–74, 1979–1983: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 624. Price: ASA members $25 (paperbound); Nonmembers $75 (clothbound).

Volumes 75–84, 1984–1988: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 625. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound).

Volumes 85–94, 1989–1993: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 736. Price: ASA members $30 (paperbound); Nonmembers $80 (clothbound).

Volumes 95–104, 1994–1998: JASA and Patents. Classified by subject and indexed by author and inventor. Pp. 632, Price: ASA members $40 (paperbound); Nonmembers $90 (clothbound).

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

ACOUSTICAL NEWS—INTERNATIONAL

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

Russian-French Joint Workshop

The 15th meeting of the Russian Acoustical Society will be held in Moscow 14–18 November 2005. One part of the program will be a workshop, organized jointly by the Russian Acoustical Society (RAS) and the French Acoustical Society (SFA). The theme of this workshop is "High Intensity Acoustic Waves in Modern Technological and Medical Applications." For further information visit http://www.akin.ru/e_main.htm

DEGA elects new board of officers

The German Acoustical Society (DEGA) has elected a new board. The President is Hugo Fastl of the Technical University Munich, the Vice President is Jens Blauert of the Ruhr University Bochum, and the Treasurer is Joachim Scheuren of the MüllerBBM Company (Planegg/Munich). Three other members were elected, they are Otto von Estorff of the University of Technology Hamburg, Helmut Fleischer of the Armed Forces University (Neubiberg), and Ulrich Widmann of AUDI AG (Ingolstadt).

WESPAC IX—A working URL

For a great number of months the web site of the 9th Western Pacific Acoustics Conference (Wesapac IX) was published but unfortunately that URL never led to the Conference organizers in Seoul, Korea. In April 2005 a working address was established. Click on http://wespac9.org for details of this Western Pacific Conference.

International Meetings Calendar

Below are announcements of meetings and conferences to be held abroad. Entries preceded by an asterisk are new or updated listings.

August 2005

6–10	**Inter-Noise**, Rio de Janeiro, Brazil (Web: www.internoise2005.ufsc.br).
28–2	**EAA Forum Acusticum Budapest 2005**, Budapest, Hungary (I. Bába, OPAKFI, Fö u. 68, Budapest 1027, Hungary; Fax: +36 1 202 0452; Web: www.fa2005.org).
28–1	**World Congress on Ultrasonics Merged with Ultrasonic International (WCU/UI'05)**, Beijing, China (Secretariat of WCU 2005, Institute of Acoustics, Chinese Academy of Sciences, P.O. Box 2712 Beijing, 100080 China; Fax: +86 10 62553898; Web: www.ioa.ac.cn/wcu-ui-05).
31–3	**6th Pan European Voice Conference**, London, UK (Web: www.pevoc6.com/home.htm).

September 2005

4–8	**9th Eurospeech Conference (EUROSPEECH'2005)**, Lisbon, Portugal (Fax: +351 213145843; Web: www.interspeech2005.org).
5–9	**Boundary Influences in High Frequency, Shallow Water Acoustics**, Bath, UK (Web: acoustics2005.ac.uk).
18–21	**IEEE International Ultrasonics Symposium**, Rotterdam, The Netherlands (Web: www.ieee-uffc.org).
20–22	**International Symposium on Environmental Vibrations**, Okayama, Japan (Web: isev2005.civil.okayama-u.ac.jp).
27–29	**Autumn Meeting of the Acoustical Society of Japan**, Sendai, Japan (Acoustical Society of Japan, Nakaura 5th-Bldg., 2-18-20 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan; Fax: +81 3 5256 1022; Web: www.asj.gr.jp/index-en.html).

October 2005

12–14	**Acoustics Week in Canada**, London, Ontario, Canada (Web: caa-aca.ca).
17–18	**Wind Turbine Noise: Perspectives for Control**, Berlin, Germany (G. Leventhall, 150 Craddocks Avenue, Ashtead Surry KT21 1NL, UK; Fax: +44 1372 273 406; Web: www.windturbinenoise2005.org).
19–21	**36th Spanish Congress on Acoustics Joint with 2005 Iberian Meeting on Acoustics**, Terrassa (Barcelona), Spain (Sociedad Española de Acústica, Serrano 114, 28006 Madrid, Spain; Fax: +34 914 117 651; Web: www.ia.csic.es/sea/index.html).
25–26	**Autumn Conference 2005 of the UK Institute of Acoustics**, Oxford, UK (Web: www.ioa.org.uk).
27–28	**Autumn Meeting of the Acoustical Society of Switzerland**, Aarau, Switzerland (Web: www.sga-ssa.ch).

November 2005

4–5	**Reproduced Sound 21**, Oxford, UK (Web: www.ioa.org.uk).
9–11	**Australian Acoustical Society Conference on "Acoustics in a Changing Environment,"** Busselton, WA, Australia (Web: www.acoustics.asn.au/divisions/2005-conference.shtml).
14–18	**XVI Session of the Russian Acoustical Society**, Moscow, Russia (Web: www.akin.ru).

December 2005

7–9	**Symposium on the Acoustics of Poro-Elastic Materials**, Lyon, France (Fax: +33 4 72 04 70 41; Web: v0.intellagence.eu.com/sapem2005).

January 2006

5–7	**First International Conference on Marine Hydrodynamics**, Visakhapatnam, India (V. B. Rao, Naval Science & Technological Laboratory, Vigyan Nagar, Visakhapatnam—530 027, India; Web: www.mahy2006.com).

March 2006

20–23	*__Meeting of the German Acoustical Society (DAGA 2006)__, Braunschweig, Germany (Web: www.daga2006.de)

May 2006

15–19	**IEEE International Conference on Acoustics, Speech, and Signal Processing**, Toulouse, France (Web: icassp2006.org).
30–1	*__6th European Conference on Noise Control (EURONOISE2006)__, Tampere, Finland (Fax: +358 9 7206 4711; Web: www.euronoise2006.org).

June 2006

26–28	*__9th Western Pacific Acoustics Conference (WESPAC 9)__, Seoul, Korea (Web: wespac9.org).

 0001-4966/2005/118(2)/565/2/$22.50

July 2006

3–7 **13th International Congress on Sound and Vibration (ICSV13)**, Vienna, Austria (Web: info.tuwienac.at/icsv13).

17–20 *__International Symposium for the Advancement of Boundary Layer Remote Sensing (ISARS13)__, Garmisch-Partenkirchen, Germany (Fax: +49 8821 73 573; Web: www.isars.org.uk).

17–19 **9th International Conference on Recent Advances in Structural Dynamics**, Southampton, UK (Web: www.isvr.soton.ac.uk/sd2006/index.htm).

September 2006

13–15 **Autumn Meeting of the Acoustical Society of Japan**, Kanazawa, Japan (Acoustical Society of Japan, Nakaura 5th-Bldg., 2-18-20 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan; Fax: +81 3 5256 1022; Web: www.asj.gr.jp/index-en.html).

18–20 *__International Conference on Noise and Vibration Engineering (ISMA2006)__, Leuven, Belgium (Fax: 32 16 32 29 87; Web: www.isma-isaac.be).

November 2006

20–22 *__1st Joint Australian and New Zealand Acoustical Societies Conference__, Christchurch, New Zealand (Web: www.acoustics.org.nz).

July 2007

9–12 **14th International Congress on Sound and Vibration (ICSV14)**, Cairns, Australia (e-mail: n.kessissoglou@unsw.edu.au).

August 2007

27–31 **Interspeech 2007**, Antwerp, Belgium (e-mail: conf@isca-speech.org).

September 2007

2–7 **19th International Congress on Acoustics (ICA2007)**, Madrid, Spain (SEA, Serrano 144, 28006 Madrid, Spain; Web: www.ica2007madrid.org).

9–12 **ICA Satellite Symposium on Musical Acoustics (ISMA2007)**, Barcelona, Spain (SEA, Serano 144, 28006 Madrid, Spain; Web: www.ica2007madrid.org).

June 2008

30–4 *(New Date) **Joint Meeting of European Acoustical Association (EAA), Acoustical Society of America (ASA), and Acoustical Society of France (SFA)**, Paris, France (E-mail: phillipe.blanc-benon@ec-lyon.fr).

July 2008

28–1 **9th International Congress on Noise as a Public Health Problem**, Mashantucket, Pequot Tribal Nation (ICBEN 9, P.O. Box 1609, Groton CT 06340-1609, USA: Web: www.icben.org).

Preliminary Announcements

October 2006

3–6 **IEEE Ultrasonics Symposium**, Vancouver, BC, Canada (TBA).

August 2010

TBA **20th International Congress on Acoustics (ICA2010)**, Sydney, Australia (Web: www.acoustics.asn.au).

Remote sensing symposium

The calendar listing for 17–20 July 2006 may need some clarification. For many years ISARS, the International Symposium on Acoustic Remote Sensing of the Atmosphere and the Oceans, was held. Some time ago the official name of the next symposium, ISARS13, was changed to International Symposium for the Advancement of Boundary Layer Remote Sensing, as shown in the calendar. However, the short form "ISARS" was retained.

OBITUARIES

Leonid Maximovich Brekhovskikh • 1917–2005

Leonid Brekhovskikh, a pioneer in wave propagation and scattering, academician, codiscoverer of the deep sound channel, and author of the classic text, *Waves in Layered Media*, died in Moscow on January 15, 2005 at the age of 87. He was active in his research, regularly working at his office at the Shirshov Institute of Oceanography up to his very last days. Brekhovskikh was born just south of the Arctic Circle near Arkhangel'sk, Russia on the White Sea, on May 6, 1917 to a humble family. He studied physics at the State University in Perm near the Ural Mountains, where he received his undergraduate degree in 1939. His talents recognized, he continued his postgraduate work at the prestigious Lebedev Physics Institute in Moscow, where he received his degree of Candidate of Science in 1941 for his dissertation on x-ray diffraction in crystals. In 1947, he was awarded his Doctor of Physics and Mathematics for his dissertation on the propagation of sound and radio waves in layered media.

Leonid Brekhovskikh began his long and productive work in underwater acoustics at the Acoustics Laboratory of the Lebedev Physics Institute in 1942. While on the staff at the Institute he made one of his most important discoveries—the deep sound channel—in an experiment that was conducted in the Sea of Japan in 1946. This experiment took place only a few months after the Ewing and Worzel experiment on the *Saluda,* of which he was unaware. Very similar to the *Saluda* experiment, Brekhorskikh and his colleagues set off charges that were recorded by a hydrophone suspended at depth. Leonid correctly ascertained the existence of the deep sound channel to explain the observations, and recognized its implications for efficient long-range propagation of sound in the ocean. The discovery of the deep sound channel laid the foundation for the exploitation of long-range acoustic propagation in the ocean for submarine detection, acoustic communications, and acoustic tomography and thermometry.

Brekhovskikh introduced the tangent plane approximation (TPA) in 1951, into the theory of wave scattering from rough surfaces. This powerful and useful technique is the second "classical" method for dealing with rough surface scattering after the small perturbation method (SPM) introduced by Rayleigh in 1907. The TPA allows one to handle problems in which the roughness is large compared to a wavelength, where SPM breaks down, as long as a smoothness criterion is met. As important as the discovery of the deep sound channel and the development of the TPA were, it could still be argued that Brekhovskikh's greatest contribution to acoustics is his classic book *Waves in Layered Media,* first published in Russian in 1957. It is currently in its second edition in English (1980), and has become a classic textbook and reference book for three generations of acoustic researchers and modelers around the world. His monograph *Fundamentals of Ocean Acoustics,* which he published in 1982 with Yuri Lysanov, a long-time colleague, has also become a much-referenced and popular book for practical applications of sound in the ocean, and is in its third edition. Brekhovskikh wrote several books on physics and acoustics with V. Goncharov and O. Godin.

The view that the ocean was dominated by a stable ocean circulation structure was beginning to wane in the late 1960s. Stommel had hypothesized in 1963 that smaller "meso" scale variability might account for a considerable portion of the ocean's kinetic energy. Brekhovskikh had come to this conclusion since he could not reconcile measured *acoustic* fluctuations with a time-invariant ocean. In 1970 he led a large-scale hydrophysical experiment to the tropical Atlantic that involved six USSR research vessels (including the *Sergei Vavilov* and the *Petr Lebedev),* the installation of 17 moorings over an area of 113×113 nautical miles for six months of temperature, salinity, and current measurements. This experiment, called "Polygon-70," established the existence of open ocean eddies and mesoscale variability that was much greater than had been previously thought, and began what some have called the "mesoscale" revolution in oceanography, one of the major oceanographic discoveries of this century. It is now known that the mesoscale in the ocean accounts for more than 90% of its kinetic energy, and its influence on acoustic propagation and fluctuations is profound.

Leonid Brekhovskikh left the Lebedev Institute in 1953 to help establish and be the first Director of the Acoustics Institute in Moscow, an institute dedicated to acoustic research and development, the first of its kind in the Soviet Union, and which still exists today. He supervised the staffing of this new institute and attracted a cadre of young gifted scientists. He formulated the main branches of research. In addition to supervising the construction of the physical plant of the Institute, he also undertook to design, and have built, two dedicated acoustics and oceanographic research ships, the *Sergei Vavilov,* and the *Petr Lebedev.* It was during this period that Brekhovskikh almost single-handedly established the importance of acoustics as a field of research in the USSR and charted its course for the future. He also had tremendous influence on the development of acoustics in China, traveling there and helping as the invited acoustics expert in formulating China's 12-year Science and Technology Development Plan in 1956. Brekhovskikh remained Director of the Acoustics Institute until 1962 when he returned to full-time research. In 1968, he was elected a Full Member (Academician) of the USSR Academy of Sciences. For many years Brekhovskikh was working on the International Commission on Acoustics and represented the International Council of Scientific Unions in the Scientific Committee on Ocean Research (SCOR).

Brekhovskikh took on more important and prestigious posts within the Soviet scientific hierarchy, including membership in the Presidium of the "Physics and Hydrocosmos" Chair of the Moscow Institute of Physics and Technology from 1980 to 1997. In the latter position he had tremendous influence in attracting some of the "best and brightest" students into the fields of acoustics and oceanography. Also in 1980, Brekhovskikh left the Acoustics Institute to found and head the new Department of Acoustics of the Shirshov Institute of Oceanology, where he worked until his death. This move reflected his understanding of the inextricable link between oceanography and acoustics, and at the Shirshov Institute his and his student's research contributed to the beginning of what we now call Acoustical Oceanography.

Brekhovskikh came from a modest but remarkable family. Among his six brothers, two became State Prize winners for their prominent achievements in science and technology: one for development of bulletproof glass and the other for his outstanding work in metallurgy. His colleagues use the words "tough," "persistent," and "focused" in describing Brekhovskikh. He spent most of his productive scientific life behind the Iron Curtain, before the fall of the Soviet Union, making active participation in our society impossible. The fact that his contributions to the international community in acoustics transcended the political divide are a testament to his persistence and resolve, and the importance of his work.

Leonid Brekhovskikh garnered numerous prestigious honors and awards in his long career. Among them are the Rayleigh Gold Medal of the Institute of Acoustics of the United Kingdom in 1977 and Foreign Member

 0001-4966/2005/118(2)/577/2/$22.50

of the U. S. National Academy of Sciences. He was awarded two USSR State Prizes and the Lenin Prize. He is probably the only Lenin Prize winner who has also won an award for his research from the United States Navy. In 1996, Brekhovskikh was awarded the Walter Munk Medal for Exploring the Seas, from the Oceanographic Society and the Office of Naval Research. In 1999, the Acoustical Society of America made him an Honorary Fellow for his pioneering contributions to wave propagation and scattering. In 2002, the Russian Acoustical Society awarded Brekhovskikh an Honorary Fellowship for his prominent contributions to acoustics.

PETER N. MIKHALEVSKY
NIKOLAI DUBROVSKY
OLEG GODIN
KONSTANTINE NAUGOLNYKH

OBITUARIES

Betty H. Goodfriend • 1919–2004

Betty (Hofstadter) Goodfriend, past Secretary and Fellow of the Acoustical Society of America (ASA), died on November 22, 2004.

Betty was born in Buffalo, New York on April 1, 1919. She received a B.A. in English Literature in 1940 from the University of Buffalo.

In 1943, she was hired by Wallace Waterfall, Secretary of ASA from 1929 to 1969, to work with him at the Columbia University Division of War Research, Office of Scientific Research and Development (OSRD). She continued to work on OSRD projects until 1948, including management of the staff preparing the OSRD Summary Technical Reports. Many of the OSRD Division 6 reports are on acoustical subjects and are still of interest, including *Principles of Underwater Sound, Physics of Sound in the Sea, Calibration of Sonar Equipment, Design and Construction of Crystal Transducers*, and *Design and Construction of Magnetostriction Transducers*.

During this period, Wallace Waterfall was Secretary of both the ASA and the American Institute of Physics (AIP), and he enlisted Betty to handle the secretarial work of the ASA on a volunteer basis. In 1948 she was hired to be the Assistant Secretary of AIP and she served in that position until 1965, while continuing to handle ASA assignments.

Betty was on a leave of absence from 1953 to 1957. Upon her return to AIP, she was named Assistant Secretary of ASA on a part-time basis and transferred to full time beginning in 1965. She was appointed Administrative Secretary of ASA in 1969 when Waterfall resigned as ASA Secretary to become Treasurer.

Betty served as Administrative Secretary from 1969 to 1982, when her title was changed to Secretary. She also served in many unofficial capacities over the years upon the deaths of ASA Treasurers Herbert Erf in 1967, Wallace Waterfall in 1974, and Editor-in-Chief R. Bruce Lindsay in 1985 while their successors were being selected.

During her 43 years as an ASA volunteer and then employee, the ASA grew in size and scope of activities. Membership grew from 1000 members to over 5000. In handling the administrative affairs of the Society, Betty worked with 41 presidents and hundreds of ASA officers, committee chairs, and other volunteer members in the conduct of ASA business and programs.

In recognition of her dedication to the Society, Betty was awarded the ASA's Distinguished Service Citation in 1973. She was elected a Fellow of the Society in 1983 "for conspicuous service to all areas of acoustics and outstanding contributions to the welfare of the Society."

Betty announced her "retirement" to part-time status in 1985. During the plenary session at the Spring 1985 meeting, she was recognized for her 42 years of continuous service to the Society. At that same meeting, a reception was held in her honor that was attended by the past Presidents and others who had worked closely with Betty through the years. [see J. Acoust. Soc. Am. **78**, 1129–1131 (1985)].

She continued as Secretary on a part-time basis until 1987 and then as an adviser to the new Secretary and others on meetings and administrative matters until her full retirement in 1989. In June 1986, Betty was appointed an Associate Editor of the *Journal of the Acoustical Society of America* (JASA) for Acoustical News USA, a position in which she served until 1992. Among her other publication-related contributions to ASA were preparation of several JASA Cumulative Indices and the Contemporary Papers section of JASA.

She is survived by her daughter and son-in-law, Karen and John Chaffee of Great Falls, Virginia.

ELAINE MORAN
MURRAY STRASBERG

 0001-4966/2005/118(2)/579/1/$22.50

OBITUARIES

Frederic L. Lizzi • 1942–2005

Frederic L. Lizzi, a Fellow of the Acoustical Society of America, died peacefully in his home in Manhattan, New York on January 8, 2005.

Dr. Lizzi was born December 11, 1942 in Brooklyn, NY. He received his Bachelor of Arts degree in electrical engineering from Manhattan College in 1963, his Master of Science degree in bioengineering from Columbia University in 1965, and his Engineering Science Doctorate degree in bioengineering from Columbia in 1971. His doctoral dissertation was titled, "Transient radiation patterns in ophthalmic ultrasound."

In 1967, Dr. Lizzi became a member of the research staff at Riverside Research Institute (RRI) in New York City. In 1973, he became an assistant manager of Optics. Dr. Lizzi led RRI's biomedical engineering studies since the early 1970s and he became the manager of the Biomedical Engineering Laboratories in 1976. He was promoted to research director in 1984. Dr. Lizzi holds several patents in medical ultrasound technology that he developed at RRI. Since 1994, he also was the chief technical adviser for Spectrasonics, Inc., of Wayne, PA. Dr. Lizzi was Adjunct Professor of Ophthalmic Physics at the Weill Medical College of Cornell University since 1979, and Adjunct Professor of Applied Physics at Columbia University since 1997.

Dr. Lizzi is internationally recognized as a pioneering and leading investigator in advanced diagnostic and therapeutic applications of ultrasound, which he investigated on theoretical, experimental, and clinical levels. In particular, Dr. Lizzi is known for his model relating ultrasonic backscatter to particle sizes and concentrations. Through a series of papers, he established a fundamental analytical model of scattering, described the statistical framework necessary to utilize the model in practice, and demonstrated its utility for *in vivo* applications.

These papers include "Theoretical framework for spectrum analysis in ultrasonic tissue characterization" in J. Acoust. Soc. Amer. **73**, 1366 (1983), "Relationship of ultrasonic spectral parameters to features of tissue microstructure" in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control (1987); "Diagnostic spectrum analysis in ophthalmology: A physical perspective" in Ultrasound Med. Biol. (1986); and "Statistical framework for ultrasonic spectral parameter imaging" in Ultrasound Med. Biol. (1997).

Dr. Lizzi was also a pioneer, starting in the 1970s, in high-frequency ultrasound in ophthalmology. In his ophthalmic investigations, Dr. Lizzi developed means of characterizing tissue for diagnostic purposes, and he studied the safety and therapeutic aspects of high-frequency ultrasound. These seminal research efforts were expanded and applied to additional organs such as the prostate, breast, and heart in a broad range of contemporary studies in Dr. Lizzi's laboratories and in research centers around the world. More recently, Dr. Lizzi returned his attention to the field of high-intensity focused ultrasound, where he had done pioneering work in ophthalmic ultrasound in the 1970s. This early work included the first FDA-approved high-intensity ultrasound therapy device.

Dr. Lizzi was an active leader in several professional societies including the American Institute of Ultrasound in Medicine (AIUM), the Acoustical Society of America (ASA), the World Federation of Ultrasound in Medicine and Biology (WFUMB), and the International Society for Diagnostic Ultrasound and Ophthalmology. Dr. Lizzi was a Fellow of the ASA and AIUM. In addition, he served on the Board of Governors of the AIUM from 1985 to 1988 and the International Society for Therapeutic Ultrasound since 2002. He helped to organize many international scientific meetings and served as a guest editor and editorial adviser for scientific journals such as Ultrasound in Medicine and Biology and Transactions of the Institute of Electrical and Electronics Engineers. Dr. Lizzi received numerous professional awards including the William J. Fry Memorial Award of the AIUM in 1986; the Presidential Recognition Award of the AIUM in 1988 and again in 1996; the Pioneer Award of the AIUM and WFUMB in 1988; the Mayneord Award of the British Institute of Radiology in 1990; and the Joseph Holmes Pioneer Award of the AIUM in 1994. The journal, Ultrasound in Medicine and Biology, gave him the Best Clinical Paper Award in 1984 and the Best Technical Paper Award in 1986.

Dr. Lizzi is survived by his wife, Mary; his son, Joseph; his daughter, Marian; his mother; and three sisters.

ERNEST J. FELEPPA
JEFFREY A. KETTERLING

 0001-4966/2005/118(2)/580/1/$22.50

OBITUARIES

Edward M. Kerwin, Jr. • 1927–2004

Edward M. Kerwin, Jr., a member of the Society since 1951 and a Fellow since 1964, died in Waltham, Massachusetts, on December 31, 2004 at the age of 77, after suffering from Alzheimer's disease for a few years and succumbing to a short bout of pneumonia.

He was born in Oak Park, Illinois, on April 20, 1927, the fourth of eight children of the late Edward M. Kerwin, Sr. and Marie Kerwin, and received his early education there. After graduation from high school in Elmhurst, Illinois, he served in the U.S. Navy and then attended the Massachusetts Institute of Technology (MIT). He received S.B. and S.M. degrees from MIT in 1950, and the Doctor of Science in Electrical Engineering degree, also from MIT, in 1954. During his doctoral studies he served as a research assistant at the MIT Acoustics Laboratory, whose director, Leo Beranek, also was his thesis advisor. Dr. Kerwin's thesis research concerned sound generation by heat sources in moving fluids, and his first presentation at an ASA meeting [JASA, **26**, 948 (1954)] concerned the related Rijke phenomenon.

In 1950, while still a student at MIT, Dr. Kerwin worked part time with the then-fledgling firm of Bolt Beranek and Newman (BBN). In 1954 he became one of BBN's earliest full-time staff members, and spent his entire 44-year post-university career with that firm, retiring in 1996 with the title of Principal Scientist, BBN's highest technical rank, equivalent to corporate vice president.

At BBN, one of his first projects concerned quieting of the two-engine propeller-driven Convair 340 aircraft, for which he devised a way to combine the dual exhausts from each engine into a single one, so that the pressure fluctuations were radically reduced. During his tenure at BBN, he lectured in summer courses at MIT and participated in numerous research, development, and consulting projects, most of which were related to ship silencing and other Navy concerns. Because of the classified nature of much of his work, including his contributions in the area of acoustical hull coatings, response and radiation from fluid-loaded structures, and sonar self-noise, some of his work may not be known as widely as it perhaps should be.

He probably is most widely recognized for his seminal work on so-called viscoelastic damping treatments—essentially, sandwich plate structures with dissipative cores—whose basic principles are delineated in the epochal paper "Damping of flexural waves by a constrained viscoelastic layer," [JASA **31**, 952 (1959)]. Dr. Kerwin's concepts, which have led to several patents, have guided a great many applications in the naval, aircraft, automotive, and consumer products industries.

Over the course of his many years at BBN, he was a technical mentor and patient educator to many colleagues. During his long involvement with our Society, he presented much of his work on structural damping first at meetings of our Society, and he published papers in this and related fields in our Society's Journal. He served on the Shock and Vibration Committee (before it changed its name to the Structural Acoustics Committee) from 1969 to 1975.

Dr. Kerwin's wife Margaret died five years previously, and he is survived by their nine children, 13 grandchildren, six of his siblings, and numerous nieces and nephews.

ERIC E. UNGAR
WILLIAM J. CAVANAUGH

 0001-4966/2005/118(2)/581/1/$22.50

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

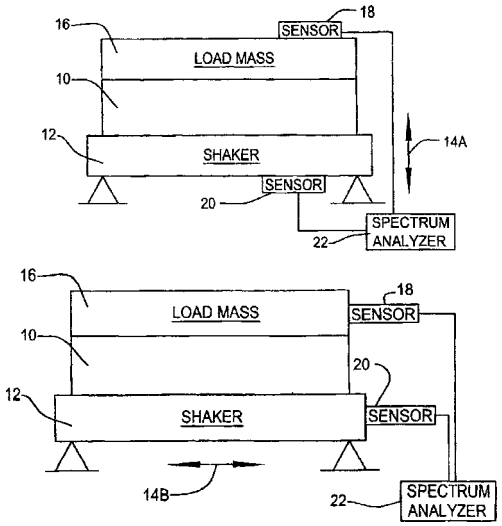
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JOHN M. EARGLE, *JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068*
SEAN A. FULOP, *California State University, Fresno, 5245 N. Backer Avenue M/S PB92, Fresno, California 93740-8001*
JEROME A. HELFFRICH, *Southwest Research Institute, San Antonio, Texas 78228*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, *Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
ROBERT C. WAAG, *University of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

6,848,311

43.20.Ye METHOD FOR ESTIMATING THE PROPERTIES OF A SOLID MATERIAL SUBJECTED TO COMPRESSIONAL FORCES

Andrew J. Hull, assignor to The United States of America as represented by the Secretary of the Navy
1 February 2005 (Class 73/579); filed 9 February 2004

A straightforward method is described that measures the complex frequency-dependent ''dilatational'' and shear wave numbers of a material **10** under a static compression force **16**. The Lamé constants, complex Young's modulus, complex shear modulus, and complex Poisson's ratio are



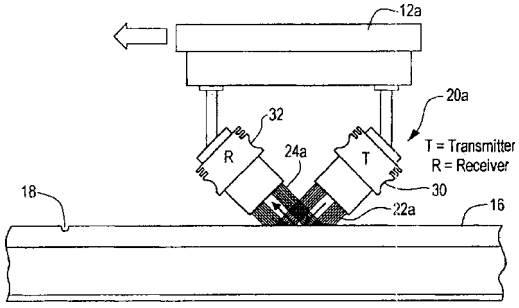
calculated by first determining the transfer functions in both the vertical and horizontal directions using the apparatus illustrated in the figures. The patent provides a very detailed and informative mathematical description of the method used in the determination of the described mechanical parameters.—NAS

6,854,333

43.20.Ye FLAW DETECTION SYSTEM USING ACOUSTIC DOPPLER EFFECT

Shi-Chang Wooh, assignor to Massachusetts Institute of Technology
15 February 2005 (Class 73/643); filed 26 January 2001

Surface flaws such as **18** in railroad tracks **16** cost money. Other industries, such as mining, use continuous linear mechanisms that are also subject to damage. Damage means downtime. ''The invention results from the realization that a truly elegant yet extremely reliable continuous and high-speed detection system for detecting a flaw in a medium such as a



conveyer belt, cable, rope, railroad track, or road can be effected by sensing a Doppler shift in a carrier signal cause by a flaw.'' The list of U.S. patent documents in the cited references is the most extensive seen by this reviewer.—NAS

6,870,792

43.28.Tc SONAR SCANNER

Mark J. Chiappetta, assignor to iRobot Corporation
22 March 2005 (Class 367/98); filed 2 August 2001

This patent discusses the sonar system for a mobile robot and, in particular, some signal processing procedures to mitigate the effects of multipath reflections, extraneous environmental noise, and the ring-down time of the projector.—WT

 0001-4966/2005/118(2)/585/17/$22.50

6,868,041

43.30.Gv COMPENSATION OF SONAR IMAGE DATA PRIMARILY FOR SEABED CLASSIFICATION

Jonathan M. Preston and Anthony C. Christney, assignors to Quester Tangent Corporation
15 March 2005 (Class 367/88); filed 30 April 2003

This patent relates to a technique for estimating the characteristics of the seabed from sonar echoes. In particular, the patent describes a method for compensating the amplitudes of these echoes received by multibeam and sidescan sonar systems to remove the effects of both range and angle of incidence, thus providing an improved assessment of the characteristics of the seabed.—WT

6,868,043

43.30.Jx BEAM BROADENING WITH MAXIMUM POWER IN ARRAY TRANSDUCERS

Evan Frank Berkman, assignor to BBNT Solutions LLC
15 March 2005 (Class 367/123); filed 20 February 2003

A technique is discussed for generating a broad, high-power beam using an array of transducers. The array, either a line or planar array, is considered to be divided into a number of adjacent subsegments which may contain only one, or perhaps more, of the elements of the array. These subsegments are individually excited to form beams. However the signals to each subsegment are sequentially phase shifted so these individual beams are tilted to different directions. The phase shifts are chosen so that the difference in the tilt direction of the beams from two adjacent array subsegments is equal to one-half the sum of the half-power beamwidths of these two beams. The ensemble of individual beams then results in one broad beam. The concept can also be used to form a broad receive beam.—WT

6,861,783

43.35.Ns STRUCTURE TO ACHIEVE HIGH-Q AND LOW INSERTION LOSS FILM BULK ACOUSTIC RESONATORS

Li-Peng Wang *et al.*, assignors to Intel Corporation
1 March 2005 (Class 310/324); filed 19 November 2003

A process is described for forming a film bulk acoustic wave resonator (FBAR) in which the bottom electrode is deposited through a hole in the bottom of the substrate wafer. This allows the use of various crystalline layers to seed the piezo material growth and allows higher Q's for the FBAR due to lack of bottom energy leakage paths.—JAH

6,864,619

43.35.Ns PIEZOELECTRIC RESONATOR DEVICE HAVING DETUNING LAYER SEQUENCE

Robert Aigner *et al.*, assignors to Infineon Technologies AG
8 March 2005 (Class 310/321); filed in Germany 18 May 2001

This patent describes a method of tuning a micromachined bulk acoustic wave resonator by using additional reflective layers of material on the bottom of a common substrate. This is apparently most useful when a number of these resonators must be linked together, as in a filter. There is little given in the way of design equations that would help a novice do this.—JAH

6,870,445

43.35.Ns THIN FILM BULK ACOUSTIC WAVE RESONATOR

Takashi Kawakubo *et al.*, assignors to Kabushiki Kaisha Toshiba
22 March 2005 (Class 333/187); filed in Japan 28 March 2002

This patent describes a novel way of fabricating high-Q acoustic resonators using a cavity beneath a film bulk acoustic wave resonator. This construction increases the resonator Q and keeps etchant materials away from the piezo material, a factor the authors cite as essential for good piezo film properties.—JAH

6,817,250

43.35.Ud ACOUSTIC GAS METER WITH A TEMPERATURE PROBE HAVING AN ELONGATED SENSOR REGION

Erik Cardelius and Lars Skoglund, assignors to Maquet Critical Care AB
16 November 2004 (Class 73/861.27); filed in Sweden 24 January 2002

This device uses the transmission of an ultrasonic beam through a portion of gas flowing in a tube to determine the average temperature of the gas within that region. In addition to the ultrasonic beam, a heated wire is placed within the path of the measured flow of gas, possibly looping around so as to expose a greater length of wire to the gas. Together, these mechanisms allow the temperature to be determined under varying conditions of flow and/or pressure.—DLR

6,870,304

43.38.Ar VIBRATORY MOTORS AND METHODS OF MAKING AND USING SAME

Bjoern Magnussen *et al.*, assignors to Elliptec Resonant Actuator AG
22 March 2005 (Class 310/323.02); filed 8 March 2001

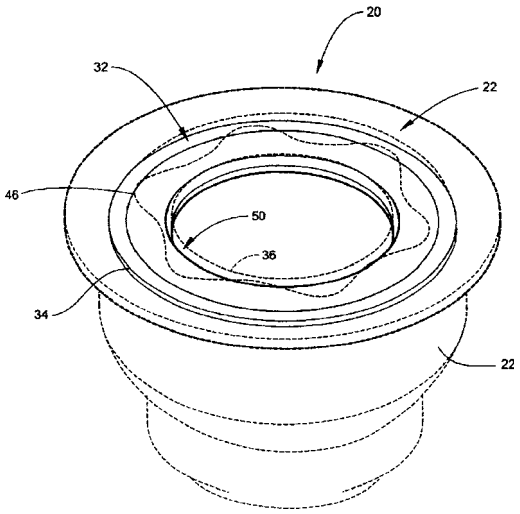
This patent describes a linear actuator (slider) driven by a motor vibrating in intermittent contact with the slider. There are 113 claims, most of them citing certain particular arrangements of the motor and actuator. The reader is advised to look elsewhere for ideas.—JAH

6,851,513

43.38.Ar TANGENTIAL STRESS REDUCTION SYSTEM IN A LOUDSPEAKER SUSPENSION

Brendon Stead *et al.*, assignors to Harvard International Industries, Incorporated
8 February 2005 (Class 181/172); filed 27 March 2002

Small multimedia loudspeakers are asked to do a lot with a little. The surround and spider can limit the excursion of these devices, at least if the distortion is to be held to a tolerable level at large (for these devices) excursions. Several interesting shapes for these support parts are described that are said to reduce the tangential and radial stresses therein. One of the



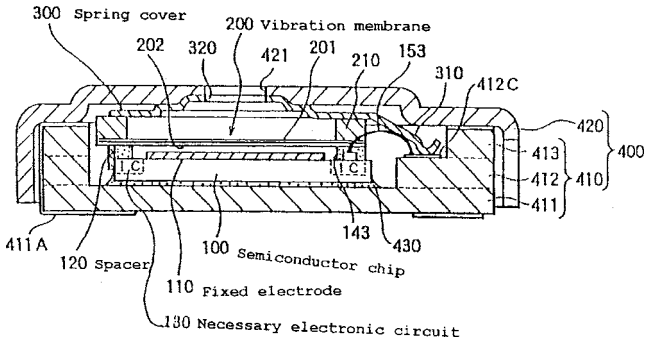
shapes, **46**, is a sinusoidal pattern for the roll of the suspension. A similar design is used for the spider, where this sinusoidal warp is applied to each concentric roll in the spider. The patent also mentions that although the height of the roll in each ridge does not vary, it could. No test results comparing prior art to the invention are included in the patent. The patent lists a ''Harvard International Industries'' as the assignee; might this be a *nom de guerre* for Harman International Industries?—NAS

6,870,938

43.38.Bs SEMICONDUCTOR ELECTRET CAPACITOR MICROPHONE

Takanobu Takeuchi *et al.*, assignors to Mitsubishi Denki Kabushiki Kaisha; Hosiden Corporation
22 March 2005 (Class 381/175); filed in Japan 26 April 2000

The patent describes mechanical details in the design of electret microphones. Both preamplification and noise canceling circuitry are integral to the structure.—JME

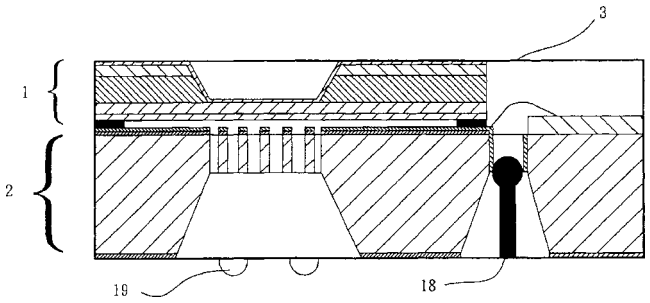


6,870,939

43.38.Bs SMT-TYPE STRUCTURE OF THE SILICON-BASED ELECTRET CONDENSER MICROPHONE

Dar-Ming Chiang and Tsung-Lung Yang, assignors to Industrial Technology Research Institute
22 March 2005 (Class 381/175); filed 28 November 2001

The patent describes an electret microphone that is largely ''grown'' in two separate layers, as are integrated circuits. One assembly includes the diaphragm portion, while the other includes backplate and preamplification. The two sections are then joined to form a very low cost microphone.—JME

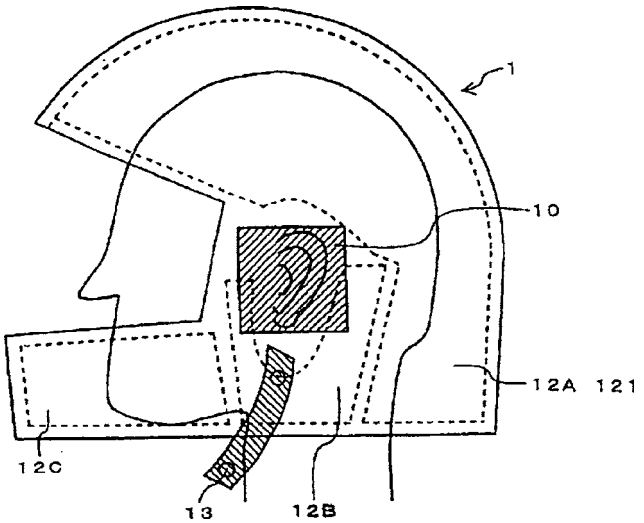


6,862,358

43.38.Fx PIEZO-FILM SPEAKER AND SPEAKER BUILT-IN HELMET USING THE SAME

Hajime Tabata, assignor to Honda Giken Kogyo Kabushiki Kaisha
1 March 2005 (Class 381/301); filed in Japan 8 October 1999

To prevent motorcyclists from becoming bored while threading through traffic at 70 miles per hour, several manufacturers now offer helmets with built-in headphones for communications or entertainment. This patent argues that a curved, piezofilm transducer **10** is ideal for this application. The patent includes frequency response curves showing smooth output from about 500 Hz to more than 10 kHz.—GLA



6,864,621

43.38.Fx PIEZOELECTRIC ELEMENT AND METHOD FOR MANUFACTURING THE SAME

Hirozumi Ogawa *et al.*, assignors to Murata Manufacturing Company, Limited
8 March 2005 (Class 310/358); filed in Japan 25 March 2002

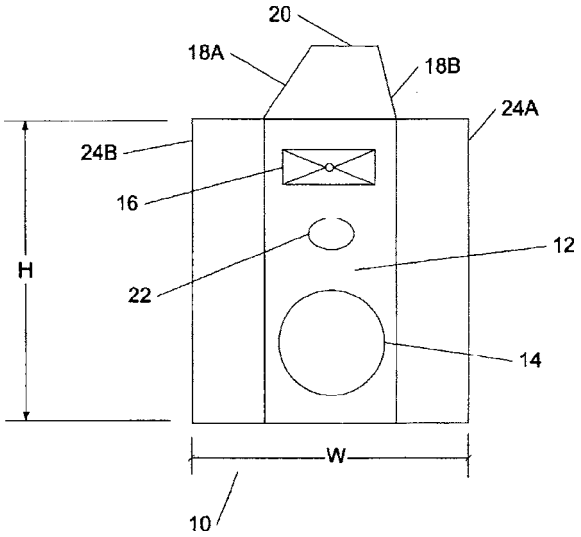
This patent describes the use of oriented piezoceramics of the NaBiTiO family to increase the coupling constant and the efficiency of energy trapping in certain layered structures. The patent is somewhat vague about what is actually novel about this. The manufacturing process is laid out in detail, but the methods involved in layering and materials do not seem to be unique. There is very little information in the patent to support the claims for enhanced electromechanical coupling in this material over the PbZrTiO family, and no mention is made of its scalability to micron dimensions.—JAH

6,860,363

43.38.Ja PLANAR ACOUSTIC WAVEGUIDE

Christopher Gardner, Hopkinton, New Hampshire and Christopher Huston, Brentwood, Tennessee
1 March 2005 (Class 181/155); filed 3 July 2002

Although the illustration looks like a pattern for a fold-up loudspeaker enclosure, wings **20**, **24A**, and **24B** are really continuations of front baffle **12**. The patent claims broaden the concept to ''at least one baffle extension comprising at least one acoustically reflective surface substantially coplanar



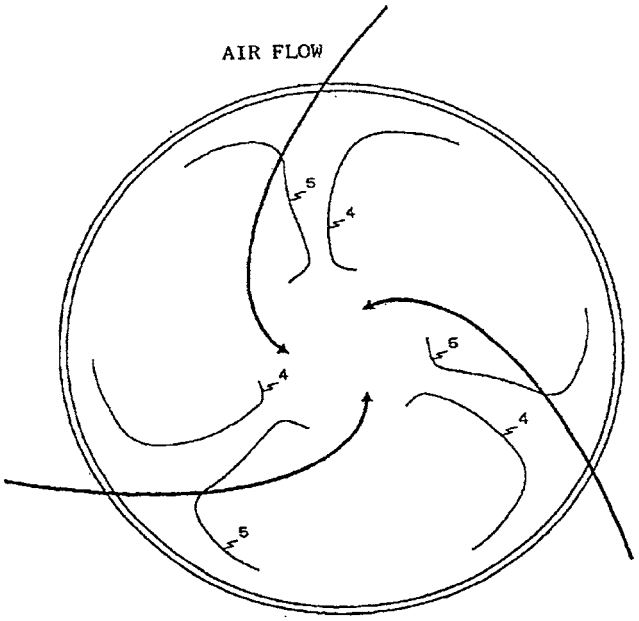
with the front surface of the baffle, and which baffle extension extends outwardly from at least one side of the baffle.'' Such loudspeaker baffle extensions have been standard practice in the motion picture industry for more than 50 years.—GLA

6,863,153

43.38.Ja LOUDSPEAKER DIAPHRAGM

Junichi Hayakawa and Masaya Kasai, assignors to Kabushiki Kaisha Kenwood
8 March 2005 (Class 181/173); filed in Japan 22 April 1999

This patent sets forth a novel thesis: by incorporating suitably shaped bumps in a loudspeaker cone a rotational force is created that redirects



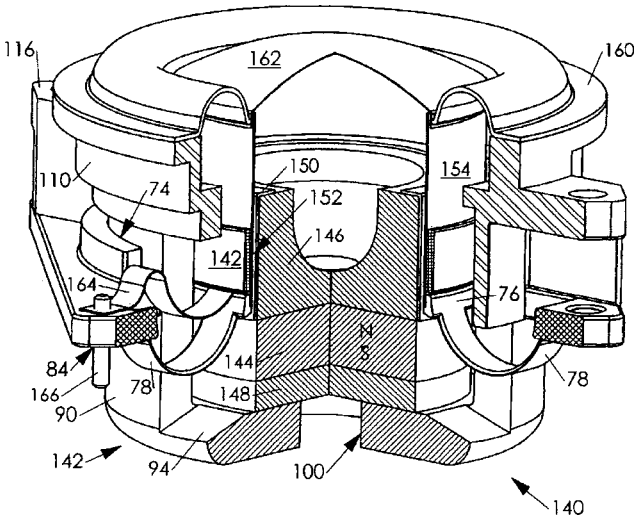
airflow, thus lowering pressure near the center of the cone and improving the quality of sound. I am not making this up.—GLA

6,865,282

43.38.Ja LOUDSPEAKER SUSPENSION FOR ACHIEVING VERY LONG EXCURSION

Richard L. Weisman, Pasadena, California
8 March 2005 (Class 381/404); filed 1 May 2003

Instead of a centering spider at the top of the voice coil, this long-throw speaker employs flexible springs **78** located at the bottom of the coil. JBL experimented with this concept in the 1970s, but in a somewhat differ-



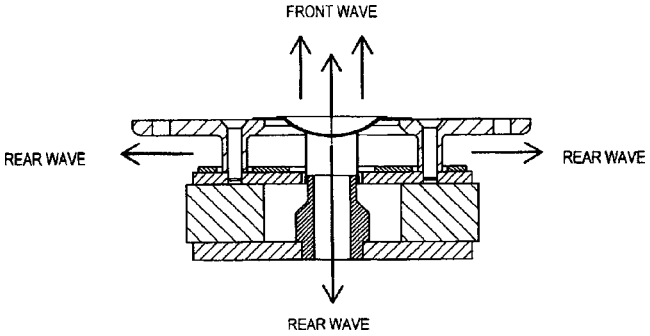
ent arrangement than is shown here. An obvious drawback is the added complexity of assembly. The inventor attempts to mitigate this by mounting the springs on a separate subassembly that locks in place.—GLA

6,870,941

43.38.Ja DIPOLE RADIATING DYNAMIC SPEAKER

Glenn A. Marnie, Oceanside, California
22 March 2005 (Class 381/337); filed 15 July 2002

This ''improved'' loudspeaker incorporates a vented pole piece to allow unrestricted radiation from the rear of the cone. Somehow, the inventor



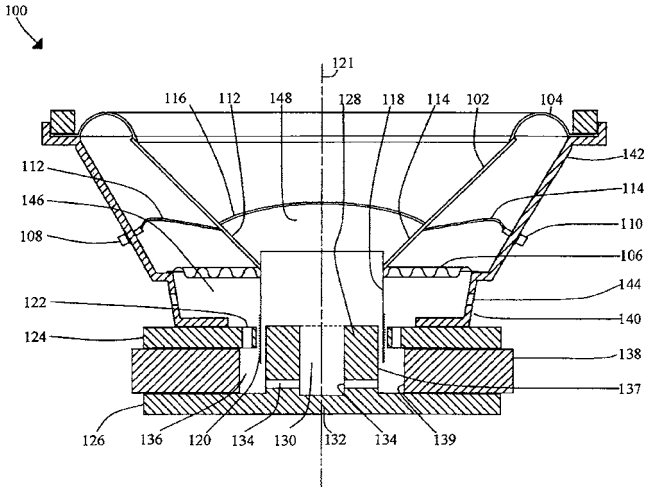
has managed to patent a well-known loudspeaker design that has been used since the 1930s. The mind boggles.—GLA

6,868,165

43.38.Ja LOUDSPEAKER

Frank W. Fabian, assignor to The Canadian Loudspeaker Corporation
15 March 2005 (Class 381/397); filed in Canada 8 September 1998

The claims of this patent are awash with terms like "distal," "longitudinal," "reciprocation," and "therein," but the concept is not all that complicated. The goal of the invention is to provide sufficient cooling to a



loudspeaker voice coil while at the same time minimizing back pressure that restricts cone motion. In the variant shown, vents **134** allow some of the air trapped under the center dome to escape via chamber **136** and vents **122**.—GLA

6,868,167

43.38.Ja AUDIO SPEAKER AND METHOD FOR ASSEMBLING AN AUDIO SPEAKER

Shiro Tsuda, assignor to Ferrotec Corporation
15 March 2005 (Class 381/415); filed 17 June 2002

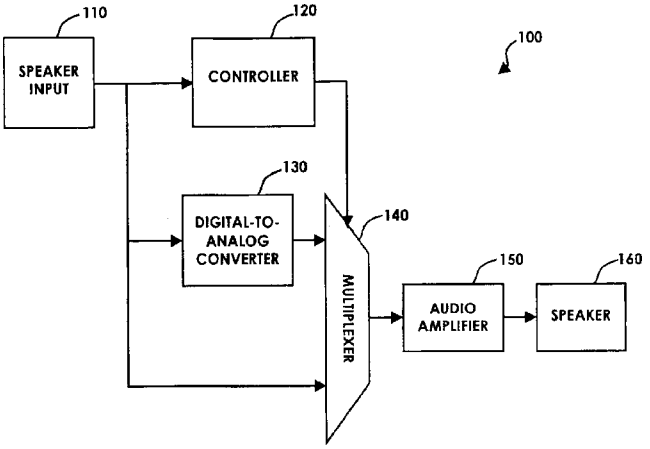
Although this invention has no direct acoustical implications, it is nonetheless interesting. Ferro-magnetic fluid is commonly injected into the gap of a moving coil loudspeaker to act as a coolant and provide additional mechanical damping. It (hopefully) remains in place for the life of the loudspeaker. In this case, a volatile magnetic fluid acts as a centering device during assembly. It subsequently evaporates, leaving behind a thin lubricating film on the pole pieces to provide protection from excessive cone excursions. During normal operation, the voice coil does not touch the film.—GLA

6,862,636

43.38.Lc MULTI-MODE SPEAKER OPERATING FROM EITHER DIGITAL OR ANALOG SOURCES

Bruce Young, assignor to Gateway, Incorporated
1 March 2005 (Class 710/69); filed 16 November 2001

In today's world of computerized electronics, the signal from an audio output jack may be either analog or digital. A "smart" self-powered loudspeaker should be able to tell the difference and adapt itself accordingly.



Possible methods for doing this are described, but the patent claims cover the basic concept as well as any conceivable method and/or hardware.—GLA

6,870,795

43.38.Pf ACOUSTIC SOURCE ARRAY SYSTEM MODULE FOR UNDERWATER OPERATION WHICH CAN BE INSTALLED ON A MOTORIZED BOAT

John V. Bouyoucos and Dennis R. Courtright, assignors to Hydroacoustics Incorporated
22 March 2005 (Class 367/144); filed 21 January 2003

An array of airguns is described. The array is realized as a ladder-type structure in which the airguns are the rungs of the ladder disposed between two parallel vertical support and umbilical lines. The airguns in adjacent rungs are laterally offset to maximize the acoustic output of the array. The umbilical cords provide the compressed air to the airguns as well as control signals to time their actuations. The whole structure is sufficiently flexible that it may be wound on a drum by a winch on the boat for deployment and retrieval.—WT

6,862,002

43.38.Si ANTENNA GROUND PLANE AND WIRELESS COMMUNICATION DEVICE WITH ANTENNA GROUND PLANE AND ACOUSTIC RESISTOR

Adam M. Demicco *et al.*, assignors to Motorola, Incorporated
1 March 2005 (Class 343/846); filed 28 May 2003

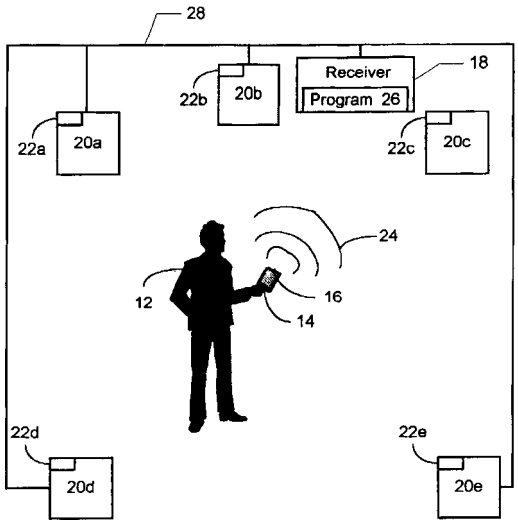
In a small cellular phone, a wire mesh ground plane can also serve as a protective grill over the receiver sound ports. Moreover, the acoustic resistance thus introduced can alter frequency response by a decibel or two. The patent text not only proclaims the "inventive principles" of the design but also takes care to define specific words such as "a" and "an."—GLA

6,856,688

43.38.Tj METHOD AND SYSTEM FOR AUTOMATIC RECONFIGURATION OF A MULTI-DIMENSION SOUND SYSTEM

Daryl Carvis Cromer *et al.*, assignors to International Business Machines Corporation
15 February 2005 (Class 381/303); filed 27 April 2001

Remote control **16** sends a signal to receiver **18** which in turn sends a signal to the speakers **20** in the sound system **10** so they will begin to count the pulses **24** emitted by the remote control. Receiver **18** then queries the



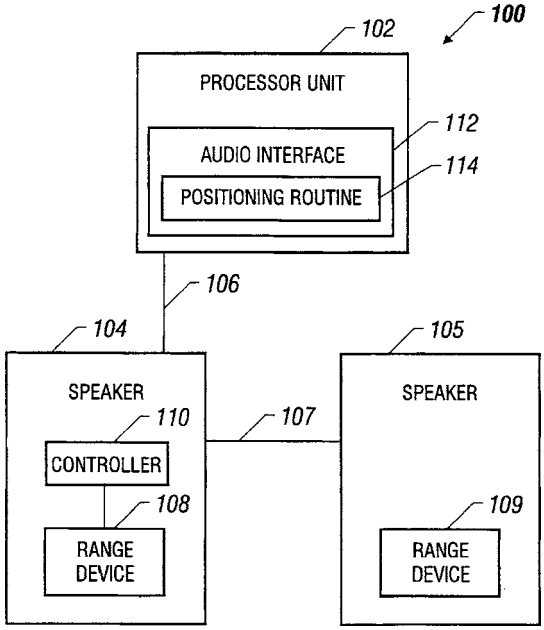
speakers which then send the pulse counts to the receiver where program **26** determines the distance between each of the speakers **20** and the remote control **16** based on the received wave counts. The computed distances are used to ''program the digital encoding system with the appropriate speaker delay times.'' This, the patent admits in the concise description of the invention, allows for the autoconfiguration of a multi-dimension sound system.—NAS

6,859,417

43.38.Tj RANGE FINDING AUDIO SYSTEM

Todd C. Houg, assignor to Micron Technology, Incorporated
22 February 2005 (Class 367/96); filed 7 May 1999

A system is described for optimizing the delay settings for a sound system, such as audio system **100**, and as might be found in a laptop computer. Range devices **108** and **109** may be used, along with controller **110**, to determine where a listener may be located at a particular time. The distance data is sent to the processor unit **102** via communication link **107**. An audio interface **112** may include a positioning routine **114** which may be used to



modify the audio output sent to speakers **104** and **105**. The patent clearly states that this method may be extended to three-dimensional audio systems.—NAS

6,861,914

43.40.Cw MONOLITHIC VIBRATION ISOLATION AND AN ULTRA-HIGH Q MECHANICAL RESONATOR

Douglas Photiadis and Angie Sarkissian, assignors to The United States of America as represented by the Secretary of the Navy
1 March 2005 (Class 331/156); filed 30 September 2002

This patent describes the attachment of ''isolation'' masses to the base of a beam that is oscillating either in a flexural (cantilever) or a torsional mode. The masses act as separate tuned mechanical resonators which, when driven near resonance, act as large impedances to block the leakage of energy down the beam into the supporting structure.—JAH

6,870,300

43.40.Cw MICRO-ELECTRICAL-MECHANICAL SYSTEM (MEMS) DEVICE HAVING A PLURALITY OF PAIRS OF REFLECTIVE ELEMENT ACTUATORS LOCATED ON OPPOSING SIDES OF A REFLECTIVE ELEMENT AND A METHOD OF MANUFACTURE THEREFOR

Cristian A. Bolle and Edward Chan, assignors to Lucent Technologies Incorporated; Agere Systems Incorporated
22 March 2005 (Class 310/309); filed 1 December 2001

This patent describes a novel kind of flexure element that can be actuated electrostatically. It appears to be well suited to mirror- and valve-type devices. A complete description of the fabrication process is given.—JAH

6,865,944

43.40.Kd METHODS AND SYSTEMS FOR DECELERATING PROOF MASS MOVEMENTS WITHIN MEMS STRUCTURES

Max C. Glenn *et al.*, assignors to Honeywell International Incorporated
15 March 2005 (Class 73/504.12); filed 16 December 2002

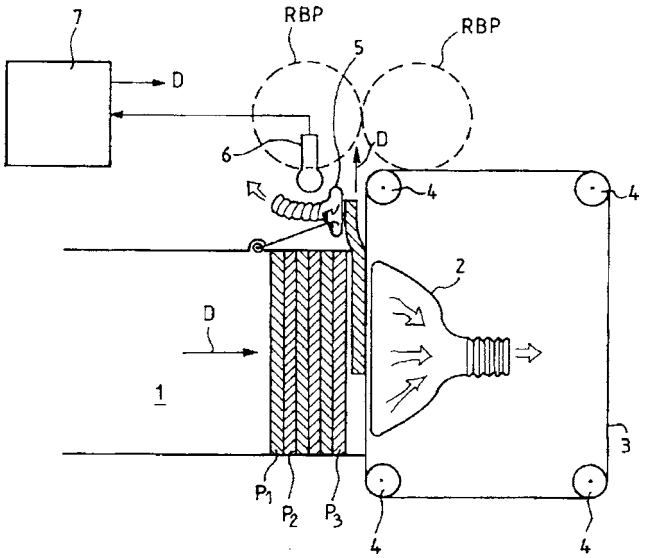
This patent describes a type of stop for moving silicon sensor elements that is called a ''deceleration extension.'' It is meant to engage certain protrusions or indentations when the device is subjected to extreme accelerations, and thereby soften the blow as the sensor element (accelerometer or gyro) comes to rest. There is not much described here that would not have occurred to a reasonably astute practitioner of the art.—JAH

6,811,034

43.40.Le ACOUSTIC METHOD FOR DISCRIMINATING PAPER AND PLASTIC ENVELOPES

François Chaume and Jean-Marc Teluob, assignors to Solystic
2 November 2004 (Class 209/591); filed in France 7 April 2000

During the processing of mail articles, it is useful to be able to separate items in paper envelopes from items in plastic envelopes. In the patented device, envelopes **P1**, **P2**, and **P3** are fed out and pass over suction nozzles



2 and **5**. The resulting sound is picked up by microphone **6** and compared with reference signals to determine the composition of the envelopes.—DLR

6,871,149

43.40.Le CONTACT SENSITIVE DEVICE

Darius Martin Sullivan and Nicholas Patrick Roland Hill, assignors to New Transducers Limited
22 March 2005 (Class 702/56); filed in the United Kingdom 6 December 2002

This patent relates to touch-sensitive screens and similar devices where the location of the contact needs to be determined. A member, such as a screen, that supports bending waves is provided with several sensors that can measure bending wave signals. The location of a contact is determined via a processor from the phase differences between the signals produced by the various sensors. Waves reflected from the edges of the wave-bearing element may be suppressed by means of absorbers mounted on the edges, with the impedance of the absorbers matched to that of the wave-bearing element, particularly in a selected limited frequency band.—EEU

6,860,369

43.40.Tm VIBRATION DAMPER WITH VARIABLE DAMPING FORCE

Raimund Weiffen and Wolfgang Hertz, assignors to Mannesmann Sachs AG
1 March 2005 (Class 188/282.4); filed in Germany 26 October 1998

This damper, intended for use as a vehicle suspension component, employs a piston that divides the fluid-filled volume of a cylinder into two working spaces. One-way-flow valves provide different flow restrictions in the rebound and compression directions. An adjustable flow restriction in series with these valves provides additional flow restrictions, which may be controlled either passively or actively.—EEU

6,863,628

43.40.Tm VIBRATION DAMPING STRIKING IMPLEMENT

Richard A. Brandt, New York, New York
8 March 2005 (Class 473/520); filed 20 March 2000

A sports implement, such as a baseball bat, golf club, or tennis racket, subjects the user's hands to shocks when the implement strikes a ball or other object. These shocks are associated with transverse vibrations of the implement, and the present patent addresses reduction of the corresponding vibration transmission to the hands. The handle end of the striking implement is tapered and enclosed in a tubular barrel handle, with an elastomeric material inserted between the implement and the barrel. The material's modulus and loss factor are selected so as to maximize the absorption of transverse vibrations.—EEU

6,863,629

43.40.Tm VIBRATION DAMPING TAPE

Thomas Falone, Mickelton, New Jersey *et al.*
8 March 2005 (Class 473/520); filed 10 September 2003

This patent pertains to cushioning of impacts, and not to viscoelastic damping tapes that provide energy dissipation. A tape according to this patent is intended to be placed around the handle of a tool or sport implement that subjects the user's hands to shock or vibration. The tape consists in essence of three layers: a resilient layer in contact with the vibrating implement, a relatively rigid layer, and an outer resilient layer grasped by the user's hand.—EEU

6,870,303

43.40.Tm MULTI-MODE VIBRATION DAMPING DEVICE AND METHOD USING NEGATIVE CAPACITANCE SHUNT CIRCUITS

Chul-hue Park, assignor to Pohang University of Science and Technology Foundation
22 March 2005 (Class 310/319); filed in the Republic of Korea 8 May 2002

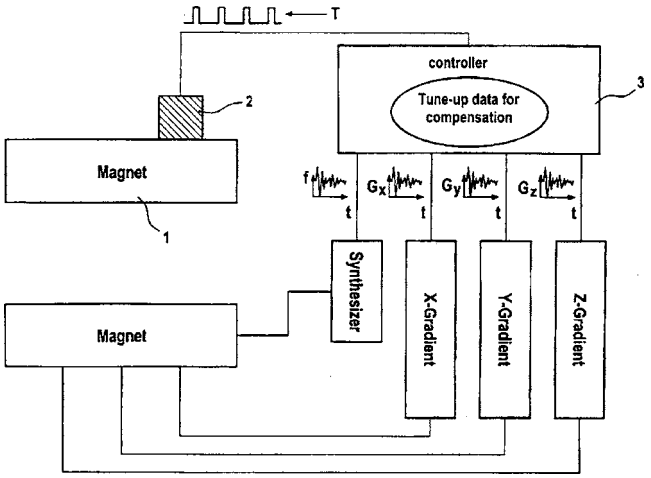
A pair of piezoelectric patches is attached to a vibrating structure, such as a beam. One member of the pair is connected to a circuit consisting of a resistance and a negative capacitance in series; the other is connected to a circuit consisting of similar elements in parallel. The first can provide significant damping at low frequencies, the second, at high frequencies. The effect of a negative capacitance can be obtained by means of a synthetic negative impedance unit based on an operational amplifier.—EEU

6,864,682

43.40.Vn METHOD FOR VIBRATION COMPENSATION IN A MAGNETIC RESONANCE TOMOGRAPHY APPARATUS

Joerg Fontius and Volker Weissenberger, assignors to Siemens Aktiengesellschaft
8 March 2005 (Class 324/309); filed in Germany 15 May 2002

The object of this device is to enable more precise compensations of the vibrational oscillations of the magnets caused by the cryo-head in a magnetic resonance tomography apparatus. This is achieved when the compensation device sets the synthesizer frequency and/or the gradient currents according to the time curve of field terms of the zeroth and first orders



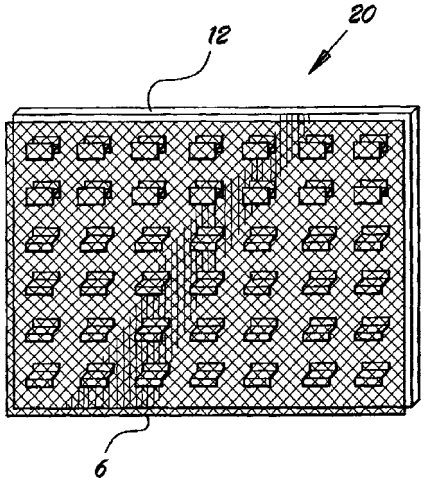
acquired in a tune-up. Motion trigger pulses from the cryo-head can be used to ensure the synchronization of its vibrations.—DRR

6,868,940

43.50.Gf SOUND ABSORBING PANEL

Julius Mekwinski, Sanford, Florida
22 March 2005 (Class 181/290); filed 29 April 2003

This sound absorbing panel is basically a sandwich with a back plate on one side and a screen or perforated plate on the other. These are separated by a honeycomb core and a layer of stand-off brackets (Z shaped or angled). There may also be a layer of felt (or fabric, canvas, fiber, or other material)



between the screen and the honeycomb core. The spacing of the standoffs is $\frac{1}{4}$ of the wavelength of the desired frequency to be absorbed. These features make the panel inexpensive to manufacture using standard sizes. Common applications for the panel are for traffic or jet engine noise absorption.—CJR

6,862,567

43.50.Ki NOISE SUPPRESSION IN THE FREQUENCY DOMAIN BY ADJUSTING GAIN ACCORDING TO VOICING PARAMETERS

Yang Gao, assignor to Mindspeed Technologies, Incorporated
1 March 2005 (Class 704/228); filed 30 August 2000

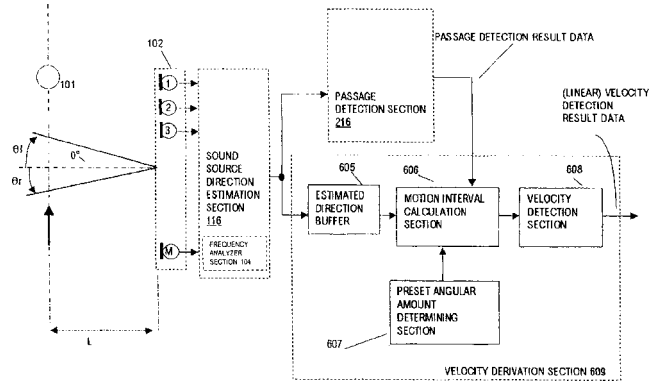
A method is described for noise suppression in a speech signal which extends the prior art technique of spectral weighting in the frequency domain, as first employed to determine the basic signal-to-noise ratio. Improvements put forth in the patent are described thoroughly, and include a silence enhancement procedure and a variable channel gain which depends upon classification of the signal segment as voiced/unvoiced speech or as speech-intrinsic/background noise. In this way the mistaking of speech components for unwanted noise by the initial simple metric is adumbrated.—SAF

6,862,541

43.60.Fg METHOD AND APPARATUS FOR CONCURRENTLY ESTIMATING RESPECTIVE DIRECTIONS OF A PLURALITY OF SOUND SOURCES AND FOR MONITORING INDIVIDUAL SOUND LEVELS OF RESPECTIVE MOVING SOUND SOURCES

Koichiro Mizushima, assignor to Matsushita Electric Industrial Company, Limited
1 March 2005 (Class 702/76); filed in Japan 14 December 1999

The patent relates to a complex set of strategies for detecting bearing angle and motion data for one or more moving sound sources, while simultaneously monitoring the levels of each of those individual sources! The



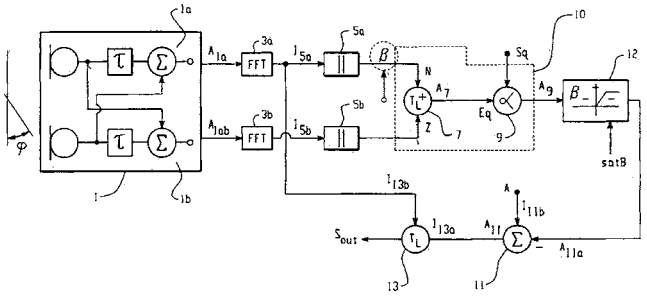
application is for use in traffic monitoring and control. The figure gives some idea of the complexities involved here, and the lengthy patent discusses eight specific embodiments of the scheme.—JME

6,865,275

43.60.Fg METHOD TO DETERMINE THE TRANSFER CHARACTERISTIC OF A MICROPHONE SYSTEM, AND MICROPHONE SYSTEM

Hans-Ueli Roeck, assignor to Phonak AG
8 March 2005 (Class 381/92); filed in Switzerland 31 March 2000

The patent describes a sophisticated method for controlling signal gain in hearing aid applications. As shown in the figure, the outputs of two microphones are processed in a nonlinear fashion so that sounds arriving from a bearing angle of 90° will be limited in amplitude gain, while those



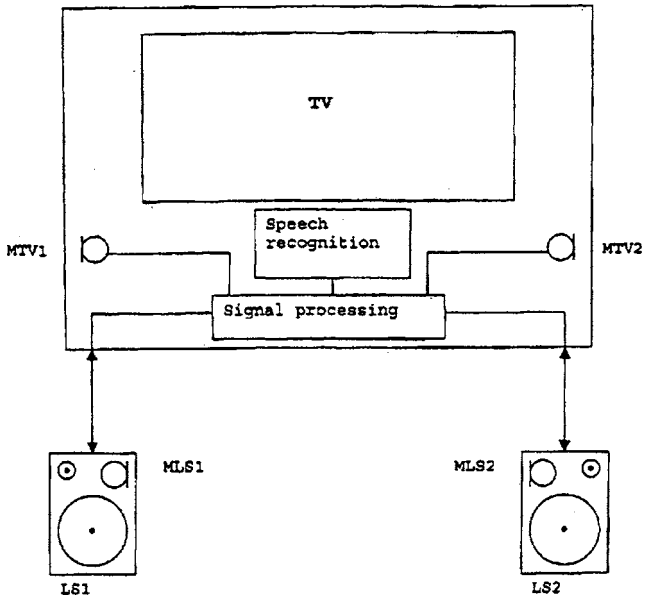
arriving from the front and back of the array (0° and 180°) will not. This should help discriminate angular pickup in favor of those signals arriving from the front.—JME

6,868,045

43.60.Fg VOICE CONTROL SYSTEM WITH A MICROPHONE ARRAY

Ernst F. Schröder, assignor to Thomson Licensing S.A.
15 March 2005 (Class 367/198); filed in Germany 14 September 1999

This simple patent proposes a microphone array that spans the width of the stereo loudspeakers in a typical home theater setup. The microphones



are used to receive voice commands for operating the systems; the wider the spread, the greater the acuity of the control system. The figure says it all.—JME

6,862,558

43.60.Gk EMPIRICAL MODE DECOMPOSITION FOR ANALYZING ACOUSTICAL SIGNALS

Norden E. Huang, assignor to The United States of America as represented by the Administrator of the National Aeronautics and Space Administration
1 March 2005 (Class 702/194); filed 13 February 2002

The signal decomposition method disclosed here attempts to move away from frequency-domain spectral analysis entirely and appears to be finding components of the signal by extensive computation in the time domain. By analyzing the signal directly for oscillation patterns around the local mean, a set of ''intrinsic mode functions'' is numerically extracted. These functions are then used, at that point in time, as a basis for numerical spectral representation by applying a Hilbert transform. The resulting time-frequency ''distribution'' is completely empirical in the components discovered in the time domain, having no closed analytic form. This is an exciting new development, previously published in part by the inventor, which in some sense returns to the earliest history of signal analysis before Fourier spectra could be computed.—SAF

6,868,162

43.60.Pt METHOD AND APPARATUS FOR AUTOMATIC VOLUME CONTROL IN AN AUDIO SYSTEM

Christopher Michael Jubien *et al.*, assignors to Mackie Designs Incorporated
15 March 2005 (Class 381/107); filed 17 November 2000

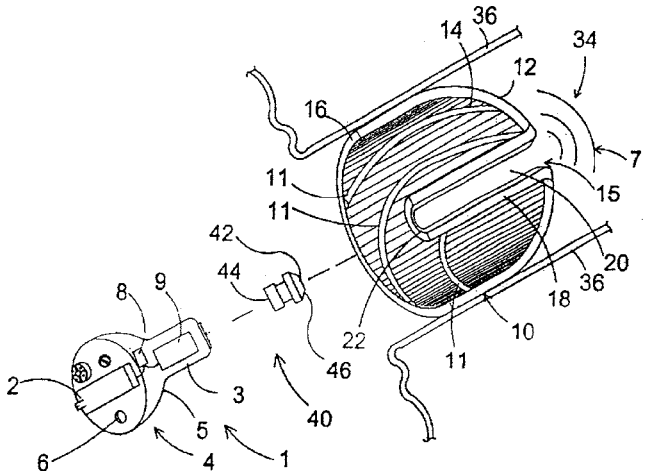
A system is described which updates an adaptive filter in real time in response to the error signal (ambient minus filtered original) detected in an audio playback environment. The adaptive filter coefficients, updated using the error signal according to a formula provided, are intended to simulate the room transfer function, so the error signal is an estimate of the ambient noise. An automatic playback gain compensator can then operate more reliably based on the hoped-for accurate measure of the ambient noise.—SAF

6,860,362

43.66.Ts HEARING AID INSTRUMENT FLEXIBLE ATTACHMENT

Oleg Saltykov, assignor to Siemens Hearing Instruments, Incorporated
1 March 2005 (Class 181/135); filed 21 February 2003

A custom hearing aid housing fits into the opening of a flexible cup that is inserted into the ear canal of a hearing aid wearer. Spiral-shaped ribs



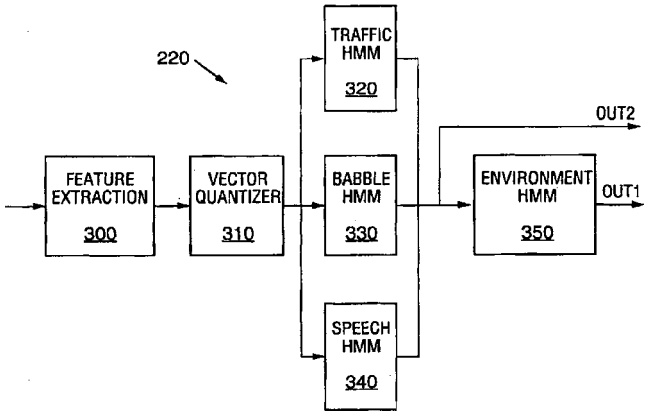
on an inner surface of the flexible cup prevent collapsing and help create a tight seal to the ear canal. An adaptor connects the housing's sound output to a cylindrical member in the flexible earpiece cup.—DAP

6,862,359

43.66.Ts HEARING PROSTHESIS WITH AUTOMATIC CLASSIFICATION OF THE LISTENING ENVIRONMENT

Nils Peter Nordqvist and Arne Leijon, assignors to GN ReSound A/S
1 March 2005 (Class 381/312); filed 29 May 2002

A methodology is proposed for a hearing aid that automatically adjusts itself to particular performance parameters when determination is made that the wearer is in a particular listening environment. Hidden Markov models are employed for input signal analysis to extract features that classify the



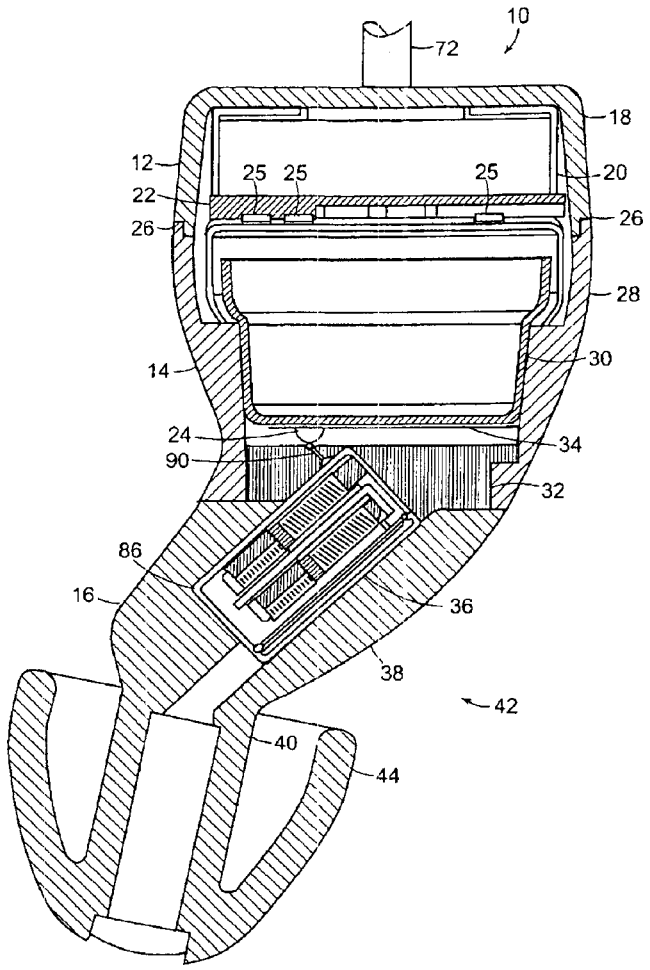
listening environment. The acoustic environment vectors extracted include differential spectral and differential temporal signal features.—DAP

6,865,279

43.66.Ts HEARING AID WITH A FLEXIBLE SHELL

Marvin A. Leedom, assignor to Sarnoff Corporation
8 March 2005 (Class 381/322); filed 13 March 2001

To streamline the assembly process, a three-piece hearing aid is described with a microphone and electronics in the first section, a battery in



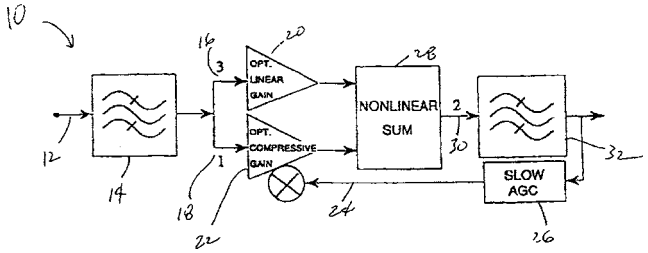
the second section, and a receiver and compliant tip in the third section. A flex circuit connects the microphone, electronics, battery, and receiver.—DAP

6,868,163

43.66.Ts HEARING AIDS BASED ON MODELS OF COCHLEAR COMPRESSION

Julius L. Goldstein, assignor to BECS Technology, Incorporated; Hearing Emulations, LLC
15 March 2005 (Class 381/321); filed 22 September 1998

A methodology is recommended for multichannel compression in hearing aids which would involve fast gain compression at intermediate sound levels and linear gain at high levels. In each channel, a slow-acting AGC, which acts on sustained high-level sounds, is utilized in a feedback



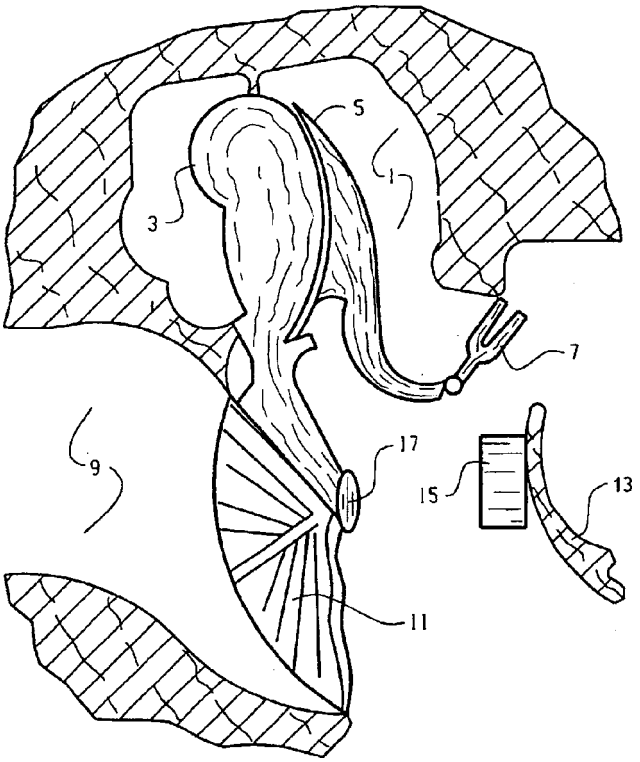
loop to control a fast-acting, compressive gain block. A linear gain block output and the compressive gain block output are summed nonlinearly using a cochlear filterbank model.—DAP

6,869,391

43.66.Ts IMPLANTED HEARING AIDS

Herbert Bächler *et al.*, assignors to Phonak AG
22 March 2005 (Class 600/25); filed 17 August 2001

To provide greater amplification than previous implanted hearing aids, a large permanent magnet is positioned on the rigid promontory to prevent loading on the ossicles. Current injected into a small coil placed behind the



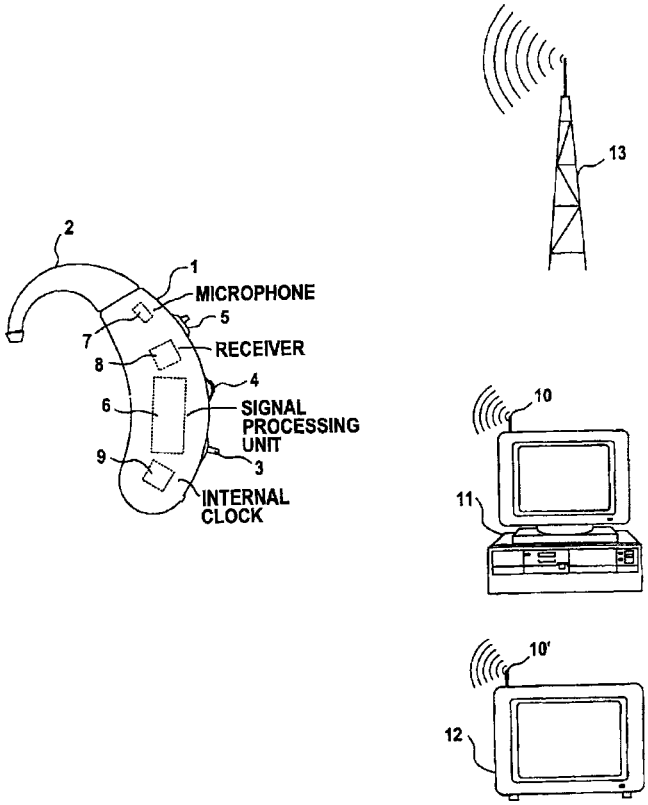
ear drum or on the ossicles is said to produce relatively large deflections and high force levels due to the large magnetic flux from the magnet.—DAP

6,870,940

43.66.Ts METHOD OF OPERATING A HEARING AID AND HEARING-AID ARRANGEMENT OR HEARING AID

Wolfram Meyer and Torsten Niederdränk, assignors to Siemens Audiologische Technik GmbH
22 March 2005 (Class 381/314); filed in Germany 29 September 2000

The most recent digital hearing aids attempt to make a classification of the wearer's current acoustic environment. This determination, which is not always accurate, may be used to switch the hearing aid circuitry automatically to an appropriate set of parameters for that environment. To improve



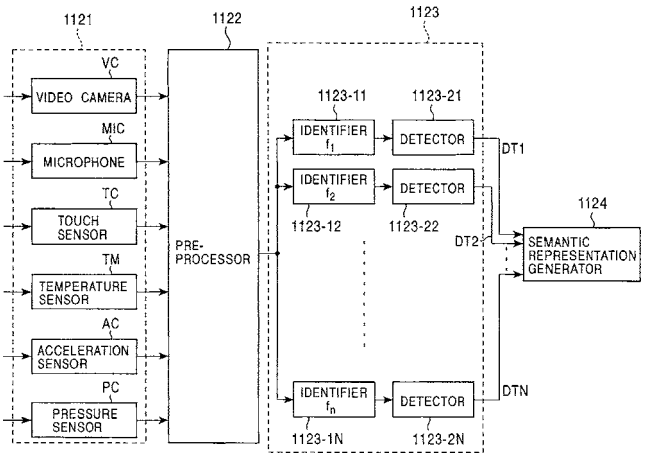
the accuracy of environmental detection, a technique is proposed which determines parameter selection automatically when the wearer comes into close proximity with an associated transmitter placed, for example, in the car, workplace, or living room.—DAP

6,816,831

43.71.Hw LANGUAGE LEARNING APPARATUS AND METHOD THEREFOR

Naoto Iwahashi, assignor to Sony Corporation
9 November 2004 (Class 704/9); filed in Japan 28 October 1999

This is not, as might be inferred from the title, a system to teach languages to humans, but, rather, is a language-learning computer system. According to the lofty goals stated here, the system would be able to construct semantic, syntactic, and phonetic models of the input based not only on audio and visual input signals, but may also have sensors for temperature, pressure, acceleration, etc. Oh, yes, and in the process, the system also learns to read and write. The figure shows what is clearly only a tiny part of the overall plan.—DLR

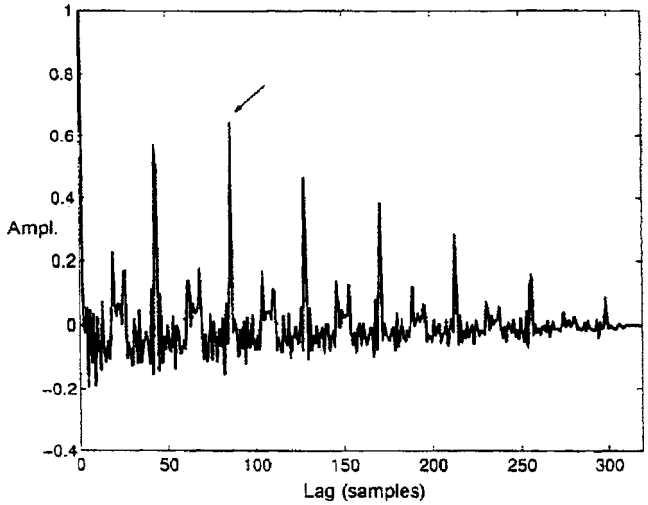


6,865,529

43.72.Ar METHOD OF ESTIMATING THE PITCH OF A SPEECH SIGNAL USING AN AVERAGE DISTANCE BETWEEN PEAKS, USE OF THE METHOD, AND A DEVICE ADAPTED THEREFOR

Cecilia Brandel and Henrik Johannisson, assignors to Telefonaktiebolaget L M Ericsson (publ)
8 March 2005 (Class 704/207); filed in the European Patent Office 6 April 2000

An improved pitch tracking method, based on time-domain conformity techniques such as autocorrelation, is described. Rather than simply selecting the supposed fundamental peak in the autocorrelation, for example, it is proposed to first detect the series of peaks in the autocorrelation and to then



estimate the average distance between the peaks. This latter value is then adopted as a chief candidate for the true pitch in a pitch tracker, combating pitch-halving as well as pitch-doubling errors which plague more naive procedures.—SAF

6,871,106

43.72.Gy AUDIO SIGNAL CODING APPARATUS, AUDIO SIGNAL DECODING APPARATUS, AND AUDIO SIGNAL CODING AND DECODING APPARATUS

Tomokazu Ishikawa *et al.*, assignors to Matsushita Electric Industrial Company, Limited
22 March 2005 (Class 700/94); filed in Japan 11 March 1998

An audio encoder/decoder is proposed in which a normalized time-to-frequency conversion is first performed and an audio signal is then coded via a first stage quantizer and subsequent encoder stages that quantize the quantization error outputs from a previous stage encoder. Frequency bands are selectively quantized if they have summed quantization errors larger than predetermined thresholds. A decision is performed on which bands to use after psychoacoustic weighting.—DAP

6,801,931

43.72.Ja SYSTEM AND METHOD FOR PERSONALIZING ELECTRONIC MAIL MESSAGES BY RENDERING THE MESSAGES IN THE VOICE OF A PREDETERMINED SPEAKER

Rajaram Ramesh *et al.*, assignors to Ericsson Incorporated
5 October 2004 (Class 709/206); filed 20 July 2000

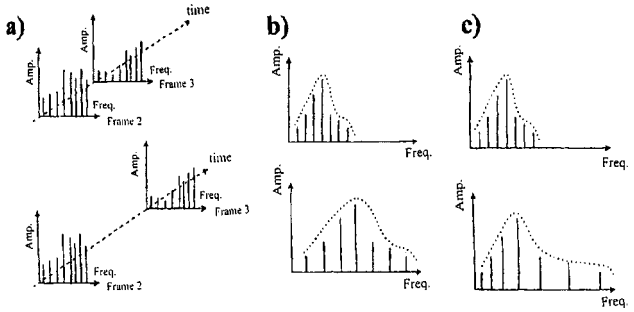
The idea here is that, if you like to have your email messages spoken aloud, you would hear the messages from each sender spoken in a specific, personalized voice. There is only a brief statement without further elaboration that this could be done by extracting appropriate ''basis vectors'' from a linear prediction representation of the speech. The basis vector for a particular speaker could then be obtained from a local database or sent as an attachment with the message.—DLR

6,804,649

43.72.Ja EXPRESSIVITY OF VOICE SYNTHESIS BY EMPHASIZING SOURCE SIGNAL FEATURES

Eduardo Reck Miranda, assignor to Sony France S.A.
12 October 2004 (Class 704/258); filed in the European Patent Office 2 June 2000

Intended for research purposes, this source/filter-style speech synthesizer design involves special emphasis on the excitation source. A glottal source pulse is built up by adding a small number of sine waves, the results of which are smoothed and concatenated by an overlap-and-add technique



before being used to drive the vocal tract resonator model. The figure illustrates some of the ways the control parameter sets may be modified before being used: (a) time expansion and (b) linear and (c) nonlinear frequency expansion. Libraries of source parameters are classified, for example, by diph one context or by emotive content.—DLR

6,813,607

43.72.Ja TRANSLINGUAL VISUAL SPEECH SYNTHESIS

Tanveer Afzal Faruquie *et al.*, assignors to International Business Machines Corporation
2 November 2004 (Class 704/276); filed 31 January 2000

As otherwise expressed in the title, the goal of this patented technique is language-independent facial animation of a speaker using only a single-language (e.g., English) speech recognizer. The argument is made that facial movements are related to sounds in a way that is independent of the language. In fact, the recognizer is, to an extent, modified in a way that is specific to the target visual display language, at least in that a mapping is constructed between the phoneme sets of the recognizer language and the target language. It is acknowledged that there may be ''holes'' left in the target phoneme set, where it does not match well to the recognizer langauge phoneme set. It is then assumed that the recognizer will function sufficiently well under those conditions.—DLR

6,816,837

43.72.Ne VOICE MACROS FOR SCANNER CONTROL

Kenneth P. Davis, assignor to Hewlett-Packard Development Company, L.P.
9 November 2004 (Class 704/275); filed 6 May 1999

Although the abstract lists additional details, such as operation on one or multiple processors, the essence of this patent is the use of macros to improve the performance of voice control of various types of devices. By a ''voice macro'' is meant that any voice input sequence may be arbitrarily associated with any sequence of one or more device control commands. In the example given, the individual device commands have already been associated with specific voice controls. In that case, a macro is created by first speaking the sequence of device commands to be performed as a set and then speaking the new word or phrase to be defined as the ''macro'' to perform the sequence of commands.—DLR

6,801,897

43.72.Ne METHOD OF PROVIDING CONCISE FORMS OF NATURAL COMMANDS

Thomas A. Kist and James R. Lewis, assignors to International Business Machines Corporation
5 October 2004 (Class 704/275); filed 28 March 2001

As the performance of speech recognizers has improved, there has been a trend toward allowing the user to speak in a more natural style. But this freedom of expression implies a proliferation of possible ways to state a given command. This recognizer will reduce all equivalent expressions to a common form, based on a natural-language command grammar. It is then presumed desirable for the system to choose, based on the minimum length, which one of the many is the ''real'' command.—DLR

6,804,644

43.72.Ne TEXT PROCESSING SYSTEM INCLUDING A SPEECH RECOGNITION DEVICE AND TEXT CHANGE MEANS FOR CHANGING A TEXT-BLOCK DATA

Gabor Janek *et al.*, assignors to Koninklijke Philips Electronics N.V.
12 October 2004 (Class 704/235); filed in the European Patent Office 3 March 1998

This patent concerns a method of text editing by the use of a speech recognition system. The cited prior patents all deal with narrow issues in editing text by voice, ignoring the extensive material on dictation by speech recognition. Accordingly, the claims here seem to deal with a narrow issue of how to control one or multiple text editors, whether in a single, or in multiple, editing applications.—DLR

6,804,645

43.72.Ne DYNAMIC PHONEME DICTIONARY FOR SPEECH RECOGNITION

Peter Kleinschmidt, assignor to Siemens Aktiengesellschaft
12 October 2004 (Class 704/243); filed in Germany 2 April 1996

This patent describes a dynamic dictionary for use in a speech recognition system. In order to limit memory usage and to prevent the degradation in recognition which occurs with a large vocabulary, words are generated and stored in the form of phonetic sequences as they are recognized, but are maintained by the system only as needed for a particular document under consideration. Methods are described for forming a permanent dictionary for the most commonly occurring words, as well as possible additional dictionaries containing certain ''preferred'' words.—DLR

6,801,896

43.72.Ne VOICE-BASED SEARCH AND SELECTION OF SPEECH RECOGNITION DATA

Koji Endo, assignor to Pioneer Corporation
5 October 2004 (Class 704/270); filed in Japan 30 June 1999

This speech recognition system is geared toward the control of electronic equipment, such as audio devices, radio, CD player, etc., in an automobile. Prior recognizers have been described which would allow an arbitrary phrase to be stored to perform any given device function. The added capability presented here addresses the problem of the user forgetting the keywords which have been stored for a specific function. During normal operation, the recognizer converts the spoken phrase into a phonetic string, which serves as the device control string. The phonetic string, along with a recorded snippet of the original input, is stored and keyed to the selected device function. A keypad on the voice unit has keys for ''forward scan'' and ''reverse scan.'' Pressing one of these keys produces a playback of the stored voice fragment for each audio device function. The keypad operations for these and other manipulations are described in some detail.—DLR

6,816,836

43.72.Ne METHOD AND APPARATUS FOR AUDIO-VISUAL SPEECH DETECTION AND RECOGNITION

Sankar Basu *et al.*, assignors to International Business Machines Corporation
9 November 2004 (Class 704/270); filed 30 August 2002

This patent describes a possible way to use video inputs of a talking face together with audio input from a microphone to achieve better performance from a speech recognizer. The discussion includes normalization of any view to a frontal facial view, detection and processing of visual features, and coordination of information obtained from audio and visual inputs.—DLR

6,868,382

43.72.Ne SPEECH RECOGNIZER

Makoto Shozakai, assignor to Asahi Kasei Kabushiki Kaisha
15 March 2005 (Class 704/254); filed in Japan 9 September 1998

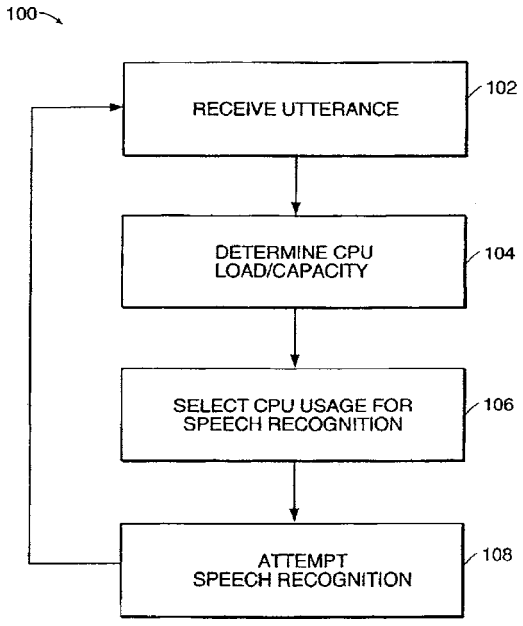
A method for speech recognition is described which is a hybrid of known speaker-independent and speaker-specific techniques. By using hidden Markov models (HMMs) of ''acoustic events'' in certain words registered to certain users in place of the general phonemic models employed for other words, a more specific overall HMM is obtained for the particular pronunciations of the registered words. The hybrid approach results in modest improvements in recognition performance for the utterances of the registered users.—SAF

6,862,570

43.72.Ne LOAD-ADJUSTED SPEECH RECOGNITION

Johan Schalkwyk, assignor to ScanSoft, Incorporated
1 March 2005 (Class 704/270); filed 28 April 2003

At the beginning of each utterance, loading of the speech recognition processor is determined to be in one of four categories: relatively idle,



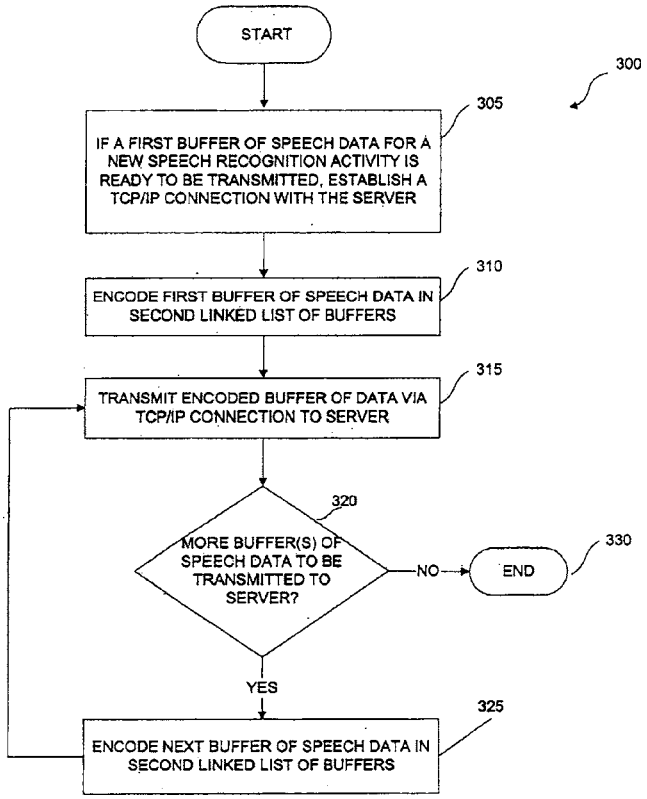
normal, busy, and pegged. A control mechanism alters the computational level limits of the processor in accordance with the loading category determined.—DAP

6,865,536

43.72.Ne METHOD AND SYSTEM FOR NETWORK-BASED SPEECH RECOGNITION

Christopher S. Jochumson, assignor to GlobalEnglish Corporation
8 March 2005 (Class 704/270.1); filed 19 July 2002

A system is described, said to be simple to install, that provides quasi-real-time speech recognition processing for anyone with a computer and Internet access. Audio speech is input by a user, compressed by the client into a compact data representation, and stored for later transmission to a



server. The server decodes the received audio packets from multiple clients and responds back to each client. Applications include interactive language learning, dictation, and voice control.—DAP

6,868,379

43.72.Ne SPEECH RECOGNITION DEVICE WITH TRANSFER MEANS

Heribert Wutte, assignor to Koninklijke Philips Electronics N.V.
15 March 2005 (Class 704/235); filed in the European Patent Office 8 July 1999

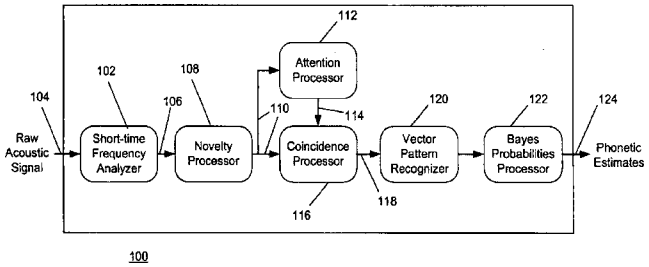
A speech recognition device and methodology are proposed in which a user can employ a speech coefficient indicator which has already been trained by the user on another speech recognition device. The speech coefficient indicator includes words that can be recognized, probabilities for sequences of words, and how phonemes are pronounced by a speaker.—DAP

6,868,380

43.72.Ne SPEECH RECOGNITION SYSTEM AND METHOD FOR GENERATING PHONOTIC ESTIMATES

John Kroeker, assignor to Eliza Corporation
15 March 2005 (Class 704/240); filed 23 March 2001

After a useful historical review, a complete speech recognition system is disclosed, together with computer code taken from the implemented software. From front end to back end, the system involves numerous inventions which appear to be unique, including a short-time frequency analyzer, pattern recognizer, and probability estimator of special design, which are fully disclosed in other patents by this inventor. Of chief importance here is the ''novelty processor,'' which attempts to separate the ''figure'' of speech from the ''ground'' of irrelevant signal variation by assuming that the speech itself



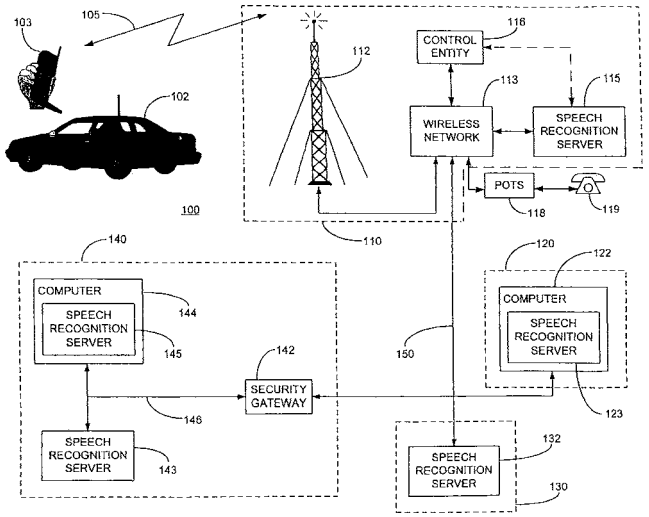
changes more rapidly than the confusing signal. The most relevant (novel) aspects of the signal are then fed to a ''coincidence processor'' whose output includes co-occurrences between samples of the novelty output. This last is then used by the specially designed back end disclosed elsewhere. The most significant achievement overall is the avoidance of standard data-reduction techniques such as mel-frequency cepstral vectors.—SAF

6,868,385

43.72.Ne METHOD AND APPARATUS FOR THE PROVISION OF INFORMATION SIGNALS BASED UPON SPEECH RECOGNITION

Ira A. Gerson, assignor to Yomobile, Incorporated
15 March 2005 (Class 704/275); filed 5 October 1999

A system is described to incorporate speech recognition into telematics systems for wireless communication applications in vehicles. Front end speech recognition processing, such as feature extraction, is performed in a



handheld client device, whereas back end processing is performed in a speech recognition server.—DAP

6,820,056

43.72.Ne RECOGNIZING NON-VERBAL SOUND COMMANDS IN AN INTERACTIVE COMPUTER CONTROLLED SPEECH WORD RECOGNITION DISPLAY SYSTEM

Shlomi Harif, assignor to International Business Machines Corporation
16 November 2004 (Class 704/275); filed 21 November 2000

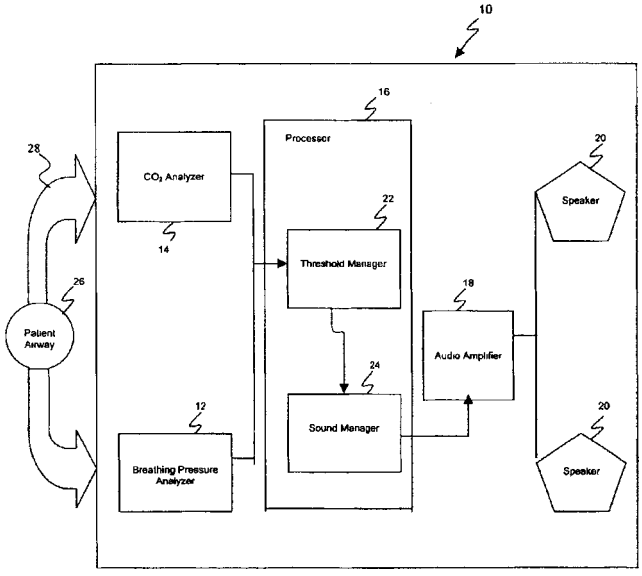
In the application of a speech recognition system to the editing of text on a computer, there is always an issue of being able to distinguish phrases intended to be placed into the edited text from phrases intended as editing control functions. The solution proposed here is to use nonverbal sounds, perhaps grunts and groans, to perform the editing commands. Some suggested command sounds include hand claps, tongue clicks, or metallic taps. There is no discussion of how the recognizer for this application might differ from any other speech recognizer, except that an unspecified test would initially determine whether an input sound is speech or not.—DLR

6,863,068

43.80.Qf VENTILATION SOUND DETECTION SYSTEM

David Thomas Jamison and Mark J. Maritch, assignors to Draeger Medical, Incorporated
8 March 2005 (Class 128/204.23); filed 25 July 2002

The purpose of this ventilation sound detection system is to monitor respiration of a patient whose breathing is assisted by a ventilator. The system includes pressure and CO_2 sensors and an audible display that emits one of two different sound patterns based on the status of the ventilator: one for inspiration and one for expiration. The inspiration sound is produced when monitored breathing pressure crosses a threshold level, thereby indicating to the anesthesiologist/clinician that sufficient pressure has been developed in the breathing circuit (i.e., the patient has made a proper inspiration breath). The exhalation sound indicates that the CO_2 level has risen a certain amount above a mean inspired CO_2 level (i.e., the patient has properly exhaled). The anesthesiologist/clinician can thus verify that the patient is being properly ventilated.—DRR

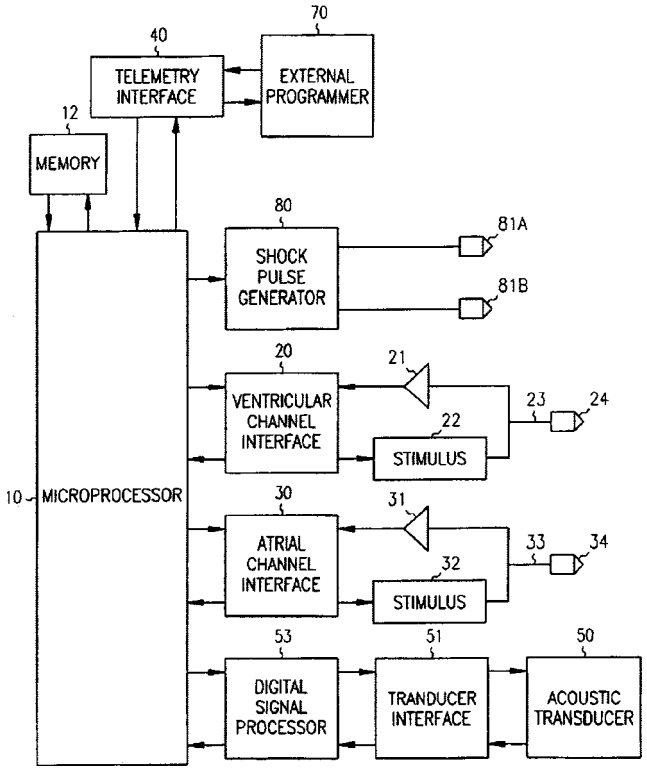


6,865,424

43.80.Qf IMPLANTABLE MEDICAL DEVICE WITH VOICE RESPONDING AND RECORDING CAPACITY

Douglas R. Daum *et al.*, assignors to Cardiac Pacemakers, Incorporated
8 March 2005 (Class 607/62); filed 8 August 2002

An implantable medical device, such as a cardiac pacemaker or cardioverter/defibrillator, is equipped with the capability for receiving communications in the form of speech spoken by the patient. An acoustic transducer is incorporated within the device, which, along with filtering circuitry, enables the voice communication to be used to affect the operation of the device or to be recorded for later playback.—DRR

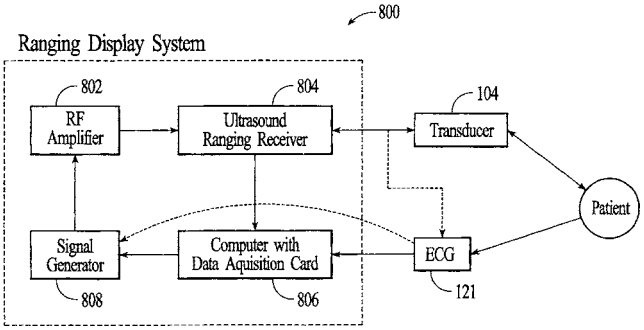


6,863,653

43.80.Qf ULTRASOUND DEVICE FOR AXIAL RANGING

Claudio I. Zanelli *et al.*, assignors to Eclipse Surgical Technologies, Incorporated
8 March 2005 (Class 600/437); filed 9 October 1998

This is a device that can be in the form of a catheter or similarly elongated medical device that includes an ultrasound transducer, making the device particularly suitable for determining the depth of dynamic tissue in beating heart laser-assisted transmyocardial revascularization. As the ultrasound transducer is activated, an acoustic wave is generated and a signal is reflected back to the transducer from targeted anatomical structures, thereby providing information on the position of the catheter or other elongated



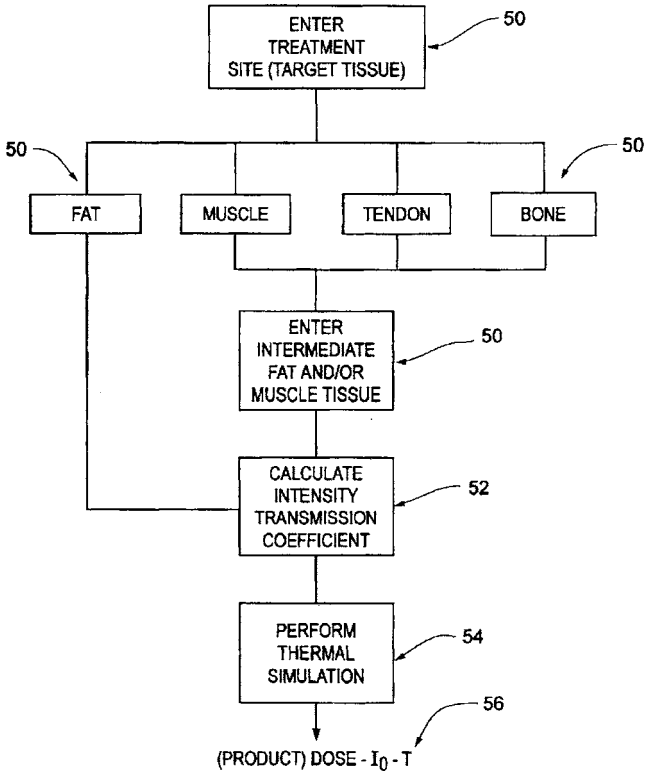
surgical apparatus in relation to the anatomical structure. Other applications are also covered by the 115 claims.—DRR

6,860,852

43.80.Sh ULTRASOUND THERAPEUTIC DEVICE

Klaus Schönenberger *et al.*, assignors to Compex Medical S.A.
1 March 2005 (Class 600/439); filed 25 October 2002

This is a therapeutic ultrasound device that is capable of automatically determining treatment dosage, controlling acoustic power efficiency, and performing adjustable calibration functions. The system consists of a generator unit, at least one transducer treatment head, and a programmable controller. The generator supplies acoustic power to the transducer treatment



head. Reprogrammable controller software controls all features and functions for the system. The controller calculates an initial optimal dose, maintains effective acoustic power transmitted to the patient through at least one transducer head, controls the output for each supported treatment mode, and provides for other features and functions.—DRR

6,866,631

43.80.Vj SYSTEM FOR PHASE INVERSION ULTRASONIC IMAGING

Glen McLaughlin and Ting-Lan Ji, assignors to Zonare Medical Systems, Incorporated
15 March 2005 (Class 600/437); filed 31 May 2001

Multiple sets of transmit pulses are used that differ in amplitude, frequency, phase, or shape. The sets of pulses are transmitted into a region of interest and the resulting echoes are combined to obtain an average signal that represents the net common-mode signal from each transmitted signal set. The combined signal set is used to form an ultrasound image.—RCW

6,866,632

43.80.Vj ADAPTIVE RECEIVE APERTURE FOR ULTRASOUND IMAGE RECONSTRUCTION

Ching-Hua Chou *et al.*, assignors to Zonare Medical Systems, Incorporated
15 March 2005 (Class 600/443); filed 18 September 2002

The size of a receive aperture is compared with the number of available parallel channels for beamformation. If the size of the receiver aperture is not greater than the number of channels, received echo signals are processed to produce an ultrasonic image. If the size of the received aperture is greater than the number of channels, received echo signals are processed to produce a number of signals equal to the number of channels.—RCW

6,866,633

43.80.Vj METHOD AND APPARATUS FOR ULTRASONIC IMAGING USING ACOUSTIC BEAMFORMING

Andrea Trucco, assignor to Esaote, S.p.A.
15 March 2005 (Class 600/443); filed in Italy 28 November 2002

Receiver beamformation is accomplished using delays that are determined as a function of the received signal frequency and as a function of receiving element position in the receiver array.—RCW

6,863,655

43.80.Vj ULTRASOUND DISPLAY OF TISSUE, TRACKING AND TAGGING

Steinar Bjaerum *et al.*, assignors to GE Medical Systems Global Technology Company, LLC
8 March 2005 (Class 600/442); filed 10 June 2002

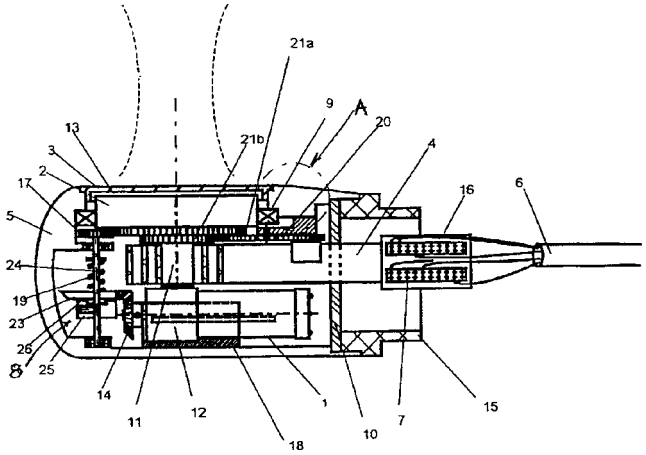
A pattern corresponding to a tracked moving structure, such as a cardiac wall, is produced by using tagging symbols related to structure movement determined from backscattered ultrasound. A Doppler processor produces a spatial set of values that represent motion within the structure. B-mode amplitudes within the structure are also obtained. The pattern is overlayed on an image to show structure motion such as expansion and contraction of the cardiac wall.—RCW

6,866,635

43.80.Vj TRANSDUCER POSITION LOCKING SYSTEM

Aimé Flesch and An Nguyen-Dinh, assignors to Vermon
15 March 2005 (Class 600/459); filed 3 September 2003

A phased array transducer within an endoscope housing is rotated around its acoustic axis by an immersed micromotorized drive. A torque-



limiting gear box couples the drive to the transducer. An automatic transducer position locking system is used to fix the transducer in a selected position.—RCW

6,860,853

43.80.Vj ULTRASONIC IMAGING APPARATUS

Hiroshi Hashimoto, assignor to GE Medical Systems Global Technology Company, LLC
1 March 2005 (Class 600/446); filed in Japan 26 April 2002

Three-dimensional image data are acquired using a hand-held ultrasound probe that also provides information about probe position. An image is formed based on the three-dimensional data.—RCW

6,860,854

43.80.Vj SYNTHETICALLY FOCUSED ULTRASONIC DIAGNOSTIC IMAGING SYSTEM FOR TISSUE AND FLOW IMAGING

Brent S. Robinson, assignor to Koninklijke Philips Electronics N.V.
1 March 2005 (Class 600/447); filed 3 October 2003

This system alternates between synthetic focus data acquisition and conventional focused beam data acquisition. Speckle in the synthetic focused ultrasound images is reduced by combining signals from subapertures that view the image field from different directions. Regions of interest within a synthetic focused image may be processed differently to emphasize various motion characteristics such as turbulent flow or differing flow velocities.—RCW

6,860,855

43.80.Vj SYSTEM AND METHOD FOR TISSUE BIOPSY USING ULTRASONIC IMAGING

Jerod O. Shelby *et al.*, assignors to Advanced Imaging Technologies, Incorporated
1 March 2005 (Class 600/459); filed 19 November 2001

Ultrasonic imaging is used to guide a biopsy needle. Three acoustically coupled chambers are employed in the imaging process. An ultrasound transducer is in one chamber, at least a portion of an ultrasound detector is in the second chamber, and the part of the patient's anatomy to be imaged is in the third chamber, located between the first and second chambers. The location of a lesion is determined in three dimensions and the lesion coordinates are used to guide the insertion of a biopsy needle. Real-time imaging is used to view both the lesion and the biopsy needle.—RCW

FORUM

Forum is intended for communications that raise acoustical concerns, express acoustical viewpoints, or stimulate acoustical research and applications without necessarily including new findings. Publication will occur on a selective basis when such communications have particular relevance, importance, or interest to the acoustical community or the Society. Submit such items to an appropriate associate editor or to the Editor-in-Chief, labeled FORUM. Condensation or other editorial changes may be requested of the author.

Opinions expressed are those of the individual authors and are not necessarily endorsed by the Acoustical Society of America.

"Transmission loss" and "propagation loss" in undersea acoustics

Michael A. Ainslie[a)]
TNO Defence, Security and Safety (Underwater Technology Department), Oude Waalsdorperweg 63, 2509 JG The Hague, The Netherlands
Christopher L. Morfey
Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom

(Received 14 February 2005; revised 22 April 2005; accepted 6 May 2005) [DOI: 10.1121/1.1960170]

In the context of sonar and ocean acoustics, the term "transmission loss" (TL) was originally defined in terms of the acoustic parameter $\overline{p^2}/Z$, where $\overline{p^2}$ is the mean square pressure (MSP) and Z is the characteristic acoustic impedance of the medium.[1–3] For a plane progressive wave $\overline{p^2}/Z$ is equal to the magnitude of the acoustic intensity, so this parameter is sometimes referred to as the "equivalent plane wave intensity" (EPWI).[2,4] Specifically TL was defined in Refs. 1–3 as the ratio of EPWI at a specified distance from the source (and under specified, idealized conditions) to its value at the receiver, expressed in decibels. Oceanic variations in Z are usually small, so it became customary to drop the Z denominator and use a MSP ratio.

The term "propagation loss" (PL) is today used synonymously with TL in the underwater context.[5,6] In the following, PL is used in preference to TL to avoid possible confusion with alternative definitions of the latter,[2,6] but the notation PL is intended to cover both propagation loss and transmission loss. More specifically the notation $\mathrm{PL_{EPWI}}$ is used here to mean "propagation loss defined in terms of a ratio of EPWI values," and similarly for $\mathrm{PL_{MSP}}$.

If the Z ratio between source and receiver locations is sufficiently close to unity, the difference between $\mathrm{PL_{MSP}}$ and $\mathrm{PL_{EPWI}}$ may be ignored. However, it has become common practice to drop the Z ratio *even when it is not close to unity*. In situations where $\mathrm{PL_{MSP}}$ and $\mathrm{PL_{EPWI}}$ are different, an ambiguity arises if it is not stated which of the two is intended. While the MSP and EPWI definitions are both defensible, the ambiguity is not. The best way to resolve this ambiguity would be for practitioners of underwater acoustics to adopt a single common definition.

We believe that some readers will be thinking "we do not need to agree on a standard definition because we already have one, namely $\mathrm{PL_{EPWI}}$"; we encourage these readers to read any one of the articles published since 1989 addressing the ASA penetrable wedge problem[7] first considered by Murphy and Chin-Bing,[8] or minor variants thereof. We are aware of 24 such articles, the most recent ones being Refs. 9 and 10. For the case involving a receiver at depth 30 m, there is a step change in impedance (a factor of 1.7) at a horizontal range of 3.4 km from the source. From the continuity of MSP it follows that $\mathrm{PL_{EPWI}}$ must change discontinuously at this point by 2.3 dB. No such step change is discernible in the relevant graphs from any of the 24 articles, so it seems reasonable to infer that $\mathrm{PL_{MSP}}$ is being used, especially as one of the articles (Ref. 11) states this definition explicitly [and the wording of another by the same authors (Ref. 12) implies it]. Publications involving geometries other than the benchmark wedge are too numerous to mention, but recent examples using $\mathrm{PL_{MSP}}$ across an impedance boundary are Refs. 13–15. Also relevant is the definition of the International Electrotechnical Commission (IEC),[5] which states explicitly that a MSP ratio should be used. Overall, the evidence against $\mathrm{PL_{EPWI}}$ as an *accepted* standard is overwhelming. Furthermore, $\mathrm{PL_{MSP}}$ has the advantage of resulting in a simpler sonar equation.[4]

Other readers are probably thinking "we do not need to agree on a standard definition because we already have one, namely $\mathrm{PL_{MSP}}$." The main case against $\mathrm{PL_{MSP}}$ comprises the explicit EPWI definitions of Urick (Ref. 16, p. 22), Jensen *et al.* (Ref. 17, pp. 11 and 12), and Hall,[18] and the implied one of Kuperman.[19] Three of these four definitions come from influential textbooks, and should not be taken lightly, but we know of no examples of the *application* of $\mathrm{PL_{EPWI}}$ later than the work of DiNapoli and Deavenport[20] in 1980, even in the textbooks that advocate this definition. Finally we mention the introduction of yet a third definition of PL in terms of a ratio of $\overline{p^2}/\rho$ values, where ρ is the local density.[21]

Perhaps a third set of readers will refer to the ANSI definition[22] and say "we already have a definition that is sufficiently flexible to accommodate both camps." They are correct, but the generality of the ANSI definition compromises its usefulness in the present context. For example, given the conflict between the IEC definition on the one hand (which requires a MSP ratio) and the statement that "The decibel … [in underwater acoustics] … denotes a ratio of intensities (not pressures)"[17,19] on the other, there is a need for clear advice on the circumstances in which each of the two definitions may be appropriate. In addition there is no mention of the need for evaluation of the field variable relative to that due to an equivalent point source in an equivalent homogeneous medium.[6] It would also be helpful to point out the synonymous use by underwater acousticians of the terms "propagation loss" and "transmission loss." In passing we note a typographical error in ANSI definition 6.33; the reference in note 3 should be to 11.43, rather than 11.42.

The question now is: what can be done to mitigate the confusion? An essential first step is for individual authors to follow note 2 of the ANSI definition by stating which characteristic of the signal (e.g., MSP or EPWI) is being used to form the propagation loss, at least in situations where ambiguity might otherwise arise. A desirable second step would be to agree on an updated standard definition, incorporating the above-mentioned refinements necessary for sonar, including a clarification of whether a MSP or EPWI ratio is preferred in this context. Such agreement is only possible after informed debate, and the main purpose of this letter is to stimulate such a debate. Which of the competing definitions is eventually chosen is less important than the act of choosing (and promulgating) one of them, although the arguments in favor of $\mathrm{PL_{MSP}}$, as described earlier and reflected in Ref. 6, seem to us the more persuasive.

[1]National Defense Research Committee (1946), "Physics of Sound in the Sea," Summary Technical Report of Division 6 NDRC, Vol. **8**, Washington DC. Reprinted in 1969 by Department of the Navy Headquarters Naval Material Command, Washington, DC 20360 (NAVMAT P-9675).
[2]J. W. Horton, *Fundamentals of SONAR*, 2nd ed. (United States Naval Institute, Annapolis, 1959).
[3]R. J. Urick, *Principles of Underwater Sound for Engineers* (McGraw-Hill, New York, 1967).
[4]M. A. Ainslie, "The sonar equation and the definitions of propagation

[a)]Electronic mail: michael.ainslie@tno.nl

loss," J. Acoust. Soc. Am. **115**, 131–134 (2004).
[5]International Electrotechnical Commission, *International Electrotechnical Vocabulary, Chapter 801: Acoustics and Electroacoustics*, 2nd ed. IEC 50(801) (IEC, Geneva, 1994).
[6]C. L. Morfey, *Dictionary of Acoustics* (Academic, San Diego, 2001).
[7]L. B. Felsen, "Benchmarks: An option for quality assessment," and seven accompanying articles by various authors, J. Acoust. Soc. Am. **87**, 1497–1545 (1990).
[8]J. E. Murphy and S. A. Chin-Bing, "A finite-element model for ocean acoustic propagation and scattering," J. Acoust. Soc. Am. **86**, 1478–1483 (1989).
[9]F. Sturm and J. A. Fawcett, "On the use of higher-order azimuthal schemes in 3-D PE modeling," J. Acoust. Soc. Am. **113**, 3134–3145 (2003).
[10]D. Mikhin, "Exact discrete nonlocal boundary conditions for high-order Padé parabolic equations," J. Acoust. Soc. Am. **116**, 2864–2875 (2004).
[11]D. Yevick and D. J. Thomson, "A hybrid split-step/finite-difference PE algorithm for variable-density media," J. Acoust. Soc. Am. **101**, 1328–1335 (1997).
[12]D. Yevick and D. J. Thomson, "Nonlocal boundary conditions for finite-difference parabolic equation solvers," J. Acoust. Soc. Am. **106**, 143–150 (1999).
[13]M. D. Collins and D. K. Dacol, "A mapping approach for handling sloping interfaces," J. Acoust. Soc. Am. **107**, 1937–1942 (2000).
[14]G. A. Athanassoulis and A. M. Prospathopoulos, "Three-dimensional acoustic scattering from a penetrable layered cylindrical obstacle in a horizontally stratified ocean waveguide," J. Acoust. Soc. Am. **107**, 2406–2417 (2000).
[15]M. A. Ainslie, A. J. Robins, and D. G. Simons, "Caustic envelopes and cusp co-ordinates due to the reflection of a spherical wave from a layered sediment," J. Acoust. Soc. Am. **115**, 1449–1459 (2004).
[16]R. J. Urick, *Principles of Underwater Sound*, 3rd ed. (Peninsula, Los Altos, 1983).
[17]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (American Institute of Physics, Woodbury, NY, 1994).
[18]M. V. Hall, "Dimensions and units of underwater acoustic parameters," J. Acoust. Soc. Am. **97**, 3887–3889 (1995); erratum: **100**, 673 (1996).
[19]W. A. Kuperman, "Propagation of sound in the ocean," in *Encyclopedia of Acoustics*, edited by M. J. Crocker (Wiley, New York, 1997), pp. 391–408.
[20]F. R. DiNapoli and R. L. Deavenport, "Theoretical and numerical Green's function field solution in a plane multilayered medium," J. Acoust. Soc. Am. **67**, 92–105 (1980).
[21]J. F. Lingevitch, M. D. Collins, and W. L. Siegmann, "Parabolic equations for gravity and acousto-gravity waves," J. Acoust. Soc. Am. **105**, 3049–3056 (1999).
[22]Acoustical Society of America, *American National Standard: Acoustical Terminology, Chapter 6: Transmission and Propagation*, definition 6.33, ANSI S1.1-1994, ASA 111-1994, (ASA, New York, 1994).

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.

Correcting the use of ensemble averages in the calculation of harmonics to noise ratios in voice signals (L)

Carlos A. Ferrer, Eduardo González, and María E. Hernández-Díaz
Center of Studies of Electronics and Information Technologies, Central University of Las Villas, C. Camajuaní, Km 5 1/2 Santa Clara, 54800 Cuba

(Received 19 November 2004; revised 2 May 2005; accepted 3 May 2005)

A correcting formula for the estimation of harmonics-to-noise ratios (HNR) based on ensemble-averaging techniques is derived. The original method yields a biased approximation which is more accurate as the number of averaged pulses (N) increases. However, the method treats gradual waveform changes incorrectly as noise, which is worsened for large values of N. The obtained formula allows the use of as few averaged pulses as desired, while allowing the complete removal of the bias from the estimate of HNR. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1940450]

PACS number(s): 43.70.Gr [AL] Pages: 605–607

I. INTRODUCTION

The determination of harmonics-to-noise ratios (HNR) in voiced speech signals is intended to yield a measure of additive noise in the acoustic waveform. For this purpose, several methods have been proposed, both in the time- and frequency (or transformed) domain. Methods in the time domain exploit similarities in successive pulses, while frequency-domain methods make use of the assumed harmonic structure. Several authors have given reasons in favor and against both approaches (see Murphy, 1999, for an extensive list of references), and there is no definitive satisfactory solution.

A much referenced time-domain method was proposed by Yumoto *et al.* (1982), who introduced the finding of a pitch epoch "template" by averaging the ensemble of individual pitch pulses. The averaged waveform is known to present a noise variance reduced by a factor of N, the number of averaged pulses. Yumoto proposed to measure HNR as the ratio of the variance of the template to the variance of the differences of the individual pulses with the template [see Eq. (4)].

It was previously criticized (Kasuya *et al.*, 1986) that a relatively large number of pulses (30–50) was needed to effectively remove noise from the template. This entails that the method treats the smooth waveform changes incorrectly as noise. Hillenbrand (1987) reported influences of jitter and shimmer in the resulting measures of HNR.

In spite of this, Yumoto's method remains "the most commonly used time-domain technique" (Murphy, 1999), and several authors have worked on its improvement. Cox *et al.* (1989) proposed a single-pass method that accounts, under certain considerations, for jitter, shimmer, and offset effects. Later, Qi (1992) used dynamic time warping (DTW) and zero-phase transforms (Qi *et al.*, 1995) of individual pulses prior to averaging to reduce waveform variability influences in the template. For the same purpose, Murphy (1999) applied the ensemble averaging technique to the spectral representations of individual glottal source pulses, although it was not recommended (p 2876) to be used with the radiated voice waveform. Dealing again with waveform variability, Lucero and Koenig (2000) used functional data analysis (FDA) to perform an optimal time alignment of pulses prior to averaging.

Almost every limitation of Yumoto's method has been addressed, and some means to overcome them have been proposed, except for the reduction of the number of pulses needed to remove the noise influence from the template. In this Letter, the possibility to completely remove the influence of the template's remaining noise in the calculation of HNR is demonstrated for any number of averaged pulses.

II. ANALYSIS OF ENSEMBLE-AVERAGING HNR CALCULATION

The model for ensemble averaging assumes each pulse representation $x_i(t)$ prior to averaging as a repetitive signal $s(t)$ plus a noise term $e_i(t)$

$$x_i(t) = s(t) + e_i(t). \tag{1}$$

Here, $x_i(t)$ can be any of the signals used in the previous references: pitch pulse waveform or its spectrum, DTW or zero-phased pulses, FDA transformed pulses, or any other

modification. The template $\bar{x}(t)$, obtained as the average of the N individual pulses, is

$$\bar{x}(t) = \frac{\sum_{i=1}^{N} x_i(t)}{N} = s(t) + \frac{\sum_{i=1}^{N} e_i(t)}{N}. \quad (2)$$

This ensemble averaging method is a common noise reduction technique in other physiological signals, as in electrocardiography (Rompelman and Ros, 1986). If $s(t)$ and $e_i(t)$ are zero mean signals with variances σ_s^2 and σ_e^2, with the noise term being stationary, white, Gaussian, ergodic, and uncorrelated with $s(t)$ and to any $e_j(t)$ such that $j\langle\ \rangle i$, it can be shown that the variance of the template is

$$\sigma_{\bar{x}}^2 = E[\bar{x}^2(t)] = E[s^2(t)] + \frac{\sum_{i=1}^{N} E[e^2(t)]}{N^2} = \sigma_s^2 + \frac{\sigma_e^2}{N}, \quad (3)$$

which means a reduction in noise variance by a factor of N. Yumoto *et al.* (1982) used this noise-reduced signal to obtain estimates of individual noise waveforms as the differences with the template, and proposed an HNR expression which can be written as

$$\mathrm{HNR}_{\mathrm{Yum}} = \frac{N \times E[\bar{x}^2(t)]}{\sum_{i=1}^{N} E\{[x_i(t) - \bar{x}(t)]^2\}}. \quad (4)$$

It can be predicted that $\mathrm{HNR}_{\mathrm{Yum}}$ will always overestimate the actual value of HNR. The numerator in (4) overrates the repetitive (harmonic) energy as seen in (3), while the denominator underrates each estimate of noise due to the presence of a fraction $(1/N)$ of the estimated noise in the template. This bias is reduced as the number of pulses increases, the reason for needing a relatively high number of averaged pulses in (4). However, it is possible to eliminate this dependency of the estimation of HNR with respect to the number of pulses. Substituting (1) and (2) in the denominator of (4) yields

$$\begin{aligned}
\mathrm{Den} &= \sum_{i=1}^{N} E\left[\left(e_i(t) - \sum_{j=1}^{N} \frac{e_j(t)}{N}\right)^2\right] \\
&= \sum_{i=1}^{N} E\left[\left(\left(e_i(t)\frac{(N-1)}{N}\right) - \sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{e_j(t)}{N}\right)^2\right] \\
\mathrm{Den} &= \sum_{i=1}^{N} E\left[e_i^2(t)\left(\frac{(N-1)}{N}\right)^2 - 2e_i(t)\frac{(N-1)}{N}\sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{e_j(t)}{N}\right. \\
&\quad \left. + \sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{e_j(t)}{N} \sum_{\substack{k=1 \\ k\neq i}}^{N} \frac{e_k(t)}{N}\right]. \quad (5)
\end{aligned}$$

Since $E[e_i(t)e_j(t)]=0$ for any $i\langle\ \rangle j$, the second term in (5) can be completely eliminated as well as all the products in the third term where $j\langle\ \rangle k$, resulting in

$$\begin{aligned}
\mathrm{Den} &= \sum_{i=1}^{N} \left(E\left[e_i^2(t)\frac{(N-1)^2}{N^2}\right] + E\left[\sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{e_j^2(t)}{N^2}\right]\right) \\
&= \sum_{i=1}^{N} \left(\sigma_{e_i}^2 \frac{(N-1)^2}{N^2} + \sigma_{e_j}^2 \frac{(N-1)}{N^2}\right) \\
\mathrm{Den} &= \sigma_e^2(N-1). \quad (6)
\end{aligned}$$

Substituting (3) and (6) in (4) yields an expression for Yumoto's formula in terms of the actual harmonic and noise variances, as well as the number of averaged pulses

$$\mathrm{HNR}_{\mathrm{Yum}} = \frac{N\sigma_s^2 + \sigma_e^2}{\sigma_e^2(N-1)} = \frac{N\sigma_s^2}{(N-1)\sigma_e^2} + \frac{1}{(N-1)}. \quad (7)$$

The actual HNR, defined as the ratio of the variances of the repetitive and noise signals, can be expressed according to (7) as

$$\mathrm{HNR} = \frac{\sigma_s^2}{\sigma_e^2} = \frac{N-1}{N}\mathrm{HNR}_{\mathrm{Yum}} - \frac{1}{N}, \quad (8)$$

which solves the problem of finding the precise value of HNR for any number of averaged pulses, departing from the one obtained using Yumoto's formula. The correction of $\mathrm{HNR}_{\mathrm{Yum}}$ in (8) is relevant mainly for small values of N, thus avoiding the criticized need of large values of N, where HNR is equivalent to $\mathrm{HNR}_{\mathrm{Yum}}$.

III. CONCLUSION

The derived formula provides the correction terms for the complete removal of a bias in the HNR estimates based on Yumoto's ensemble-averaging method. It allows the determination of the correct value of HNR for any number of averaged pulses. The correction is valid for any of the approaches described that use Yumoto's ensemble averaging formula, i.e., Yumoto *et al.* (1982), Qi (1992), Qi *et al.* (1995), Murphy (1999), and Lucero and Koenig (2000). The main contribution of this formula consists in the possibility to use as few pulses as desired to obtain the unbiased HNR estimate.

Cox, N. B., Ito, M. R., and Morrison, M. D. (**1989**). "Data labeling and sampling effects in harmonics-to-noise ratios," J. Acoust. Soc. Am. **85**, 2165–2178.

Hillenbrand, J. (**1987**). "A methodological study of perturbation and additive noise in synthetically generated voice signals," J. Speech Hear. Res. **30**, 448–461.

Kasuya, H., Ogawa, S., Kazuhiko, M., and Ebihara, S. (**1986**). "Normalized noise energy as an acoustic measure to evaluate pathologic voice," J. Acoust. Soc. Am. **80**, 1329–1334.

Lucero, J. C., and Koenig, L. L. (**2000**). "Time normalization of voice signals using functional data analysis," J. Acoust. Soc. Am. **108**, 1408–1420.

Murphy, P. J. (**1999**). "Perturbation-free measurement of the harmonics-to-noise ratio in voice signals using pitch synchronous harmonic analysis," J. Acoust. Soc. Am. **105**, 2866–2881.

Qi, Y. (**1992**). "Time normalization in voice analysis," J. Acoust. Soc. Am. **92**, 2569–2576.

Qi, Y., Weinberg, B., Bi, N., and Hess, W. J. (**1995**). "Minimizing the effect

of period determination on the computation of amplitude perturbation in voice," J. Acoust. Soc. Am. **97**, 2525–2532.
Rompelman, O., and Ros, H. H. (**1986**). "Coherent averaging technique: A tutorial review. I. Noise reduction and the equivalent filter," J. Biomed. Eng. **8**(1), 24–29.
Yumoto, E., Gould, W. J., and Baer, T. (**1982**). "The harmonic-to-noise ratio as an index of the degree of hoarseness," J. Acoust. Soc. Am. **71**, 1544–1550.

Supplementary notes on the Gaussian beam expansion (L)

Desheng Ding[a)]
Department of Electronic Engineering, Southeast University, Nanjing 210096, and State Key Laboratory of Modern Acoustics, Nanjing University, Nanjing 210093, People's Republic of China

Xiangjie Tong
Department of Electronic Engineering, Southeast University, Nanjing 210096, People's Republic of China

Peizhong He
Department of Biomedical Engineering, Shanghai Jiaotong University, Shanghai, 200030, People's Republic of China

(Received 16 September 2004; revised 3 May 2005; accepted 18 May 2005)

The letter provides alternatively a simple way of computing the Fresnel field integral, a further extension to the Gaussian-beam expansion. The zeroth-order Bessel function of the first kind is expanded into an approximate sum of Gaussian functions. The field integral is then expressible in terms of these simple functions. The approach is useful in treatment of the field radiation problem for a large and important group of piston sources in acoustics. As examples, the calculation results for the uniform and the simply-supported piston sources are presented, in a good agreement with those evaluated by numerical integration. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1953187]

PACS number(s): 43.20.Rz, 43.20.Bi, 43.20.El [MO] Pages: 608–611

I. INTRODUCTION

A powerful Gaussian beam (or generally, basis function) expansion technique has wide applications in computation of the Fresnel field integral for the sound beam of the ultrasonic transducer.[1–20] A remarkable advantage of it is to simplify the field integrals of this kind into the calculation of certain simple functions, such as the Gaussian, Gaussian-Laguerre, and Gaussian-Hermite functions. The purpose of this letter is to present a further extension to Wen and Breazeale's Gaussian expansion for ultrasonic fields.[3] We expand the zeroth-order Bessel function of the first kind into a sum of Gaussian functions. Correspondingly, the Fresnel field integral is then expressed in terms of the Gaussian functions. On comparison with Breazeale's original work, this approach provides a fairly satisfactory agreement with the results directly from numerical integration or their results. Its limits are also discussed.

II. THEORY AND RESULTS

In the Fresnel approximation, the field radiated by an ultrasonic transducer with a circularly symmetric distribution is described by the Fresnel field integral[2,5]

$$\bar{q}(\xi,\eta)=\int_0^\infty\left[\frac{2}{i\eta}\exp\left(i\frac{\xi^2+\xi'^2}{\eta}\right)J_0\left(\frac{2\xi\xi'}{\eta}\right)\right]\bar{q}(\xi')\xi'\,d\xi'$$
$$=\frac{2}{i\eta}\int_0^\infty\exp\left(i\frac{\xi^2+\xi'^2}{\eta}\right)J_0\left(\frac{2\xi\xi'}{\eta}\right)\bar{q}(\xi')\xi'\,d\xi' \quad (1)$$

with the dimensionless coordinates $\xi=r/a$ and $\eta=z/r_0$. Respectively, r and z are the radial and axial coordinates. Furthermore, the Rayleigh distance $r_0=ka^2/2$ is equal to π times the Fresnel distance $z_0=a^2/\lambda$, k is the wave number, λ is the wavelength, and a is a characteristic radius of the source. The propagating factor $\exp[-i(\omega t-kz)]$ is omitted.

In many previous papers, usually the source function $\bar{q}(\xi)$ is expanded into the superposition of the Gaussian-beam functions, namely,

$$\bar{q}(\xi)=\sum_{k=1}^{N}A_k\exp(-B_k\xi^2). \quad (2)$$

For a given source function, the coefficients A_k and B_k, the latter also referred to as the Gaussian coefficient, are obtained by computer optimization[3,13] or by the other numerical methods.[5,7,14] In what follows we shall present an alternative extension of this expansion approach.

We first consider a simple case of the field radiated from a uniform piston transducer. The source function is defined by

$$\bar{q}(\xi)=\mathrm{circ}(\xi)=\begin{cases}1, & 0\leqslant\xi<1,\\ 0, & \xi>1.\end{cases} \quad (3)$$

The field integral Eq. (1) is then

$$\bar{q}(\xi,\eta)=\frac{2}{i\eta}\int_0^1\exp\left(i\frac{\xi^2+\xi'^2}{\eta}\right)J_0\left(\frac{2\xi\xi'}{\eta}\right)\xi'\,d\xi'. \quad (4)$$

As is well known, there is no analytic expression of this integral, except at the acoustic axis and in the far-field (Fraunhofer) region.

We now note the fact that the one of the simplest integral formulas

a)Electronic mail: dds@seu.edu.cn

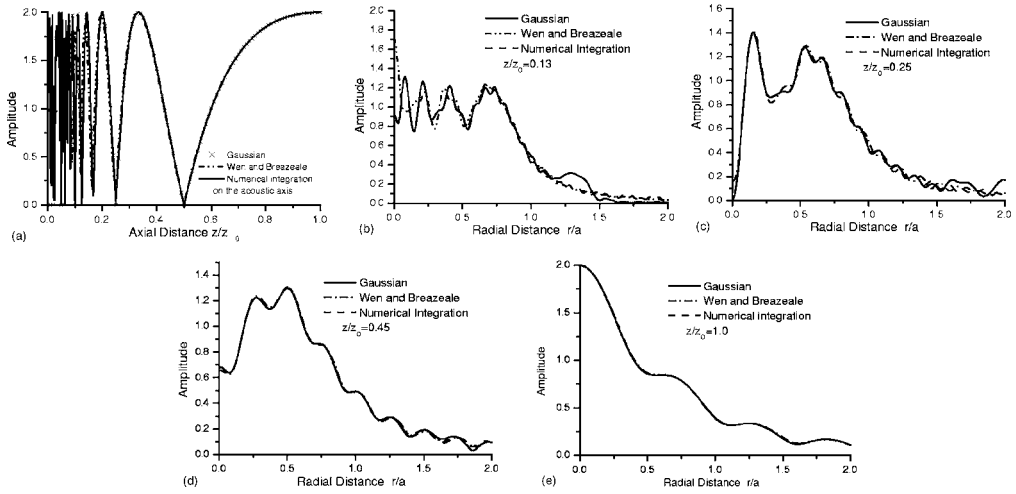


FIG. 1. Comparison of the field distributions of the uniform piston source, computed by use of the present Gaussian expansion and numerical evaluation of Fresnel field integral. The results of Ref. 3 are also graphed here (a) on the acoustic axis ($\xi=0$) and (b)–(e) at the different axial distances. Note that the axial distance is normalized to the Fresnel distance $z_0=a^2/\lambda$.

$$\int e^{ax}dx = e^{ax}/a \tag{5}$$

and that if the Bessel function, which appears in Eq. (4), is expressible in terms of Gaussian functions, then Eq. (4) is easily integrated out, also in terms of Gaussian functions, and the field calculation becomes very simple. Hence, a key problem we must solve is how to expand the Bessel function into a sum of the Gaussian functions. Of course, a very natural manner to determine the Gaussian expansion parameters of this function is use of the computer optimization approach that Wen and Breazeale have adopted successfully. But the shortest way is directly from their result for the radiation field of a uniform piston transducer. In order to calculate this field distribution, they decomposed the circ function into a series of Gaussian functions; mathematically, that is,[3]

$$\mathrm{circ}(x) = \sum_{k=1}^{N} A_k \exp(-B_k x^2). \tag{6}$$

They published two sets of the expansion coefficients of Eq. (6). One, consisting of ten pairs of A and B, is listed in in Table I of Ref. 3; the other, consisting of 15 pairs, is in Table I of Ref. 10. The precision and the usefulness of these data are fully demonstrated in many examples. At the present, we take Eq. (6) as *a known result, an approximation to the circ function*, and we directly give a Gaussian expansion of the Bessel function from this Eq. (6). With the help of the formulas presented in the Appendix, we get the relation

$$J_0(x) = \sum_{k=1}^{N} A_k \exp\left(-\frac{x^2}{4B_k}\right). \tag{7}$$

The coefficients A_k and B_k here and in the next section are always defined by Eq. (6).

Substitution of Eq. (7) in Eq. (4) and making use of the formula (5) yields

$$\begin{aligned} G_0(\xi,\eta;B) &= \frac{2}{i\eta}\int_0^1 \exp\left(i\frac{\xi^2+\xi'^2}{\eta}\right)\exp\left[-\frac{1}{4B}\left(\frac{2\xi\xi'}{\eta}\right)^2\right]\xi'\,d\xi' \\ &= -e^{i\xi^2/\eta}\left[\exp\left(\frac{i}{\eta}\left(1+\frac{i\xi^2}{B\eta}\right)\right)-1\right]\Big/\left(1+\frac{i\xi^2}{B\eta}\right). \end{aligned} \tag{8}$$

Accordingly the radiated field of a uniform piston transducer is expressed as

$$\bar{q}(\xi,\eta) = \sum_{k=1}^{N} A_k G_0(\xi,\eta;B_k). \tag{9}$$

It should be noted that Eq. (8) has not as an intuitive physical interpretation as that in the original work of Ref. 3, but it is useful in the treatment of some problems of diffraction.[21] In their paper, every expansion term there represents a Gaussian beam field.[3] The exception for $\xi=0$ or $B\to\infty$ may be explained as the on-axis description of the uniform piston field under the Fresnel approximation. For this reason, we simply call Eq. (8) of this kind the expansion function of the field.

As the first test of the approach, we give a comparison of the results for the field of the uniform circular piston transducer by using three different methods [the present, Breazeale's, and numerical integration of Eq. (4)]. The results are shown in Fig. 1 and agree well with each other. For the field distribution on the acoustic axis, the results display a surprisingly exact agreement from both the present method and direct integration of Eq. (4). It is not strange that if one remarks that on the axis of $\xi=0$, Eq. (9) is reduced to

$$\begin{aligned} \bar{q}(0,\eta) = \sum_{k=1}^{N} A_k G_0(0,\eta;B_k) &= [e^{i\xi^2/\eta}(1-e^{i/\eta})] \\ &\times \left(\sum_{k=1}^{N} A_k\right), \end{aligned} \tag{10}$$

the term in the square bracket is just the exact on-axis description for the uniform piston field. The real part of the

sum $\Sigma_{k=1}^{N} A_k$ is equal to 0.9989, by Table I of Ref. 10, extremely close to 1!

The above approach is applicable to not only the case of the uniform piston field, but also the case of the other source distributions, such as the piston sources simply supported or clamped at the edge.[22] More generally, the radiation field of a series of the piston sources with the distribution function $\bar{q}(\xi)=\Sigma_{k=0}^{n} a_k \xi^{2k}$, the polynomial of the even order, can be successfully treated in similar manners. The general integral formulas

$$\int x^m e^{ax}\, dx = \frac{x^m e^{ax}}{a} - \frac{m}{a}\int x^{m-1} e^{ax}\, dx, \tag{11a}$$

or

$$\int x^n e^{ax} dx = e^{ax}\left(\frac{x^n}{a} + \sum_{k=1}^{n} (-1)^k \times \frac{n(n-1)\cdots(n-k+1)}{a^{k+1}} x^{n-k}\right) \tag{11b}$$

are used in this treatment.

As our second example, we calculate the sound field of a simply supported piston. Its source function is given by

$$\bar{q}(\xi) = \begin{cases} 1-\xi^2, & 0 \leqslant \xi \leqslant 1, \\ 0, & \xi > 1, \end{cases} \tag{12}$$

and the expansion function $G'_1(\xi,\eta;B)$ of the field is divided into two parts, $G'_1 = G_0 - G_1$. G_0 is described by Eq. (8) and G_1 by

$$G_1(\xi,\eta;B) = \frac{2}{i\eta}\int_0^1 \exp\left(i\frac{\xi^2+\xi'^2}{\eta}\right) \times \exp\left[-\frac{1}{4B}\left(\frac{2\xi\xi'}{\eta}\right)^2\right]\xi'^3\, d\xi' = \frac{e^{i\xi^2/\eta}}{i\eta}\frac{e^a}{a} - \frac{1}{a}G_0 \tag{13}$$

with the abbreviation $a = i(1+i\xi^2/B\eta)/\eta$. Thus the field distribution is calculated according to the same relation as Eq. (9). The results are not graphed here. A good agreement is again obtained.

III. CONCLUDING REMARKS

In our treatment, there has been only one approximation, i.e., an approximate Gaussian expansion of $J_0(x)$. Apparently, the accuracy of field calculation depends on the degree of approximation of this function expansion. From Fig. 2 in the Appendix, the 15-term Gaussian expansion approximates very well $J_0(x)$ in the interval of about 0 to 30. Therefore, as the above results show, the validity domain of the present calculation is the field region up to $\xi/\eta \approx 15$.

A considerable advantage of our approach is that with only one set of 15-term Gaussian expansions, we can solve the radiation problem of a large and important group of the piston sources of interest in acoustics. It avoids numerous computation of the Gaussian expansion coefficients.[3,13] On the other hand, however, this approach does not have as wide a range of applications as Breazeale's original method, that is, in principle, applicable to the case of an arbitrary source distribution. Besides, the field expansion function is not universal for the case of the simply supported and the clamped-edge piston sources with the different order number n. This is seemingly a small shortcoming. For a small n, however, the expansion function may be analytically obtained by using Eq. (11b). For a large n, an easy and solvable means in practical computations is to use the recursive relation of Eq. (11a). In this program, what we need know is only G_0, the analytical expansion function of a uniform piston field.

A natural try is to extend the present Gaussian expansion to the nonlinear case of the second-order sound field. But such an extension has no obvious advantages over whatever we did based on Breazeale's work.[15–20] It seems to be superfluous.

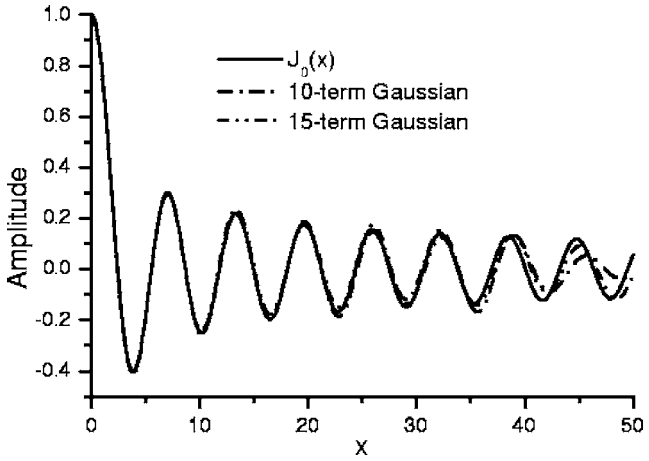


FIG. 2. Real part of the complex Gaussian expansion for the Bessel function.

ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China under Grand No. 10274010.

APPENDIX: GAUSSIAN EXPANSION OF THE BESSEL FUNCTION

Some necessary formulas are listed below:

$$\int x J_0(x)\, dx = x J_1(x), \tag{A1}$$

$$\int_0^\infty e^{-pt^2} J_0(bt) t\, dt = \frac{1}{2p}\exp\left(-\frac{b^2}{4p}\right), \quad \mathrm{Re}(p) > 0, \tag{A2}$$

$$J_0(x) = -\int J_1(x)\, dx. \tag{A3}$$

From Eq. (A1) it follows

$$\int_0^1 xJ_0(\xi x)\,dx = J_1(\xi)/\xi. \tag{A4}$$

The left-hand side of Eq. (A4) is indeed a Bessel-Fourier transform of the circ function. After substitution of Eq. (6) and use of (A2), one obtains

$$\int_0^\infty J_0(\xi x)\mathrm{circ}(x)x\,dx = \sum_{k=1}^{N} A_k \int_0^\infty \exp(-B_k x^2) J_0(\xi x) x\,dx = \sum_{k=1}^{N} \frac{A_k}{2B_k} \exp\left(-\frac{\xi^2}{4B_k}\right), \tag{A5}$$

that is,

$$J_1(\xi)/\xi = \sum_{k=1}^{N} \frac{A_k}{2B_k} \exp\left(-\frac{\xi^2}{4B_k}\right). \tag{A6}$$

Integration, with multiplying ξ on both sides of Eq. (A6), yields the required Eq. (7) in the text.

It is necessary to see first to what extent the Gaussian function expansion of Eq. (7) matches the Bessel function. Figure 2 shows the real part of Eq. (7), evaluated respectively by two sets of the coefficients from Refs. 3 and 10. In our calculation, we cite this set of 15 pairs of the coefficients with better accuracy. Due to the error, the imaginary part of the Gaussian expansion, although very close to zero, is not exactly zero. It is easy to cancel it through the following relation, i.e., by taking only the real part of the expansion

$$J_0(x) = \frac{1}{2}\left[\sum_{k=1}^{N} A_k \exp\left(-\frac{x^2}{4B_k}\right) + \sum_{k=1}^{N} A^*_k \exp\left(-\frac{x^2}{4B^*_k}\right)\right]. \tag{A7}$$

Here the asterisk designates complex conjugate. In the calculation procedure, we replace Eq. (7) of the text with Eq. (A7) here.

[1] B. D. Cook and W. J. Arnoult III, "Gaussian-Laguerre/Hermite formulation for the nearfield of an ultrasonic transducer," J. Acoust. Soc. Am. **59**, 9–11 (1976).

[2] E. Cavanagh and B. D. Cook, "Gaussian-Laguerre description of ultrasonic fields-numerical example: Circular piston," J. Acoust. Soc. Am. **67**, 1136–1140 (1980).

[3] J. J. Wen and M. A. Breazeale, "A diffraction beam field expressed as the superposition of Gaussian beams," J. Acoust. Soc. Am. **83**, 1752–1756 (1988).

[4] R. B. Thompson, T. A. Gray, J. H. Rose, V. G. Kogan, and E. F. Lopes, "The radiation of elliptical and bicylindrically focused piston transducers," J. Acoust. Soc. Am. **82**, 1818–1828 (1987).

[5] D. Ding, J. Lin, Y. Shui, G. Du, and D. Zhang, "An analytical description of ultrasonic field produced by circular piston transducer," Acta Acust. (Beijing) **18**, 249–255 (1993).

[6] D. Ding and Z. Lu, "A simplified method to calculate the sound field of pistonlike source," Chin. J. Acoust. **15**, 213–222 (1996), also appearing in Acta Acust. (Beijing) **21** (Suppl. 4), 421–428 (1996).

[7] M. D. Prange and R. G. Shenoy, "A fast Gaussian beam description of ultrasonic fields based on Prony's method," Ultrasonics **34**, 117–119 (1996).

[8] D. Ding and X. Liu, "Approximate description for Bessel, Bessel-Gauss and Gaussian beams with finite aperture," J. Opt. Soc. Am. A **16**, 1286–1293 (1999).

[9] M. Spies, "Transducer field modeling in anisotropic media by superposition of Gaussian base function," J. Acoust. Soc. Am. **105**, 633–38 (1999).

[10] D. Huang and M. A. Breazeale, "A Gaussian finite-element method for description of sound diffraction," J. Acoust. Soc. Am. **106**, 1771–1781 (1999).

[11] Y. Zhang, J. Liu, and D. Ding, "Sound field calculations of elliptical pistons by the superposition of two-dimensional Gaussian beams," Chin. Phys. Lett. **19**, 1825–1827 (2002).

[12] D. Ding, Y. Zhang, and J. Liu, "Some extensions of the Gaussian beam expansion: Radiation fields of the rectangular and the elliptical transducer," J. Acoust. Soc. Am. **113**, 3043–3048 (2003).

[13] K. Sha, J. Yang, and W. Gan, "A complex virtual source approach for calculating the diffraction beam field generated by a rectangular planar source," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 890–897 (2003).

[14] D. Ding and Y. Zhang, "Notes on the Gaussian beam expansion," J. Acoust. Soc. Am. **116**, 1401–1405 (2004).

[15] D. Ding, Y. Shui, J. Lin, and D. Zhang, "A simple calculation approach for the second harmonic sound field generated by an arbitrary axial-symmetric source," J. Acoust. Soc. Am. **100**, 727–733 (1996).

[16] D. Ding and Z. Lu, "A simplified calculation for the second-order fields generated by axial-symmetric sources at bifrequency," in *Nonlinear Acoustics in Perspective: Proceedings of the 14th International Symposium on Nonlinear Acoustics*, edited by R. Wei (Nanjing U. P., Nanjing, 1996), pp. 183–188.

[17] O. B. Matar, J. P. Rernenieras, C. Bruneel, A. Roncin, and F. Patat, "Ultrasonic sensing of vibrations," Ultrasonics **36**, 391–396 (1998).

[18] D. Ding, "A simplified algorithm for the second-order sound fields," J. Acoust. Soc. Am. **108**, 2759–2764 (2000).

[19] D. Ding and Y. Zhang, "A simple calculation approach for the second-harmonic sound beam generated by an arbitrary distribution source," Chin. Phys. Lett. **21**, 503–506 (2004).

[20] D. Ding, "A simplified algorithm for second-order sound beams with arbitrary source distribution and geometry (L)," J. Acoust. Soc. Am. **115**, 35–37 (2004).

[21] E. M. Drege, N. G. Skinner, and D. M. Byme, "Analytical far-field divergence angle of a truncated Gaussian beam," Appl. Opt. **39**, 4918–4925 (2000). Remarkably, the term in the square bracket of Eq. (1) in the text is just the beam field of an annular source (exactly speaking, that with infinitesimal width). From Huygens' principle, Eq. (1) is interpreted as the superposition of a series of annular sources with radial amplitude distribution $\bar{q}(\xi')$. Mathematically, this term represents a Green function of the annular source. Therefore, the Gaussian expansion of Bessel functions physically means that the radiation field of an annular source is decomposed into the linear superposition of a series of Gaussian beams. Accordingly, a relatively complex Green function is replaced by some simpler functions, so that the field integral may be analytically performed. The same explanation is true to the far-field result of this reference.

[22] D. L. Dekker, R. L. Piziali, and E. Dong, Jr., "Effect of boundary conditions on the ultrasonic-beam characteristics of circular disks," J. Acoust. Soc. Am. **56**, 87–93 (1974).

Application of the phase gradient method to the study of the resonances of a water-loaded anisotropic plate (L)

L. Guénégou[a] and O. Lenoir
Laboratoire d'Acoustique Ultrasonore et d'Electronique (LAUE), UMR CNRS 6068, Université du Havre, Place Robert Schuman, 76610, Le Havre, France

(Received 22 November 2004; revised 4 May 2005; accepted 5 May 2005)

The phase gradient method is applied to cubic and orthotropic plates. It consists in simply obtaining the positions and the widths of their frequency and angular resonances from the plots of the frequency and angular derivatives of the phase of the reflection coefficient of the immersed plate. There is a good match with the results obtained from the calculation of the modes of the immersed plate in the corresponding complex planes. Moreover, these two derivatives allow us to obtain frequency and angular quality factors. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1942367]

PACS number(s): 43.20.Ks, 43.20.Fn [YHB] Pages: 612–615

I. INTRODUCTION

The study of the propagation of transient and leaky Lamb type waves in isotropic or anisotropic targets has been the subject of numerous papers.[1–6] In the acoustic field, the link between frequency or angular resonances, on the one hand, and transient or leaky waves, on the other hand, was established by Überall, Gaunaurd *et al.*,[7,8] as a result of the resonant scattering theory (RST) applied to cylinders and plates. Some of these studies were devoted to the comparison of reflection coefficient minima with dispersion curves of transient or leaky modes of immersed plates[2,4,5] which correspond to the frequency or angular complex poles of the reflection coefficient. These works lead one to conclude that resonance and wave propagation are two aspects of the same phenomenon. The phase gradient method (PGM) recalled in this letter is an improvement of the RST in order to simply relate modes and resonances: the real parts of the modes correspond to the locations of the resonances and the imaginary parts to the resonance widths. The PGM deals with the study of the partial derivatives of the phase of the reflection coefficient with respect to the frequency, to the incidence angle, and to the parameters of both the plate and the surrounding fluid (densities, phase velocities, or stiffness components). In the isotropic case, the study of the frequency and angular phase derivatives was proved efficient to obtain simply and accurately both the location and the width of the frequency and angular resonances of the plate.[9] Knowing the location and the width of a resonance, its radiation quality factor (frequency or angular one) can be defined.[10,11] The only other way to obtain the modes is to perform calculations of roots in the complex frequency or angular planes. It is known that these calculations are difficult to perform, mainly if no initial guesses are known. Even the most accurate algorithm to find complex poles may be unable to converge to an expected one, or to separate poles whose real parts are very close and imaginary parts are clearly distinct, for instance. On the contrary, the PGM gives all kinds of poles with a very good precision. Indeed, the plots of the derivatives versus frequency or incidence angle exhibit maximums located at the real parts of the frequency and angular roots. The maximum amplitudes are inversely proportional to the imaginary parts.

However, this root finding aspect is not the main interest of the method. Indeed, it also enables an easy separation of the resonances associated with symmetric waves (*S* waves) from those associated with antisymmetric waves (*A* waves). The PGM clearly shows the difference between the reflection coefficient minima and Lamb modes and makes easier the study of both the transient guided waves and the leaky guided waves propagating in plates or cylinders. Moreover, by using the radiation frequency and angular quality factors obtained by the PGM, it is possible to obtain the energy velocity of a given mode.[14] In the isotropic case, we have also shown that the PGM is efficient to determine the prevailing polarization of a Lamb wave by studying the partial phase derivatives with respect to the phase velocities. Finally, all the partial derivatives of the phase are linked by a relation which can be read as an energy balance.[11] In conclusion, this method is not only an efficient numerical root finding method, but, more essentially, leads to a better understanding of the physical link between resonances and modes.

Due to anisotropy, the study of the phase and its derivatives becomes more complicated. In this letter, not all the points that have been developed for an isotropic plate[9,11,14] are examined. Attention is focused on the root finding aspect of the method and only results about the frequency and angular phase derivatives are presented. The aim of this letter is to show that the method still works simply in the anisotropic case, even though many more mechanical parameters are involved than in the isotropic case. Its physical aspects will be developed later in a longer paper.

The phase gradient method is thus applied to anisotropic plates of cubic and orthotropic symmetry. Due to the higher number of independent stiffness components, the difference between such materials and isotropic ones is the possible generation of quasi-shear horizontal (SH) waves in addition to the quasi-longitudinal (L) and quasi-shear vertical (SV)

[a]Electronic mail: lionel.guenegou@univ-lehavre.fr

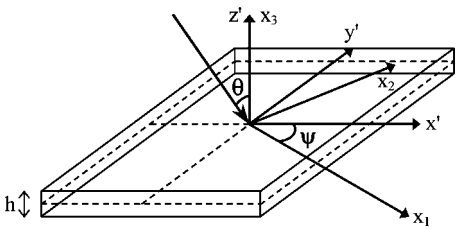


FIG. 1. Geometry of the problem.

waves observed in each case. The phase gradient method can be applied to such water-loaded anisotropic plates because the expression of their reflection coefficient was shown to be formally identical to that of an isotropic plate.[12,13]

Section II recalls the factorized expression of the reflection coefficient of an anisotropic plate. The bases of the phase gradient method are explained in Sec. III. Numerical results for both a cubic (GaAs) plate and an orthotropic (carbon-epoxy) one are given and discussed in Sec. IV.

II. FACTORIZED EXPRESSION OF THE REFLECTION COEFFICIENT OF AN ANISOTROPIC PLATE

The factorized expression of the reflection coefficient is needed in the phase gradient method to separate the resonances associated with symmetric modes from those associated with antisymmetric modes. Consider a water-loaded plate of thickness h. The parameters for water are: density $\rho_F=1000\ \mathrm{kg/m^3}$ and sound speed $c_F=1470\ \mathrm{m/s}$. The geometry of the problem is given in Fig. 1. The coordinate system (x',y',z') is the crystallographical one. The x'-y', x'-z', and y'-z' planes are symmetry planes. The z' axis is normal to the plate-water boundaries. In the coordinate system (x',y',z'), the stiffness tensor is denoted as $c'_{\alpha\beta}$ in abbreviated subscript notation. In a transformed coordinate system $(x_1,x_2,x_3=z')$ obtained by a counterclockwise rotation through an angle ψ about the z' axis the stiffness tensor is $c_{\alpha\beta}$.

The quantity of interest is the reflection coefficient of an incident monochromatic plane wave of time dependence $e^{-j\omega t}$, propagating in the x_1–x_3 plane, which insonifies the upper plate surface at $x_3=h/2$. To obtain the reflection coefficient R, the continuity of the normal displacement and stress, and the nullity of the tangential stresses are derived at the interfaces $x_3=\pm h/2$. After lengthy simplifications and manipulations of these boundary equations, the following expression for the reflection coefficient is obtained:

$$R=\frac{C_AC_S-\tau^2}{(C_A+j\tau)(C_S-j\tau)}=\frac{N_R}{AS}. \tag{1}$$

It is formally identical to the one used by Fiorito *et al.* in their application of the (RST) to isotropic elastic plates.[7] The expression of R is also similar to the one obtained by Nayfeh.[12] The roots of the $C_{A,S}$ functions correspond to the antisymmetric and symmetric modes of the plate in vacuum. Those of functions A and S correspond to the antisymmetric and symmetric modes of the plate immersed in water.

III. THE PHASE GRADIENT METHOD

The PGM consists in studying the partial derivatives of the phase of the reflection coefficient R. In this letter, only results derived from the frequency and angular derivatives are presented. Numerator N_R in Eq. (1) is real, so the global phase of the reflection coefficient R can be written as

$$\varphi=\arctan(\tau/C_S)-\arctan(\tau/C_A)=\varphi_S+\varphi_A. \tag{2}$$

The phase terms due to symmetric and antisymmetric modes can be separated easily. From Eq. (1), the analytical expressions of the differentials of the phases $\varphi_{A,S}$ are given by

$$\partial\varphi_{A,S}=(C_{A,S}\partial\tau-\tau\partial C_{A,S})/(C_{A,S}^2+\tau^2). \tag{3}$$

In the following, the plots of interest are those of the $\pm z\partial\varphi_{A,S}/\partial z$ functions versus z (z is either a frequency variable, generally the frequency-thickness product fh in MHz mm, or an angular variable, classically sine y of the incident angle). A frequency resonance related to an A or S mode is denoted as $\underline{fh}_P=fh_{\mathrm{Res}}-j\Gamma_{A,S}/2$ and corresponds to a frequency pole of R. An angular resonance is denoted as $\underline{y}_P=y_{\mathrm{Res}}+j\gamma_{A,S}/2$ [associated with the complex angle $\underline{\theta}_P=\arcsin(\underline{y}_P)$] and corresponds to an angular pole of R. In the vicinity of an isolated resonance, the $fh\partial\varphi_{A,S}/\partial fh$ and $y\partial\varphi_{A,S}/\partial y$ functions have the following approximate Breit–Wigner resonant type expressions:

$$(fh\partial\varphi_{A,S}/\partial fh)_{\mathrm{app}}=fh_{\mathrm{Res}}\Gamma_{A,S}/2/((fh-fh_{\mathrm{Res}})^2+(\Gamma_{A,S}/2)^2), \tag{4}$$

$$(-y\partial\varphi_{A,S}/\partial y)_{\mathrm{app}}=y_{\mathrm{Res}}\gamma_{A,S}/2/((y-y_{\mathrm{Res}})^2+(\gamma_{A,S}/2)^2). \tag{5}$$

Their maximums are obtained when $fh=fh_{\mathrm{Res}}$ or $y=y_{\mathrm{Res}}$ and are given by

$$(fh\partial\varphi_{A,S}/\partial fh)_{\mathrm{app}}=2fh_{\mathrm{Res}}/\Gamma_{A,S}=\mathrm{Re}(\underline{fh}_P)/\mathrm{Im}(\underline{fh}_P)=2Q_x, \tag{6}$$

$$(-y\partial\varphi_{A,S}/\partial y)_{\mathrm{app}}=2y_{\mathrm{Res}}/\gamma_{A,S}=\mathrm{Re}(\underline{y}_P)/\mathrm{Im}(\underline{y}_P)=2Q_y. \tag{7}$$

Equations (6) and (7) define the frequency and angular quality factors Q_x and Q_y.

For an isolated resonance, Eqs. (4) and (5) provide very good approximate expressions of the exact $\pm z\partial\varphi_{A,S}/\partial z$ functions. Hence, the plots of the exact functions give very good estimates of the quality factors, with no need of calculations of the poles.

The exact expressions of the frequency and angular phase derivatives are too long to be given in this letter. In the following, the results are obtained from the plots of these exact functions.

IV. NUMERICAL RESULTS

A. Study of a cubic plate

The cubic material considered is gallium arsenide (GaAs) whose density is $\rho=5307\ \mathrm{kg/m^3}$. The values of the components of the stiffness tensor are $c'_{11}=118.8$ GPa, c'_{12}

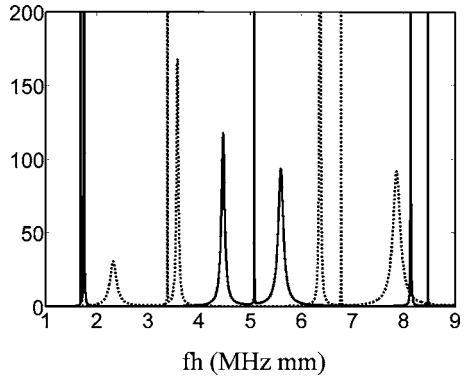


FIG. 2. GaAs cubic plate: plots of the exact $fh\partial\varphi_{A,S}/\partial fh$ functions vs fh (solid line: A modes, dotted line: S modes). $\psi=30°$, $\theta=5°$.

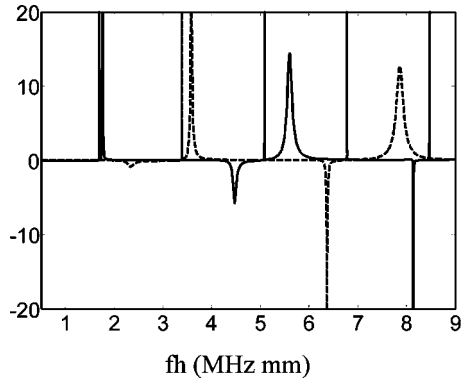


FIG. 4. GaAs cubic plate: plots of the exact $-y\partial\varphi_{A,S}/\partial y$ functions vs fh (solid line: A modes, dotted line: S modes). $\psi=30°$, $\theta=5°$.

$=53.8$ GPa, and $c'_{44}=59.4$ GPa. In the following, only plots for $\psi=30°$ are presented at the incidence angle $\theta=5°$.

First, results for the frequency phase derivative functions are presented and the plots of the exact $fh\partial\varphi_{A,S}/\partial fh$ functions are compared to those of the approximate ones. For $\psi\neq 0°$, quasi-SH waves are detected in addition to quasi-L and quasi-SV waves. Figure 2 exhibits the plots of the frequency phase derivatives in the range 1–9 MHz mm. The thinnest peaks are associated with modes which are mainly quasi-SH, while the other ones are associated with modes which are mainly either quasi-L or quasi-SV. In Fig. 3, the plots of the exact and approximate $fh\partial\varphi_A/\partial fh$ functions are compared in the vicinity of isolated resonances due to antisymmetric modes, either quasi-L ($\underline{f_A}h=4.47-j0.04$) in Fig. 3(a) or quasi-SH ($\underline{f_A}h=5.085-j1.25\times10^{-4}$) in Fig. 3(b). The curves match very well in both cases. Each maximum is located at the real part of the corresponding complex root $\underline{f_A}h$ and its amplitude is nearly equal to the ratio of the real part to the imaginary part: the error on the value of the frequency quality factor Q_x obtained from the exact plot is less than 0.25%. The amplitude of the peak corresponding to the SH mode is very high: it is due to the very small imaginary part of the assigned root, contrary to the other one. Now, the angular phase derivatives are studied. Figure 4 presents the frequency evolution of the exact phase derivatives $-y\partial\varphi_{A,S}/\partial y$. These functions can take either positive or negative values. Each negative peak is associated with a mode whose group velocity is negative.[5] Let us focus on the negative peak at $fh=4.47$ associated with the root $\underline{y_P}=0.087-j0.015$ ($\underline{\theta_P}=(5.01-j0.86)°$) and on the positive one at $fh=5.60$ associated with the root $\underline{y_P}=0.087+j6.02\times10^{-3}$ ($\underline{\theta_P}=(4.99+j0.35)°$). In Fig. 5, the angular evolution of the exact and approximate $-y\partial\varphi_A/\partial y$ functions are compared at those frequencies. The curves exhibit Breit–Wigner shapes, whose maximum (or minimum) is located around 5°. The amplitudes are nearly equal to the ratio of the real part to the imaginary part of $\underline{y_P}$: the values of the angular quality factors Q_y are obtained from the exact plot with less than 2% error.

Whatever the frequency or angular resonance considered, the frequency and angular quality factors are obtained with less than 2% error. This is due to the fact that all the resonances are well separated at small incidence angles for the cubic plate considered.

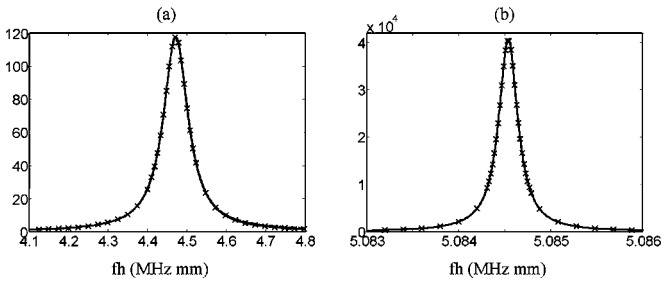


FIG. 3. (a) GaAs cubic plate: plots of the exact (solid line) and approximate (cross line) $fh\partial\varphi_A/\partial fh$ functions in the vicinity of a quasi-L antisymmetric mode. $\psi=30°$, $\theta=5°$; (b) GaAs cubic plate: plots of the exact (solid line) and approximate (cross line) $fh\partial\varphi_A/\partial fh$ functions in the vicinity of a quasi-SH antisymmetric mode. $\psi=30°$, $\theta=5°$.

B. Study of an orthotropic plate

The resonances of a carbon-epoxy plate are studied in this section. The density material is $\rho=1560$ kg/m^3. The values of the components of the stiffness tensor are: $c'_{11}=143.8$ GPa, $c'_{22}=c'_{33}=13.3$ GPa, $c'_{44}=3.6$ GPa, $c'_{55}=c'_{66}=5.7$ GPa, $c'_{12}=c'_{13}=6.2$ GPa, and $c'_{23}=6.5$ GPa.

Contrary to the previous case, all the resonances are not well separated at small incidence angles, as shown in Fig. 6. This figure shows the frequency evolution of the exact $fh\partial\varphi_{A,S}/\partial fh$ functions (solid line: A modes, dotted line: S modes). The unfavorable case of two close resonances asso-

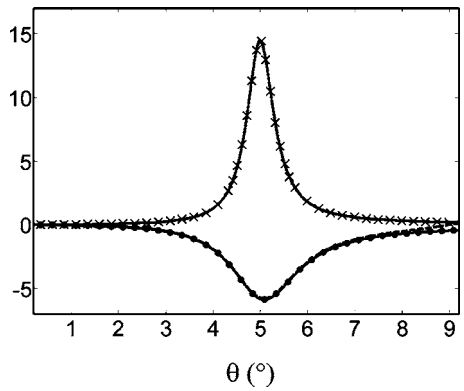


FIG. 5. GaAs cubic plate: plots of the exact (solid line) and approximate $-y\partial\varphi_A/\partial y$ functions vs incidence angle θ. $\psi=30°$. $fh=4.47$ for the negative peak (point line) and $fh=5.6$ for the positive one (cross line).

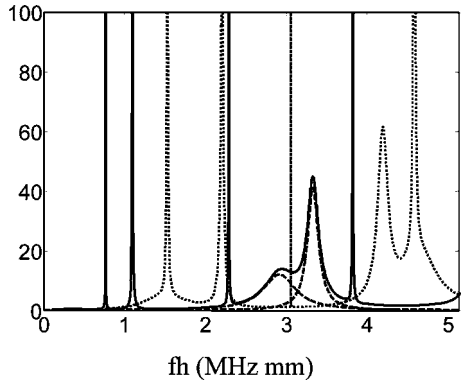


FIG. 6. Carbon-epoxy orthotropic plate: plots of the exact $fh\partial\varphi_{A,S}/\partial fh$ functions vs fh (solid line: A modes, dotted line: S modes). Plots of the approximate $fh\partial\varphi_A/\partial fh$ functions vs fh (dashed lines). $\psi=30°$, $\theta=5°$.

ciated with A modes ($\underline{fh}_P=2.92-j0.24$, $\underline{fh}_P=3.33-j0.08$) is treated. For these modes, the plot of the approximate $fh\partial\varphi_A/\partial fh$ functions are superimposed (dashed lines) to the exact plot (solid line) in Fig. 6. The approximate curves present maximums located at the real parts of the roots, but the frequency quality factors Q_x are obtained with 10% error. It is due to the overlapping of the resonances, but this lack of precision can be improved by using a nonlinear fitting algorithm. Now, the $-y\partial\varphi_S/\partial y$ function is studied for $fh=4.2$. Two relatively close symmetric modes are considered; they are related to the following roots: $\underline{y}_P=0.159+j8.91\times10^{-3}$ ($\underline{\theta}_P=(9.15+j0.52)°$) and $\underline{y}_P=0.176+j1.08\times10^{-3}$ ($\underline{\theta}_P=(10.13+j0.06)°$). In Fig. 7, the angular evolutions of the exact and approximate $-y\partial\varphi_S/\partial y$ functions are plotted. The curves still exhibit Breit–Wigner shapes. The maximums are located at the real parts of the roots and the maximum amplitudes are nearly equal to the angular quality factors Q_y. The errors are less than 3%.

In the orthotropic case considered here, even if it may happen that the accuracy on the values of the quality factors is not as good as in the cubic case of the previous section, in case of resonance overlapping phenomena, quite good estimates are still found by means of the phase gradient method.

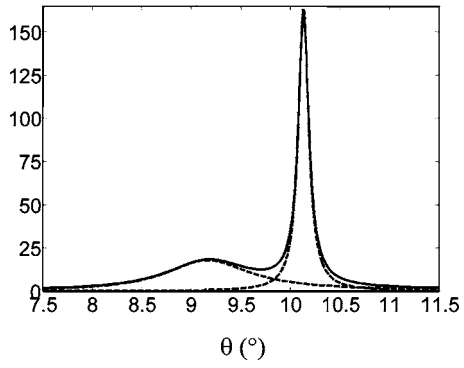


FIG. 7. Carbon-epoxy orthotropic plate: plots of the exact (solid line) and approximate (dashed lines) $-y\partial\varphi_S/\partial y$ functions vs incidence angle θ. $\psi=30°$, $fh=4.2$.

V. CONCLUSION

Just as well as for isotropic plates, the phase gradient method is an efficient tool to characterize, via their quality factors, the frequency and angular resonances of anisotropic plates. The interest of simply obtaining those quality factors by the PGM is to estimate the component of the energy velocity parallel to the sagittal plane as it was shown in the isotropic case.[14] The study of the phase derivative with respect to the rotation angle ψ leads to the determination of the component of the energy velocity normal to the sagittal plane. Now that the applicability of the PGM has been verified, these physical aspects of the method will be the subject of a further paper.

[1]S. I. Rokhlin, D. E. Chimenti, and A. H. Nayfeh, "On the topology of the complex wave spectrum in a fluid-coupled elastic plate," J. Acoust. Soc. Am. **85**, 1074–1080 (1989).
[2]D. E. Chimenti and S. I. Rokhlin, "Relationship between leaky Lamb modes and reflection coefficient zeroes for a fluid-coupled elastic layer," J. Acoust. Soc. Am. **88**, 1603–1611 (1990).
[3]M. Deschamps and O. Poncelet, "Transient Lamb waves: Comparison between theory and experiment," J. Acoust. Soc. Am. **107**, 3120–3129 (2000).
[4]M. Lowe and P. Cawley, "Comparison of reflection coefficient minima with dispersion curves for ultrasonic waves in embedded layers," in *Review of Progress in Quantitative NDE* (Plenum, New York, 1994).
[5]A. Bernard, M. Deschamps, and M. J. S. Lowe, "Comparison between the dispersion curves calculated in complex frequency and the minima of the reflection coefficient for an embedded layer," J. Acoust. Soc. Am. **107**, 793–800 (2000).
[6]O. Poncelet and M. Deschamps, "Lamb waves generated by complex harmonic inhomogeneous plane waves," J. Acoust. Soc. Am. **102**, 292–300 (1997).
[7]R. Fiorito, W. Madigosky, and H. Überall, "Resonance theory of acoustic waves interacting with an elastic plate, J. Acoust. Soc. Am. **66**, 1857–1866 (1979).
[8]L. Flax, G. Gaunaurd and H. Überall, "Theory of resonance scattering," *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1981), pp. 15, 191–294.
[9]O. Lenoir, J. Duclos, J. M. Conoir, and J. L. Izbicki, "Study of Lamb waves based upon the frequential and angular derivatives of the phase of the reflection coefficient," J. Acoust. Soc. Am. **94**, 330–343 (1993).
[10]B. A. Auld, "Acoustic fields and waves in solids," in *Acoustic Resonators, Part E* (Krieger, Malbar, FL, 1990), Vol. **II**, Chap. 11, pp. 260–268.
[11]O. Lenoir, J. M. Conoir, and J. L. Izbicki, "The radiation Q factors obtained from the partial derivatives of the phase of the reflection coefficient of an elastic plate.," J. Acoust. Soc. Am. **114**, 651–665 (2003).
[12]A. H. Nayfeh and D. E. Chimenti, "Propagation of guided waves in fluid-coupled plates of fiber-reinforced composite.," J. Acoust. Soc. Am. **83**, 1736–1743 (1988).
[13]O. Lenoir, L. Guénégou, and J. L. Izbicki, "Factorized expression of the reflection coefficient of a monoclinic plate immersed in water," Seventh European Conference on Underwater Acoustics ECUA 2004, Delft, The Netherlands, 5–8 July 2004, Proceedings of the Seventh European Conference on Underwater Acoustics ECUA 2004, TNO, TU Delft, edited by D. G. Simons, Vol. **1**, pp. 117–122.
[14]O. Lenoir, J. M. Conoir, and J. L. Izbicki, "Determination of the energy velocity of a Lamb wave by means of frequency and angular quality factors," on the CD-ROM Official Publication of the Forum Acusticum Sevilla 2002, 16–20 September 2002 (ISBN 84-87985-07-6).

Time reversal processing for source location in an urban environment (L)[a)]

Donald G. Albert[b)]
US Army Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory, 72 Lyme Road, Hanover, New Hampshire 03755-1290

Lanbo Liu
US Army Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory, 72 Lyme Road, Hanover, New Hampshire 03755-1290 and Department of Civil and Environmental Engineering, University of Connecticut, 261 Glenbrook Road, Storrs, Connecticut 06269-2037

Mark L. Moran
US Army Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory, 72 Lyme Road, Hanover, New Hampshire 03755-1290

(Received 17 June 2004; revised 8 April 2005; accepted 8 April 2005)

A simulation study is conducted to demonstrate in principle that time reversal processing can be used to locate sound sources in an outdoor urban area with many buildings. Acoustic pulse propagation in this environment is simulated using a two-dimensional finite difference time domain (FDTD) computation. Using the simulated time traces from only a few sensors and back propagating them with the FDTD model, the sound energy refocuses in the vicinity of the true source location. This time reversal numerical experiment confirms that using information acquired only at non-line-of-sight locations is sufficient to obtain accurate source locations in a complex urban terrain. [DOI: 10.1121/1.1925849]

PACS number(s): 43.28.En, 43.28.Js, 43.20.El, 43.50.Vt [DKW] Pages: 616–619

I. INTRODUCTION

An urban environment introduces many effects on acoustic propagation that are not present in the more idealized medium—a flat open area with finite impedance ground—that has often been used in studies of outdoor sound propagation. One of the major changes is the effect of buildings that act as obstacles to acoustic wave propagation and introduce multiple propagation paths, reflections, diffractions, and scattering into the propagation.

Acoustic sensors can be used to locate noise sources,[1–4] and because acoustic waves can diffract around obstacles, these sensors can potentially locate sources even when they are not directly in view of the source. (Sensors in this non-line-of-sight situation are called NLOS sensors, and those in view of the source LOS sensors.) These NLOS situations may be caused by ground topography[5] (hills) or by other obstacles like buildings.

Traditional methods of sound source location, including various beamforming or triangulation algorithms, rely on an accurate determination of the azimuthal direction of arrival of the acoustic waves. These methods may work in a NLOS situation if the acoustic wave arrival direction is unaffected by the obstacles, for example when a wave propagates over a broad hill. However, the methods will give erroneous results if the azimuthal direction of the wave arrival is different from the source direction, a situation that often occurs in both LOS and NLOS situations in an urban area with buildings present.

A method known as time reversal processing has demonstrated the ability to focus acoustic waves at the original source location in highly reverberant or scattering environments.[6–8] Relying on the time symmetry of the wave equation, the method reverses the signatures' time sequence and rebroadcasts them from the sensor locations. Remarkably, the signals propagate back through the medium and ultimately reconverge at the original source location. This method is often applied in the physical medium itself,[6,9] for example to focus waves on the stone in lithotripsy[10] or remove the reverberation in underwater communication.[11] Investigations of the time reversal method conducted using numerical models, as will be done in this paper, is less common.[6,12,13]

In this paper, time reversal processing is applied to acoustic source location in a small urban or suburban area containing a number of closely spaced buildings. The goal is to demonstrate that source location in this complex propagation environment is feasible using NLOS sensors. The finite difference method that is used to simulate acoustic propagation in the urban area is briefly discussed in Sec. II. Then, the time reversal processing method is described and is shown to provide source locations in this environment using only NLOS sensors.

[a)]An earlier version of this work was presented in L. Liu and D. G. Albert, "Time reversal for source detection in urban environment," at the 147th Meeting of the Acoustical Society of America, New York, May 2004 (J. Acoust. Soc. Am. **115**, 2596).

[b)]Electronic mail: donald.g.albert@erdc.usace.army.mil

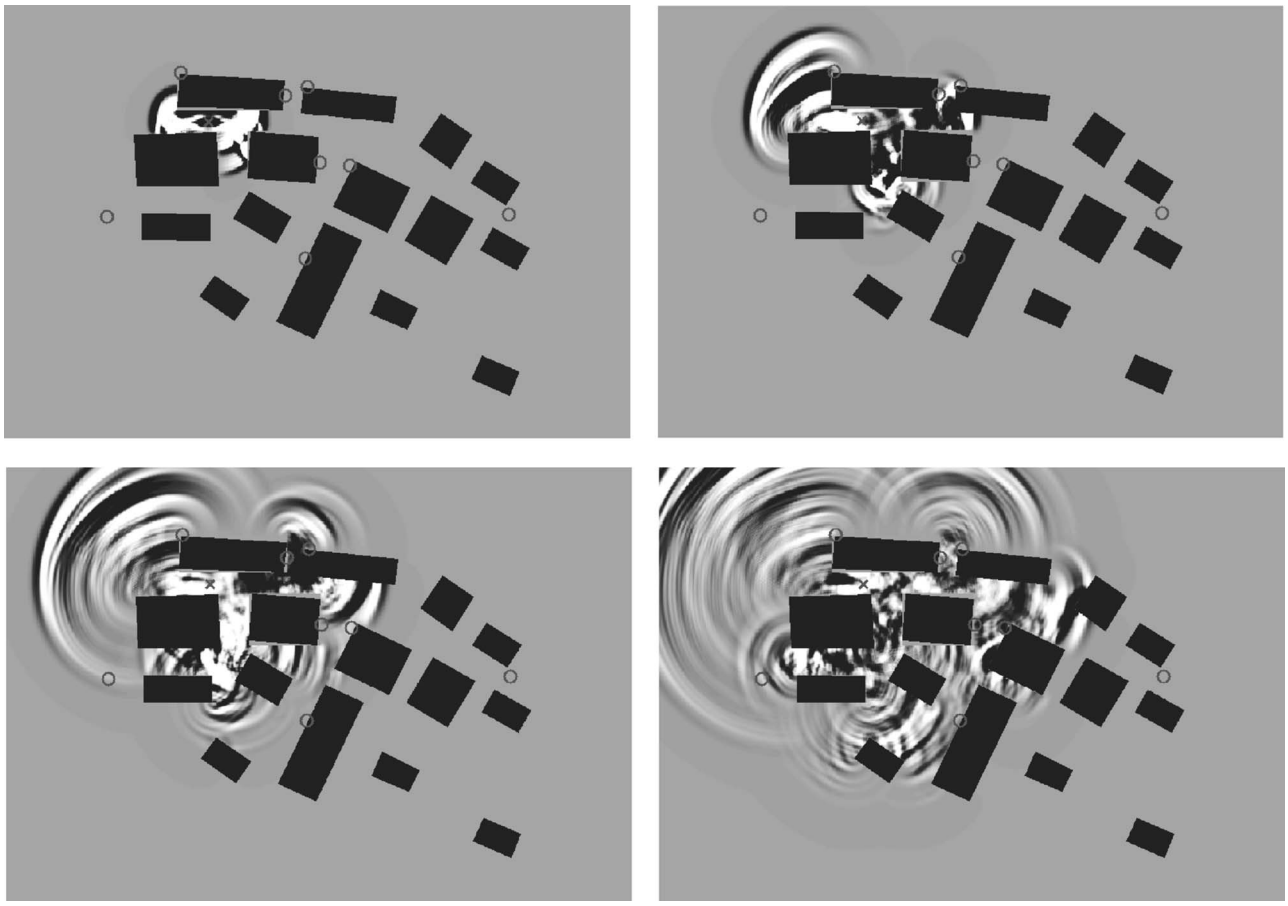


FIG. 1. Calculated acoustic pulse propagation in an urban area with 15 buildings. The area shown is 200×140 m and the building geometry is based on a full-scale artificial training village. Snapshots of the acoustic pressure are shown every 50 ms. The source was an explosion wave form located between two of the buildings and marked with an "*x*." Circles are acoustic sensor locations that will be used in the time reversal processing. (Positive pressure is black, negative pressure is white.)

II. FINITE DIFFERENCE TIME DOMAIN (FDTD) PROPAGATION MODELING

While acoustic propagation in urban areas has been studied for a long time,[14] and ray tracing or other methods for predicting sound levels in urban areas have been developed,[15–17] it is still a topic of research interest. Some recent studies have used statistical methods to measure or model traffic noise, or diffusion approaches to predict noise levels in street canyons with complex building facades.[18–21] The approach used here differs from most previous work since it attempts to model all of the waves and interactions with buildings produced by an impulsive point source. The finite difference time domain (FDTD) propagation model applied in this paper has been discussed in detail elsewhere,[22] so only a brief summary will be given.

The finite difference method is based on the expression of acoustic propagation as a set of first-order, velocity-pressure coupled differential equations, similar to the motion and continuity expressions used by other authors.[23,24] To approximate the derivatives in the acoustic wave equation with finite differences, a staggered difference algorithm proposed by Yee[25] is used in a two-dimensional spatial domain. The marching is also staggered between the computations of the air pressure and the particle velocity in the time domain. The perfectly matched layer technique[26] was adapted for the absorption boundary condition and achieved highly effective suppression of reflections from the domain boundaries.

To reduce the computational effort and make the problem tractable on a desktop computer, the real three-dimensional world is represented by a simplified two-dimensional model. Buildings are treated as solid blocks in the calculations to speed up the geometric input to the model. Corrections for the two dimensional geometric spreading (an additional factor of $r^{-1/2}$, where r is the propagation distance) and the effect of the ground surface (an additional factor of 2, assuming a rigid ground surface and that the direct and reflected path lengths are nearly identical) are applied in the calculations. With these corrections the model yields surprisingly accurate results.[22] Because of the two-dimensional approximation, acoustic energy outside the plane of the propagation model is ignored, i.e., propagation over the top of the buildings or diffraction from upper edges is not included in the calculations.

The geometry of a full-scale artificial training village is used in the calculations. This flat area has 15 closely spaced concrete buildings arranged in a 200 by 140 m area. Field measurements conducted at this location[27] will be discussed in a later publication. The computational method described

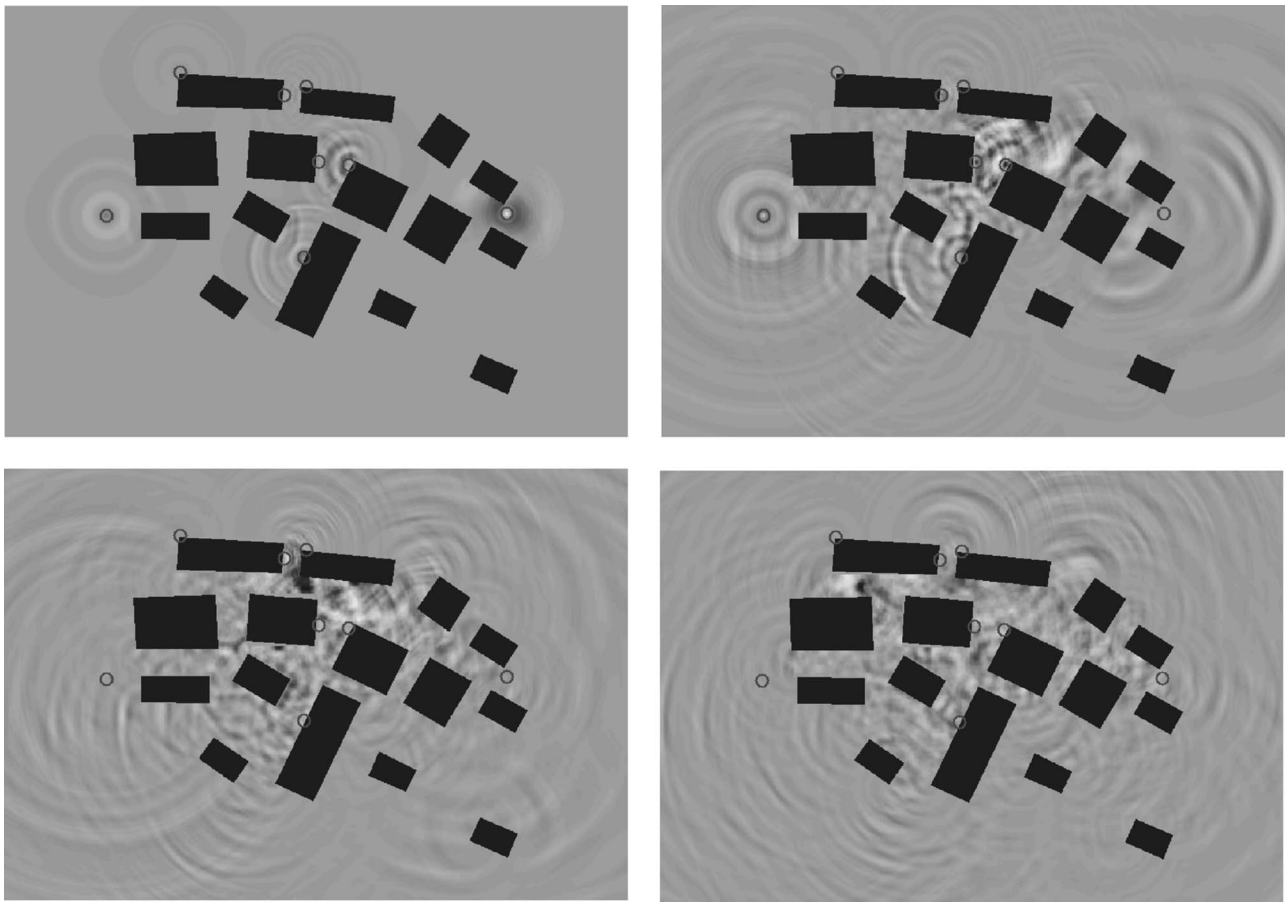


FIG. 2. Time reversal processing using eight non-line-of-sight (NLOS) sensors. The time series that had been calculated using the forward FDTD model at each of the sensor locations (circles) were time reversed and used as sources in the FDTD propagation model. Pressure wave snapshots are shown at 50, 200, 300, and 400 ms elapsed times. In the final panel (lower right), the acoustic energy has focused at the original source location (dark blob).

here should apply to any type of urban area, subject to enough computer memory and time to be able to complete the calculations.

For all of the calculations, a source function representing an explosion with a peak frequency of about 100 Hz was used. A grid spacing of 0.3 m (12 node points per wavelength) was selected, along with a time step of 0.2 ms to insure that the Courant stability criterion[24] was met. The 200 m $\times$ 140 m propagation area was divided into about 300 000 grid points. Each computation was performed in MATLAB running on a 4 GHz desktop personal computer in less than 1 h.

Figure 1 shows snapshots of the pressure wave field calculated for the small village using the FDTD method. Figure 1 shows that the many reflections, diffractions, and scattered waves caused by wave interaction with the buildings produces a very complex acoustic wave field.

III. SOURCE LOCATION USING TIME REVERSAL PROCESSING

Time reversal processing involves the following steps: First, the sound signature produced by a source is recorded at a number of sensor locations after propagation through the complex medium; next, the time series signatures are reversed in time; finally, the reversed time series are emitted from the sensor locations and propagate back through the complex medium. Because of the symmetry of the wave equation, this procedure will refocus acoustic energy at the original source location.

The time reversal steps were performed using the FDTD model for the urban situation shown in Fig. 1, and Fig. 2 shows wave field snapshots of the process using eight NLOS sensors. In the final panel, acoustic energy can be seen focusing at the original source location (shown with an x in Fig. 1). In Fig. 3, the final results of time reversal processing are compared for the same case with eight NLOS sensors, and for a case where only three NLOS sensors are used. Both cases find the correct source location, although the result is stronger for the case with eight sensors compared to the case with three sensors.

To apply the time reversal method in an actual urban area, the building locations, sensor locations, and time signatures are required. With this information the method can be applied to find unknown source locations using the FDTD model as demonstrated earlier. However, before this method could be used in practice, for example to locate gunshots in near real time, the calculation time will need to be decreased by about three orders of magnitude. In addition, further study of this technique is needed to determine the resistance of the method to errors in sensor or building locations and to ambient noise. Despite these requirements, the method could

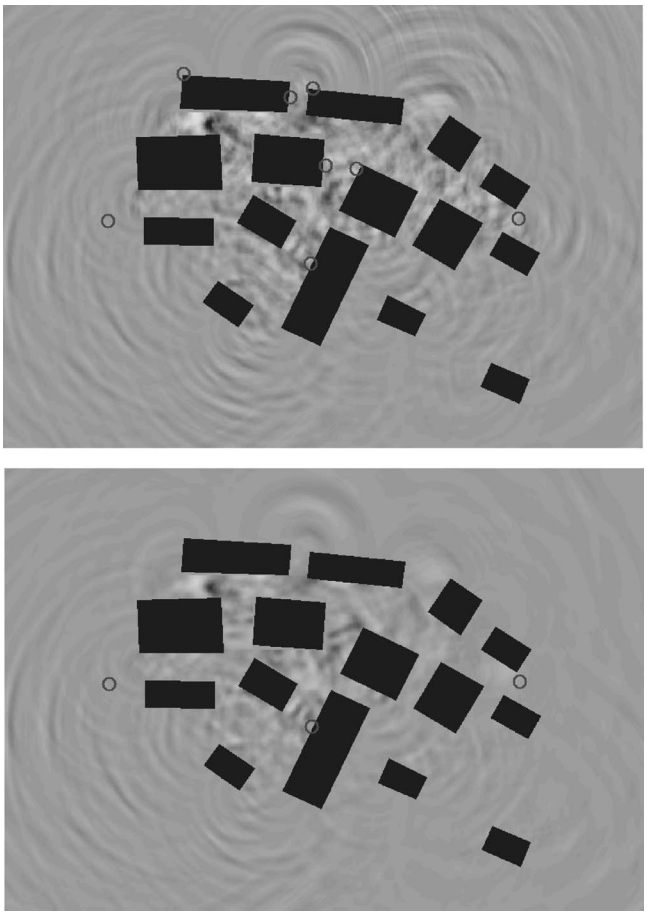


FIG. 3. Results of time reversal processing to determine the sound source location in an urban area. The top panel shows the results from eight NLOS sensors (circles), while the lower panel shows the results using only three NLOS sensors. Both give the correct source location, but the result for eight sensors is more focused and displays lower sidelobes than the result found using three sensors.

become feasible for application to a specific location with fixed and known sensor and building locations, if a precomputing strategy similar to Ref. 28 was used.

IV. CONCLUDING REMARKS

The two-dimensional finite difference time domain method has been used to calculate acoustic wave propagation in a small urban area with a number of closely spaced buildings. The simulation study presented here shows that time reversal processing on the acoustic signatures from a few non-line-of-sight sensors to determine a source location is conceptually possible.[29]

ACKNOWLEDGMENTS

The authors thank the many colleagues who have helped with this project, and the U.S. Army PM-CCS and the U.S. Army Corps of Engineers for funding this research. In addition, thanks are given to the Associate Editor, Dr. Keith Wilson, for many helpful suggestions to improve the manuscript.

[1]A. De Ciccio, "Sound ranging and locating system AN-TNS-3, used in 1950," (private communication).
[2]G. C. Carter, "Time delay estimation for passive sonar signal processing," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-29**, 463–470 (1981).
[3]B. G. Ferguson, "Variability in the passive ranging of acoustic sources in air using a wavefront curvature technique," J. Acoust. Soc. Am. **108**, 1535–1554 (2000).
[4]W. R. Hahn, "Optimum signal processing for passive sonar range and bearing estimation," J. Acoust. Soc. Am. **58**, 201–207 (1975).
[5]C. You, "Non-line-of-sight sound propagation outdoors," Ph.D dissertation, *Department of Physics and Astronomy*, University of Mississippi: Oxford, MS, 1993 109 pp.
[6]A. Derode, P. Roux, and M. Fink, "Robust acoustic time reversal with high order multiple scattering," Phys. Rev. Lett. **75**, 4206–4209 (1995).
[7]M. Fink, "Time reversal of ultrasonic fields I. Basic principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–566 (1992).
[8]M. Fink, "Time-reversed acoustics," Phys. Today **50**, 34–40 (1997).
[9]G. Montaldo, P. Roux, A. Derode, C. Negreira, and M. Fink, "Generation of very high pressure pulses with 1-bit time reversal in a solid waveguide," J. Acoust. Soc. Am. **110**, 2849–2857 (2001).
[10]J.-L. Thomas, F. Wu, and M. Fink, "Time reversal mirror applied to lithotripsy," Ultrason. Imaging **18**, 106–121 (1996).
[11]W. A. Kuperman, W. S. Hodgkiss, H. C. Song, T. Akal, C. Ferla, and D. R. Jackson, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," J. Acoust. Soc. Am. **103**, 25–40 (1998).
[12]S. R. Khosla and D. R. Dowling, "Time-reversing array retrofocusing in noisy environments," J. Acoust. Soc. Am. **109**, 538–546 (2001).
[13]P. Roux and M. Fink, "Time reversal in a waveguide: Study of the temporal and spatial focusing," J. Acoust. Soc. Am. **107**, 2418–2429 (2000).
[14]F. M. Wiener, C. I. Malme, and C. M. Gogos, "Sound propagation in urban areas," J. Acoust. Soc. Am. **37**, 738–747 (1965).
[15]K. P. Lee and H. G. Davies, "Nomogram for estimating noise propagation in urban areas," J. Acoust. Soc. Am. **57**, 1477–1480 (1975).
[16]D. J. Oldham and M. M. Radwan, "Sound propagation in city streets," Build. Acoust. **1**, 65–88 (1994).
[17]K. Heutschi, "A simple method to evaluate the increase of traffic noise emission level due to buildings, for a long straight street," Appl. Acoust. **44**, 259–274 (1995).
[18]J. Kang, "Sound propagation in street canyons: Comparison between diffusively and geometrically reflecting boundaries," J. Acoust. Soc. Am. **107**, 1394–1404 (2000).
[19]J. Picaut, L. Simon, and J. Hardy, "Sound field modeling in streets with a diffusion equation," J. Acoust. Soc. Am. **106**, 2638–2645 (1999).
[20]A. Garcia and L. J. Faus, "Statistical analysis of noise levels in urban areas," Appl. Acoust. **34**, 227–247 (1991).
[21]A. L. Brown and K. C. Lam, "Levels of ambient noise in Hong Kong," Appl. Acoust. **20**, 85–100 (1987).
[22]L. Liu and D. G. Albert, "Acoustic pulse propagation near a right-angle wall," J. Acoust. Soc. Am. (submitted).
[23]J. Virieux, "SH-wave propagation in heterogeneous media: Velocity-stress finite difference method," Geophysics **49**, 1933–1942 (1984).
[24]S. Wang, "Finite-difference time-domain approach to underwater acoustic scattering problems," J. Acoust. Soc. Am. **99**, 1924–1931 (1996).
[25]K. S. Yee, "Numerical solution of initial boundary value problems involving Maxwell″s equations in isotropic media," IEEE Trans. Antennas Propag. **14**, 302–307 (1966).
[26]J. P. Berenger, "A perfectly matched layer for the absorption of electromagnetic waves," J. Comput. Phys. **114**, 185–200 (1994).
[27]D. G. Albert and L. Liu, "Sound propagation in an urban environment I. Preliminary analysis of measurements," J. Acoust. Soc. Am. **114**, 2442 (abstract) (2003).
[28]R. K. Ing, N. Quieffin, S. Catheline, and M. Fink, "Time reversal interactive objects," J. Acoust. Soc. Am. **115**, 2589 (abstract) (2004).
[29]Movies of urban propagation and time reversal processing are available at http://www.acoustics.org/press/147th/liu-albert.html.

Transducer hysteresis contributes to "stimulus artifact" in the measurement of click-evoked otoacoustic emissions (L)

Sarosh Kapadia,[a),b)] Mark E. Lutman,[a)] and Alan R. Palmer
MRC Institute of Hearing Research, University of Nottingham, Nottingham NG7 2RD, United Kingdom

(Received 14 January 2005; revised 6 May 2005; accepted 9 May 2005)

Click-evoked otoacoustic emissions from the human ear are typically several orders of magnitude smaller than the stimuli that elicit them—a measurement technique that attempts to cancel the stimulus signal from the recorded waveform is therefore typically employed. In practice, an imperfect cancellation of the stimulus is achieved, leaving a "stimulus artifact" that obscures the early part of the emission. Input-output nonlinearities of the transducers used in recording emissions are acknowledged as one source of the stimulus artifact. Here an additional source of this artifact, related to hysteresis in the magnetic "receivers" (loudspeakers) used in such recordings, is identified and discussed. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1944547]

PACS number(s): 43.64.Jb, 43.64.Yp [BLM] Pages: 620–622

I. INTRODUCTION

The measurement of click-evoked otoacoustic emissions (CEOAEs) requires the separation within the recorded waveform of the stimulus click from the cochlear response. The majority of published data attempt to achieve this separation by the "differential nonlinear" (DNL) cancellation technique (e.g., Kemp *et al.*, 1986). In this technique, the click stimuli are divided into groups of four, consisting of three clicks of identical amplitude and polarity, followed by a single click of three times that amplitude and of inverted polarity. The responses to all four clicks (as recorded by the probe microphone) are averaged into a single buffer as the recording is made. The net result of this DNL scheme is that linear components of the response (which are exactly three times as large in response to the larger click as in response to the smaller) exactly cancel, leaving only a nonlinear residual, which is taken to be entirely of cochlear origin. However, complete cancellation of the stimulus click component of the waveform does not occur in practice, leaving a "stimulus artifact" that dominates the first few milliseconds of the record. Consequently, the separation of the click from the earliest-latency (i.e., highest-frequency) components of the CEOAE remains a near-intractable problem.

Amplitude nonlinearity of the miniature transducers used within CEOAE probes undoubtedly contributes to the error in cancellation in the stimulus clicks referred to above. However, a research study conducted by the authors (Kapadia and Lutman, 2000a, b), which required near-perfect cancellation of clicks of opposite polarity but equal amplitude has identified a second source of this error, which is reported here.

[a)]Now at the Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom.
[b)]Electronic mail: sk@isvr.soton.ac.uk

II. NATURE AND SOLUTION OF PROBLEM

A. Method

Full details of the CEOAE measurement paradigm developed for the above studies, based on one first described by Kemp and Chum (1980), are described by Kapadia and Lutman (2000a). In summary, a pair of stimulus clicks, denoted "T" (test) and "S" (suppressor), was presented within each averaging epoch (Fig. 1). Test clicks were inverted between successive epochs, while suppressor clicks were of fixed polarity. Every other recorded epoch was then inverted before adding to the average, thus the successive suppressor clicks were added in opposite polarity while the test clicks were added in the same polarity. The result of averaging pairs of such epochs was that test clicks and CEOAEs evoked by them were preserved, while suppressor clicks cancelled. The paradigm was designed to allow a detailed characterization of the effect of the suppressor clicks on the OAEs evoked by the test clicks, as a function primarily of the levels of the test and suppressor clicks and the delay of the suppressor click relative to the test, labeled Δt in Fig. 1. The test clicks were always taken to occur at time zero.

In the authors' original implementation of this paradigm, test, and suppressor clicks were digitally generated via a single channel digital-to-analog converter (DAC), and applied to the miniature loudspeaker (sometimes termed "receiver") (Knowles BK1851) within the CEOAE probe.[1] However, tests revealed a significant artifact at the position of the suppressor clicks (dotted line in Fig. 1), indicating imperfect cancellation.

B. Investigation of artifact

Figure 2 shows a pair of replicate waveforms recorded using the basic test paradigm and the original implementation described above, with the probe inserted into a 0.5 cm^3 test cavity. Test and suppressor clicks were both at a level of 70 dB peak-equivalent (pe) SPL. A highly repeatable artifact caused by incomplete suppressor cancellation, of peak-to-peak amplitude of approximately 500 μPa, is evident in the traces at about 10 ms. Note that the click amplitudes as re-

 0001-4966/2005/118(2)/620/3/$22.50

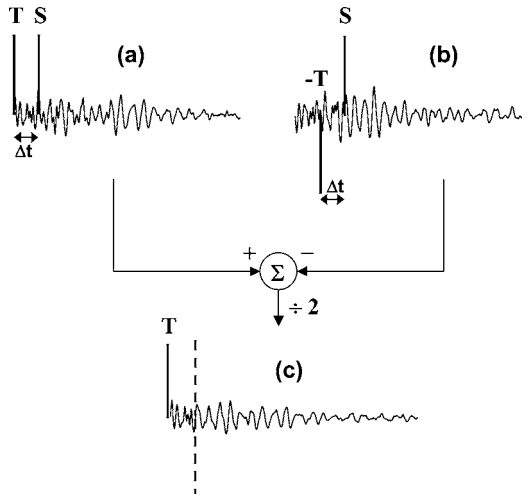


FIG. 1. Paradigm used by Kapadia and Lutman (2000a, b) to measure the effect of a suppressor click on the CEOAE elicited by a test click. Pairs of epochs (a) and (b) are averaged, each containing test (T) and suppressor (S)clicks with polarities as marked. Epoch (b) is inverted in the averaging process, thus cancelling the suppressor clicks (dotted line) in the resultant, (c).

corded in the same traces (not shown in these plots) had peak-to-peak amplitudes of the order of 500 mPa—the artifacts here are therefore approximately 60 dB smaller than the clicks themselves. This "suppressor artifact" was also evident (though somewhat smaller) in the CEOAE waveforms obtained from ear-canal recordings, particularly for lower amplitude CEOAEs.

Exhaustive testing of the system revealed a number of points. Such artifacts were not observed if the output electrical click stream was coupled directly to the measurement system input amplifier, indicating that they arose from the probe transducer section. Furthermore, they were observed when the clicks delivered into a test cavity were measured using a separate reference microphone, rather than the microphone housed in the probe itself, indicating that they arose from the loudspeaker within the probe.

The stimulus signals delivered to the loudspeaker were therefore modified in a variety of manners, none of which eliminated the suppressor artifact. These included the following.

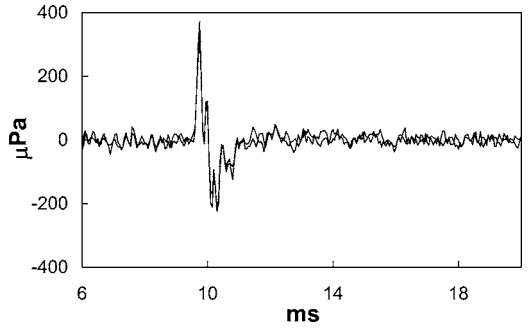


FIG. 2. Replicate recordings obtained in a test cavity using the cancellation paradigm of Fig. 1 and the original system hardware. Test and suppressor click levels were both 70 dB pe SPL, $\Delta t = +9$ ms. A relatively large, repeatable artifact is evident at the position of the suppressor clicks, indicating imperfect cancellation of these clicks.

(i) Adding and varying a dc bias to the loudspeaker signal.
(ii) Doubling the duration of the electrical clicks delivered (from 50 μs to 100 μs), and halving their amplitudes to compensate for this. (It was verified that the amplitude of the acoustic clicks output from the loudspeaker was unchanged by this manipulation of the electrical click waveforms.)
(iii) Utilizing more complex eight-click sequences, with equal numbers of positive and negative clicks, in the cancellation paradigm, rather than the basic four-click sequence illustrated in Fig. 1.

None of these modifications to the stimulus signals materially affected the magnitude of the artifact. However, it was noted that the artifact disappeared if the test clicks were not included in the presentation sequence—the suppressor clicks now cancelled completely into the noise floor. It appeared, therefore, that the artifact was caused by the presentation of the test click prior to each suppressor click. Surprisingly, the artifact was completely *insensitive* to the delay between each test click and the subsequent suppressor click, at least up to the maximum delay tested, which was 200 ms. Rather, the significant aspect of the preceding test clicks seemed simply that they alternated in polarity. In other words, suppressor clicks of very slightly, but consistently, different amplitudes were obtained if they were preceded by positive as opposed to negative test clicks, *irrespective* of the interval between these test and suppressor clicks. There therefore appeared to be a transducer "memory" of the polarity of the test click that did not reduce with time, at least over the time scales for which it was tested. As the miniature loudspeaker used operates on a magnetic principle, it was concluded that magnetic hysteresis left it in a different resting state following a positive as opposed to a negative test click.

C. Modified implementation

Tests were therefore conducted using a two-channel DAC and a pair of similar loudspeakers (in different probes), to separately deliver the test and the suppressor clicks into a reference cavity. This was expected to remove the hypothesized source of the artifact: as all suppressor clicks were positive, the "suppressor" transducer would always return to the same resting state. No noncancellation artifacts were now observed, consistent with the explanations for the effects described above. A modified CEOAE probe that housed a pair of BK1851 loudspeakers was therefore constructed and used with a dual-channel output system that delivered test and suppressor clicks separately to each loudspeaker. (The acoustic paths of the two loudspeakers within the probe assembly were closely matched.)

Figure 3 shows a pair of replicate recordings using this modified "dual-loudspeaker" probe, under exactly the same conditions as for the single-loudspeaker waveforms presented in Fig. 2. It is clear from the figure that any residual noncancellation artifact, if present, is below the measurement noise floor.

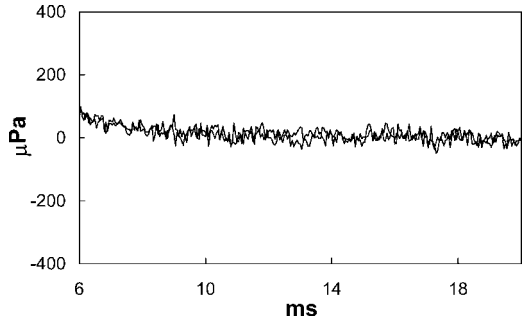


FIG. 3. Replicate recordings obtained in a test cavity using the cancellation paradigm of Fig. 1 and the modified (dual-channel) system hardware. Test and suppressor click levels were both 70 dB pe SPL, $\Delta t = +9$ ms. No artifact due to imperfect suppressor cancellation is detectable, at least down to the measurement noise floor.

III. DISCUSSION AND CONCLUSION

Communications with the manufacturer of the loudspeakers used confirmed the likelihood of a residual magnetization effect arising from hysteresis as being responsible for the artifacts observed in the original implementation of our test paradigm.[2] Such effects do not appear to have been identified in past literature on CEOAEs. However, CEOAE measurement systems almost universally utilize loudspeakers that operate on the same principle (and typically from the same manufacturer) as that used in the present study. It is therefore likely that the effects observed here would also be present, to a lesser or greater extent, in all such systems.

First, it is likely that hysteresis or residual magnetization effects also contribute to the residual stimulus artifact observed in the DNL cancellation technique, which is widely used in the measurement of CEOAEs. Once again, a single magnetic loudspeaker[3] is used to deliver both positive and negative clicks, and the lack of exact cancellation between such clicks can be partially attributed to the fact that the transducer is not returned to a constant resting state during such measurements.

Some other studies using nonstandard CEOAE measurement paradigms have also suffered from system artifacts that are likely to be related to the effects discussed here. For example, Picton *et al.* (1993) utilized a standard CEOAE probe (housing a similar Knowles loudspeaker to ours) for the recording of CEOAEs using the maximum length sequence (MLS) technique. These authors conducted measurements using two different types of MLS'—one type containing clicks of only one polarity (unipolar MLSs) and the other containing a mixture of click polarities (bipolar MLS'). While bipolar MLS click sequences offer some advantages, the authors encountered substantial stimulus-related artifacts, which were far more significant for bipolar MLS' than unipolar ones, and subsequent work on MLS CEOAEs have almost universally been restricted to unipolar sequences. Once again, the MLS technique relies on the near-perfect cancellation of stimulus clicks, i.e. on all clicks being of near-identical amplitude. However, the alternation in polarity of such clicks would leave the loudspeaker in a different resting state, which the present study demonstrates can have a material effect on the amplitude of the subsequent click.

Finally, Keefe (1998) also refers to CEOAE stimulus noncancellation artifacts in the DNL cancellation technique. (Keefe refers to these artifacts as "probe distortion.") However, although Keefe also suggests that such artifacts may (in some cases) be reduced by the use of a pair of output transducers (his "double-source variant double-evoked OAE"), he does not refer to the hysteresis effects that are the subject of this Letter. Keefe's CEOAE measurements, unlike ours and the standard DNL cancellation paradigm, did not involve an alternation in polarity of the stimulus clicks. Furthermore, our findings somewhat disagree with the assertion by Keefe (1998) that a single source would produce negligible distortion with a sufficiently large interclick delay (p. 3491), if his double-evoked OAE technique were utilized with alternating clicks (his $\epsilon < 0$). Our data, in contrast, demonstrate the existence of a significant noncancellation artifact (which we attribute to transducer hysteresis) that is *independent* of the delay between test and suppressor clicks (at least over practical time scales), and that cannot therefore be eliminated by increasing the interclick delay.

ACKNOWLEDGMENTS

The equipment used for the measurements described here, including both the standard and modified CEOAE probe, was designed and constructed by Dave Bullock and colleagues at the MRC Institute of Hearing Research, Nottingham, UK. We are grateful to Otodynamics Ltd. for the supply of technical information.

[1]The probe used was a standard "POEMS" (Programmable Otoacoustic Emission Measurement System) probe, designed and constructed at the MRC Institute of Hearing Research, Nottingham, UK, and used in a number of research studies within the UK.

[2]S. C. Ewens (personal communication).

[3]The most commonly used CEOAE measurement system (Otodynamics Limited's ILO88 and its variants) utilizes Knowles ED1913 loudspeakers.

Kapadia, S., and Lutman, M. E. (**2000a**). "Nonlinear temporal interactions in click-evoked otoacoustic emissions. I. Assumed model and polarity-symmetry," Hear. Res. **146**, 89–100.

Kapadia, S., and Lutman, M. E. (**2000b**). "Nonlinear temporal interactions in click-evoked otoacoustic emissions. II. Experimental data," Hear. Res. **146**, 101–20.

Keefe, D. H. (**1998**). "Double-evoked otoacoustic emissions. I. Measurement theory and nonlinear coherence," J. Acoust. Soc. Am. **103**, 3489–3498.

Kemp, D. T., and Chum, R. A. (**1980**). "Properties of the generator of stimulated acoustic emissions," Hear. Res. **2**, 213–232.

Kemp, D. T., Bray, P., Alexander, L., and Brown, A. M. (**1986**). "Acoustic emission cochleography—Practical aspects," Scand. Audiol. Suppl. **25**, 71–95.

Picton, T. W., Kellett, A. J. C., Vezsenyi, M., and Rabinovitch, D. E. (**1993**). "Otoacoustic emissions recorded at rapid stimulus rates," Ear Hear. **14**, 299–314.

Place-pitch discrimination of single- versus dual-electrode stimuli by cochlear implant users (L)[a)]

Gail S. Donaldson[b)] and Heather A. Kreft
Clinical Psychoacoustics Laboratory, Department of Otolaryngology, University of Minnesota, MMC 396, Rm 8-323 PWB, 420 Delaware Street SE, Minneapolis, Minnesota 55455

Leonid Litvak
Advanced Bionics Corporation, 12740 San Fernando Road, Sylmar, California 91342

(Received 13 July 2004; revised 27 April 2005; accepted 27 April 2005)

Simultaneous or near-simultaneous activation of adjacent cochlear implant electrodes can produce pitch percepts intermediate to those produced by each electrode separately, thereby increasing the number of place-pitch steps available to cochlear implant listeners. To estimate how many distinct pitches could be generated with simultaneous dual-electrode stimulation, the present study measured place-pitch discrimination thresholds for single- versus dual-electrode stimuli in users of the Clarion CII device. Discrimination thresholds were expressed as the proportion of current directed to the secondary electrode of the dual-electrode pair. For 16 of 17 electrode pairs tested in six subjects, thresholds ranged from 0.11 to 0.64, suggesting that dual-electrode stimuli can produce 2–9 discriminable pitches between the pitches of single electrodes. Some subjects demonstrated a level effect, with better place-pitch discrimination at higher stimulus levels. Equal loudness was achieved with dual-electrode stimuli at net current levels that were similar to or slightly higher than those for single-electrode stimuli. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1937362]

PACS number(s): 43.66.Ts, 43.66.Fe [AJO] Pages: 623–626

I. INTRODUCTION

Cochlear implant (CI) listeners have access to a limited number of pitches associated with place of stimulation in the cochlea. For single-electrode stimulation, place pitch is constrained by the number of electrode contacts along the implanted array, typically 12–22 in contemporary devices. Additional factors such as poor neural survival can result in "indiscriminable electrodes," further reducing the number of available pitches related to place of stimulation.

For some CI users, weighted stimulation of two adjacent electrodes can produce one or more intermediate pitches, thus increasing the total number of place-pitch steps available. This phenomenon was first demonstrated by Townshend *et al.* (1987), for simultaneous stimulation of two distant electrodes. Later, Wilson *et al.* (1993, 1994, 2003) used simultaneous stimulation of adjacent electrodes to produce intermediate pitches in four subjects with the Ineraid device. Finally, McDermott and McKay (1994) studied five subjects with the Nucleus-22 implant and showed that intermediate pitches could be generated when pulses on two electrodes were interleaved with a brief temporal separation rather than presented simultaneously.

None of the earlier studies specifically measured the number of discriminable pitches that could be generated for a given dual-electrode pair. However, one subject tested by Wilson *et al.* (2003) was able to distinguish 25% increments in the current weighting between electrodes 4 mm apart and one subject tested by McDermott and McKay could distinguish six dual-electrode stimuli between electrodes separated by 0.75 mm. Other data from the study by McDermott and McKay indicated considerable variability in place-pitch discrimination across individuals and electrode positions.

The purpose of the present study was to further evaluate place-pitch discrimination for simultaneous, dual-electrode stimulation of closely spaced electrodes. In addition to obtaining discrimination thresholds for single- versus dual-electrode stimuli, we sought to obtain preliminary information on the effects of stimulus level. Previous studies have shown that place-pitch discrimination improves with level for single-electrode stimuli (Pfingst *et al.*, 1999; McKay *et al.*, 1999) and a similar effect was anticipated for the task involving dual-electrode stimuli. We also wished to evaluate the effect of dual-electrode stimulation on loudness. In particular, we sought to determine whether a constant level of current produced the same loudness when the current was apportioned between two adjacent electrodes (dual-electrode stimulation) as when it was directed entirely to one electrode (single-electrode stimulation). Loudness summation was demonstrated in the McDermott and McKay study for non-simultaneous dual-electrode stimuli, but loudness effects were not evaluated in earlier studies using simultaneous, dual-electrode stimulation.

II. METHODS

Subjects were six postlingually-deafened adults with a Clarion CII cochlear implant. Relevant subject information is provided in Table I. Each subject had a HiFocus or HiFocus II electrode array, with 16 flat-plate electrode contacts arranged in a line with center-to-center distances of approxi-

[a)]Portions of these data were presented at the VIII Annual International Cochlear Implant Conference, May 2004 (Indianapolis, IN).

[b)]Present address: Deptartment of Communication Sciences and Disorders, University of South Florida, Tampa, FL; electronic mail: gdonalds@cas.usf.edu

TABLE I. Description of subjects. Subject code, gender, age, duration of implant use, duration of deafness prior to implantation, and phonemes-correct score on NU-6 words in quiet.

Subj	M/F	Age (yrs)	Cl use (yrs)	Deaf (yrs)	NU-6 % phon
D01	M	56.1	2.9	26	65
D02	F	54.7	2.9	1	61
D05	F	73.9	2.3	3	40
D08	F	52.8	1.9	13	45
D10	F	50.4	1.9	8	87
D11	M	73.4	1.1	32	39
D18	F	65.4	1.1	19	5

mately 1 mm. All six subjects used the HiResolution speech processing strategy, which employs high-rate, nonsimultaneous pulsatile stimulation.

Experiments were controlled by a personal computer running custom programs written for the Bionic Ear Data Collection System (Advanced Bionics, 2003). Stimuli were 200 ms trains of 32.2 μs/ph, 1000 pulse/s, cathodic-first biphasic pulses, presented in monopolar mode. Pairs of adjacent electrodes in the apical, middle and basal regions of the electrode array were tested (electrodes 2-3, 7-8 and 12-13, respectively). For the single-electrode stimulus, the more apical electrode in the pair was stimulated alone. For the dual-electrode stimulus, both the apical and basal electrodes of the pair were stimulated simultaneously. The proportion of the total current directed to the more basal electrode for the dual-electrode stimulus was denoted as α, with α ranging from 0 (all current to the more apical electrode) to 1 (all current to the more basal electrode). The single- and dual-electrode stimuli used for measuring psychometric functions were balanced in loudness to a perceptual level of medium loud (ML) or medium soft (MS). Equal loudness levels were determined using a double staircase procedure (Jesteadt, 1980) with a reference stimulus that produced a loudness of ML or MS on the more apical electrode of the dual-electrode pair. Two or three equal-loudness estimates were averaged to obtain a final equal-loudness level for each dual-electrode stimulus.

Psychometric functions relating α to pitch discrimination sensitivity were obtained with a two-alternative forced-choice (2AFC) procedure. Functions were initially obtained using relatively large increments (α=0, 0.25, 0.50, 0.75, and 1.0); however, when preliminary threshold estimates indicated good place-pitch discrimination ($\alpha < 0.25$), they were repeated using finer increments (α=0, 0.1, 0.2, 0.3, 0.4, and 0.5). On each trial, the single-electrode stimulus and dual-electrode stimulus were presented in random order, and the subject selected the interval with the higher-pitched sound. A correct response was scored when the subject chose the interval containing the dual-electrode stimulus. No feedback was given. Stimuli were presented in blocks of 50 or 60 trials, comprised of ten trials for each value of α in random order. At least five blocks were obtained for each condition, so that 50 or more comparisons were incorporated in the mean percent-correct score for each value of α. The mean scores were converted to d' values (Hacker and Ratcliffe, 1979) and linear interpolation was used to compute the value of α producing performance of $d'=1.16$ (equivalent to 79.4% correct).

Adaptive place-pitch discrimination thresholds were obtained for comparison to the psychometric function estimates using a 2AFC, 3-down, 1-up procedure that also estimated 79.4% correct performance ($d'=1.16$). The adaptive variable, α, was initially set to a value that allowed the single- and dual-electrode stimuli to be easily discriminated (typically, 0.5 or 0.7). It was then altered in steps of 0.05 for the first three reversals of the track and in steps of 0.025 for the remaining seven reversals. The place-pitch threshold was computed as the mean value of α for the final six reversals. Linear interpolation was used to estimate equal-loudness levels for values of α encountered during the adaptive track that were intermediate to those measured for the psychometric functions.

III. RESULTS AND DISCUSSION

Figure 1 shows psychometric functions and adaptive threshold estimates for the ML stimuli. For most subjects, data are shown for apical, middle, and basal electrode pairs.[1] Psychometric functions were generally well-behaved, showing monotonic increases in performance (d') with increasing values of α. In one case (D18, apical pair), the subject could not reliably discriminate any of the dual-electrode stimuli from the reference, single-electrode stimulus: The psychometric function was nearly flat and performance never reached the threshold criterion value of $d'=1.16$. Not surprisingly, this subject was unable to perform the corresponding adaptive pitch-discrimination task.

In general, there was good agreement between the threshold estimates based on the psychometric functions and the thresholds obtained with the adaptive procedure. The only exception occurred for subject D02 on the basal electrode pair, where the adaptive threshold (α=0.87) was significantly larger than the threshold based on the psychometric function (α=0.64). There was no obvious explanation for this discrepancy.

It is evident from Fig. 1 that place-pitch thresholds varied considerably across subjects and electrodes. Subjects D01 and D08 demonstrated small thresholds for all three electrodes ($\alpha < 0.22$). Subjects D02, D05, and D11 demonstrated larger thresholds, on average, and their thresholds

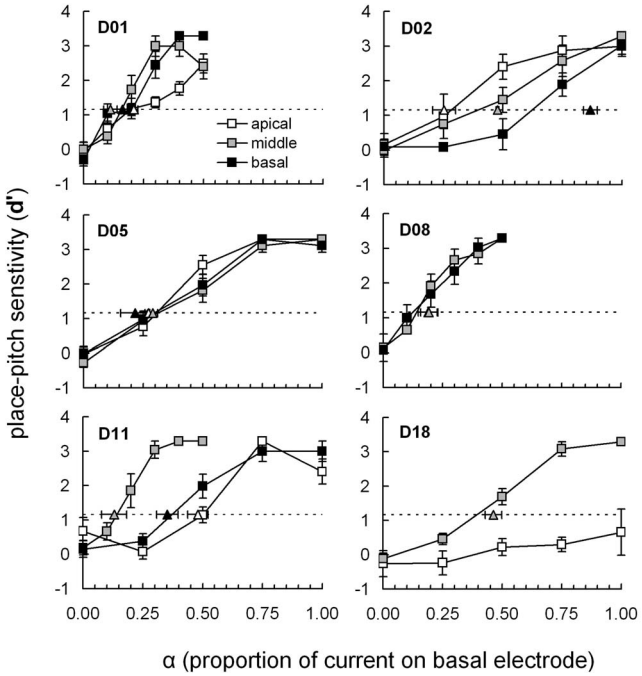


FIG. 1. Psychometric functions and adaptive thresholds for discrimination of single- versus dual-electrode stimuli for apical, middle and basal electrode pairs in six subjects. Stimuli are ML. Adaptive thresholds are represented by the filled triangles. The dashed line indicates the threshold criterion of $d'=1.16$.

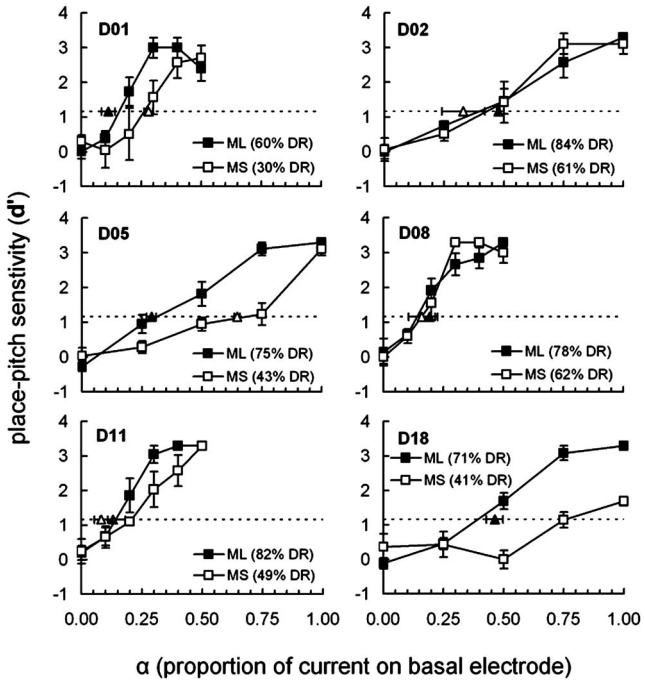


FIG. 2. Psychometric functions and adaptive thresholds for discrimination of single- versus dual-electrode stimuli for one middle electrode pair in each of six subjects. Stimuli are ML or MS. Adaptive thresholds are represented by filled triangles. The dashed line indicates the threshold criterion of $d'=1.16$.

were more variable across electrodes. The remaining subject, D18, achieved a moderate threshold ($\alpha=0.35$) for the middle electrode but, as indicated earlier, was unable to reliably discriminate pitch differences on the apical electrode pair. There was no systematic relation between place-pitch sensitivity and the word recognition scores shown in Table I.

Figure 2 shows psychometric functions and adaptive threshold estimates at two loudness levels (ML and MS) for each subject's middle electrode pair. Two subjects (D05, D18) showed a clear level effect with better performance for the higher-level stimulus. The remaining subjects showed a smaller level effect (D01, D11) or no effect of level (D02, D08). On average, place-pitch sensitivity was significantly better for the ML stimulus than for the MS stimulus, as expected on the basis of previous single-electrode studies (one-tailed paired-comparison t test for thresholds estimated from the psychometric functions, $t=-2.22$, $df=5$, $p<0.05$).

Figure 3 shows representative loudness balance data for two subjects (D02 and D05) obtained for the ML and MS conditions depicted in Figs. 1 and 2. Current levels are expressed in the clinical units used by the Clarion device. Each data point reflects the average current level computed from two or three individual loudness-balance estimates. Differences among the individual estimates for a given condition were generally very small, averaging 1.74% (0.15 dB) across 19 loudness-balance functions in six subjects.

The current required to produce a medium loud (or medium soft) percept was generally similar for the two single electrodes of a given electrode pair ($\alpha=0$ and $\alpha=1$ conditions in Fig. 3). For about half of the loudness balance functions, the net current levels producing equal loudness for the intermediate, dual-electrode conditions fell along an imaginary line connecting the two end points. This indicates that the current requirements for the dual-electrode stimuli were equivalent to those for the single-electrode stimuli. Examples of this are seen in Fig. 3 for all of D05's loudness-balance functions and for D02's functions for the apical and basal electrode pairs. For the other half of the functions, data points for the dual-electrode conditions fell slightly above the imaginary line connecting the end points of the function, indicating that the dual-electrode stimuli required a higher net current level than the single-electrode stimuli. Examples of this occur for D02's middle electrode pair (MS and ML conditions). A one-way, repeated measures analysis of variance on ranks was applied to the data for 19 functions in six subjects after normalization to adjust for current differences in the single-electrode stimuli. This analysis showed that the dual-electrode stimuli required significantly higher current

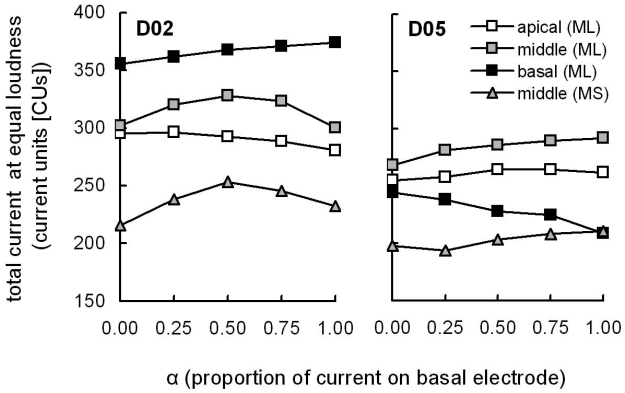


FIG. 3. Loudness balance functions for two subjects.

levels, on average, than the corresponding single-electrode stimuli (chi-squared=28.8, $df=4$, $p<0.001$). Post-hoc tests indicated that current levels for the dual-electrode stimuli were not significantly different for the three values of α (0.25, 0.50, and 0.75). Although the dual-electrode stimuli required higher net current levels on average than the single-electrode stimuli, the absolute magnitude of these differences was small. The largest difference observed between the measured value and the value expected from the single-electrode data was 1.1 dB (D02, middle electrode pair, MS) and the difference was greater than 0.5 dB in only one other instance (D02, middle electrode pair, ML).

The loudness effects observed here for simultaneous, dual-electrode stimulation contrast with those reported by McDermott and McKay (1994) for nonsimultaneous stimulation. In their study, each pulse of the dual-electrode stimulus required a 0.76–1.1 dB reduction in current amplitude to achieve equal loudness with the corresponding single-electrode stimulus; however, this corresponds to a net increase in total charge of approximately 5 dB for the dual-electrode stimulus (~6 dB increase for presentation of two stimuli less ~1 dB reduction). Simultaneous dual-electrode stimulation requires less total charge than nonsimultaneous stimulation because it involves the direct summation of field currents as compared to the summation of neural responses or loudness.

Comparison of the present results with those of McDermott and McKay (1994) suggest that average place-pitch discrimination is similar for simultaneous and nonsimultaneous dual-electrode stimulation. This appears to be true even though the underlying mechanisms are different: Nonsimultaneous stimulation involves integration of responses at the neural membranes or more centrally in the auditory system, whereas simultaneous stimulation involves summation of intracochlear current fields.

Although the present study evaluated the discrimination of single-electrode versus dual-electrode stimuli, it is likely that similar thresholds would be obtained for the discrimination of dual-electrode stimuli with different values of α. Further research is needed to confirm this assumption and to extend the present findings to a larger sample of subjects.

IV. CONCLUSIONS

(1) Dual-electrode stimulation can increase the number of place-pitch steps available to cochlear implant patients with contemporary devices. In the present study, place-pitch discrimination of adjacent electrodes was possible for 16 of 17 electrode pairs evaluated in six subjects. Thresholds for single- versus dual-electrode stimulation ranged from 0.11 to 0.64, suggesting that a two- to ninefold increase in the number of place-pitch steps is possible with dual-electrode stimuli.

(2) Some subjects demonstrate a level effect in which place-pitch discrimination of dual-electrode stimuli improves with stimulus level. This effect is similar to the level effects observed for place-pitch discrimination with single-electrode stimulation.

(3) Equal loudness can be achieved with simultaneous, dual-electrode stimuli at net current levels that are similar to or only slightly higher than those for single-electrode stimuli.

ACKNOWLEDGMENTS

This research was supported by NIDCD Grant Nos. P01-DC00110 and R01-DC006699. Edward Overstreet and Lakshmi Mishra contributed to the conceptual development of this research. Andrew Oxenham and two anonymous reviewers provided helpful comments on an earlier version of the manuscript. The authors thank the six subjects who participated in this study.

[1]Data for the apical electrode pair of subject D08 are not shown because the subject demonstrated a pitch reversal for these electrodes. Despite the pitch reversal, the psychometric function and adaptive thresholds showed good place-pitch resolution (thresholds of $\alpha=0.7$ and $\alpha=0.9$, respectively). Subject D18's basal electrode pair was not tested because the subject could not tolerate moderate or loud stimuli for electrodes in the basal portion of her array.

Advanced Bionics (**2003**). Bionic Ear Data Collection System, Version 1.16 Users Manual.

Hacker, M. J., and Ratcliff, R. (**1979**). "A revised table of d' for M-alternative forced choice," Percept. Psychophys. **26**, 168–170.

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophys. **28**, 85–88.

McDermott, H. J., and McKay, C. M. (**1994**). "Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea," J. Acoust. Soc. Am. **96**, 155–162.

McKay, C. M., O'Brien, A., and James, C. J. (**1999**). "Effect of current level on electrode discrimination in electrical stimulation," Hear. Res. **136**, 159–164.

Pfingst, B. E., Holloway, L. A., Zwolan, T. A., and Collins, L. M. (**1999**). "Effects of stimulus level on electrode-place discrimination in human subjects with cochlear implants," Hear. Res. **134**, 105–115.

Townshend, B., Cotter, N., van Compernolle, D., and White, R. L. (**1987**). "Pitch perception by cochlear implant subjects," J. Acoust. Soc. Am. **82**, 106–115.

Wilson, B. S., Zerbi, M., and Lawson, D. (1993). Speech processors for auditory prostheses, NIH Contract N01-DC-2-2401, 3rd Quarterly Progress Report, February 1–April 30, 1993.

Wilson, B. S., Lawson, D. T., Zerbi, M., and Finley, C. C. (**1994**). "Recent developments with the CIS strategies," in *Advances in Cochlear Implants, Proceedings of the Third International Cochlear Implant Conference, Insbruck, Austria, April 1993*, edited by I. J. Hochmair-Desoyer and E. S. Hochmair (Manz, Wien), pp. 103–112.

Wilson, B. S., Wolford, R., Schatzer, R., Sun, X., and Lawson, D. (2003). Speech processors for auditory prostheses, NIH Project N01-DC-2-1002, 7th Quarterly Progress Report, October 1–December 31, 2003.

The energy method for analyzing the piezoelectric electroacoustic transducers. II. (With the examples of the flexural plate transducer)

Boris Aronov
Acoustic Research Laboratory, Advanced Technology and Manufacturing Center and Department of Electrical and Computer Engineering, The University of Massachusetts, Dartmouth, 151 Martine Street, Fall River, Massachusetts 02723 and BTech Acoustics, LLC, 1445 Wampanoag Trail, Suite 115, East Providence, Rhode Island 02915

(Received 14 August 2004; revised 30 January 2005; accepted 11 May 2005)

The energy method for analyzing piezoelectric electro-acoustic transducers, described in the previous paper [B. S. Aronov, J. Acoust. Soc. Am. **117**, 210–220 (2005)] is used here in application to transducers operating in the receive mode. Changes to the equivalent circuit of a transducer are introduced, arising from the action of the sound field considered as the external source of energy. The application of the energy method is demonstrated by an example of the circular flexural plate transducer. The validity of the single degree of freedom approximation for calculating the transducer parameters is considered, and estimation is made of inaccuracies arising from the simplifying assumptions regarding the mode of transducer vibration. The acceptable tolerance levels for accuracy in the calculation of the transducer parameters are also discussed. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1953087]

PACS number(s): 43.38.Ar, 43.38.Fx, 43.38.Pf, 43.30.Yj [AJZ] Pages: 627–637

I. INTRODUCTION

The energy method of analyzing the piezoelectric electro-acoustic transducers was considered in a previous paper[1] mainly for the application in the transmit mode of operation. Though the conclusions on the operation in the receive mode can be made by reciprocity, it is instructive to consider the application of the method for the receive mode in detail separately and especially in terms of interaction between the mechanical system of the transducer and the acoustical field. In the course of application of the energy method, some simplifying assumptions were suggested for determining the modes of vibration, such as one degree of freedom approximation and substitution of a real mechanical system by an analogous system made of an isotropic material. As the assumptions made may result in a possible reduction of accuracy, a technique has to be developed for estimation of the accuracy of the proposed assumptions. The objective of this paper is to further consider the energy method for analyzing transducers in both the transmit and receive modes, and to give an example of estimation of possible inaccuracies arising in the process of calculating the transducer parameters.

In Sec. II the energy balance in the receive mode of operation and related interaction of the mechanical system of the transducer and the acoustic field as a source of energy are considered. Equivalent circuits of the transducers in the receive mode are introduced. In Sec. III the validity of a fixed vibration mode approximation to the real velocity distribution for the mechanical system of a transducer and in particular the approximation by the static deflection curve is considered. This is done with an example of a rectangular flexural plate transducer. In Sec. IV the theory of calculating the parameters of the circular plate flexural transducer is presented, and it is shown that the results obtained for the case of the static deflection curve are in very good agreement with those obtained by employing the first fundamental mode of vibration. The assumption of the static deflection curve being used as the mode of vibration is not applicable to some hydrophone designs, for example, for the pressure gradient hydrophones of the motion type, in which case the mechanical system of the transducer is allowed to move as a rigid body. Under this circumstance an additional degree of freedom has to be taken into account related to the piston-like motion of the body, and the transducer has to be treated as having two mechanical degrees of freedom. This approach is also illustrated in Sec. IV, with the same example of the circular plate transducer under the condition that its boundary is allowed to move. Estimation of inaccuracies of the calculation of the circular plate transducer parameters arising due to assuming that the Poisson's ratio is $\sigma=0.3$ instead of its real value for the entire variety of ceramic compositions commonly used for the hydrophones designing is done in Sec. V. The resulting errors are qualitatively compared with possible inaccuracies due to approximate knowledge of the real boundary conditions for the circular plates in the practical transducer designs. Also, some considerations are discussed on a reasonably acceptable inaccuracy for calculation of the transducer's properties in exchange for the improvement in the physical interpretation and ease of application of the results.

II. ENERGY BALANCES AND THE EQUIVALENT ELECTROMECHANICAL CIRCUITS IN THE RECEIVE MODE

We consider the use of the energy method for transducers in the receive mode and the changes that have to be made to the equivalent circuits originally developed for the trans-

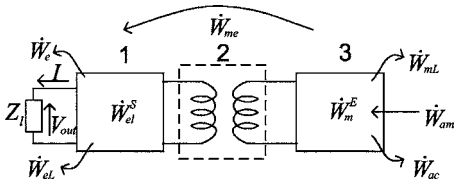


FIG. 1. The block diagram of a transducer as an energy-converting system operating in the receive mode.

mit mode as published in the previous paper.[1] The block diagram of acousto-electrical transduction system is shown in Fig. 1. The block diagram for energy conversion in the receive mode differs from those introduced for the transmit mode by the direction of energy flow. In the receive mode the acoustic field constitutes the source of energy, W_{am}, supplied to the transducer, and the electrical load, Z_l, is applied to the transducer output. In order to characterize the acoustic field as a source of energy, consider the energy of interaction between acoustic field and the transducer mechanical system, W_{am}. Under the action of acoustic field [sound pressure $P_\Sigma(\bar{r})$ on the transducer surface] the transducer surface vibrates. The mode of vibration can be represented in the same way as in the transmit mode by the formula $U(\bar{r}_\Sigma)=U_o\theta(\bar{r}_\Sigma)$, where U_o is the magnitude of velocity of the reference point on the transducer surface, and the function $\theta(\bar{r}_\Sigma)$ is the distribution of vibration over the transducer surface. This situation is illustrated in Fig. 2. The acousto-mechanical power in the complex form can be expressed as

$$\bar{\dot{W}}_{\text{am}} = \int_\Sigma P(\bar{r}_\Sigma)U^*(\bar{r}_\Sigma)d\Sigma = U_o^*\int_\Sigma P(\bar{r}_\Sigma)\theta(\bar{r}_\Sigma)d\Sigma. \quad (1)$$

The sound pressure on the transducer surface may be represented as

$$P(\bar{r}_\Sigma) = P^U(\bar{r}_\Sigma) - P_{\text{br}}(\bar{r}_\Sigma), \quad (2)$$

where $P^U(\bar{r}_\Sigma)$ is the sound pressure on the blocked transducer surface (at $U=0$) and $P_{\text{br}}(\bar{r}_\Sigma)$ is the sound pressure due to the back radiation generated by vibration of the transducer surface. Upon substitution of expression (2) for $P(\bar{r}_\Sigma)$ into Eq. (1), we obtain

$$\bar{\dot{W}}_{\text{am}} = U_o^*\int_\Sigma P^U(\bar{r}_\Sigma)\theta(\bar{r}_\Sigma)d\Sigma - U_o^*\int_\Sigma P_{\text{br}}(\bar{r}_\Sigma)\theta(\bar{r}_\Sigma)d\Sigma \quad (3)$$

or

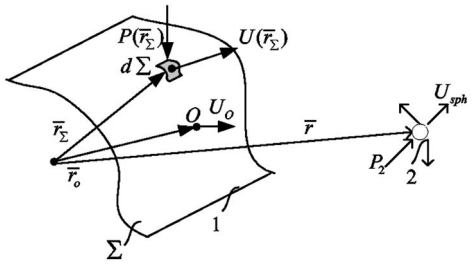


FIG. 2. Illustration of the mechano-acoustic system consisting of the surface of a transducer 1 and the pulsating sphere of a small radius 2.

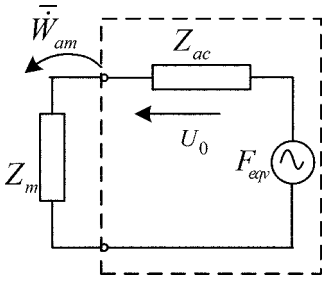


FIG. 3. Interpretation of acoustical field as a source of mechanical energy supplied to a transducer.

$$\bar{\dot{W}}_{\text{am}} = \bar{\dot{W}}_{\text{am}}^U - \bar{\dot{W}}_{\text{ac.br}}. \quad (4)$$

The second term on the right side of the relation (3) is the acoustic energy of the back radiation, and it can be represented as $\bar{\dot{W}}_{\text{ac.br}}=Z_{\text{ac}}|U_o|^2$, where Z_{ac} was previously defined.[1]

We denote the integral in the first term as the equivalent force, F_{eqv}

$$F_{\text{eqv}} = \int_\Sigma P^U(\bar{r}_\Sigma)\theta(\bar{r}_\Sigma)d\Sigma. \quad (5)$$

Thus

$$\bar{\dot{W}}_{\text{am}}^U = U_o^*F_{\text{eqv}}. \quad (6)$$

Now, the expression for the acousto-mechanical power, $\bar{\dot{W}}_{\text{am}}$, can be rewritten in the form

$$\bar{\dot{W}}_{\text{am}} = (F_{\text{eqv}} - Z_{\text{ac}}U_o)U_o^*. \quad (7)$$

On the other hand, to denote the input impedance of the mechanical system as Z_m, the power supplied to the mechanical system is

$$\bar{\dot{W}}_{\text{am}} = Z_mU_oU_o^*. \quad (8)$$

Comparing relations (7) and (8), we arrive at

$$(Z_m + Z_{\text{ac}})U_o = F_{\text{eqv}}. \quad (9)$$

This relation can be interpreted by the circuit shown in Fig. 3. Both Eq. (9) and Fig. 3 show that the acoustic field as a source of energy for a transducer can be considered as the equivalent acousto-mechanical generator with "mechanomotive" force F_{eqv} and internal impedance Z_{ac}. Calculation of the equivalent force, F_{eqv}, is the subject of radiation theory. In the case where the dimensions of a transducer are small compared with the wavelength of sound, we have $P^U(\bar{r}_\Sigma)\doteq P_0$, where P_0 is the sound pressure of the propagating wave, and

$$F_{\text{eqv}} = P_0\int_\Sigma \theta(\bar{r}_\Sigma)d\Sigma = P_0S_{\text{av}}. \quad (10)$$

In formula (10) the average transducer surface S_{avi} is introduced as

$$S_{\mathrm{av}} = \int_{\Sigma} \theta(\bar{r}_{\Sigma}) d\Sigma. \tag{11}$$

If the dimensions of the transducer are comparable with the wavelength, an equivalent force may be represented as

$$F_{\mathrm{eqv}} = P_0 k_{\mathrm{dif}} S_{\Sigma}, \tag{12}$$

where k_{dif} is the diffraction coefficient and S_{Σ} is the surface area.

The diffraction coefficient k_{dif} has to be calculated by equating the formulas (12) and (5) after the sound-pressure distribution $P^{U}(\bar{r}_{\Sigma})$ is found by solving the diffraction problem for the blocked transducer surface. But, in fact, if the radiation problem is already solved, the diffraction coefficient k_{dif} can be determined by applying the reciprocity principle to the mechano-acoustic system, consisting of two transducers: transducer 1 with surface Σ, on which the distribution of velocity is specified as $U(\bar{r}_{\Sigma}) = U_o \theta(\bar{r}_{\Sigma})$, and the pulsating sphere 2 of small radius a located at a large distance from the transducer (see Fig. 2). One of the formulations of the reciprocity principle is as follows:

$$\left. \frac{P_2}{U_{\tilde{V}1}(\bar{r}_{\Sigma})} \right|_{U_{\tilde{V}2}=0} = \left. \frac{P_1^{U}(\bar{r}_{\Sigma})}{U_{\tilde{V}2}} \right|_{U_{\tilde{V}1}=0}. \tag{13}$$

In expression (13) the notation is adopted such that P_2 is the sound pressure acting on clamped surface of the sphere 2, but generated by vibrations of an element $d\Sigma$ of transducer 1 with the volume velocity $U_{\tilde{V}1}(\bar{r}_{\Sigma}) = U_o \theta(\bar{r}_{\Sigma}) d\Sigma$; $P_1^{U}(\bar{r}_{\Sigma})$ is the sound pressure acting on the same element $d\Sigma$ of the clamped surface Σ, when sphere 2 vibrates with the volume velocity $U_{\tilde{V}2} = U_{\mathrm{sph}} 4\pi a^2$.

Taking into account a well-known expression for the sound pressure P_0 generated by a pulsating sphere of small radius in the free field

$$P_0 = \frac{\rho c}{r} U_{\mathrm{sph}} k a^2 e^{-j(kr-\pi/2)}, \tag{14}$$

we obtain the relationship between the volume velocity of the sphere and P_0

$$U_{\tilde{V}2} = \frac{2r}{\rho c} P_0 \lambda e^{j(kr-\pi/2)}, \tag{15}$$

where $\lambda = 2\pi/k$ is the wavelength.

The sound pressure generated in the free field by an elementary source $d\Sigma$ located on a clamped surface Σ, which we denote as $P_2 = P_{d\Sigma}(\bar{r}, \bar{r}_{\Sigma}) = P_{\delta}(\bar{r}, \bar{r}_{\Sigma}) d\Sigma$, where P_{δ} is the sound pressure generated by a point source, can be considered as known from the solution of the problem of radiation. As the clamped sphere of a small radius does not disturb the acoustic field, it should be $P_2|_{U_{\tilde{V}2}=0} = P_{d\Sigma}$. After substitution of the values $P_2 = P_{\delta}(\bar{r}, \bar{r}_{\Sigma}) d\Sigma$, $U_{\tilde{V}1} = U_o \theta(\bar{r}_{\Sigma}) d\Sigma$, and $U_{\tilde{V}2}$ from formula (15) into expression (13) written as $P_1 U_{\tilde{V}1} = P_2 U_{\tilde{V}2}$, we obtain

$$P_1(\bar{r}_{\Sigma}) U_o \theta(\bar{r}_{\Sigma}) d\Sigma = \frac{2\lambda r}{\rho c} P_0 e^{-j(kr-\pi/2)} P_{\delta}(\bar{r}, \bar{r}_{\Sigma}) d\Sigma.$$

Now, we integrate both parts of this expression over the surface Σ and take into account formula (5) for F_{eqv} and expression (7) from Ref. 1 for the sound pressure generated by the vibrating surface Σ

$$P(\bar{r}, \omega) = \int_{\Sigma} P_{\delta}(\bar{r}, \bar{r}_{\Sigma}) d\Sigma = \frac{\rho c}{r} U_o e^{-j(kr-\pi/2)} \chi(\bar{r}, \omega).$$

In the result we obtain the following relation:

$$F_{\mathrm{eqv}} = 2\lambda \chi(\bar{r}, \omega) P_0. \tag{16}$$

As the distance r from the pulsating sphere is arbitrarily large, here P_0 can be considered to be the sound pressure of the plane acoustic wave in the free field at the transducer location. The function $\chi(\bar{r}, \omega)$ is the known diffraction function obtained from the solution of the radiation problem. By comparing expressions (16) and (12), we obtain the diffraction coefficient for the transducer in the receive mode in the form

$$k_{\mathrm{dif}} = \frac{2\lambda \chi(\bar{r}, \omega)}{S_{\Sigma}}, \tag{17}$$

which coincides with the difraction coefficient, $k_{\mathrm{dif}.t}$, introduced for the transducer in the transmit mode [formula (12) in Ref. 1]. This result is valid for an arbitrary mode of vibration, whereas the analogous result obtained in Ref. 2 was related only to transducers with uniformly vibrating surfaces.

After considering the acoustic field as a source of energy, the energy balance associated with block 3 in Fig. 1 for the transducers with one mechanical degree of freedom (i.e., at $\theta(\bar{r}_{\Sigma})$ independent of frequency) can be represented as

$$\bar{\dot{W}}_m = \bar{\dot{W}}_{\mathrm{am}}^{U} - \bar{\dot{W}}_{\mathrm{mL}} - \bar{\dot{W}}_{\mathrm{ac.br}} - \bar{\dot{W}}_{\mathrm{me}}. \tag{18}$$

The only incoming (positive) energy flux is $\dot{W}_{\mathrm{am}}^{U}$. The energy fluxes of mechanical loss, $\dot{W}_{\mathrm{mL}}$, of acoustic back radiation, $\dot{W}_{\mathrm{ac.br}}$, and the flux of mechanoelecrical energy, $\dot{W}_{\mathrm{me}}$, that is, the part of mechanical energy transformed to the electrical block 1, are outgoing (negative).

The energy balance associated with block 1 may be represented as

$$\bar{\dot{W}}_{\mathrm{el}} = \bar{\dot{W}}_{\mathrm{me}} - \bar{\dot{W}}_{l}, \tag{19}$$

where $\dot{W}_{\mathrm{el}}$ is the electric energy accumulated in the block, $\dot{W}_{\mathrm{me}}$ is the incoming mechano-electrical energy flux, and $\dot{W}_{l}$ is the energy flux outgoing in the electrical load. In Ref. 3 it is shown that the mechano-electrical energy can be defined as

$$\bar{\dot{W}}_{\mathrm{me}} = U_o V_{\mathrm{out}}^{*} n, \tag{20}$$

where V_{out} is the output voltage of the transducer. In addition to the expressions for the energies involved in re-

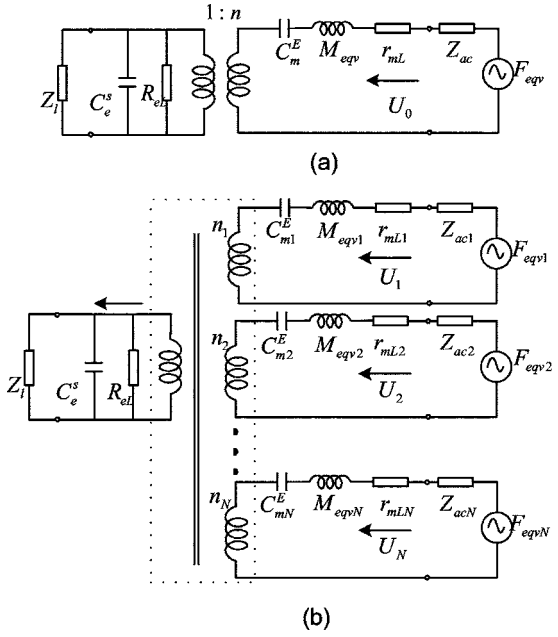


FIG. 4. The equivalent circuits of a transducer in the receive mode: (a) in the case of a single mechanical degree of freedom; (b) in the case of multiple degrees of freedom.

lations (18) and (19), which are given in Ref. 1, note that $\bar{W}_l = V_{\text{out}} V_{\text{out}}^* / Z_l$.

After substituting the expressions for energies in relation (19), the following equation for the electrical subsystem of the transducer will be obtained:

$$V_{\text{out}}\left(j\omega C_e^S + \frac{1}{R_{\text{eL}}} + \frac{1}{Z_l}\right) = U_o n. \tag{21}$$

The term in the parentheses is the electrical admittance of the "clamped" transducer, Y_{el}^U. Now, expression (20) for the mechano-electrical power can be transformed to

$$\bar{W}_{\text{me}} = U_o U_o^* \frac{n^2}{Y_{\text{el}}^U}. \tag{22}$$

After substituting in relation (18) the corresponding expressions for the energies involved, we arrive at the equation for the mechanical subsystem of the transducer

$$U_o\left(j\omega M_{\text{eqv}} + \frac{1}{j\omega C_m^E} + r_{\text{mL}} + Z_{\text{ac}} + \frac{n^2}{Y_{\text{el}}^U}\right) = F_{\text{eqv}}. \tag{23}$$

Equations (21) and (23) can be considered the Kirchhoff's equations for the circuit presented in Fig. 4(a), which differs from the equivalent circuit of the transducer in the transmit mode by introducing the equivalent force, F_{eqv}, as an external driver and by applying the electrical load to the receiver output. After this is shown, the corresponding changes to the equivalent circuit of the transducer having multiple degrees of freedom can be made by observation, as is illustrated in Fig. 4(b). The multicontour equivalent circuit can be proved in the same way as the equivalent circuit for the transmit mode was derived.[1] In this case the Lagrangian has to be used in the form

$$L = W_{\text{kin}} - W_m^E + (W_{\text{am}}^U - W_{\text{mL}} - W_{\text{ac.br}} - W_{\text{me}}). \tag{24}$$

The equivalent force, F_{eqvi}, in the circuit of Fig. 4(b) has to be determined by the formula

$$F_{\text{eqvi}} = \frac{\partial \bar{\dot{W}}_{\text{am}}^U}{\partial U_i}, \tag{25}$$

where $\bar{\dot{W}}_{\text{am}}^U$ should be represented according to formula (1) after substituting $U(\bar{r}_\Sigma) = \Sigma_1^N U_i \theta_i(\bar{r}_\Sigma)$ in the form

$$\bar{\dot{W}}_{\text{am}}^U = \int_\Sigma P^U(\bar{r}_\Sigma)\left[\sum_{i=0}^{N} U_i \theta_i(\bar{r}_\Sigma)\right]^* d\Sigma. \tag{26}$$

Thus, the equivalent force is

$$F_{\text{eqvi}} = \int_\Sigma P^U(\bar{r}_\Sigma)\,\theta_i(\bar{r}_\Sigma) d\Sigma. \tag{27}$$

And, for the transducers small when compared with wavelength, the force is

$$F_{\text{eqvi}} = P_0 S_{\text{avi}}, \tag{28}$$

where $S_{\text{av.i}}$ is the average area of the transducer radiating surface, which is expressed as

$$S_{\text{avi}} = \int_\Sigma \theta_i(\bar{r}_\Sigma) d\Sigma. \tag{29}$$

The equivalent circuits of Fig. 4 are applicable to transducers both in the transmit and receive modes. They are the most general tools for calculation of the parameters of transducers intended to be used as either transmitters and/or receivers. For the transducers that are specialized as receivers (hydrophones in particular), the most common operation is below the first resonance frequency. In this case the mechanical system of a receiver is usually approximated as having a single degree of freedom, with the mode of vibration corresponding to the first resonance mode or to the deflection curve under action of hydrostatic pressure (the latter is typical for application of Rayleigh's principle). The validity of this approximation can be illustrated qualitatively by considering the multimode frequency response of a receiver and estimated quantitatively with an example of a transducer for which the exact analytical solution is known.

III. ON THE JUSTIFICATION OF A SINGLE-MODE APPROXIMATION FOR THE TRANSDUCERS IN THE RECEIVE MODE

According to the equivalent circuit of Fig. 4(b), the open circuit output voltage of a receiver in the case of the multimode approach may be obtained as

$$V_{\text{out}} = \frac{1}{C_e^S}\sum_{i=1}^{N} n_i \dot{\xi}_i(\omega), \tag{30}$$

where $\xi_i(\omega)$ are the generalized coordinates in the expansion of the transducer surface displacement

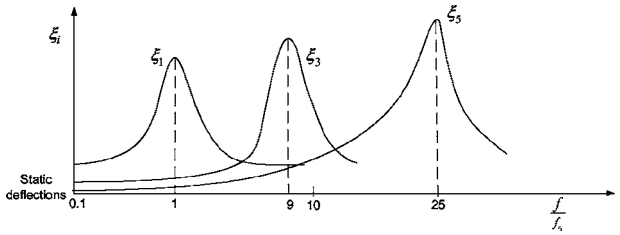


FIG. 5. Qualitative representation of the generalized coordinates, $\xi_i(\omega)$, as functions of frequency. The separation between resonant frequencies corresponds to the case of a rectangular flexural plate simply supported on the two opposite edges.

$$\xi(\bar{r}_\Sigma,\omega)=\sum_{i=1}^{N}\xi_i(\omega)\,\theta_i(\bar{r}_\Sigma). \tag{31}$$

The contribution of different modes into the output voltage can be illustrated by Fig. 5, where the generalized coordinates are displayed qualitatively as functions of frequency. It can be concluded that the contribution of higher modes in the total response at low frequencies is relatively small compared with contribution of the first mode. And, the static deflection mode, which can be imagined as the superposition of all the modes at $f=0$, provides an even better approximation to the real displacement distribution at low frequencies.

In order to qualitatively estimate the accuracy of the one mechanical degree of freedom approximation for a transducer in the frequency range below the first resonance frequency, consider the example of a receiver with mechanical system in the shape of a rectangular plate simply supported at the two opposite edges. The multimode flexural-type electromechanical transducer of this kind was considered in Ref. 4. The electromechanical parameters of this transducer are summarized as follows:

$$K_m^E=\frac{1}{C_{mi}^E}\doteq\frac{i^4\pi^4wt^3}{24s_{11}^E l^3},\quad M_{\mathrm{eqv.i}}=\frac{1}{2}\rho lwt,$$
$$n_i=(-1)^{k-1}i\frac{\pi}{2}\frac{wtd_{31}}{ls_{11}^E},\quad S_{\mathrm{av.i}}=\frac{2wl}{i\pi}, \tag{32}$$

where $i=(2k-1)$, $k=1,2,\ldots$ and l, w, and t are the length, width, and thickness of the plate, respectively. It is assumed that piezoelectric half plates are connected in parallel and that, in the expression for the equivalent rigidity, $1/C_m^E$, a small additional term due to electrical interaction in the deformed piezoelectric plate[3] is neglected (the inaccuracy constitutes about 2.5% in the case where PZT-4 is used). In the frequency range below the fundamental frequency the transducer vibrations can be considered as rigidity controlled in all the high modes. This means that the approximations can be made $\omega M_{\mathrm{eqvi}}\ll 1/\omega C_m^E$ and $Z_{\mathrm{ac.i}}\ll 1/\omega C_{mi}^E$ for $i>1$. The equivalent force can be expressed as $F_{\mathrm{eqvi}}=P_0S_{\mathrm{avi}}$, because typically the dimensions of the transducer are small compared with the wavelength. Thus, the equivalent circuit in Fig. 4(b) is simplified and the output voltage V_{out} of the receiver may be estimated as

$$V_{\mathrm{out}}\doteq\frac{P_0}{j\omega C_e^S}\left[\frac{S_{\mathrm{av.}1}n_1}{Z_{m1}^D}+\sum_{k=2}^{N}S_{\mathrm{av}(2k-1)}n_{2k-1}j\omega C_{m(2k-1)}^D\right]$$
$$=V_{\mathrm{out1}}\left[1+Z_{m1}^D\sum_{p=2}^{N}\frac{S_{\mathrm{av}(2k-1)}}{S_{\mathrm{av1}}}\frac{n_{2k-1}}{n_1}j\omega C_{m(2k-1)}^D\right], \tag{33}$$

where Z_{m1}^D is the mechanical impedance corresponding to the first mode of vibration. In this expression V_{out1} is the output voltage due to the first mode of vibration, and the second term in the brackets characterizes contribution of higher modes (the superscripts D are due to taking into account the reaction of the electrical side). After noting that $|Z_{m1}^D|<1/\omega C_{m1}^D$ and calculating the parameters involved by formulas (32), we arrive at

$$Z_{m1}^D\sum_{k=2}^{N}\frac{S_{\mathrm{av}(2k-1)}}{S_{\mathrm{av1}}}\frac{n_{2k-1}}{n_1}j\omega C_{m(2k-1)}^D<\sum_{k=2}^{N}\frac{S_{\mathrm{av}(2k-1)}}{S_{\mathrm{av1}}}\frac{n_{2k-1}}{n_1}\frac{C_{m(2k-1)}^D}{C_{m1}^D}$$
$$=\sum_{k=2}^{N}\frac{1}{(2k-1)^4}.$$

It is known that $\Sigma_{k=1}^{\infty}[1/(2k-1)^4]=\pi^4/96\doteq 1.01$, and therefore $\Sigma_{k=2}^{N}[1/(2k-1)^4]=\Sigma_{k=1}^{\infty}[1/(2k-1)^4]-1\doteq 0.01$. Thus, the second term in the brackets in expression (33) that represents the contribution of all the higher modes may be neglected. Definitely this result is valid at $f\to 0$, i.e., in the case where the mode shape of vibration becomes the static deflection curve under the action of uniformly distributed force.

The most typical frequency range of operation of transducers in the transmit mode is the range around the resonant frequency. It is qualitatively clear from Fig. 5 that the relative contribution of the higher order modes to the vibration of the transducer surface in this frequency range is even smaller than it is below the resonant frequency. In order to illustrate how the contribution due to these modes can be estimated, we will consider the same example of the flexural rectangular plate transducer (although this transducer type cannot be recognized as a typical projector design). As the sound pressure generated by a transducer is proportional to the velocity of the reference point, the contribution of the high modes can be estimated as follows:

$$P<P_1\left[1+\sum_{k=2}^{\infty}(U_{2k-1}/U_1)\right].$$

The inequality is due to the fact that radiation of higher order modes is reduced due to out-of-phase vibration of the parts of transducer surface. After substituting $U_1=Vn/Z_{m1}^E$ and $U_{2k-1}=Vn_{2k-1}/Z_{m(2k-1)}^E\doteq Vn_{2k-1}j\omega C_{m(2k-1)}^E$, where Z_{mi}^E is the modal mechanical impedance of the transducer, into this inequality, we arrive at

$$P<P_1\left[1+\sum_{k=2}^{\infty}(-1)^{k-1}i|Z_{m1}^E\omega C_{m(2k-1)}^E|\right].$$

In the worst case, when deviation from the resonant frequency is of such a magnitude that $Q_m^2(2\Delta f/f_r)^2\gg 1$, the mechanical impedance Z_{m1}^E can be represented as $|Z_{m1}^E|$

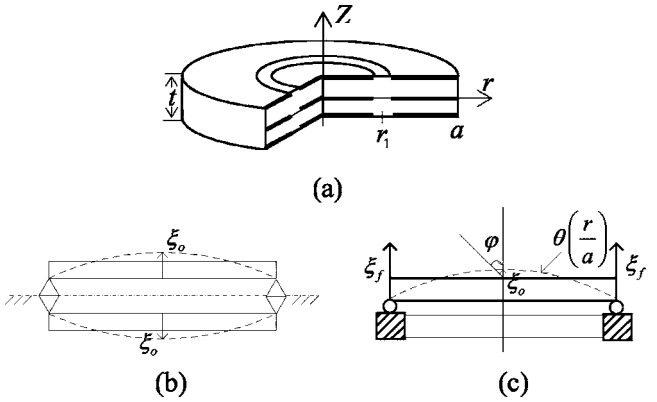


FIG. 6. Mechanical systems of the flexural-type circular plate transducer. ξ_0 is the displacement of the reference point at $r=0$; $\theta(r/a)$ is the mode of vibration.

$\doteq(2\Delta f/f_r)/\omega C_{m1}^E$. Also taking into account the expression for C_{mi}^E from formulas (32), we finally obtain

$$P < P_1\left[1+\frac{2\Delta f}{f_r}\sum_{k=2}^{\infty}\frac{(-1)^{k-1}}{(2k-1)^3}\right].$$

It is known that $\Sigma_{k=1}^{\infty}[(-1)^{k-1}/(2k-1)^3]=\pi^3/32\doteq 0.97$; therefore, the second term in brackets, which characterizes the relative contribution of higher order modes, Δ, is $\Delta=-0.03(2\Delta f/f_r)$. This contribution is less than 1% even in the case where the deviation from the resonant frequency is as large as ±15%.

Though the above estimations are made with the particular example of a rectangular plate with simply supported edges, the analogous treatment can be applied to plates of different configurations and with different boundary conditions. The general conclusion can be made that, in the frequency range below the fundamental resonant frequency, the first normal mode and the static deflection curve may be used interchangeably. The static deflection curve for the plate with simply supported two opposite edges under uniform load is[5]

$$\Theta(x)=\frac{16}{5l}\left(x-\frac{2x^3}{l^2}+\frac{x^4}{l^3}\right). \tag{34}$$

The equivalent parameters determined using this mode of vibration are

$$M_{\text{eqv}}=0.49M,\quad K_m^E=4.0\frac{wt^3}{s_{11}^E l^3},$$
$$n=1.6\frac{wtd_{31}}{ls_{11}^E},\quad S_{\text{avi}}=0.64wl. \tag{35}$$

Obviously, the difference between these values of the equivalent parameters and those determined using the fundamental mode and represented by formulas (32) at $i=1$ is negligible.

IV. THE FLEXURAL CIRCULAR PLATE TRANSDUCER

As an example of a typical transducer for application in the receive mode, consider the flexural circular bilaminar plate transducers shown in Fig. 6. The plate is assumed to be thin compared with the radius $t\ll a$ and free of stress on the major surfaces. Thus, stress in the axial direction throughout the thickness is neglected, and $T_3=0$. The electrodes and the plate vibrations are axially symmetric. Therefore, in the polar coordinates the strains are

$$S_1=S_{rr}=\frac{\partial\xi_r}{\partial r},\quad S_2=S_{\varphi\varphi}=\frac{\xi_r}{r}, \tag{36}$$

where ξ_r is displacement in the radial direction. On the other hand, from the elementary theory of bending[5]

$$S_{rr}=-z\frac{\partial^2\xi_z}{\partial r^2}, \tag{37}$$

where ξ_z is the displacement in the axial direction, which we represent as

$$\xi(r/a)=\xi_0\theta(r/a). \tag{38}$$

Combining (36)–(38), we have

$$S_1=-z\xi_0\cdot\frac{\partial^2\theta}{\partial r^2},\quad S_2=-z\xi_0\frac{1}{r}\frac{\partial\theta}{\partial r}. \tag{39}$$

The piezoelectric equations in common notations[1] are

$$S_1=s_{11}^E T_1+s_{12}^E T_2+d_{31}E_3, \tag{40}$$

$$S_2=s_{12}^E T_1+s_{22}^E T_2+d_{32}E_3, \tag{41}$$

$$D_3=d_{31}T_1+d_{32}T_2+\varepsilon_{33}^T E_3. \tag{42}$$

Given that, because of symmetry of properties of piezoelectric ceramic at normal conditions $s_{22}^E=s_{11}^E$ and $d_{32}=d_{31}$, upon substituting T_1 and T_2 from Eqs. (40) and (41) into Eq. (42), we obtain

$$D_3=\frac{d_{31}}{s_{11}^E+s_{12}^E}(S_1+S_2)+\varepsilon_{33}^{S_{1,2}}E_3, \tag{43}$$

where $\varepsilon_{33}^{S_{1,2}}=\varepsilon_{33}^T(1-k_p^2)$, $k_p^2=2d_{31}^2/\varepsilon_{33}^T(s_{11}^E+s_{12}^E)$.

It follows from Eqs. (40) and (41) that at $E_3=0$

$$T_1^E=\frac{s_{11}^E}{(s_{11}^E)^2-(s_{12}^E)^2}S_1-\frac{s_{12}^E}{(s_{11}^E)^2-(s_{12}^E)^2}S_2, \tag{44}$$

$$T_2^E=-\frac{s_{12}^E}{(s_{11}^E)^2-(s_{12}^E)^2}S_1+\frac{s_{11}^E}{(s_{11}^E)^2-(s_{12}^E)^2}S_2. \tag{45}$$

The mechanical energy of the plate vibration is

$$W_m^E=\frac{1}{2}\int_{\tilde V}(S_1T_1^E+S_2T_2^E)d\tilde V$$
$$=\frac{1}{2}\xi_0^2\frac{2\pi t^3}{12s_{11}^E[1-(\sigma^E)^2]}\int_0^a\left[\left(\frac{\partial\theta}{\partial r}\right)^2+2\sigma^E\frac{1}{r}\frac{\partial\theta}{\partial r}\frac{\partial^2\theta}{\partial r^2}+\left(\frac{1}{r}\frac{\partial\theta}{\partial r}\right)^2\right]r\,dr=\frac{1}{2}\xi_0^2K_m^E, \tag{46}$$

where $\sigma^E=-s_{12}^E/s_{11}^E$ is denoted as the analog of the Poisson's ratio for piezoelectric ceramic material.

The kinetic energy of the vibrating plate is

$$W_{\text{kin}} = \tfrac{1}{2}\int_{\tilde{V}} \rho \dot{\xi}^2(\bar{r}) d\tilde{V} = \tfrac{1}{2}\dot{\xi}_0 2\pi\rho t \int_0^a \theta^2\left(\frac{r}{a}\right) r\, dr = \tfrac{1}{2}\dot{\xi}_0 M_{\text{eqv}}. \tag{47}$$

We denote the equivalent mass, M_{eqv}, as $M_{\text{eqv}} = \rho t S_{\text{eff}}$, where

$$S_{\text{eff}} = 2\pi \int_0^a \theta^2\left(\frac{r}{a}\right) r\, dr \tag{48}$$

is the effective surface area of a plate.

The electromechanical energy is

$$W_{\text{em}} = \frac{1}{2}\int_{\tilde{V}} \frac{d_{31}}{s_{11}^E + s_{12}^E}(S_1 + S_2) E_3 d\tilde{V} = \frac{1}{2}\xi_0 \upsilon n. \tag{49}$$

If we assume that the piezoelectric ceramic plates are connected in series, then $E_3 = (\upsilon/t)\Omega(z)$. In the case where conditionally positive direction of the electric field coincides with direction of polarization $\Omega(z) = 1$, and in the case where they are opposite, $\Omega(z) = -1$. As to the configuration of the electrodes in the radial direction, suppose that the electrodes may be divided in two parts: at $0 \leqslant r \leqslant r_1$ and at $r_1 < r \leqslant a$, as it is shown in Fig. 6(a). The particular case of fully electroded plates corresponds to $r_1 = a$. If only one part of the electrodes is active, we assume that the remaining part is short circuited, in which case the rigidity of the transducer remains unchanged, as the condition of the constant electric field is fulfilled throughout the volume of the plate. At first we will consider that $r = r_1$. After substituting expressions (39) for S_1 and S_2 into formula (49) we obtain

$$\begin{aligned} W_{\text{em}} &= \frac{1}{2}\xi_0 \upsilon \frac{4\pi d_{31}}{s_{11}^E + s_{12}^E}\frac{1}{t}\int_0^{t/2} (-) z\, dz \int_0^{r_1}\left(\frac{\partial^2\theta}{\partial r^2} + \frac{1}{r}\frac{\partial\theta}{\partial r}\right) r\, dr \\ &= \frac{1}{2}\xi_0 \upsilon \frac{(-)\pi d_{31} t}{s_{11}^E + s_{12}^E}\int_0^{r_1} \frac{\partial}{\partial r}\left(r\frac{\partial\theta}{\partial r}\right) dr \\ &= \frac{1}{2}\xi_0 \upsilon \frac{(-)\pi d_{31} t}{s_{11}^E + s_{12}^E}\left(r\frac{\partial\theta}{\partial r}\right)\Bigg|_0^{r_1} = \frac{1}{2}\xi_0 \upsilon n, \end{aligned} \tag{50}$$

and the electromechanical transformation coefficient is

$$n = -\pi \frac{d_{31} t}{s_{11}^E + s_{12}^E}\left(r\frac{\partial\theta}{\partial r}\right)\Bigg|_0^{r_1}. \tag{51}$$

If both parts of electrodes are used and connected in opposite phase, we have

$$n = -\pi \frac{d_{31} t}{s_{11}^E + s_{12}^E}\left(2r_1 \left.\frac{\partial\theta}{\partial r}\right|_{r=r_1} - a \left.\frac{\partial\theta}{\partial r}\right|_{r=a}\right). \tag{52}$$

Thus, the electromechanical transformation coefficient is determined by the slope of the mode of vibration on the borders of electrodes. The mode of vibration of the plate is defined by the boundary conditions, which may in practice vary significantly depending on the transducer design.

At first we will assume that the boundary is simply supported, which presupposes that $\xi(r/a)|_{r=a} = 0$ and $d^2\xi/dr^2|_{r=a} = 0$. These conditions may be closely achieved in the symmetrical double-sided transducer design shown in Fig. 6(b). For the simply supported plate the static deflection curve is[5]

$$\theta\left(\frac{r}{a}\right) = \left(1 - \frac{r^2}{a^2}\right)\left(1 - \frac{1+\sigma}{5+\sigma}\frac{r^2}{a^2}\right), \tag{53}$$

and for the piezoelectric plate $\sigma = \sigma^E$.

The peculiarity of the above expressions for defining the equivalent parameters in the case of a simply supported circular plate by comparison with a rectangular plate is that they depend on the material parameter σ^E, which affects the mode of vibration (53). Because of this, strictly speaking, the general formulas for the transducer equivalent parameters cannot be obtained in a closed form, and calculation of the integrals involved must be completed for each particular ceramic composition having different Poisson's ratio σ^E. However, the values of these integrals for all the modern ceramic materials do not deviate significantly from those obtained with an approximately average value of $\sigma^E = 0.3$, as will be shown in Sec. V. If we neglect these small deviations, then after substituting the mode of vibration by formula (53) into the expressions (10), (46), (48), and (51), the equivalent parameters for the circular, simply supported plate may be expressed as follows:

$$\begin{aligned} &K_m^E = \frac{23}{a^2}\frac{t^3}{12(s_{11}^{E2} - s_{12}^{E2})}, \quad S_{\text{eff}} = 0.29\pi a^2, \\ &S_{\text{av.i}} = 0.46\pi a^2, \quad n = 1.5\pi \frac{d_{31} t}{s_{11}^E + s_{12}^E}. \end{aligned} \tag{54}$$

We recall that the piezoelectric elements are assumed to be connected in series. In the case of the parallel connection, the coefficient in formula (54) for the electromechanical transformation coefficient, n, would be 3π.

The additional rigidity ΔK due to nonuniform distribution of strain along the electric field, which was introduced in Ref. 1 and was shown to be $\Delta K = 0.25 K_m^E k_{31}^2/(1 - k_{31}^2)$ for the flexural plates employing the transverse piezoelectric effect, amounts to about 2.5% of K_m^E in the case where PZT-4 is used. And, it was estimated in Ref. 1 that the discrepancy is even smaller for practical trilaminar plate transducer designs, in which case the thickness of the piezoelectric material in combination with passive materials is optimized. Thus, to the first approximation this term may be neglected.

In the single-plate transducer design variant, such as may be used for the pressure gradient diffraction-type hydrophones, the boundary situation can be modeled as shown in Fig. 6(c). The plate is fixed in a hinge-like manner to a uniformly distributed circular foundation having a finite total mass M_f. Analytically these conditions correspond to $\xi(r/a)|_{r=a} = \xi_f$ and $d^2\xi/dr^2|_{r=a} = 0$. If the mass of the foundation M_f was infinitely large, the displacement of the foundation would be $\xi_f = 0$, and the plate might be considered as simply supported. Note that this condition is achieved even with a finite foundation mass in the double-sided transducer design shown in Fig. 6(b), as the middle plane of the transducer does not vibrate because of symmetry.

In the case under consideration, the static deflection curve cannot represent the mode of vibration since the entire mechanical system can move as a whole under the action of the static pressure. In terms of resonant frequencies, the fre-

quency $\omega=0$ has to be formally considered as the first resonant frequency of this mechanical system, and the piston-like motion has to be considered as the corresponding normal mode of vibration. Though the piston-like motion may dominate at low frequencies, it does not produce an output effect itself, because the piezoelectric plate does not deform under this motion. Therefore, the next mode of the plate vibration has to be taken into account, and the total distribution of displacement may be represented as the superposition of two modes as follows:

$$\xi(r/a) = \xi_f + \xi_0\theta(r/a), \tag{55}$$

where the first term is the displacement of the mechanical system as a whole (as if the plate could not deform) and the second term is the displacement of the simply supported plate (as if the foundation were "clamped"). We assume that $\theta(r/a)$ is the normalized static deflection curve expressed by formula (53) at $\sigma=0.3$.

The kinetic energy of the mechanical system will be

$$W_{\text{kin}} = \tfrac{1}{2}M_f\dot{\xi}_f^2 + \pi\rho t\int_0^a [\dot{\xi}_f + \dot{\xi}_0\theta(r/a)]^2 r\,dr$$

$$= \tfrac{1}{2}(M_f + M_{\text{pl}})\dot{\xi}_f^2 + \rho t S_{\text{av}}\dot{\xi}_f\dot{\xi}_0 + \tfrac{1}{2}\rho t S_{\text{eff}}\dot{\xi}_0^2. \tag{56}$$

Here, $M_{\text{pl}}=\pi a^2\rho t$ is the total mass of the plate, S_{av} and S_{eff} have values defined in expressions (54), namely, $S_{\text{av}}=0.46\pi a^2$ and $S_{\text{eff}}=0.29\pi a^2$. The term $\rho t S_{\text{av}}$ characterizes the mutual inertia coupling between generalized coordinates, and it will be denoted as $\rho t S_{\text{av}}=M_{\text{of}}$. Obviously, $M_{\text{of}}=0.46M_{\text{pl}}$ and $M_{\text{eqvo}}=M_{\text{eff}}=0.29M_{\text{pl}}$.

The potential energy, W_m^E, and electromechanical energy, W_{em}, remain the same as defined by Eq. (46) and (49), because they depend only on strain distribution and do not include ξ_f. For the acousto-mechanical power $\bar{\dot{W}}_{\text{am}}^U$, which is defined as the first term in Eq. (1), in the case where sound pressure is acting only on one side of the plate we obtain

$$\bar{\dot{W}}_{\text{am}}^U = 2\pi\int_0^a P^U(r)[U_f + U_0\theta(r/a)]^* r\,dr$$

$$= F_{\text{eqvf}}U_f^* + F_{\text{eqv0}}U_0^*. \tag{57}$$

If the diameter of the plate is small compared with wavelength, i.e., $P^U(r)\doteq P_0$, then

$$F_{\text{eqvf}} \doteq P_0 S = \pi a^2 P_0, \quad F_{\text{eqv0}} \doteq P_0 S_{\text{av}} = 0.46\pi a^2 P_o. \tag{58}$$

In the case where both sides of the plate are exposed to the sound field, the diffraction on the circular plate, considered as oscillating, has to be taken into account, and the energy $\bar{\dot{W}}_{\text{am}}^U$ has to be calculated by integrating sound pressure over both sides of the plate. As the result of this, the diffraction coefficient[6] $k_{\text{dif}}\doteq -j(4/3\pi)ka\cos\varphi$ will appear in the expressions for the equivalent forces (58) and the surface area of the plate will be doubled. Thus, we arrive at

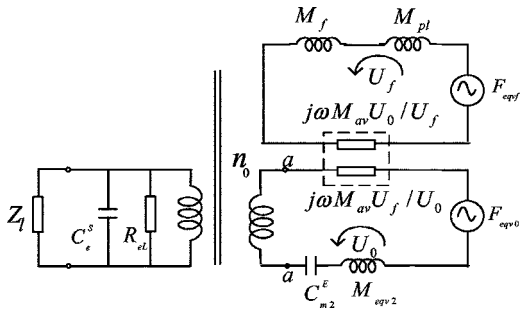


FIG. 7. The equivalent circuit of the transducer with mechanical system supported by a finite mass foundation.

$$F_{\text{eqwf}} \doteq -j2.7ka^3P_0\cos\varphi, \quad F_{\text{eqv0}} \doteq -j1.2ka^3P_o\cos\varphi, \tag{59}$$

where φ is the angle shown in Fig. 6(c).

The Lagrangian (24) for the system will include the energies W_{kin}, W_m^E, W_{me}, and W_{am}^U defined by expressions (56), (46), (20), and (57) respectively (the two latter expressions represent the corresponding powers in the complex form); the energies W_{ac} and W_{mL} may be neglected so far as operation in the range below the first resonant frequency is concerned. And, the following coupled Lagrange's equations for the generalized velocities $\dot{\xi}_f$ and $\dot{\xi}_o$ (in the complex form U_f and U_o) will be obtained:

$$j\omega(M_f + M_{pl})U_f + j\omega M_{0f}U_0 = F_{\text{eqvf}}, \tag{60a}$$

$$j\omega M_{0f}U_f + \left(j\omega M_{\text{eqv}} + \frac{K_m^E}{j\omega} + nV_{\text{out}}\right)U_0 = F_{\text{eqv0}}. \tag{60b}$$

The parameters $M_{\text{eqv}}=\rho t S_{\text{eff}}$, K_m^E, and n in Eq. (60b) are given by formulas (54). For the electrical side of the transducer Eq. (23) is valid. The term nV_{out} in Eq. (60b) can be rewritten as n^2/Y_{el}^U according to Eq. (21). This term represents the reaction of the electrical side of the transducer to vibration of its mechanical system. The set of equations (60) and Eq. (23) may be considered the Kirchhoff's equations describing the equivalent electromechanical circuit presented in Fig. 7. The coupling between the mechanical contours of the circuit is accounted for by introducing the coupled reactances $X_{f0}=j\omega M_{0f}U_0/U_f$ and $X_{\text{of}}=j\omega M_{0f}U_f/U_0$.

It would be interesting to estimate the accuracy of the results obtained by employing the approximate expression (55) for the displacement distribution by comparison with the results of the exact solutions for the case where they are actually known. Thus, it is known that the resonant frequency of a passive circular plate with a free boundary, f_{free}, is related to the resonant frequency of the same plate simply supported, f_{ss}, as $f_{\text{free}}=1.82\ f_{\text{ss}}$, and the radius of the nodal circle of a free plate is $r_{\text{nl}}=0.68a$. In the case under consideration the free-plate conditions can be fulfilled if the mass of foundation is decreased to $M_f=0$. Likewise, when the mass of foundation is increased to $M_f\to\infty$, the conditions for a simply supported plate are achieved.

The solution for the resonance frequency of the plate with a finite foundation mass, f_{fm}, can be derived from the condition where the mechanical impedance between points

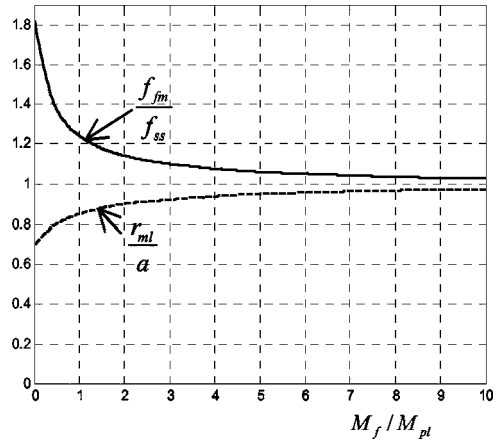


FIG. 8. Relative resonant frequency and relative nodal circle radius for a circular plate supported by a finite mass foundation vs relative mass of the foundation.

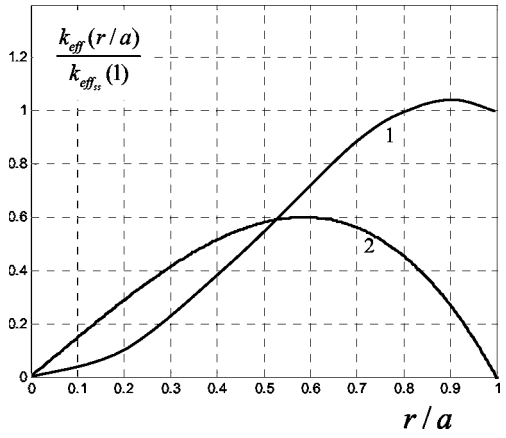


FIG. 9. The effect of reduction of electrodes on the coupling coefficients of a circular plate in the case of the simply supported (1) and clamped (2) boundary conditions.

"a,a" in the equivalent circuit has to be zero, $Z_{a,a}(M_f/M_{\rm pl})=0$. The values for the relative radius of the nodal line, $r_{\rm nl}/a$, can be found using the relation

$$\xi(r_{\rm nl}/a)=\xi_f(M_f/M_{\rm pl})+\xi_0(M_f/M_{\rm pl})\theta(r_{\rm nl}/a)=0,$$

where the ratio ξ_f/ξ_0 may be determined from Eq. (60a). In both cases the calculations must be completed at $F_{\rm eqvf}=0$ and $F_{\rm eqv0}=0$, as the free vibrations of the plate are considered. After some manipulations the following equations to calculate the resonant frequency and the nodal circle radius may be obtained:

$$f_{\rm fm}=f_{\rm ss}\sqrt{\frac{(M_f/M_{\rm pl})+1}{(M_f/M_{\rm pl})+0.3}}, \tag{61}$$

$$\theta(r_{\rm nl}/a)=\frac{0.46}{(M_f/M_{\rm pl})+1}. \tag{62}$$

The function $\theta(r/a)$ in Eq. (62) is expressed by formula (53). For the plate with the free boundary (at $M_f=0$) from Eqs. (61) and (62) will be found $f_{\rm fm}=1.82f_{\rm ss}$ and $r_{\rm nl}=0.69a$. These are nearly the exact values predicted for these quantities. The plots of the relative resonant frequency and the nodal line radius vs the relative mass of foundation are shown in Fig. 8.

In terms of designing practical flexural plate transducers, it is also instructive to consider the plate having a clamped boundary, i.e., under the condition that $(d\theta/dr)=0$ at $r=a$. The static deflection curve in this case is[5]

$$\theta\left(\frac{r}{a}\right)=\left(1-\frac{r^2}{a^2}\right)^2, \tag{63}$$

and the equivalent parameters calculated from relations (46), (48), and (11) are

$$K_m^E=\frac{67}{a^2}\frac{t^3}{12s_{11}^E[1-(\sigma^E)^2]},\quad S_{\rm eff}=0.2\pi a^2,\quad S_{\rm av}=0.33\pi a^2. \tag{64}$$

The electromechanical transformation coefficient, n, in the case where the plates are fully electroded, appears to be zero, as it follows from the boundary conditions and correspondingly from formula (51) at $r_1=a$. Physically this fact is explained in Ref. 7, and it is due to the changing of the sign of the charge density on either side of the plate at radius $r_1=a/\sqrt{2}$. However, if the electrodes are divided concentrically in two parts at this nodal circle and connected in opposite phases, then we will obtain from formula (52)

$$n=2\pi\frac{d_{31}t}{s_{11}^E+s_{12}^E}. \tag{65}$$

Even in this ideal case the effective coupling coefficient for the clamped plate, $k_{\rm eff.cl}$, is still smaller than for the simply supported plate, $k_{\rm eff.ss}$, namely, $k_{\rm eff.cl}=0.77\ k_{\rm eff.ss}$. Also of note, in the real transducer designs it is unlikely to have an ideally clamped boundary conditions; therefore, the dividing of the electrodes at the expected nodal location may not be considered as a very common or practical solution. It is always desirable to achieve the simply supported conditions. However, in reality the actual boundary conditions will be somewhere in between these limits. In this case, in order to avoid the compromising effect of the clamped boundary it is often desirable to reduce the diameter of the electrodes. The comparative effect of reducing the diameter of the electrodes on the effective coupling coefficient of the circular plate transducer in the case of the ideally supported and ideally clamped boundary is shown in Fig. 9, in which the coupling coefficients normalized to its value for the fully electroded simply supported plate are represented for both cases. The calculations were done as described in Ref. 7.

V. ON THE ESTIMATION OF THE ACCURACY OF CALCULATION OF THE TRANSDUCER PARAMETERS

A number of simplifying approximations were suggested in the course of the application of the energy method to the analysis of the circular flexural disk transducer in order to make the solution more physically clear and to obtain usable results in an analytically closed form. The question arises how to determine the level of accuracy achieved with the present analysis.

First, we estimate the errors of calculations of the transducer parameters, which result from substituting the approximate Poisson's ratio of $\sigma=0.3$ for their actual values under the integral in expression (46) directly and in formula (53) for the displacement curve $\theta(r/a)$. Expressions (11), (46), (48), and (51), from which the equivalent parameters are determined, all include the factors depending on the mode shape, $\theta(r/a)$. If to denote these factors as $X_i(\sigma)$, then the relative errors arising due to changing the value of σ may be calculated as

$$\frac{X_i(\sigma^E)-X_i(0.3)}{X_i(\sigma^E)} \doteq \frac{X_i'(0.3)}{X_i(\sigma^E)}(\sigma^E-0.3). \tag{66}$$

In the case where $0.25<\sigma^E<0.35$, the results of calculations show that these errors do not exceed 2.4% for K_m^E and they are less than 0.7% for S_{eff}, less than 0.2% for S_{av}, and less than 0.9% for n.

The assumption that the thickness of the plate is much smaller than its diameter permitted the application of the theory of thin plates, which significantly simplified the calculation of transducer parameters. But, there are no quantitative criteria on how small the t/a ratio should be and how the magnitude of this ratio is related to the accuracy of the results. The energy approach makes it relatively easy to estimate the corrections, which can be made to the elementary solution for the thin plate due to its finite thickness, by including the additional kinetic and potential energies accounting for the effects of rotary inertia and shearing deformations.[8] This is demonstrated in Ref. 9, where it is shown, in particular, that for the simply supported circular plates of finite thickness the equivalent mass,M', stiffness, K', and the transformation coefficient, n', can be determined to the second-order approximation as

$$M' \doteq M_{\text{eqvo}}(1+0.5t^2/a^2), \quad K' \doteq K_{\text{mo}}^E\left(1-0.75\frac{t^2}{a^2}\right),$$
$$n' \doteq n(1-0.75t^2/a^2). \tag{67}$$

The acceptability of the above computational inaccuracies becomes clear, if we take into account other limitations of accuracy encountered in the transducer's design and development. Thus, the comparison of parameters of transducers with simply supported and clamped boundaries considered in Sec. IV shows that the actual mechanical boundary conditions have the most critical influence upon the electromechanical properties of the circular plate transducer. However, in a real transducer design the boundary conditions are not known beforehand with great accuracy. It is hardly worth applying significant efforts to calculate properties of the plate-supporting structures theoretically, as it is much simpler and more reliable to get the actual results in this direction in the course of the transducer prototyping. Moreover, the prototyping is usually needed anyway, because the properties of materials used in the transducer's design and possible variation of the transducer parameters as a result of manufacturing are not known precisely prior to theoretical analysis. Another limitation of the accuracy of calculation of acoustic-related parameters of transducers, such as the radiation impedance and diffraction constants, is that the actual acoustic boundary conditions usually are not familiar. The diffraction effects for the finite-size elastic bodies, i.e., real transducers of various shapes and attachments, often cannot be evaluated with great accuracy, especially in the case where the transducers operate as members of arrays. In addition to the above, the accuracy of the measurement of the transducer acoustical parameters does not insure their very precise estimation (commonly it is considered to be about ±1 dB).

Thus, it seems quite appropriate to accept a reasonable level of accuracy in the theoretical analysis in order to achieve physically clearer analytical results that assist with the intuitive design process and help with the understanding of the technology and modifications necessary to improve the transducer performance.

VI. CONCLUDING REMARKS

In the previous paper[1] some of the advantages of the energy approach, in comparison with the partial differential equations approach, for solving the piezoelectric transducer problems were considered, and among them the relative simplicity and clarity of the results were obtained. One of the reasons for the relative simplicity of the method may be qualitatively explained with reference to Fig. 5. When employing the energy method the solution can be restricted beforehand to a particular frequency range of operation and, correspondingly, to a particular mode of vibration, which may be acquired from the solution of the purely mechanical problem by means of the theory of mechanical vibrations. When employing the partial differential equations approach, one deals with all the contributing generalized coordinates illustrated in Fig. 5 regardless of their significance, and they involve all aspects of the problem including the electromechanical conversion and the radiation or reception of sound. This complicates the solution to the problem without contributing practically to the final result. The simplifying assumptions made in the course of application of the energy method give rise to some level of approximation in calculation of the transducer parameters. But, the accuracy of these approximations can be easily estimated in the framework of the energy approach, as was shown by the example of the circular plate transducer. Usually this level of accuracy may be considered acceptable in comparison with other uncertainties encountered in transducer design and development such as real mechanical and acoustical boundary conditions, actual properties of the materials used, variations in manufacturing and measurement conditions.

ACKNOWLEDGMENTS

The author wishes to thank Dr. David A. Brown for his cooperation and assistance in revising, preparing, and editing the paper for publication. This work was supported in part by ONR 321SS Lindberg and BTech Acoustics.

[1] B. S. Aronov, "The energy method for analyzing the piezoelectric electroacoustic transducers," J. Acoust. Soc. Am. **117**, 210–220 (2005).
[2] J. R. Bobber, "Diffraction constants of transducers," J. Acoust. Soc. Am. **37**, 591–595 (1965).

[3]B. S. Aronov, "Energy analysis of a piezoelectric body under nonuniform deformation," J. Acoust. Soc. Am. **113**, 2638–2646 (2003).
[4]F. J. Rosato, "Lagrange equations applied to flexural mode transducers," J. Acoust. Soc. Am. **57**, 1397–1401 (1975).
[5]S. P. Timoshenko, *Strength of Materials*, 3rd ed. (Van Nostrand, Princeton, NJ, 1955), Vol. 2.
[6]For more details on the diffraction coefficients, see the paper by R. S. Woollett, "Diffraction constants for pressure gradient transducers," J. Acoust. Soc. Am. **72**(4), 1105–1113 (1982).
[7]B. S. Aronov, "On the optimizaton of the effective coupling coefficients of a piezoelectric body," J. Acoust. Soc. Am. **114**(2), 792–800 (2003).
[8]S. P. Timoshenko and D. H. Young, *Vibration Problems in Engineering*, 3rd ed. (Van Nostrand, New York, 1955).
[9]B. S. Aronov and L. B. Nikitin, "Calculation of the flexural modes of piezoceramic plates," Sov. Phys. Acoust. **27**(5), 382–387 (1981).

Field impact insulation testing: Inadequacy of existing normalization methods and proposal for new ratings analogous to those for airborne noise reduction

John J. LoVerde and D. Wayland Dong
Veneklasen Associates, Inc., 1711 Sixteenth Street, Santa Monica, California 90404

(Received 29 March 2004; revised 9 April 2005; accepted 12 May 2005)

The field test method for determining the Field Impact Isolation Class (FIIC) rating of a floor/ceiling assembly, prescribed in ASTM E 1007, requires an estimation of the receiving room absorption and the normalization of the receiving room sound pressure levels based on a standard quantity of room absorption. Normalization is intended to remove the effects of receiving room acoustical characteristics, but an analysis of field impact testing indicates that this method has a limited engineering value. The normalization correction is strongly dependent on receiving room volume, does not correlate with the normalization used in airborne sound isolation testing, and can be unreasonably large under certain conditions. The current test method leads to potential errors in engineering judgment, as illustrated by example field tests. By contrast, airborne noise reduction ratings (prescribed in ASTM E 336) are either non-normalized or normalized to a standard reverberation time, which provides significant advantages over normalizing to a standard amount of absorption. New impact noise metrics, the non-normalized impact sound rating (ISR) and the normalized impact sound rating (NISR), analogous to the airborne noise reduction metrics NIC and NNIC, are proposed for incorporation into ASTM E 1007. Revised building code requirements using the proposed metrics are suggested. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1946267]

PACS number(s): 43.55.Ti, 43.55.Vj, 43.50.Pn, 43.50.Jh [NX] Pages: 638–646

I. INTRODUCTION

The American Society for Testing and Materials (ASTM) standards include test methods for field measurements of the airborne and impact sound insulation provided by building elements. These field tests involve measuring the sound pressure level in the receiving room while a noise source or tapping machine is active in the source room. The sound field in the receiving room is assumed to be diffuse and reverberant and therefore dependent on the amount of absorption in the receiving room. A receiving room with carpet, drapery, furniture, and other sound absorbing elements will have lower levels and hence better ratings than an otherwise identical unfurnished receiving room. In order to remove the variable of receiving room absorption and increase the comparability of the test results, the both airborne and impact standards include procedures to normalize the receiving room levels to those that would be measured in a standard receiving room. There have historically been two procedures for this normalization. The first procedure is to normalize to a standard amount of absorption; the second procedure is to normalize to a standard reverberation time. Both procedures have been adopted by governing bodies throughout the world,[1] and the current ASTM standards includes the latter procedure within the airborne noise field test standard[2] and the former within the impact noise field test standard.[3]

Since the introduction of these standards, it has been assumed both implicitly and explicitly that the two normalization methods yielded substantially similar results for typical receiving rooms.[4] However, a number of field impact noise tests with peculiar results appear to indicate otherwise, which prompted the authors to undertake a systematic analysis of the normalization methods and the effects of the choice of normalization method on field test results. In this paper we present our analysis, which indicates that in many cases the choice of normalization method can cause surprisingly large differences in reported test results. In this paper we will show that the normalization to reverberation time is highly preferred to normalization to absorption.

New metrics and the consequent modification of the ASTM standards are proposed as a solution to the limitations described. Since one important application of the standards is reference to them within building codes and other legally binding documents, we also discuss the use of standards in building codes, and recommend changes to improve the clarity of the codes.

II. ANALYSIS OF NORMALIZATION PROCEDURES

A. General comparison

The normalization of receiving room sound pressure levels to a standard reverberation time T_0 is calculated by

$$L'_{RT} = L + 10 \log\left(\frac{T_0}{T}\right), \tag{1}$$

where L is the non-normalized (measured) receiving room sound pressure level, L'_{RT} is the receiving room sound pressure level normalized to T_0, and T is the reverberation time in seconds. The value of T_0 is 0.5 s in ASTM E336 and ISO 140. The reverberation time is calculated from

 0001-4966/2005/118(2)/638/9/$22.50

the measured decay rate of the receiving room, d, in decibels per second, since $T=60/d$ seconds. The decay rate or reverberation time is the only parameter required for this normalization procedure.

The normalization to a standard amount of absorption A_0 is calculated by

$$L'_{ABS} = L + 10\log\left(\frac{A}{A_0}\right), \tag{2}$$

where L refers to the non-normalized receiving room sound pressure level, L'_{ABS} is the receiving room sound pressure level normalized to A_0, and A is the receiving room absorption. The value of A_0 is 10 m^2 or 107.6 sabins in ASTM E 1007 and ISO 140. The reverberation time, T, of the receiving room is determined by measuring the decay rate, as above, and the receiving room absorption is calculated from the Sabine equation,

$$A = \frac{4V}{cT}\ln 10^6, \tag{3}$$

where V is the volume of the room and c is the speed of sound. Compared to Eq. (1), normalization to a standard absorption requires the additional measurement of the receiving room volume and temperature (needed to calculate c).

Equations (1) and (2) apply to the measurement of impact sound pressure levels. For airborne sound isolation testing, Eqs. (1) and (2) are modified by replacing the receiving room sound pressure level with the noise reduction (the difference in sound pressure levels between source and receiving rooms) and reversing the sign of the normalization term. The sign reversal is necessary because better insulation results in lower impact levels but higher noise reductions. To be as general as possible, the authors have chosen to define "positive" normalization adjustments as an adjustment that will improve the resultant rating; we do not mean that the absolute sign of the correction is positive. Where it is necessary to compare airborne and impact tests, we use only the single number ratings defined in the ASTM standards, where higher numbers imply better insulation for both airborne and impact testing.

The difference in the normalization methods is found by subtracting Eq. (1) from Eq. (2), and then substituting Eq. (3) for A, yielding

$$L'_{ABS} - L'_{RT} = 10\log\left(\frac{A}{A_0}\frac{T}{T_0}\right) \tag{4}$$

$$= 10\log\left(\frac{4\ln 10^6}{cA_0T_0}\right) + 10\log V. \tag{5}$$

As the first term in Eq. (5) is constant, the difference in normalized levels varies directly with the common logarithm of the receiving room volume. For V in cubic feet, c at 20 °C (=1126 ft per second) and the typical values of A_0 =107.6 square feet and T_0=0.5 s, Eq. (5) is

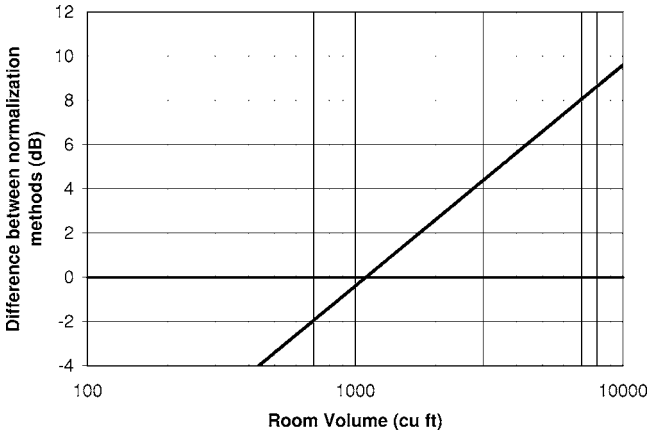


FIG. 1. The difference between the two methods of normalizing the receiving room sound pressure levels, Eq. (6), in the text. For a receiving room of volume V, sound pressure levels normalized to a standard amount of absorption (108 sabins) will exceed the levels normalized to a standard reverberation time (0.5 s) by the amount shown. For impact noise insulation testing, higher sound levels result in lower ratings. If normalizing to absorption, rooms greater than 1100 cubic feet are penalized, and rooms under 1100 cubic feet are benefited, compared to normalizing to reverberation time. It has long been assumed that most receiving rooms are sufficiently close to 1100 cubic feet in volume so that the normalization method is inconsequential; however, this is not the case (after Ref. 1).

$$L'_{ABS} - L'_{RT} = 10\log V - 30.4 = 10\log\left(\frac{V}{1096}\right). \tag{6}$$

The relationship is graphed in Fig. 1 (after Ref. 1). The difference vanishes for a receiving room volume of about 1100 cubic feet. For receiving rooms larger than this value, normalizing to absorption will yield higher normalized impact sound pressure levels than normalizing to reverberation time, and *vice versa* for rooms smaller than 1100 cubic feet.

B. ASTM standards

1. Airborne noise insulation field tests

ASTM E 336 defines three field test metrics, shown in Table I. Noise Isolation Class (NIC) is calculated from the noise reduction between two spaces, including all flanking paths and regardless of receiving room absorption, and is non-normalized. Normalized Noise Isolation Class (NNIC) is the same as NIC, except that the measured noise reduction is normalized to a reverberation time of 0.5 s in the receiving room per Eq. (1). The standard indicates that 0.5 s is the typical reverberation time when a space is "ordinarily furnished" for occupancy,[5] so that NNIC presents what the NIC would be if the receiving room were ordinarily furnished. Studies have confirmed that residential spaces have about a 0.5 s reverberation time, largely independent of frequency and room volume.[6]

Field sound transmission class (FSTC) is calculated from field sound transmission loss, which requires measurements of the receiving room absorption, radiating area of the assembly, and evaluation and control of flanking paths. This calculation uses a term $10\log(S/A)$, where S is the radiating area of the assembly being tested. This term is not (in our view) a normalization procedure in the sense of Eqs. (1) and (2). It is part of the definition of the transmission loss coef-

TABLE I. Field test metrics defined by ASTM E 336 and E 1007. The intended use of the three airborne noise insulation metrics is well defined in the standard. By contrast, the intent of the single impact insulation metric is less clear.

Airborne noise insulation	Impact noise insulation	Intent of standard
NIC		Actual noise isolation between rooms including multiple acoustic paths
NNIC		Noise isolation that would exist between ordinarily furnished rooms, including multiple acoustic paths
FSTC		Demonstrate sound attenuation provided by a construction is comparable to laboratory rating Evaluate field acoustical performance of nominally similar specimens
	FIIC	A comparison of closely similar assemblies, including multiple structure-borne paths Not generally comparable to laboratory rating

ficient, and is not referred to as a receiving room normalization nor used for any other metric. (The FSTC procedure for evaluating flanking paths is cumbersome and rarely performed. The apparent sound transmission class is sometimes measured, which is calculated in the same way as FSTC but includes any flanking paths. The radiating area is often taken to be one of the walls or ceiling of the receiving room common with the source room, although it is not clear how meaningful it is to assign a radiating area to an assembly where significant flanking may be present. In this case, S will be of the order $V^{2/3}$ for normally proportioned rooms, and the term $10\log(S/A)$ reduces to Eq. (1) with a value for T_0 that varies with the one-third power of the volume and is 0.4–0.8 s for receiving room volumes from 500 to 5000 cubic feet. Therefore, in most conditions the result of the $10\log(S/A)$ calculation will be similar to normalizing to a reverberation time of 0.5 s, although the rationale is very different.)

2. Impact noise insulation field tests

ASTM E 1007 defines a sole metric, Field Impact Insulation Class (FIIC), shown in Table I. The impact sound pressure levels are normalized to a standard receiving absorption of 10 m^2 or 107.6 sabins. The ASTM standard does not indicate the reasoning behind the adoption of this normalization procedure.

C. Problems with normalizing to absorption

From an inspection of Eqs. (1) and (2) and the discussion above, normalization to reverberation time seems preferable for several reasons. Defining the standard reverberation time T_0 as the typical reverberation time of a furnished residential room is intuitive and practical, and the value $T_0 = 0.5$ s is well supported by studies.[5,6] By contrast, there is no indication in the literature of the rationale behind the chosen value for A_0. Normalization to reverberation time requires only the measurement of the receiving room decay rate (the reciprocal of the reverberation time), which is a directly measurable parameter of the room acoustics. Using other descriptors of the room acoustics such as absorption require additional assumptions such as Eq. (3) and the measurement of additional parameters such as room volume and temperature, which introduce unnecessary additional sources of error.

More importantly, however, an analysis of Eqs. (1)–(6) and our field testing experience indicate normalizing to absorption instead of reverberation time creates potential problems that are substantive and affect our engineering judgment in the design of assemblies. These are examined below, and illustrated with examples from actual field tests performed recently by the authors. The illustrated cases do not represent anomalies, but were chosen to be representative of a sizable fraction of our field testing experience.

1. Determination of receiving room volume

As described above, if normalizing to a standard absorption, the room volume must be estimated to calculate the absorption using Eq. (3). For a given measurement of L and T, the resultant difference $\Delta L'_{ABS}$ in the normalized level between a measured volume $V+\Delta V$ and a measured volume V is, from Eqs. (2) and (3),

$$\Delta L'_{ABS} = 10\log(V+\Delta V) - 10\log V = 10\log\left(1+\frac{\Delta V}{V}\right). \tag{7}$$

For example, a 10% uncertainty in estimating the receiving room volume will result in a 0.4 dB uncertainty in the normalized level. Therefore, volume uncertainties typically will not cause a large change in the test result.[7]

Even though high precision in the measurement of room volume is not necessary, *defining* the volume of the receiving room in the field is a nontrivial exercise that often requires the judgment of the test engineer. Many floor plans consist of a living room, dining room, kitchen, and entry, open to each other to varying degrees. The receiving room is not well defined in this condition. We recently witnessed an impact test where the hard surface floor was installed within only

the kitchen of a large combined space, as described above. One test engineer estimated the entire combined space as the receiving room volume, and a different engineer used only the volume directly below the hard surface area. Both interpretations are consistent with the intent of the ASTM standard. The estimated volumes were different by a factor of 3, resulting in a difference of 5 dB from Eq. (7) and a large discrepancy in the reported results. One could argue that the standard should better define the volume, but a better solution is to normalize to the reverberation time instead of absorption. This eliminates the measurements and computational steps related to the volume and reduces the subjective judgments required of the test engineers.

It is true that in an ambiguous space it is not clear where to measure the room decays, but this is less problematic in part because normalization to reverberation time is less sensitive to room volume than normalization to absorption. This is discussed in detail below.

2. Strong dependence on receiving room volume

It is evident from Eq. (6) and Fig. 1 that for receiving rooms larger than 1100 cubic feet, normalizing to a fixed amount of absorption penalizes the test assembly (i.e., reduces the reported rating), requiring greater isolation than for the same assembly in a smaller room. Although this has been evident for at least 40 years,[8] the prevailing assumption has been that such "large rooms" are not generally encountered in typical residential impact test situations. For example, it has been stated that "the typical range of room volumes encountered in multifamily dwellings" is between about 775 and 2100 cubic feet.[1] In receiving room volumes in this range, the difference between the two types of normalization ranges from -1.5 to $+2.8$ dB, which is "no greater than the uncertainty of typical field measurements."[1]

In our experience, however, typical receiving room volumes encompass a much greater range than 775–2100 cubic feet. In impact field testing performed in buildings across the country by Veneklasen Associates, Inc., and Western Electro-Acoustic Laboratory during 2003, approximately 40% of the receiving rooms exceeded 2100 cubic feet, while about 15% were below 775 cubic feet. In the current architectural design, it is common to see floor plans where the living room, dining room, kitchen, and entry of a multifamily residential unit are one combined space with volume typically much greater than 2100 cubic feet. Loft-type residences have become popular, where almost the entire habitable area is a single acoustic space. On the other extreme, bathrooms are often very small (less than 775 cubic feet). Although we have not performed a nationwide survey of residential spaces, our tests include units in many parts of the country and at many different market levels. With so many of the receiving rooms we tested having volumes outside of the 775–2100 cubic foot range, there is strong reason to doubt that this range encompasses "typical" receiving rooms.

Further, ASTM E 1007 indicates that the receiving room volume should ideally be greater than 2100 cubic feet to create an approximately diffuse sound field at all required frequencies. However, any increased accuracy due to a more diffuse low-frequency field is more than offset by the minimum 2.8 dB reduction in measured isolation caused by normalizing to absorption rather than reverberation time (see Fig. 1). In other words, for large spaces, the choice of normalization becomes very important and cannot be dismissed as an error of similar order to ordinary experimental uncertainties.

It may be objected that the measurement is not appropriate for all receiving spaces, especially in very small rooms where the sound field is not diffuse at low frequencies. However, many such rooms exist and are required to meet standards set by regulatory agencies, legal entities, and homeowners associations, which do not exempt a floor from testing simply because of the volume of the room. Regardless of the room volume, tests will continue to be performed in all types of habitable spaces.

It is apparent from inspection that the reason for this penalization of large rooms [the dependence of Eq. (6) on room volume] is that a fixed amount of absorption will create a specific reverberant field only for a specific room volume. If the volume is allowed to vary over a wide range, as in field testing, this amount of absorption can become highly inappropriate. This can be made more explicit by examining the differences in the normalization method in terms of the average sound absorption coefficient for the room, $\bar{\alpha}$, defined by the total absorption in the room divided by the total surface area S:

$$\bar{\alpha} = \frac{A}{S}. \tag{8}$$

Consider a normally shaped rectangular room with proportions 1:1.5:2, yielding a total surface area of about[9]

$$S = 6.25V^{2/3}. \tag{9}$$

Then the average sound absorption coefficient required to meet the standard amount of absorption is, from Eqs. (8) and (9),

$$\bar{\alpha}_{ABS} = \frac{A_0}{6.25}V^{-2/3}, \tag{10}$$

where the *ABS* in the subscript indicates normalization to standard absorption. The average sound absorption coefficient required to achieve the standard reverberation time, using Eqs. (3), (8), and (9), is

$$\bar{\alpha}_{RT} = \frac{4\ln 10^6}{6.25cT_0}V^{1/3}, \tag{11}$$

where the *RT* in the subscript indicates normalization to standard reverberation time.

It is clear from the powers of V in Eqs. (10) and (11) that the normalization to absorption is much more sensitive to receiving room volume than normalization to reverberation time. Also, the sign of the exponent is opposite between the two methods. The result of these differences may be seen more easily when Eqs. (10) and (11) are plotted. See Fig. 2.

To interpret this graph, we can plot any given receiving room by its volume and average absorption coefficient. (All of the rooms we examine here were empty rectangular rooms so the calculation of total surface area and hence $\bar{\alpha}$ was

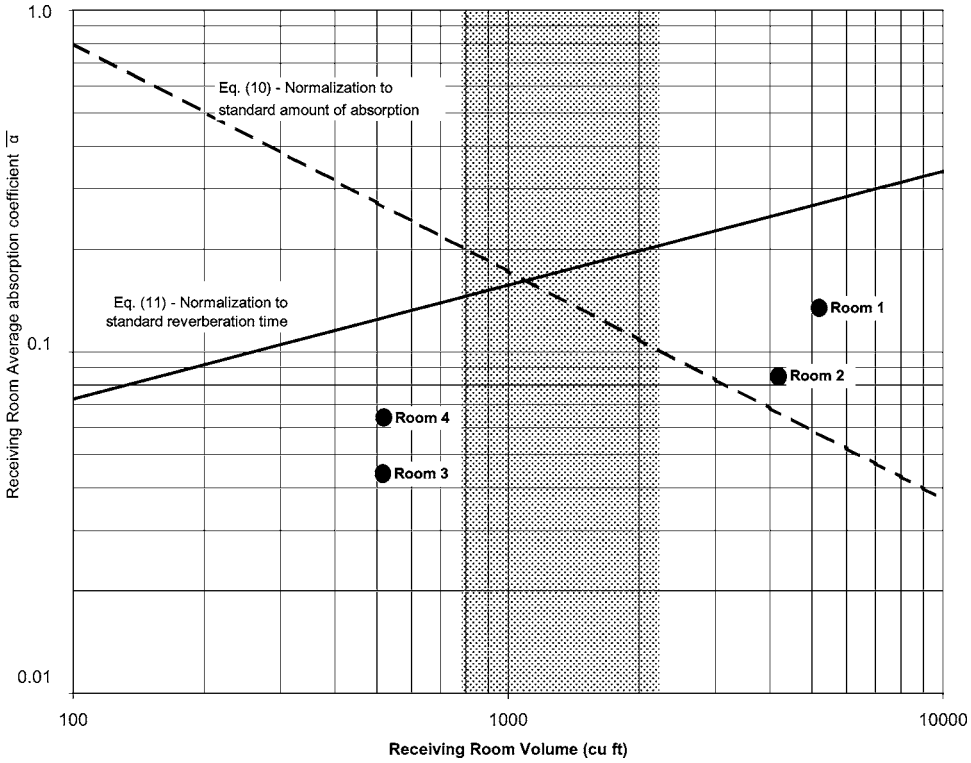


FIG. 2. The average absorption coefficient $\bar{\alpha}$ for a receiving room plotted against its volume. The two normalization procedures are solved for $\bar{\alpha}$ in Eqs. (10) (dotted line) and (11) (solid line), assuming an empty rectangular room of typical proportions. Receiving rooms are plotted by their volumes and absorption coefficients. Receiving rooms above the curve for a normalization procedure have more absorption than the standard and are negatively corrected by the normalization, and the opposite for rooms below the curve. Receiving room volumes have historically been assumed to fall within 775–2100 cubic feet; the four receiving rooms presented in the text are indicated by the circled numerals and fall far outside this region. Large rooms such as 1 and 2 are between the curves and experience different corrections, depending on the choice of the normalization procedure. The steep negative slope of normalization to absorption, Eq. (9), has many unfortunate consequences. Rooms with similar acoustics such as 2 and 4 are not only normalized in different directions (solely because of differing volumes), but rooms such as 4 are so far below the curve that the magnitude of the correction is excessive. In this case, the correction for Room 4 is +7, while that for Room 2 is −2. A normalization procedure that results in a 9 dB difference in correction for rooms with similar acoustics is clearly inadequate. It misrepresents assemblies, prevents accurate rank ordering, hampers design, and causes much needless confusion.

straightforward. Obviously it is not practical to measure total surface area, in general, for field tests, especially in furnished rooms.) Any such receiving room point that lies above the normalization line has more absorption than the normalization calls for and is negatively adjusted, and the opposite for receiving rooms that lie below the line.

The slight positive slope of the curve of Eq. (11) is expected because the volume-to-surface area ratio of a room of given shape increases with volume. All other things being equal, a larger room will require surfaces with higher average absorption to achieve the standard reverberation time. In contrast, the curve of Eq. (10) has the opposite sign and twice the slope. It requires very low average absorption ($\bar{\alpha} < 0.08$) for large rooms to meet the standard but allows small rooms to have large amounts of absorption ($\bar{\alpha} > 0.25$). In other words, while the standard reverberation time maintains a consistent acoustic character over a range of room volumes, the standard amount of absorption corresponds to very dead small rooms and very live large rooms, which may or may not occur in the field.

3. Direction of normalization correction

For any receiving room plotted in Fig. 2 that lies between the two curves, the two normalization procedures will adjust the levels in opposite directions. This is not just an academic concern, as the current ASTM standards require normalization to absorption for impact testing and normalization to the reverberation time (or not at all) for airborne testing. One would expect that the normalization for airborne and impact testing into the same receiving room would be similar, as the purpose of normalization (to remove the effects of the receiving room absorption from the results) is the same for both tests.

Because the line for normalization to absorption in Fig. 2 runs to very small values of $\bar{\alpha}$ for large rooms, large rooms will almost always be between the curves and hence experience this problem. Consider the following test case in which airborne and impact testing was performed on the floor/ceiling assembly between loft-type residences. The receiving room was two stories high with an estimated volume of 5500 cubic ft (156 m^3). The room was unfinished, with a measured T60 of 1.1 s at 500 Hz and a resultant $\bar{\alpha}=0.13$. This point is shown as "Room 1" in Fig. 2. The floor/ceiling system was an exposed concrete slab. The results are shown in Table II. The FIIC rating calculated using non-normalized levels was 23. When using levels normalized to a standard reverberation time per Eq. (1), the rating was 25. When using levels normalized to a standard amount of absorption per Eq. (2), as defined in the standard, the rating was 18.

Airborne tests were also conducted on this assembly.

TABLE II. Field test results for loft receiving room calculated using various normalization methods.

	Normalization method	Standard	Loft
Impact	Non-normalized	None	23
	Normalized to 108 sabins	FIIC per ASTM E 1007, 989	18
	Normalized to 0.5 s	None	25
Airborne	Non-normalized	NIC per ASTM E 336, 413	46
	Normalized to 0.5 s	NNIC per ASTM E 336, 413	49

The non-normalized NIC rating was 46, while the normalized NNIC rating was 49. The receiving room was the same as for the impact tests.

Normalization to a fixed amount of absorption (the FIIC test) resulted in a five-point reduction from the non-normalized value, while normalization to a standard reverberation time (the NNIC test) resulted in a three-point *increase*. This situation where the normalization is in opposite directions for impact and airborne testing with the same receiving room is counterintuitive and creates confusion among those who view the data. It would obviously be preferable for normalization procedures to be consistent between airborne and impact tests with the same receiving room.

Even when using a single normalization method, the normalization to absorption can result in corrections to different directions, depending on the volume of the receiving room, even when the rooms have similar reverberation characteristics. Consider the following set of three tests performed on three different floor/ceiling assemblies within the same apartment unit. The assemblies were installed in the living room, master bathroom, and second bathroom of the apartment. All of the rooms were unfurnished. The volumes and reverberation times of the three spaces and the results of FIIC tests are shown in Table III, and graphed in Fig. 2 as Rooms 2, 3, and 4, respectively. Note that the average absorption coefficients are similar between the rooms, even though the volumes are very different. This is as expected, because all three rooms had similar finishes (gypsum board, concrete, and glass) and were unfurnished. Room 2 contained some miscellaneous construction materials, where Rooms 3 and 4 were empty.

Normalizing to reverberation time results in a small positive increase in rating of two to three points for all three rooms, which is as expected, because the rooms presented similar acoustical environments. Normalizing to absorption, in contrast, results in a *decrease* in rating for the living room (2) but an *increase* for the baths (3 and 4), solely because of the differing volumes. Looking at Fig. 2, it is evident that the flatter the curve of a normalization procedure, the more consistent the results will be between similarly reverberant spaces with differing volumes. The steep slope of the curve of Eq. (9) cuts between the room points and results in a different direction of the normalization corrections, which is counterintuitive and needlessly confusing.

TABLE III. FIIC values calculated using various normalization methods as measured in multiple receiving rooms. The normalization correction using a standard amount of absorption (−2 to +7 points) is large and in both directions, even though the receiving rooms have similar reverberation characteristics. By contrast, the normalization correction using a standard reverberation time is small and consistent (two to three points).

	Living room (2)	Master bath (3)	Second bath (4)
Volume	4235 cubic feet	490 cubic feet	500 cubic feet
RT	1.4 sec	1.4 s	1.0 s
$\bar{\alpha}$	0.08	0.04	0.06
Non-normalized	61	53	50
Normalized to 108 sabins	59	59	57
Normalized to 0.5 s	64	55	53

4. Magnitude of normalization correction

Perhaps more importantly than the counterintuitive direction of the normalization correction is the magnitude of the correction, which is excessive when using a standard amount of absorption. As shown in Table III, the correction ranges from −2 to +7 dB. This is further illustrated by Fig. 2. Because of the steep negative slope of Eq. (9), rooms 3 and 4 are much farther below this curve, and the resultant correction is much larger. None of these receiving rooms had an acoustical environment so removed from the ordinary furnished condition as to require a correction of this order. In any case, one would question the validity of extrapolating test results to an environment that was so different as to require a correction on the order of 7 dB! Even if such a large correction were necessary, rooms 3 and 4 had similar acoustics as room 2, which has a small correction in the *opposite* direction!

5. Assembly design engineering

A large part of the work of many acoustical consultants is the rank ordering of the various floor/ceiling assemblies under consideration for a project. Published laboratory testing of the proposed materials is a useful starting point, but because the building structure and construction are important components to the overall impact noise insulation, the evaluation of mockup assemblies installed in the building is often the best way to perform a meaningful comparison.

Such an evaluation was the purpose of the tests shown in Table III. Because the three receiving rooms had similar reverberant characteristics, the non-normalized levels should be quite useful for comparison purposes. One difference is that the large room (2) would have better diffusion at the low frequencies; however, in all three cases the levels were controlled at the mid-band frequencies, so this was not a factor. In the field, the assemblies had clearly differentiable impact insulation in terms of perceived loudness of the tapping noise levels. Subjectively, the impact noise in the living room was considerably less intrusive than either of the other two assemblies, while the impact noise levels in the master bath were slightly but noticeably quieter compared to that in the second bath. For that specific impact source and building, the three assemblies could be clearly ranked in terms of subjective impact insulation, (i.e., the protection afforded the occupants). The ranking of assemblies by both the non-normalized ratings and the ratings normalized to reverberation time per Eq. (1) are in agreement with subjec-

TABLE IV. Existing and proposed (boldface) metrics for field testing of airborne and impact noise insulation.

Airborne noise insulation	Impact noise insulation	Intent of standard
FSTC	FIIC	
NIC	**ISR**	Actual noise isolation or impact sound level between rooms including multiple acoustic paths
NNIC	**NISR**	Noise isolation or impact sound level that would exist between ordinarily furnished rooms, including multiple acoustic paths

tive observations. This is as expected, as all three rooms had similar reverberant characteristics, and this comparison is not distorted by normalization to reverberation time. As described above, in contrast, the normalization to absorption per Eq. (2) had corrections in different directions and of large magnitudes.

A consultant who viewed a published test report of these three tests (normalized to absorption) might reasonably conclude that all three assemblies would result in similar impact sound pressure levels and hence provide similar protection from impact noise to occupants. An inspection of the non-normalized levels or levels normalized to standard reverberation time, on the other hand, would accurately rank order the assemblies (for this source and this building) and provide a more accurate assessment. The normalization procedure is intended to remove the misleading effects of room absorption. In this case, the normalization itself generated results far more misleading than the use of non-normalized values.

D. Summary

It should be clear that normalization to a fixed amount of absorption has many undesirable characteristics compared to normalization to a standard reverberation time. It requires an estimation of the receiving room volume, an avoidable exercise that sometimes requires subjective judgment. Assemblies with "large" receiving rooms are penalized, where "large" refers to a volume ($>$2100 cubic feet) that appears to be encountered quite often in field testing. The direction of the correction is different between airborne and impact insulation tests on the same room, which is needlessly confusing. The direction of the correction is also different between receiving rooms of differing volumes, even when reverberant conditions are similar. The magnitude of the correction can be excessive. This large and variable correction can mask real differences between assemblies that are readily apparent using other metrics. For these reasons, the usefulness of normalizing field measurements to absorption is of a questionable value. All of these issues are avoided by normalizing to a standard reverberation time instead of a standard amount of absorption.

III. PROPOSED NEW METRICS

A. General

We have treated the two methods in Eqs. (1) and (2) as alternative procedures to normalize for the acoustical characteristics of the receiving room, which is the common view.[1,8] However, a different perspective is available by noting that if the sound field in the receiving room is assumed to be diffuse, the receiving room sound pressure level will depend only on the sound power transmitted into the room and the total room absorption. Therefore, normalizing to a standard amount of absorption gives a normalized power level radiated from the test specimen. This may be the appropriate procedure for laboratory testing, where flanking is insignificant, the receiving room sound field is highly diffuse, and the sample sizes are all approximately the same. In the field, however, these assumptions are often invalid, and it is not generally practical to make an accurate assessment of the sound power radiated from a specimen in the field.

This is reinforced by a comparison with the airborne noise testing metrics. Field transmission loss (from which FSTC is computed) is a measurement of the sound power level transmitted across the specimen, and includes the term 10 log(S/A). In impact testing it is implicitly assumed that the power radiated by the assembly is independent of its size, so that Eq. (2) is equivalent (up to a constant) to 10 log(S/A). In this view, Eq. (2) is not a normalization of the receiving room acoustical character as is Eqs. (1), but a method of calculating sound power, and the similarity between Eqs. (1) and (2) is superficial. Because the results of Eqs. (1) and (2) are similar for a certain range of receiving room volumes, both have been used to correct for the differing acoustical characteristics of receiving rooms, but from this perspective, only Eq. (1) is suitable for this purpose.

More importantly, perhaps, in most field impact noise situations, the sound power radiated by the specimen is not relevant. The more common goal is an attempt to quantify "impact sound insulation," or the protection from impact noise provided to the building occupants. The occupants' concern is obviously the actual impact sound pressure levels within the occupied space, regardless of radiated area, room volume, or flanking paths. Some measure of the actual impact sound pressure levels within the space is needed. (Doubts about the suitability of the standard tapping machine as the source for such an assessment are widespread, but this discussion is independent of the source used to generate the impact noise.) Again, by comparison, for airborne testing the NIC and NNIC metrics measure the actual airborne insulation (noise reduction) between spaces, regardless of radiated area, room volume, or flanking paths.

B. New metrics

To address these issues, we propose two new metrics for field impact measurements analogous to the NIC and NNIC

TABLE V. Impact and airborne isolation values calculated using various metrics. The proposed metrics are in boldface.

		Loft	Living room	Master bath	Second bath
Impact	FIIC	18	59	59	57
	ISR	23	61	53	50
	NISR	25	64	55	53
Airborne	NIC	46	--	--	--
	NNIC	49	--	--	--

metrics for airborne noise reduction field measurements, which are shown in Table IV. The measurement procedures would remain the same (as described in ASTM E 1007 and E 989), only the calculation and reporting would be affected. The first new metric is the Impact Sound Rating or ISR. This is a single number rating calculated as FIIC except derived from the non-normalized impact sound pressure levels. ISR would measure the actual impact sound pressure levels that result in the receiving space from a tapping machine on the source floor, including flanking paths and any other peculiarities of the assembly. As such, it would be a measure of the actual impact noise isolation of the assembly, just as NIC measures the actual amount of airborne sound isolation between any two spaces.

The second new metric is Normalized Impact Sound Rating (NISR), which is the same as ISR, except derived from the impact pressure levels normalized to 0.5 s reverberation time as in Eq. (1). NISR would therefore measure the impact isolation that would exist if the receiving rooms were "ordinarily furnished," similar to NNIC.

FIIC (AIIC, where the A stands for "apparent," has been proposed instead) would remain in the standard, and its purpose clarified to be used where comparisons are desired between nominally similar assemblies, supporting structures, and receiving room conditions, similar to FSTC. Table V shows the test cases described above with the new metrics, showing that the proposed metrics correct the problems of the FIIC metric raised in this paper.

These proposed impact noise metrics are comparable to those in ISO standards 717/II and 140/VII, which define the "normalized impact sound pressure level" similar to FIIC as currently defined and also a "standardized impact sound pressure level" similar to NISR introduced above.

IV. BUILDING CODES

The unfortunate situation exists where most legally binding documents such as building codes and condominium codes, covenants, and regulations (CC&R's) are less than clear as to the metric and procedure to be used for field testing. For example, the International Building Code (IBC) for multifamily dwellings, Sec. 1207 (sound transmission) sets forth minimum airborne and impact ratings when assemblies are field tested, but does not specify the metric or a standard. Many CC&R's reference inappropriate metrics or none at all. The acoustical engineer performing the test or the regulatory agent (rarely an expert in acoustics) must choose the metric and standard to use.

TABLE VI. Suggested metrics for field testing for various conditions. Currently, most building codes and condominium CC&R's reference an inappropriate metric or none at all in their standards. Such confusing and arbitrary conditions would be greatly reduced if the codes referenced the metrics below in their standards.

Condition	Airborne noise isolation	Impact noise isolation
Occupied	NIC	ISR
Unoccupied	NNIC	NISR
Comparison to laboratory or nominally similar assemblies	FSTC	FIIC

The California Building Code (CBC) is a notable exception. For airborne isolation, it references NIC for occupied units and NNIC for unoccupied units, which is in accordance with the purposes of the metrics, as described in the ASTM standards. For impact isolation, the CBC references FIIC, but with a modification to the ASTM standard that no normalization is carried out.[10] Hence, in California we have been using the ISR metric since 1988, although it has been still been called FIIC, adding to the confusion.

We suggest that the goal of most field measurements is the determination of the noise insulation afforded to the occupants, i.e., the actual noise reduction between spaces or impact sound pressure levels (using a standard impact source) in a space. Therefore, FSTC and FIIC are of limited use. For unfurnished receiving rooms, the use of Eq. (1) with 0.5 s for the standard reverberation time is reasonable (Refs. 5 and 6) for improving comparability between tests. Therefore, we suggest that the recommended metrics for field measurements are NIC and ISR for occupied units and NNIC and NISR for unoccupied units (see Table VI). It is of course preferable that these metrics be included in the building codes; at least, however, they should be defined in the ASTM standard so that it is clear to acoustical professionals that are the proper metrics for each condition.

V. CONCLUSIONS

Receiving room normalization can be useful for increasing the comparability of field test measurements by compensating for the effect of receiving room absorption. We contend that such normalization should be consistent between airborne and impact field testing. We have shown that normalization to a fixed amount of absorption as in the current field impact metric (FIIC) is inappropriate. We recommend that ASTM modify E1007 to adopt the new impact noise metrics ISR and NISR. Further, building codes and other legal requirements should specify ISR and NISR rather than FIIC for field impact noise tests. In the meantime, the codes remain subject to interpretation, and we suggest that the consultants choose the proper metric for the situation and educate the owners and regulatory agencies accordingly.

[1]T. J. Schultz, "Impact-noise recommendations for the FHA," J. Acoust. Soc. Am. **36**, 729–739 (1964).
[2]ASTM E 336-97e1, "Standard test method for measurement of airborne sound insulation in buildings," ASTM International.
[3]ASTM E 1007-04e1, "Standard test method for field measurement of tap-

ping machine impact sound transmission through floor–ceiling assemblies and associated support structures," ASTM International.
[4]For example, Ref. 1. A more detailed discussion is in Sec. II C 2.
[5]Reference 2, Sec. 3.2.2 and Note 1.
[6]J. S. Bradley, "Acoustical measurements in some Canadian homes," Can. Acoust. **14**, 14 (1986).
[7]Reference 2, Sec. 11.4, Note 12.
[8]O. Brandt, "European experience with sound-insulation requirements," J. Acoust. Soc. Am. **36**, 719–724 (1964).
[9]L. L. Beranek, *Acoustics* (Acoustical Society of America through the American Institute of Physics, Inc., Woodbury, NY, 1996), Chap. 10, pp. 315–316.
[10]Uniform Building Code with California Building Code modifications, Appendix Chap. 12, Division IIA, Sound Transmission Control.

Acoustic axes in triclinic anisotropy

Václav Vavryčuk[a)]
Geophysical Institute, Academy of Sciences, Boční II/1401, 141 31 Praha 4, Czech Republic

(Received 12 November 2004; revised 9 May 2005; accepted 23 May 2005)

Calculation of acoustic axes in triclinic elastic anisotropy is considerably more complicated than for anisotropy of higher symmetry. While one polynomial equation of the 6th order is solved in monoclinic anisotropy, we have to solve two coupled polynomial equations of the 6th order in two variables in triclinic anisotropy. Furthermore, some solutions of the equations are spurious and must be discarded. In this way we obtain 16 isolated acoustic axes, which can run in real or complex directions. The real/complex acoustic axes describe the propagation of homogeneous/inhomogeneous plane waves and are associated with a linear/elliptical polarization of waves in their vicinity. The most frequent number of real acoustic axes is 8 for strong triclinic anisotropy and 4 to 6 for weak triclinic anisotropy. Examples of anisotropy with no or 16 real acoustic axes are presented. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1954587]

PACS number(s): 43.20.Bi, 43.35.Gk [ANN] Pages: 647–653

I. INTRODUCTION

Acoustic axes (singularities, degeneracies) in anisotropic media are the directions in which phase velocities of two or three plane waves coincide.[1–4] These directions are important in studying acoustic, seismic, or electromagnetic waves, because they cause anomalies in the field of polarization vectors,[5,6] and they are associated with energy focusing due to caustics in the wave surface.[7–14] Acoustic axes also pose complications in tracing rays[15,16] and in wave field modeling because of coupling of waves.[17–22]

Acoustic axes form single isolated points on the slowness surface, or they join into lines. The isolated acoustic axes can exist in all anisotropy symmetries, the line acoustic axes typically occur in transverse isotropy. The maximum number of isolated acoustic axes is:[23–31] 16 in monoclinic, orthorhombic, and trigonal symmetry, 13 in tetragonal symmetry, 7 in cubic symmetry, and 1 in transverse isotropy. The directions of the acoustic axes in the mentioned symmetries are calculated by solving polynomial equations in one variable. The highest degree of the polynomials is 6 for monoclinic symmetry, or less for other symmetries, hence solving the polynomials numerically does not pose any difficulty. However, as regards triclinic anisotropy, the problem is more involved. Khatkevich[23] proved that the acoustic axes in triclinic anisotropy can be calculated by solving two polynomial equations of the 6th order in two variables. This implies that the number of acoustic axes in triclinic anisotropy cannot exceed 36. However, Khatkevich[23] did not discuss whether the actual maximum number of acoustic axes is less than 36 or not. Later on, Darinskii[28] proved that typical triclinic anisotropy (when $S1$ and $S2$ waves are degenerate) cannot have more than 16 acoustic axes, and Holm[27] proved that generic triclinic anisotropy (anisotropy with stable acoustic axes) also possess no more than 16 acoustic axes. Here it is proved that the maximum number of acoustic axes is 16 under no restrictions on triclinic anisotropy. Several approaches to determining the acoustic axes in triclinic anisotropy are presented and it is discussed which scheme is optimum for numerical calculations. Examples of triclinic anisotropy with no and 16 acoustic axes are presented, and the most frequent number of acoustic axes in triclinic media and its dependence on anisotropy strength is investigated.

II. THEORY

The Christoffel tensor $\mathbf{\Gamma}(\mathbf{n})$ is defined as[20,24,32]

$$\Gamma_{jk}(\mathbf{n}) = a_{ijkl} n_i n_l, \tag{1}$$

where a_{ijkl} are the density-normalized elastic parameters and $\mathbf{n}$ is the unit vector defining the slowness direction. The Einstein summation convention over repeated subscripts is applied. The elastic parameters a_{ijkl} must satisfy the stability conditions, if the medium is to be physically realizable:

$$a_{11} > 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} > 0,$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{vmatrix} > 0, \ldots, \det(a_{\alpha\beta}) > 0, \tag{2}$$

where the two-index Voigt notation $a_{\alpha\beta}$, $\alpha,\beta = 1,\ldots,6$ has been used [see Musgrave,[32] Eq. (3.13.4)]. These conditions correspond to the requirement that the strain energy of the medium must be positive [see Payton,[33] Eqs (1.4.3.) and (1.4.4)].

The Christoffel tensor $\mathbf{\Gamma}(\mathbf{n})$ has three eigenvalues $G^{(M)}$ and three unit eigenvectors $\mathbf{g}^{(M)}$, which are calculated from

$$\Gamma_{jk} g_k^{(M)} = G^{(M)} g_j^{(M)}, \quad M = 1,2,3, \tag{3}$$

where M denotes the type of the wave (P, $S1$, or $S2$). The eigenvalues correspond to the squared phase velocities, $G = c^2$, and the eigenvectors describe the polarization vectors of the waves.

Acoustic axes are defined as the directions in which two eigenvalues of the Christoffel tensor coincide

a)Electronic mail: vv@ig.cas.cz

$$G^{(1)}(\mathbf{n}) = G^{(2)}(\mathbf{n}) \neq G^{(3)}(\mathbf{n}), \tag{4}$$

and thus the Christoffel tensor is degenerate. Exceptionally, all three eigenvalues can coincide

$$G^{(1)}(\mathbf{n}) = G^{(2)}(\mathbf{n}) = G^{(3)}(\mathbf{n}), \tag{5}$$

but this type of the acoustic axis is unstable and very rare.

In the following, I will present three different approaches to determining acoustic axes in anisotropic media: the Fedorov approach[24] based on solving one 12th-order multivariate polynomial equation, the Khatkevich approach[23] based on solving two 6th-order multivariate polynomial equations, and the Darinskii approach[28] based on solving six multivariate quadratic equations.

A. The Fedorov equation

Calculating the eigenvalues G of the Christoffel tensor $\mathbf{\Gamma}(\mathbf{n})$ from

$$\det(\Gamma_{jk} - G\delta_{jk}) = 0, \tag{6}$$

we obtain the following cubic equation:

$$G^3 - PG^2 + QG - R = 0, \tag{7}$$

where P, Q, and R are invariants of $\mathbf{\Gamma}(\mathbf{n})$,

$$P = \Gamma_{11} + \Gamma_{22} + \Gamma_{33}, \tag{8}$$

$$Q = \Gamma_{11}\Gamma_{22} + \Gamma_{11}\Gamma_{33} + \Gamma_{22}\Gamma_{33} - \Gamma_{12}^2 - \Gamma_{13}^2 - \Gamma_{23}^2, \tag{9}$$

$$R = \Gamma_{11}\Gamma_{22}\Gamma_{33} + 2\Gamma_{12}\Gamma_{13}\Gamma_{23} - \Gamma_{11}\Gamma_{23}^2 - \Gamma_{22}\Gamma_{13}^2 - \Gamma_{33}\Gamma_{12}^2. \tag{10}$$

The cubic equation (7) has three real roots, of which at least two are equal, if [see Fedorov,[24] Eq. (18.18)]

$$4P^3R - P^2Q^2 - 18PQR + 4Q^3 + 27R^2 = 0. \tag{11}$$

Equation (11) is the 12th-order homogeneous polynomial equation in three unknowns n_1, n_2, and n_3. It has an infinite number of complex-valued solutions, but the number of real-valued solutions is finite.

B. The Darinskii equations

Using the spectral decomposition of $\mathbf{\Gamma}(\mathbf{n})$

$$\Gamma_{jk} = G^{(1)}g_j^{(1)}g_k^{(1)} + G^{(2)}g_j^{(2)}g_k^{(2)} + G^{(3)}g_j^{(3)}g_k^{(3)}, \tag{12}$$

and applying the condition for the acoustic axis, $G^{(2)}=G^{(3)}$, we obtain

$$\Gamma_{jk} = (G^{(1)} - G^{(2)})g_j^{(1)}g_k^{(1)} + G^{(2)}\delta_{jk}, \tag{13}$$

where δ_{jk} is the Kronecker delta and the following identity was used:

$$\delta_{jk} = g_j^{(1)}g_k^{(1)} + g_j^{(2)}g_k^{(2)} + g_j^{(3)}g_k^{(3)}. \tag{14}$$

If $G^{(1)} > G^{(2)}=G^{(3)}$, the $S1$ and $S2$ phase velocities coincide at the acoustic axis, if $G^{(1)} < G^{(2)}=G^{(3)}$, the P and $S1$ phase velocities coincide at the acoustic axis. Equation (13) can be expressed as follows [see Darinskii,[28] Eq. (4)]:

$$a_{ijkl}s_is_l = g_jg_k + \delta_{jk}, \tag{15}$$

where $\mathbf{s}=\mathbf{n}/\sqrt{G^{(2)}}$ is the slowness vector of the degenerate wave and $\mathbf{g}=\mathbf{g}^{(1)}\sqrt{(G^{(1)}-G^{(2)})/G^{(2)}}$ is an eigenvector of the nondegenerate wave of a generally nonunit length. The vectors $\mathbf{s}$ and $\mathbf{g}$ may be real- or complex-valued. Equation (15) is a system of six quadratic equations for six unknowns: $\mathbf{s}=(s_1,s_2,s_3)^T$ and $\mathbf{g}=(g_1,g_2,g_3)^T$. The number of solutions is $2^6=64$. If we take into account that solutions of different signs: $\pm\mathbf{s}$, $\pm\mathbf{g}$, correspond to the same acoustic axis, the maximum number of acoustic axes is reduced from 64 to 16. This number comprises acoustic axes with real-valued as well as complex-valued slowness vector $\mathbf{s}$.

C. The Khatkevich equations

Eliminating eigenvalues and eigenvectors in Eq. (13), we obtain:[28]

$$\Gamma_{11} - \frac{\Gamma_{12}\Gamma_{13}}{\Gamma_{23}} = \Gamma_{22} - \frac{\Gamma_{12}\Gamma_{23}}{\Gamma_{13}} = \Gamma_{33} - \frac{\Gamma_{13}\Gamma_{23}}{\Gamma_{12}}, \tag{16}$$

and subsequently [see Khatkevich,[23] Eq. (11)]

$$(\Gamma_{11} - \Gamma_{22})\Gamma_{13}\Gamma_{23} - \Gamma_{12}(\Gamma_{13}^2 - \Gamma_{23}^2) = 0, \tag{17a}$$

$$(\Gamma_{11} - \Gamma_{33})\Gamma_{12}\Gamma_{23} - \Gamma_{13}(\Gamma_{12}^2 - \Gamma_{23}^2) = 0, \tag{17b}$$

$$(\Gamma_{22} - \Gamma_{33})\Gamma_{12}\Gamma_{13} - \Gamma_{23}(\Gamma_{12}^2 - \Gamma_{13}^2) = 0. \tag{17c}$$

Equations (17a)–(17c) represent a system of 6th-order equations for three unknown components of the unit direction vector $\mathbf{n}$: n_1, n_2, and n_3. The three equations [(17a)–(17c)] are not independent, hence we solve only two of them. We obtain 72 solutions, which are generally complex-valued. Taking into account that $\pm\mathbf{n}$ describes the same direction, the number of directions reduces from 72 to 36.

Since Eq. (15) yields only 16 acoustic axes, 20 of the 36 directions calculated from Eqs. (17) must be spurious and do not describe acoustic axes. In fact, the spurious directions were incorporated into the solution, when Eq. (16) was multiplied by terms $\Gamma_{12}\Gamma_{13}$, $\Gamma_{12}\Gamma_{23}$ or $\Gamma_{13}\Gamma_{23}$ in order to derive Eqs. (17). Therefore, we should eliminate from the solutions of Eqs. (17a)–(17c), the directions for which

$$\Gamma_{13} = 0, \quad \Gamma_{23} = 0, \tag{18a}$$

or

$$\Gamma_{12} = 0, \quad \Gamma_{23} = 0, \tag{18b}$$

or

$$\Gamma_{12} = 0, \quad \Gamma_{13} = 0. \tag{18c}$$

Equations (18a)–(18c) describe three systems of quadratic equations, each of them having 8 solutions which reduce to 4 directions when omitting different signs of $\mathbf{n}$. Hence, we obtained a total of 12 spurious directions. Furthermore, 8 of the 12 spurious directions appear in Eqs. (17) twice. Which 8 spurious directions are doubled, depends on the pair of Eqs. (17a)–(17c) we actually solve. For example, when solving

Eqs. (17a) and (17b), the solutions of Eqs. (18a) and (18b) are doubled, when solving Eqs. (17b) and (17c), the solutions of Eqs. (18b) and (18c) are doubled. Hence, the total number of spurious directions in Eqs. (17) is 20. This confirms that only 16 acoustic axes can exist in triclinic anisotropy.

III. NUMERICAL CALCULATION OF ACOUSTIC AXES

In principal, all three above-mentioned approaches can be used for determining acoustic axes in triclinic anisotropy. However, they are differently efficient from the point of view of programming and numerics. The numerical tests show that the acoustic axes in triclinic anisotropy can be conveniently calculated using the Khatkevich approach, because it represents a standard problem of solving two polynomial equations in two unknowns. The Darinskii approach requires elaborating with higher number of equations, and the Fedorov approach yields real-valued acoustic axes but not complex-valued acoustic axes (for details, see Sec. IV). Moreover, the Fedorov approach is rather untypical, because we have to solve one polynomial equation in two unknowns.

The Khatkevich equations (17a)–(17c) are homogeneous polynomial equations of the 6th order in three unknown components n_1, n_2, and n_3 of the unit direction vector $\mathbf{n}$. Using the substitutions $u=n_1/n_3$ and $v=n_2/n_3$, we obtain inhomogeneous polynomial equations of the 6th order in unknowns u and v. The roots of the equations can be calculated, for example, using Gröbner bases,[34] implemented in symbolic algebra packages. Solving Eqs. (17a)–(17c) we obtain 36 solutions, from which we have to exclude 20 spurious solutions defined by Eqs. (18a)–(18c). To identify and eliminate the spurious solutions, we can either solve Eqs. (18a)–(18c) or we can simply check at which of the 36 directions calculated from Eqs. (17a)–(17c) the Christoffel tensor is nondegenerate. Using this approach, we obtain real as well as complex acoustic axes in triclinic anisotropy. For anisotropy of higher symmetry, it is not convenient to use this approach, because it is too complex and it may even fail when the true or spurious solutions are not isolated. Instead, much simpler systems of algebraic equations designed for each specific symmetry are used.[23,29] Equations (17a)–(17c) can also fail when triclinic anisotropy is extremely weak. In this case, the left-hand sides of Eqs. (17a)–(17c) are very close to zero for all directions $\mathbf{n}$, hence the solution can be distorted by numerical errors.

Finally, I should also mention a possibility to calculate acoustic axes using a direct numerical approach. This approach is based on minimizing the square of the difference between numerically calculated eigenvalues of the Christoffel tensor. The minimization can be performed using some standard inversion technique like the gradient method. Since the misfit function has several minima, we have to invert repeatedly for varying initial guesses of the position of the acoustic axis. Not to skip some solutions, the initial positions of the acoustic axes should cover the whole hemisphere in a regular grid and the grid should be sufficiently dense. This approach is applicable to any type of anisotropy and is also reasonably fast. However, it does not yield complex acoustic axes and sometimes it may skip some solutions, for example, in situations when two acoustic axes are very close each to the other.

IV. REAL- AND COMPLEX-VALUED ACOUSTIC AXES

As mentioned earlier, the 16 acoustic axes in triclinic anisotropy can lie along a real-valued or complex-valued direction $\mathbf{n}$. The real-valued acoustic axis corresponds to the propagation of a homogeneous plane wave:

$$u_k(\mathbf{x},t) = A g_k \exp\lfloor -i\omega(t - s_j x_j)\rfloor, \tag{19}$$

where $\mathbf{u}$ is the real-valued displacement, A is the scalar real- or complex-valued amplitude, and $\mathbf{s}=\mathbf{n}/\sqrt{G}$ is the real-valued slowness vector. Since the Christoffel tensor $\mathbf{\Gamma}$ and its eigenvalues G and eigenvectors $\mathbf{g}$ are real-valued, the polarization near the acoustic axis is linear. Strictly at the singularity, the polarization is not defined because of the degeneracy of $\mathbf{\Gamma}$.

If the acoustic axis is complex-valued, the corresponding plane wave solution describes an inhomogeneous wave:

$$\begin{aligned} u_k(\mathbf{x},t) &= A g_k \exp\lfloor -i\omega(t - s_j x_j)\rfloor \\ &= A g_k \exp(-a_j x_j)\exp\lfloor -i\omega(t - p_j x_j)\rfloor, \end{aligned} \tag{20}$$

where A is the scalar real- or complex-valued amplitude, $\mathbf{s}$ is the complex-valued slowness vector, $\mathbf{s}=\mathbf{n}/\sqrt{G}=\mathbf{p}+i\mathbf{a}$, and $\mathbf{p}$ and $\mathbf{a}$ are the real-valued propagation and attenuation vectors.[35] Subsequently, the Christoffel tensor $\mathbf{\Gamma}$ and its eigenvalues G and eigenvectors $\mathbf{g}$, calculated by Eqs. (1) and (3), are complex-valued, and the polarization near the acoustic axis is elliptical. The ellipticity depends on the direction and magnitude of attenuation vector $\mathbf{a}$. Similarly as for homogeneous waves, the polarization is not defined at the acoustic axis because of the degeneracy of $\mathbf{\Gamma}$. Since the problem of degeneracy is more involved for complex-valued tensors than for real-valued tensors, Eqs. (15) and (17) do not describe all acoustic axes for inhomogeneous waves, but only the so-called "semisimple" axes (degeneracies). The "nonsemisimple" degneracies cannot be investigated by these equations. For more details about the semisimple and nonsemisimple degeneracies, see Ting[36] and Shuvalov.[37]

The polarization properties of homogeneous and inhomogeneous waves near real and complex acoustic axes are exemplified for triclinic anisotropy generated from an isotropic medium with $\lambda=1$, $\mu=1$, by adding small perturbations. The elastic parameters are as follows:

$$\mathbf{A} = \begin{bmatrix} 3.004 & 1.004 & 1.004 & 0.004 & 0.002 & -0.001 \\ & 3.004 & 1.002 & -0.002 & -0.001 & 0.001 \\ & & 3.000 & 0.003 & -0.001 & 0.004 \\ & & & 1.001 & 0.001 & -0.001 \\ & & & & 1.000 & 0.003 \\ & & & & & 1.000 \end{bmatrix}. \tag{21}$$

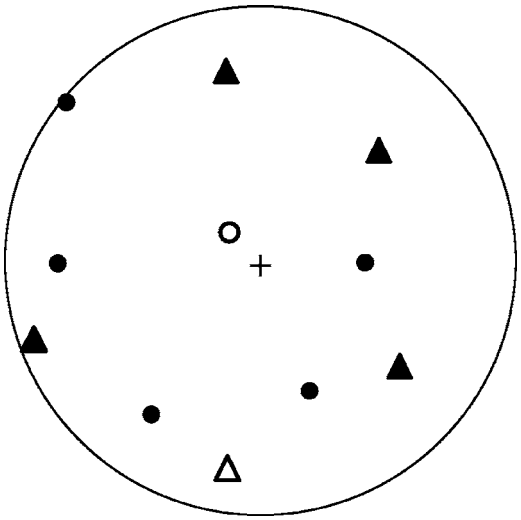


FIG. 1. Positions of real (circles) and complex (triangles) acoustic axes for triclinic anisotropy described by Eq. (21). The plus sign marks the vertical axis. The open circle and open triangle show the acoustic axes, for which the polarization field in their vicinities is shown in Figs. 2 and 3. Equal-area projection is used.

The positions of the acoustic axes over the unit sphere are shown in Fig. 1. The medium contains 6 real and 10 complex acoustic axes. For complex axes, their positions on the sphere are calculated from the real parts of complex directions **n**. Since the complex axes always appear in complex conjugate pairs, Fig. 1 shows 5 instead 10 positions. Figure 2 shows the polarization of homogeneous waves near the real acoustic axis marked in Fig. 1 by an open circle. The field of polarization vectors displays a singularity with the topological charge $-1/2$. Figure 3 shows the polarization of inhomogeneous waves near the complex acoustic axis marked in Fig. 1 by an open triangle. The field of polarization vectors displays a singularity with the topological charge $+1/2$.

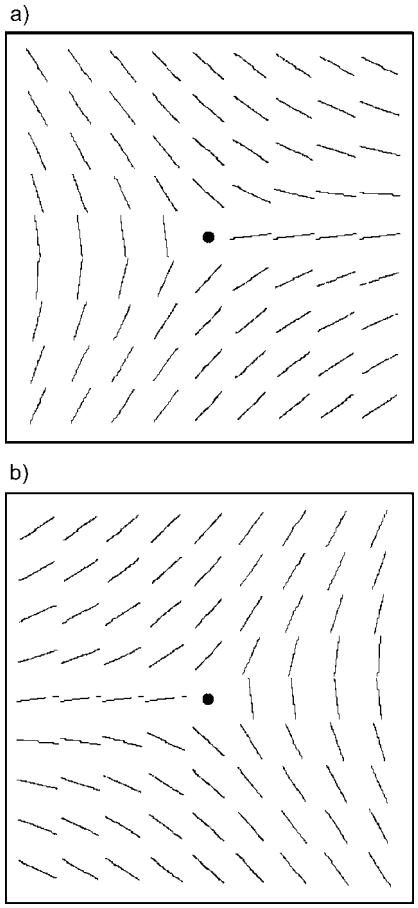


FIG. 2. Polarization field near a real acoustic axis for (a) $S1$ wave, and (b) $S2$ wave. The topological charge is $-1/2$. The dot marks the position of the acoustic axis.

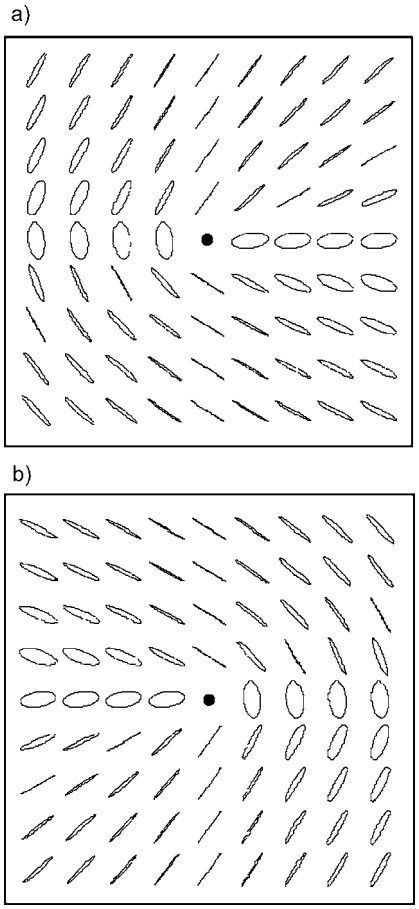


FIG. 3. Polarization field near a complex acoustic axis for (a) $S1$ wave, and (b) $S2$ wave. The topological charge is $+1/2$. The dot marks the position of the acoustic axis.

V. STABLE AND UNSTABLE ACOUSTIC AXES

Acoustic axes can be either single or multiple. The single/multiple acoustic axis corresponds to a nondegenerate/degenerate solution of Eqs. (17a)–(17c). The single axis is stable, because it cannot split or disappear under a small perturbation of elastic parameters. The axis only slightly changes its direction. The multiple axis is unstable, because a small perturbation of elastic parameters removes its degeneracy and splits the axis into two or more single axes. The real-valued multiple axis can split into real-valued and/or complex-valued single axes. The topological charge of the multiple real acoustic axis is equal to the sum of the topological charges of split single real axes.[28]

The properties of the multiple acoustic axes will be exemplified on a transition from cubic to triclinic anisotropy by a small perturbation of elastic parameters. The cubic anisotropy is characterized by 4 single real-valued axes in the directions $\langle \pm 1, \pm 1, \pm 1 \rangle$ and three multiple real-valued axes $\langle \pm 1, 0, 0 \rangle$, $\langle 0, \pm 1, 0 \rangle$, and $\langle 0, 0, \pm 1 \rangle$. The multiple axes are 4 times degenerate. The total number of real-valued acoustic axes is 7. However, if we consider multiplicities, we obtain 16 real-valued axes, which is the maximum number of acoustic axes in anisotropy. This implies that no other complex-valued axis can exist in cubic anisotropy. The topological charge of each multiple axis is +1. If we perturb elastic parameters from cubic to triclinic anisotropy, the single axes slightly change their directions and each multiple axis splits into two real-valued and two complex-valued single axes. Each split real-valued single axis has a topological charge of +1/2.

VI. FREQUENCY OF ACOUSTIC AXES

Now let us address the problem of the balance between the numbers of real- and complex-valued acoustic axes, and how this balance depends on the strength of anisotropy. The number of real and complex acoustic axes will be studied numerically on randomly generated triclinic anisotropy. The triclinic anisotropy with elastic parameters a_{ijkl} is obtained by perturbing an isotropic medium in the following way:

$$a_{ijkl} = a^0_{ijkl} + \varepsilon a^1_{ijkl}, \tag{22}$$

where

$$a^0_{ijkl} = \lambda \delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \tag{23}$$

Parameters λ and μ define the Lamé coefficients of an isotropic medium, tensor a^1_{ijkl} defines perturbations into triclinic anisotropy, and ε is a measure of the anisotropy strength. Parameters λ and μ are fixed at values $\lambda = 1$, $\mu = 1$, perturbations a^1_{ijkl} are generated randomly with a uniform nonzero probability in the interval $\langle -3, 3 \rangle$, and ε is 0.01 or 1. The generated triclinic media were checked to stability using Eq. (2), and the unstable media were discarded. For $\varepsilon = 0.01$, all the media were stable, for $\varepsilon = 1$, approximately one hundredth of the media were stable. To obtain statistically relevant results, the number of randomly generated stable triclinic media is 1000 for each ε.

Figure 4 shows the frequency of occurrence of acoustic axes in the studied triclinic media. The figure shows that the number of acoustic axes depends on the strength of anisotropy. For weak anisotropy, defined by $\varepsilon = 0.01$, the randomly generated triclinic anisotropy contains most frequently 4 to 6 real and 10 to 12 complex axes. For strong anisotropy, defined by $\varepsilon = 1$, the most frequent number is 8 real and 8 complex axes. Media with no real or no complex axis are admissible, but they are very rare. To demonstrate their existence, elastic parameters of two media are presented in Table I: anisotropy I with 16 real axes, and anisotropy II with 16 complex axes. Figure 5 shows the distribution of axes over the sphere. For complex axes, their positions on the sphere are calculated from the real parts of complex directions **n**. Since the complex axes always appear in complex conjugate pairs, Fig. 5(b) shows 8 instead 16 positions. Anisotropy I was generated in a similar way as other randomly generated triclinic media in the above-described numerical tests. Anisotropy II was derived from orthorhombic anisotropy with no real acoustic axes, presented by Boulanger and Hayes[29] by perturbing it into the triclinic anisotropy.

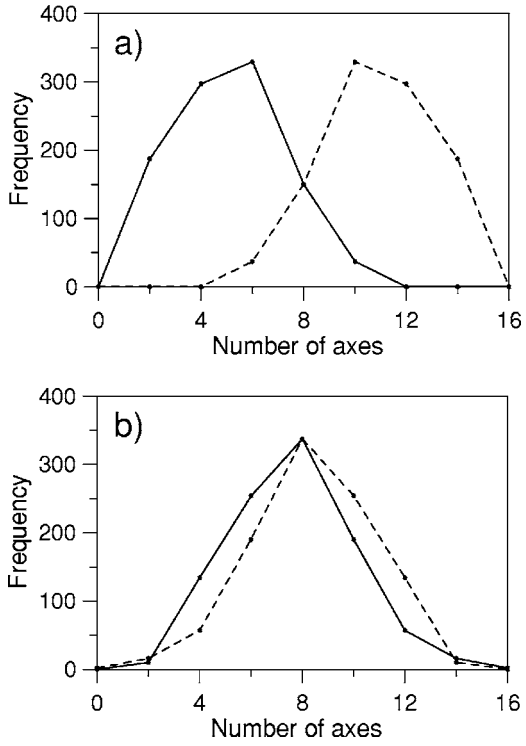


FIG. 4. Frequency of occurrence of real (solid line) and complex (dashed line) acoustic axes in randomly generated triclinic anisotropy with strength: (a) $\varepsilon = 0.01$, and (b) $\varepsilon = 1$.

VII. POSITIONS OF ACOUSTIC AXES AS A FUNCTION OF ANISOTROPY STRENGTH

Here, variations in positions of acoustic axes will be studied in dependence on anisotropy strength. The variations will be shown on two examples of triclinic anisotropy de-

TABLE I. Examples of triclinic anisotropy with 16 real and 16 complex acoustic axes.

	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{22}	a_{23}	a_{24}	a_{25}	a_{26}	a_{33}	a_{34}	a_{35}	a_{36}	a_{44}	a_{45}	a_{46}	a_{55}	a_{56}	a_{66}
I[a]	137	52	57	−13	32	−20	147	18	−6	20	−9	100	22	−15	5	52	26	−7	75	−40	30
II[b]	31	−5	−32	2	−2	−3	9	−11	2	5	4	90	5	−1	−2	12	1	−3	35	−2	10

[a]I—anisotropy with 16 real acoustic axes.
[b]II—anisotropy with 16 complex acoustic axes.

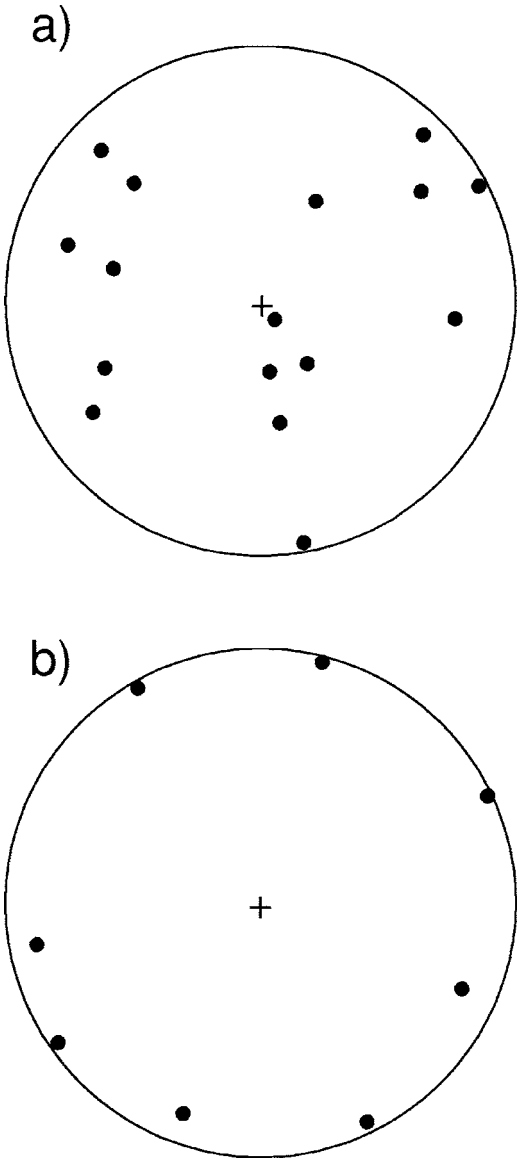


FIG. 5. Positions of acoustic axes in: (a) anisotropy I with 16 real and no complex acoustic axes, and (b) anisotropy II with 16 complex and no real acoustic axes. The acoustic axes are marked by dots, the vertical axis is marked by the plus sign. Equal-area projection is used. Note that the near-horizontal directions of all complex acoustic axes in (b) are rather untypical and not frequently observed.

fined by the Lamé coefficients of the isotropic background medium $\lambda=1$, $\mu=1$, and by the perturbations a^1_{ijkl}:

$$\mathbf{A}^1=\begin{bmatrix} 1.083 & 1.200 & 0.966 & 0.954 & 0.483 & -0.474 \\ & 1.062 & 0.435 & -0.630 & -0.474 & 0.105 \\ & & 0.000 & 0.684 & -0.570 & 1.017 \\ & & & 0.282 & 0.207 & -0.387 \\ & & & & -0.009 & 0.609 \\ & & & & & 0.000 \end{bmatrix}, \tag{24}$$

$$\mathbf{A}^1=\begin{bmatrix} -0.114 & 0.309 & -1.347 & -0.585 & 1.125 & -1.452 \\ & 0.204 & -0.252 & 0.804 & 1.413 & 1.473 \\ & & 0.000 & 0.867 & -0.183 & -0.003 \\ & & & 0.885 & -0.858 & 0.432 \\ & & & & -1.320 & -0.537 \\ & & & & & 0.000 \end{bmatrix}. \tag{25}$$

The anisotropy strength ε ranged from -1000 to 1000 with a varying step. The step was small enough to map densely changes in the directions of acoustic axes. Figure 6 shows the positions of the real acoustic axes in the equal-area projection. The blue/red points mark the axes for stable/unstable media. The figure shows that: (1) The positions of the acoustic axes depend on strength of anisotropy. (2) The acoustic axes are not distributed randomly around a sphere but form a complicated line which can intersect itself. Some segments of the line correspond to the acoustic axes of stable media, the other segments correspond to the axes of unstable media. (3) For high absolute values of ε, the directions of acoustic axes become insensitive to anisotropy strength, and for ε close to ± 1000, the positions of the acoustic axes are almost constant. (4) Since the acoustic axes for $\varepsilon \to \pm\infty$ coincide, the line is closed.

The complex acoustic axes display a similar pattern. Obviously, for anisotropy of higher symmetry, the form of the line simplifies.

VIII. CONCLUSIONS

The acoustic axes in triclinic anisotropy can be conveniently calculated by solving two coupled polynomial equations of the 6th order in two unknowns. From the 36 directions obtained, 20 of them are spurious and must be eliminated. The spurious directions are solutions of three systems of quadratic equations in two unknowns. Hence, the maximum number of isolated acoustic axes in triclinic anisotropy is 16. These axes can be real or complex, and single or multiple. If we count both real and complex acoustic axes and their multiplicities, the total number of isolated axes is always 16 regardless of symmetry or strength of anisotropy. The real axes correspond to the degeneracy directions of the real-valued Christoffel tensor, which describes the propagation of homogeneous plane waves. The complex axes correspond to the semisimple degeneracy directions of the complex-valued Christoffel tensor which describes the propagation of inhomogeneous plane waves. The inhomogeneous waves can also possess other types of acoustic axes called nonsemisimple. The real-valued acoustic axis is associated with a linear polarization in its vicinity, which is singular at the axis. The complex-valued (semisimple) acoustic axis is associated with an elliptical polarization, which is also singular at the axis. Numerical simulations indicate that the most frequent number of real acoustic axes is only 4 to 6 for weak anisotropy and 8 for strong anisotropy. A medium with no or 16 real acoustic axes is admissible, but it is very rare. Positions of the acoustic axes depend on strength of anisotropy. If we fix the perturbation matrix a^1_{ijkl} and change anisotropy strength ε, the positions of acoustic axes form a one closed curve.

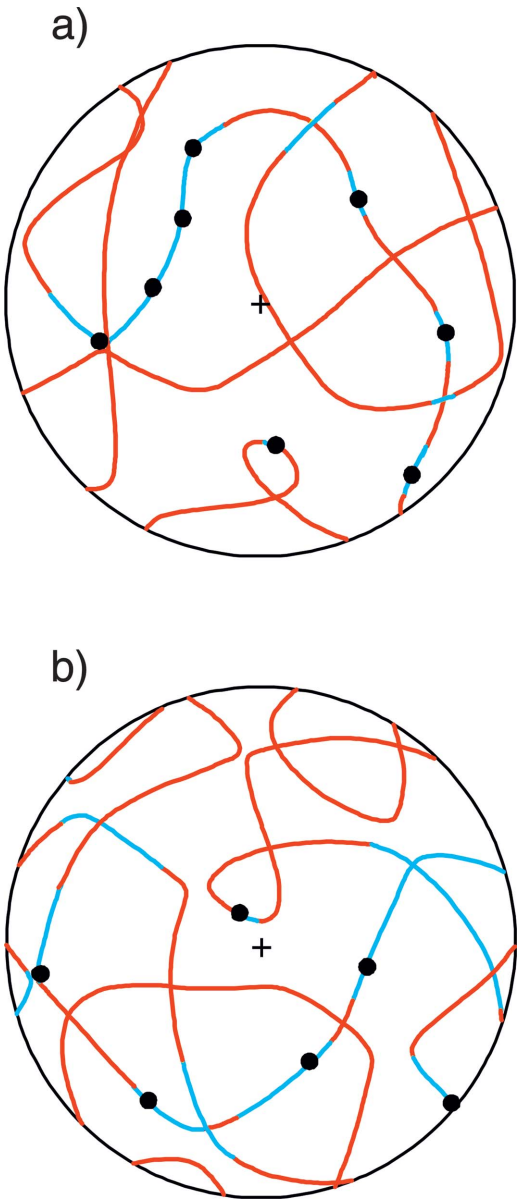


FIG. 6. Directions of real acoustic axes as a function of anisotropy strength for two types of triclinic anisotropy. (a) Anisotropy is described by elasticity tensor (25). (b) Anisotropy is described by elasticity tensor (24). Blue/red lines denote the acoustic axes of stable/unstable media. Black dots mark the positions of acoustic axes for infinitesimally weak anisotropy ($\varepsilon \to 0$), the plus sign marks the vertical axis. Equal-area projection is used.

ACKNOWLEDGMENTS

I thank Vlastislav Červený, Klaus Helbig, and Ivan Pšenčík for stimulating discussions on the subject. The work was supported by the Consortium Project SW3D "Seismic waves in complex 3-D structures," and by the Grant Agency of the Academy of Sciences of the Czech Republic, Grant No. A3012309.

[1]V. I. Alshits and J. Lothe, "Elastic waves in triclinic crystals. I. General theory and the degeneracy problem," Sov. Phys. Crystallogr. **24**, 387–392 (1979).
[2]V. I. Alshits and J. Lothe, "Elastic waves in triclinic crystals. II. Topology of polarization fields and some general theorems," Sov. Phys. Crystallogr. **24**, 393–398 (1979).
[3]V. I. Alshits, A. V. Sarychev, and A. L. Shuvalov, "Classification of degeneracies and analysis of their stability in the theory of elastic waves in crystals" (in Russian), Zh. Eksp. Teor. Fiz. **89**, 922–938 (1985).
[4]A. N. Norris, "Acoustic axes in elasticity," Wave Motion **40**, 315–328 (2004).
[5]A. L. Shuvalov, "Topological features of the polarization fields of plane acoustic waves in anisotropic media," Proc. R. Soc. London, Ser. A **454**, 2911–2947 (1998).
[6]V. I. Alshits and J. Lothe, "Some basic properties of bulk elastic waves in anisotropic media," Wave Motion **40**, 297–313 (2004).
[7]A. G. Every, "Formation of phonon-focusing caustics in crystals," Phys. Rev. B **34**, 2852–2862 (1986).
[8]D. C. Hurley, J. P. Wolfe, and K. A. McCarthy, "Phonon focusing in tellurium dioxide," Phys. Rev. B **33**, 4189–4195 (1986).
[9]A. G. Every, "Classification of the phonon-focusing patterns of tetragonal crystals," Phys. Rev. B **37**, 9964–9977 (1988).
[10]M. R. Hauser, R. L. Weaver, and J. P. Wolfe, "Internal diffraction of ultrasound in crystals: Phonon focusing at long wavelengths," Phys. Rev. Lett. **68**, 2604–2607 (1992).
[11]K. Y. Kim, W. Sachse, and A. G. Every, "Focusing of acoustic energy at the conical point in zinc," Phys. Rev. Lett. **70**, 3443–3446 (1993).
[12]A. L. Shuvalov and G. Every, "Shape of the acoustic slowness surface of anisotropic solids near points of conical degeneracy," J. Acoust. Soc. Am. **101**, 2381–2383 (1997).
[13]J. P. Wolfe, *Imaging Phonons. Acoustic Wave Propagation in Solids* (Cambridge University Press, Cambridge, 1998).
[14]V. Vavryčuk, "Parabolic lines and caustics in homogeneous weakly anisotropic solids," Geophys. J. Int. **152**, 318–334 (2003).
[15]V. Vavryčuk, "Ray tracing in anisotropic media with singularities," Geophys. J. Int. **145**, 265–276 (2001).
[16]V. Vavryčuk, "Behavior of rays near singularities in anisotropic media," Phys. Rev. B **67**, 054105 (2003).
[17]C. H. Chapman and P. M. Shearer, "Ray tracing in azimuthally anisotropic media. II. Quasi-shear wave coupling," Geophys. J. **96**, 65–83 (1989).
[18]R. T. Coates and C. H. Chapman, "Quasi-shear wave coupling in weakly anisotropic 3-D media," Geophys. J. Int. **103**, 301–320 (1990).
[19]Yu. A. Kravtsov and Yu. I. Orlov, *Geometrical Optics of Inhomogeneous Media* (Springer, Berlin, 1990).
[20]V. Červený, *Seismic Ray Theory* (Cambridge University Press, Cambridge, 2001).
[21]P. Bulant and L. Klimeš, "Comparison of quasi-isotropic approximations of the coupling ray theory with the exact solutions in the 1-D anisotropic oblique twisted crystal model," Stud. Geophys. Geod. **48**, 97–116 (2004).
[22]G. Rümpker and C. J. Thomson, "Seismic-waveform effects of conical points in gradually varying anisotropic media," Geophys. J. Int. **118**, 759–780 (1994).
[23]A. G. Khatkevich, "Acoustic axes in crystals," Sov. Phys. Crystallogr. **7**, 601–604 (1963).
[24]F. I. Fedorov, *Theory of Elastic Waves in Crystals* (Plenum, New York, 1968).
[25]A. G. Khatkevich, "Classification of crystals by acoustic properties," Sov. Phys. Crystallogr. **22**, 701–705 (1977).
[26]M. J. P. Musgrave, "Acoustic axes in orthorhombic media," Proc. R. Soc. London, Ser. A **401**, 131–143 (1985).
[27]P. Holm, "Generic elastic media," Phys. Scr., T **44**, 122–127 (1992).
[28]B. M. Darinskii, "Acoustic axes in crystals," Crystallogr. Rep. **39**, 697–703 (1994).
[29]Ph. Boulanger and M. Hayes, "Acoustic axes for elastic waves in crystals: Theory and applications," Proc. R. Soc. London, Ser. A **454**, 2323–2346 (1998).
[30]V. G. Mozhaev, F. Bosia, and M. Weihnacht, "Oblique acoustic axes in trigonal crystals," J. Comput. Acoust. **9**, 1147–1161 (2001).
[31]A. Duda and T. Paszkiewicz, "Number of longitudinal normals and degenerate directions for triclinic and monoclinic media," Eur. Phys. J. B **31**, 327–331 (2003).
[32]M. J. P. Musgrave, *Crystal Acoustics* (Holden-Day, San Francisco, 1970).
[33]R. G. Payton, *Elastic Wave Propagation in Transversely Isotropic Media* (Nijhoff, The Hague, 1983).
[34]R. Fröberg, *An Introduction to Gröbner Bases* (Wiley, New York, 1997).
[35]V. Červený, "Inhomogeneous harmonic plane waves in viscoelastic anisotropic media," Stud. Geophys. Geod. **48**, 167–186 (2004).
[36]T. C. T. Ting, *Anisotropic Elasticity. Theory and Applications* (Oxford University Press, New York, 1996).
[37]A. L. Shuvalov, "On the theory of plane inhomogeneous waves in anisotropic elastic media," Wave Motion **34**, 401–429 (2001).

Parametrization of acoustic boundary absorption and dispersion properties in time-domain source/receiver reflection measurement

Adrianus T. de Hoop,[a] Chee-Heun Lam,[b] and Bert Jan Kooij
Laboratory of Electromagnetic Research, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 4 Mekelweg, 2628 CD Delft, The Netherlands

(Received 30 September 2004; revised 8 March 2005; accepted 25 May 2005)

Closed-form analytic time-domain expressions are obtained for the acoustic pressure associated with the reflection of a monopole point-source excited impulsive acoustic wave by a planar boundary with absorptive and dispersive properties. The acoustic properties of the boundary are modeled as a local admittance transfer function between the normal component of the particle velocity and the acoustic pressure. The transfer function is to meet the conditions for linear, time-invariant, causal, passive behavior. A parametrization of the admittance function is put forward that has the property of showing up explicitly, and in a relatively simple manner, in the expression for the reflected acoustic pressure. The partial fraction representation of the complex frequency domain admittance is shown to have such a property. The result opens the possibility of constructing inversion algorithms that enable the extraction of the relevant parameters from the measured time traces of the acoustic pressure at different offsets, parallel as well as normal to the boundary, between source and receiver. Illustrative theoretical numerical examples are presented. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1954567]

PACS number(s): 43.20.Bi, 43.20.El, 43.20.Px, 43.55.Ev [JJM] Pages: 654–660

I. INTRODUCTION

In a variety of applications in acoustics (for example, in outdoor sound propagation, traffic noise analysis, jet-engine sound absorption in aircraft engineering and architectural acoustics), the analysis of the point-source excited reflection of sound waves by a boundary surface with certain absorptive and dispersive properties is of interest. In all these cases, the absorptive and dispersive properties of the boundary need characterization by a judiciously chosen set of parameters. Following the pioneering paper by Ingard (1951), such a characterization goes via a local acoustic admittance, i.e., via a linear, time-invariant, causal, passive transfer function that links the normal component of the particle velocity on the boundary to the local acoustic pressure. For the canonical configuration consisting of a planar boundary, a monopole acoustic (volume injection) source and a monopole acoustic (pressure) point receiver, we derive closed-form time-domain expressions for the received signal. For the same configuration and along similar lines, a recent paper (Lam *et al.*, 2004) discusses some ad-hoc cases, where the boundary's properties are expressed via a complex-frequency domain Padé representation, the coefficients in which are matched to experimental data available in the literature. The approach via the Padé representation appears, however, to be limited to at most the Padé (2,2) one. In the present paper, a more systematic approach is followed where the complex-frequency domain characterization of the boundary admittance goes via a partial-fraction representation that allows the incorporation of an arbitrary number of terms, each of them with an interpretable influence on the received signal. Amongst others, it is shown that, when source and receiver are both close to the boundary and the terms in the partial-fraction admittance representation meet a certain condition, large-amplitude oscillatory surface effects can occur. Their amplitudes can even exceed the acoustic pressure values associated with the reflection against a perfectly rigid boundary, a phenomenon that has also been reported elsewhere in the literature (Wenzel, 1974; Thomasson, 1976; Donato, 1976a, 1976b) and is confirmed by pertaining experiments (Daigle *et al.*, 1996) as well as by computational finite-difference time-domain and finite element method studies (Ju and Fung; 2002; Van den Nieuwenhoff and Coyette, 2001).

The analysis is carried out with the aid of an extension (De Hoop, 2002) of the senior (first) author's modification of the Cagniard method (Cagniard, 1962; De Hoop, 1960; De Hoop and Van der Hijden, 1984). It yields closed-form analytic expressions for the time-domain acoustic pressure in the model configuration under investigation. Not only do these expressions reveal how the parameters governing the absorption and dispersion properties of the reflecting boundary show up in the measured acoustic pressure, but they can also serve as benchmarks in further computational studies based on the numerical discretization of the acoustic wave equations.

The methodology leans heavily on the use of the Schouten–Van der Pol theorem of the unilateral Laplace transformation (Schouten, 1934, 1961; Van der Pol, 1934; Van der Pol and Bremmer, 1950). This theorem interrelates two (causal) functions of time whose (unilateral) Laplace

[a]Electronic mail: a.t.dehoop@ewi.tudelft.nl
[b]Presently at Laboratory of Circuits and Systems, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, 4 Mekelweg, 2628 CD Delft, The Netherlands.

transforms are related such that the Laplace transform of the latter arises out of the Laplace transform of the former upon replacing the transform parameter s with a certain function $\phi(s)$, where $\phi(s)$ belongs to the class of functions for which a causal time function corresponding to $\exp[-\phi(s)\tau]$, with $\tau \geqslant 0$, exists.

The analysis can be carried out for an arbitrary number of terms in the partial-fraction characterization of the boundary's acoustic admittance, each of them provided with its associated two parameters. This implies that a rather accurate tuning of the parameters to match the measured values of the admittance (a procedure that is usually carried out in the frequency domain) can be achieved by incorporating a sufficient number of terms.

Some theoretical numerical examples illustrate how some physical phenomena can be attributed to certain ranges of the values of the parameters involved.

II. FORMULATION OF THE PROBLEM

Position in the configuration is specified by the coordinates $\{x,y,z\}$ with respect to an orthogonal, Cartesian reference frame with the origin $\mathcal{O}$ and the three mutually perpendicular base vectors $\{\boldsymbol{i}_x,\boldsymbol{i}_y,\boldsymbol{i}_z\}$ of unit length each; they form, in the indicated order, a right-handed system. The position vector is $\boldsymbol{r}=x\boldsymbol{i}_x+y\boldsymbol{i}_y+z\boldsymbol{i}_z$. The vectorial spatial differentiation operator is $\boldsymbol{\nabla}=\boldsymbol{i}_x\partial_x+\boldsymbol{i}_y\partial_y+\boldsymbol{i}_z\partial_z$. The time coordinate is t; differentiation with respect to time is denoted by ∂_t.

The acoustic wave motion is studied in the half-space $\mathcal{D}=\{-\infty<x<\infty,-\infty<y<\infty,0<z<\infty\}$, which is filled with a fluid with volume density of mass ρ_0 and compressibility κ_0. The speed of sound waves in it is $c_0=(\rho_0\kappa_0)^{-1/2}$. The acoustic wave motion is excited by an acoustic monopole point source with volume injection rate $Q_0(t)$ and located at $\boldsymbol{r}_0=\{0,0,h\}$, with $h\geqslant 0$. We assume that $Q_0(t)=0$ for $t<0$. The acoustic pressure $p(\boldsymbol{r},t)$ and the particle velocity $\boldsymbol{v}(\boldsymbol{r},t)$ then satisfy the first-order acoustic wave equations (De Hoop, 1995, p. 44)

$$\boldsymbol{\nabla}p+\rho_0\partial_t\boldsymbol{v}=\boldsymbol{0}, \tag{1}$$

$$\boldsymbol{\nabla}\cdot\boldsymbol{v}+\kappa_0\partial_t p=Q_0(t)\,\delta(\boldsymbol{r}-\boldsymbol{r}_0). \tag{2}$$

Causality entails that $p(\boldsymbol{r},t)=0$ and $\boldsymbol{v}(\boldsymbol{r},t)=\boldsymbol{0}$ for $t<0$ and all $\boldsymbol{r}\in\mathcal{D}$. The acoustic properties of the planar boundary are modeled via the local, linear, time-invariant, causal, passive acoustic admittance relation

$$v_z(x,y,0,t)=-(\rho_0c_0)^{-1}Y(t)\overset{(t)}{*}p(x,y,0,t), \tag{3}$$

where $\overset{(t)}{*}$ denotes time convolution and $Y(t)$ is the boundary's acoustic time-domain admittance transfer function, normalized with respect to the acoustic plane-wave admittance $(\rho_0c_0)^{-1}$ of the fluid. Figure 1 shows the configuration.

The acoustic wave field in the fluid is written as the superposition of the incident wave field to be denoted by the superscript i and the reflected wave field to be denoted by the superscript r. The incident wave field is the wave field that is

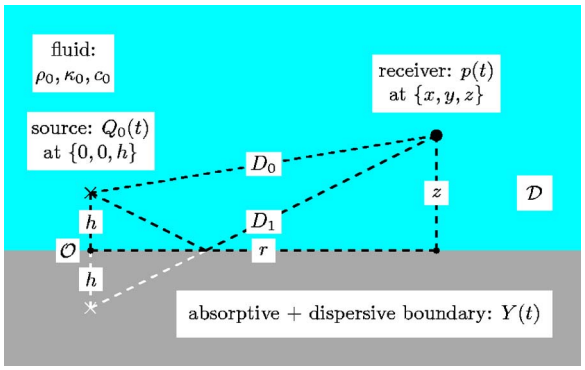


FIG. 1. Fluid-filled half-space with volume injection point source, acoustic pressure point receiver, and reflecting absorptive and dispersive boundary.

generated by the source and would be the total wave field in the absence of the boundary. Its acoustic pressure satisfies the scalar wave equation

$$\boldsymbol{\nabla}^2p^i-c_0^{-2}\partial_t^2p^i=-\rho_0\partial_tQ_0(t)\,\delta(x,y,z-h). \tag{4}$$

From this equation we obtain (see, for example, De Hoop, 1995, pp. 93–97)

$$p^i(\boldsymbol{r},t)=\rho_0\partial_t^2Q_0(t)\overset{(t)}{*}G^i(\boldsymbol{r},t), \tag{5}$$

in which the incident-wave Green's function is

$$G^i(\boldsymbol{r},t)=\frac{H(t-T_0)}{4\pi D_0}\quad\text{for } D_0>0, \tag{6}$$

with

$$D_0=[x^2+y^2+(z-h)^2]^{1/2}\geqslant 0 \tag{7}$$

as the distance from the source to the receiver,

$$T_0=D_0/c_0 \tag{8}$$

as the travel time from source to receiver and $H(t)$ as the Heaviside unit step function.

III. THE COMPLEX SLOWNESS REPRESENTATION FOR THE ACOUSTIC WAVE FIELDS

The time invariance of the configuration and the causality of the sound waves are taken into account by the use of the unilateral Laplace transform:

$$\{\hat{p},\hat{\boldsymbol{v}}\}(\boldsymbol{r},s)=\int_{t=0}^{\infty}\exp(-st)\{p,\boldsymbol{v}\}(\boldsymbol{r},t)dt. \tag{9}$$

The Laplace transform parameter s is taken positive and real. Then, according to Lerch's theorem (Widder, 1946) a one-to-one mapping exists between $\{p,\boldsymbol{v}\}(\boldsymbol{r},t)$ and their time-Laplace transformed counterparts $\{\hat{p},\hat{\boldsymbol{v}}\}(\boldsymbol{r},s)$. The fluid is initially at rest, which has the consequence that the transformation property $\partial_t\rightarrow s$ holds. Next, the complex slowness representations for $\{\hat{p},\hat{\boldsymbol{v}}\}(\boldsymbol{r},s)$ are introduced as

$$\{\hat{p},\hat{\boldsymbol{v}}\}(x,y,z,s)=\frac{s^2}{4\pi^2}\int_{\alpha=-\infty}^{\infty}d\alpha\int_{\beta=-\infty}^{\infty}\{\tilde{p},\tilde{\boldsymbol{v}}\}(\alpha,\beta,z,s)\times\exp[-is(\alpha x+\beta y)]d\beta, \tag{10}$$

where α and β are the wave slownesses in the x and y directions, respectively. This representation entails the properties $\partial_x\rightarrow -is\alpha$, $\partial_y\rightarrow -is\beta$. Use of the transforms in Eqs. (1)–(4) yields for the incident wave

$$\tilde{p}^i(\alpha,\beta,z,s)=\frac{\rho_0\hat{Q}_0(s)}{2\gamma_0}\exp(-s\gamma_0|z-h|), \tag{11}$$

while for the reflected wave we write

$$\tilde{p}^r(\alpha,\beta,z,s)=\frac{\rho_0\hat{Q}_0(s)}{2\gamma_0}\tilde{R}\exp[-s\gamma_0(z+h)], \tag{12}$$

in which

$$\gamma_0(\alpha,\beta)=(c_0^{-2}+\alpha^2+\beta^2)^{1/2}\quad\text{with Re}(\gamma_0)\geqslant 0 \tag{13}$$

is the wave slowness normal to the boundary and $\tilde{R}$ denotes the slowness-domain reflection coefficient. Use of Eqs. (11) and (12) in the complex slowness domain counterpart of the admittance boundary condition (3), together with the property [cf. Eq. (1)]

$$\tilde{v}_z=-(s\rho_0)^{-1}\partial_z\tilde{p}, \tag{14}$$

Eqs. (11), (12), and (14) lead to

$$(\gamma_0/\rho_0)(1-\tilde{R})=(\rho_0 c_0)^{-1}\hat{Y}(s)(1+\tilde{R}), \tag{15}$$

from which it follows that

$$\tilde{R}=\frac{c_0\gamma_0-\hat{Y}(s)}{c_0\gamma_0+\hat{Y}(s)}=1-\frac{2\hat{Y}(s)}{c_0\gamma_0+\hat{Y}(s)}. \tag{16}$$

IV. SPACE-TIME EXPRESSIONS FOR THE ACOUSTIC WAVE FIELD CONSTITUENTS

The expressions for the time Laplace transformed reflected wave field quantities are written as

$$\hat{p}^r(\boldsymbol{r},s)=\rho_0 s^2\hat{Q}_0(s)\hat{G}^r(\boldsymbol{r},s), \tag{17}$$

$$\hat{\boldsymbol{v}}^r(\boldsymbol{r},s)=-s\hat{Q}_0(s)\,\boldsymbol{\nabla}\,\hat{G}^r(\boldsymbol{r},s), \tag{18}$$

in which

$$\hat{G}^r(\boldsymbol{r},s)=\frac{1}{4\pi^2}\int_{\alpha=-\infty}^{\infty}d\alpha\int_{\beta=-\infty}^{\infty}\tilde{R}\frac{1}{2\gamma_0}\exp\{-s[i(\alpha x+\beta y)+\gamma_0(z+h)]\}d\beta \tag{19}$$

is the time Laplace transformed reflected-wave Green's function. The time-domain counterparts of Eqs. (17)–(19) are determined with the aid of an extension (De Hoop, 2002) of the standard modified Cagniard method (De Hoop, 1960; De Hoop and Van der Hijden, 1984). First, upon writing $x=r\cos(\theta)$, $y=r\sin(\theta)$, the transformation

$$\alpha=ip\cos(\theta)-q\sin(\theta),$$

$$\beta=ip\sin(\theta)+q\cos(\theta), \tag{20}$$

is carried out, which for the slowness normal to the boundary leads to $\bar{\gamma}_0(q,p)=[\Omega(q)^2-p^2]^{1/2}$, with $\Omega(q)=(c_0^{-2}+q^2)^{1/2}$. Next, the integrand in the integration with respect to p is continued analytically into the complex p plane, away from the imaginary axis and, under the application of Cauchy's theorem and Jordan's lemma, the integration along the imaginary p axis is replaced by one along the hyperbolic path (modified Cagniard path) consisting of $pr+\bar{\gamma}_0(q,p)(z+h)=\tau$, together with its complex conjugate, for $T_1(q)<\tau<\infty$, where $T_1(q)=\Omega(q)D_1$ and $D_1=[x^2+y^2+(z+h)^2]^{1/2}>0$ is the distance from the image of the source to the receiver, while τ is introduced as the variable of integration. In the relevant Jacobian, the relation $\partial p/\partial\tau=i\bar{\gamma}_0/[\tau^2-T_1^2(q)]^{1/2}$ is used. Next, Schwarz's reflection principle of complex function theory is used to combine the integrations in the upper and lower halves of the complex p plane, the orders of integration with respect to τ and q are interchanged, and in the resulting integration with repect to q, that extends over the interval $0<q<(\tau^2/D_1^2-c_0^{-2})^{1/2}$, the variable of integration q is replaced with ψ defined through $q=(\tau^2/D_1^2-c_0^{-2})^{1/2}\sin(\psi)$, with $0\leqslant\psi\leqslant\pi/2$. This procedure leads to

$$\hat{G}^r(r,s)=\frac{1}{4\pi D_1}\int_{\tau=T_1}^{\infty}\exp(-s\tau)\hat{K}^r(\boldsymbol{r},\tau,s)d\tau, \tag{21}$$

in which

$$\hat{K}^r(\boldsymbol{r},\tau,s)=\frac{2}{\pi}\int_{\psi=0}^{\pi/2}\text{Re}\left[1-\frac{2\hat{Y}(s)}{c_0\bar{\gamma}_0+\hat{Y}(s)}\right]d\psi, \tag{22}$$

with

$$c_0\bar{\gamma}_0=\Gamma_1(\boldsymbol{r},\tau)-i\Gamma_2(\boldsymbol{r},\tau)\cos(\psi), \tag{23}$$

$$\Gamma_1(\boldsymbol{r},\tau)=c_0\tau(z+h)/D_1^2, \tag{24}$$

$$\Gamma_2(\boldsymbol{r},\tau)=c_0(\tau^2-T_1^2)^{1/2}r/D_1^2, \tag{25}$$

is the reflected-wave kernel function and

$$T_1=T_1(0)=D_1/c_0 \tag{26}$$

is the travel time from the image of the source to the receiver. Evaluation of the integral in the right-hand side of Eq. (22) yields (see the Appendix)

$$\hat{K}^r(\boldsymbol{r},\tau,s)=1-\frac{2\hat{Y}(s)}{\{[\Gamma_1(\boldsymbol{r},\tau)+\hat{Y}(s)]^2+\Gamma_2^2(\boldsymbol{r},\tau)\}^{1/2}}. \tag{27}$$

Since the right-hand side of Eq. (27) is an analytic function of s in the right half $\{\text{Re}(s)>0\}$ of the complex s plane, it has a causal time-domain counterpart $K^r(\boldsymbol{r},\tau,t)$ that vanishes for $t<0$. In terms of the latter, Eq. (21) leads to the time-domain expression

$$G^r(\boldsymbol{r},t)=\left[\frac{1}{4\pi D_1}\int_{\tau=T_1}^{t}K^r(\boldsymbol{r},\tau,t-\tau)d\tau\right]H(t-T_1). \tag{28}$$

To further separate in the second term on the right-hand side of Eq. (27) the influence of the configurational parameters of the measurement setup from the influence of the parameters

associated with the boundary's acoustic admittance on the reflected field acoustic pressure, we make use of the Schouten–Van der Pol theorem of the unilateral Laplace transformation (Schouten, 1934, 1961; Van der Pol, 1934; Van der Pol and Bremmer, 1950) and employ the Laplace-transform integral Formula (29.3.55) from Abramowitz and Stegun (1968, p. 1024), together with some elementary rules of the Laplace transformation to obtain

$$\frac{\hat{Y}(s)}{\{[\Gamma_1(\boldsymbol{r},\tau)+\hat{Y}(s)]^2+\Gamma_2^2(\boldsymbol{r},\tau)\}^{1/2}} = 1-\int_{w=0}^{\infty} K_F(\boldsymbol{r},\tau,w)\hat{K}_Y(w,s)dw, \tag{29}$$

in which

$$\hat{K}_Y(w,s)=\exp[-\hat{Y}(s)w]H(w) \tag{30}$$

and

$$K_F(\boldsymbol{r},\tau,w)=\exp[-\Gamma_1(\boldsymbol{r},\tau)w]\{\Gamma_1(\boldsymbol{r},\tau)J_0[\Gamma_2(\boldsymbol{r},\tau)w] + \Gamma_2(\boldsymbol{r},\tau)J_1[\Gamma_2(\boldsymbol{r},\tau)w]\}H(w), \tag{31}$$

where J_0 and J_1 are the Bessel functions of the first kind and orders zero and one, respectively. Use of this result in Eq. (27) yields

$$\hat{K}^r(\boldsymbol{r},\tau,s)=-1+2\int_{w=0}^{\infty} K_F(\boldsymbol{r},\tau,w)\hat{K}_Y(w,s)dw. \tag{32}$$

In terms of the (causal) time-domain counterpart $K_Y(w,t)$ of $\hat{K}_Y(w,s)$ we end up with

$$K^r(\boldsymbol{r},\tau,t)=-\delta(t)+2\left[\int_{w=0}^{\infty} K_F(\boldsymbol{r},\tau,w)K_Y(w,t)dw\right]H(t). \tag{33}$$

Note that in this expression the space-time configurational parameters of the fluid only occur in the kernel function $K_F(\boldsymbol{r},\tau,w)$, while the parameters of $Y(t)$ only occur in the kernel function $K_Y(w,t)$. The space-time expressions for the reflected acoustic wave field quantities are from Eqs. (17) and (18) finally obtained as

$$p^r(\boldsymbol{r},t)=\rho_0\partial_t^2 Q_0(t)\overset{(t)}{*}G^r(\boldsymbol{r},t), \tag{34}$$

$$\boldsymbol{v}^r(\boldsymbol{r},t)=-\partial_t Q_0(t)\overset{(t)}{*}\boldsymbol{\nabla}G^r(\boldsymbol{r},t). \tag{35}$$

In Sec. V, an expression for $K_Y(w,t)$ is obtained for the case where a partial fraction parametrization of the complex frequency domain acoustic admittance $\hat{Y}(s)$ is used to specify the boundary's acoustic dispersion and absorption properties.

V. PARTIAL-FRACTION PARAMETRIZATION OF THE COMPLEX FREQUENCY DOMAIN ACOUSTIC ADMITTANCE AND ITS COROLLARIES

In this section an expression for the kernel function $K_Y(w,t)$, introduced via Eq. (30), is constructed for the case where $\hat{Y}(s)$ is parametrized through a partial fraction representation. Let

$$\hat{Y}(s)=\sum_{n=0}^{N}\hat{Y}^{(n)}(s), \tag{36}$$

with

$$\hat{Y}^{(0)}(s)=Y^{\infty}, \tag{37}$$

$$\hat{Y}^{(n)}(s)=\frac{A_n}{s+\alpha_n} \quad \text{for } n=1,\ldots,N. \tag{38}$$

Since the underlying assumption of such a representation is that $\hat{Y}(s)$ arises as the causal response from a rational time differentiation operator with real-valued coefficients and a finite number of degrees of freedom, a number of properties hold (Kwakernaak and Sivan, 1991). First, $\hat{Y}(s)$ has to be real and positive for s real and positive, which entails that Y^{∞} is real and $\geqslant 0$. Furthermore, $\hat{Y}(s)$ has, in general, simple poles at $s=-\alpha_n$ $(n=1,\ldots,N)$ that should be located in the left half of the complex s-plane. As to the terms $\hat{Y}^{(n)}$ $(n=1,\ldots,N)$ two possibilities arise: either α_n $(n=1,\ldots,N)$ is real and $\geqslant 0$ and the residues A_n $(n=1,\ldots,N)$ at the poles $s=-\alpha_n$ $(n=1,\ldots,N)$ are real, or pairs of α_n $(n=1,\ldots,N)$ are complex conjugate with positive real parts and the residues A_n $(n=1,\ldots,N)$ at such pair of poles $s=-\alpha_n$ $(n=1,\ldots,N)$ are each other's complex conjugate. (The case of higher-order poles is most easily handled by a limiting confluence procedure.) Equation (36) entails a representation of $\hat{K}_Y(w,s)$ of the form

$$\hat{K}_Y(w,s)=\prod_{n=0}^{N}\hat{K}_Y^{(n)}(w,s), \tag{39}$$

with

$$\hat{K}_Y^{(0)}(w,s)=\exp(-Y^{\infty}w)H(w), \tag{40}$$

$$\hat{K}_Y^{(n)}(w,s)=\exp[-Y^{(n)}(s)w]H(w) \text{ for } n=1,\ldots,N. \tag{41}$$

The time-domain counterpart of Eq. (40) is

$$K_Y^{(0)}(w,t)=\exp(-Y^{\infty}w)H(w)\delta(t). \tag{42}$$

To construct the time-domain counterpart of Eq. (41) we again use the Schouten–Van der Pol theorem and employ Formula (29.3.75) of Abramowitz and Stegun (1968, p. 1026), together with some elementary rules of the time Laplace transformation to obtain

$$K_Y^{(n)}(w,t)=H(w)\delta(t)-\exp(-\alpha_n t)\times(A_n w/t)^{1/2}J_1[2(A_n wt)^{1/2}]H(w)H(t) \quad \text{for } n=1,\ldots,N. \tag{43}$$

In terms of Eq. (43) (that also holds for complex values of

the parameters), the time-domain counterpart of Eq. (39) follows as

$$K_Y(w,t) = K_Y^{(0)}(w,t) \overset{(t)}{*} K_Y^{(1)}(w,t) \overset{(t)}{*} \cdots \overset{(t)}{*} K_Y^{(N)}(w,t). \tag{44}$$

In this expression each of the factors contains only two parameters, a property that can facilitate the parameter sensitivity analysis of the reflection measurement setup.

VI. PLANE-WAVE ADMITTANCE PARAMETRIZATION OF THE COMPLEX FREQUENCY DOMAIN ACOUSTIC ADMITTANCE AND ITS COROLLARIES

In this section an expression for the kernel function $K_Y(w,t)$, introduced via Eq. (30), is constructed for the case where $\hat{Y}(s)$ is parametrized through a plane-wave admittance expression, applying to a fluid with volume density of mass ρ_1, compressibility κ_1, normalized inertia relaxation function $\hat{\alpha}_1(s)$, and normalized compressibility relaxation function $\hat{\beta}_1(s)$. Accordingly, we write (De Hoop, 1995, p. 42)

$$\hat{Y}_W(s) = Y_1^{\infty}[\hat{X}(s)]^{1/2}, \tag{45}$$

in which

$$Y_1^{\infty} = \rho_0 c_0 \left(\frac{\kappa_1}{\rho_1}\right)^{1/2} = \frac{\rho_0 c_0}{\rho_1 c_1}, \tag{46}$$

with $c_1=(\rho_1\kappa_1)^{-1/2}$ as the corresponding wave speed, is representative for the instantaneous response and

$$\hat{X}(s) = \frac{s+\hat{\alpha}_1(s)}{s+\hat{\beta}_1(s)} \tag{47}$$

is representative for the absorptive and dispersive properties. To construct the time-domain counterpart $K_W(w,t)$ of the corresponding kernel function

$$\hat{K}_W(w,s) = \exp[-\hat{Y}_W(s)w] \tag{48}$$

we again use the Schouten–Van der Pol theorem and employ Formula (29.3.82) of Abramowitz and Stegun (1968, p. 1026) to obtain:

$$\hat{K}_W(w,s) = \int_{u=0}^{\infty} \exp[-\hat{X}(s)u]\Upsilon(w,u)du, \tag{49}$$

where

$$\Upsilon(w,u) = \frac{Y_1^{\infty} w}{(4\pi u^3)^{1/2}} \exp\left[-\frac{(Y_1^{\infty} w)^2}{4u}\right] H(w)H(u). \tag{50}$$

Since $\hat{\alpha}_1(s)$ and $\hat{\beta}_1(s)$ are system's response functions of the linear, time-invariant, causal, passive type, $\hat{X}(s)$ admits a partial-fraction parametrization of the type (36)–(38) and the time-domain counterpart of $\exp[-\hat{X}(s)u]$ follows from Eq. (44).

VII. SOME ILLUSTRATIVE NUMERICAL EXAMPLES

In the following, some illustrative numerical examples are presented. The source is placed at the boundary ($h=0$). Two receiver positions are considered, viz. one at the boundary ($r>0$, $z=0$), i.e., the propagation takes place parallel to the boundary, and one at the normal to the boundary through the source ($r=0$, $z>0$), i.e., the propagation takes place normal to the boundary. With regard to the boundary's acoustic admittance, two examples are discussed: (A) the zero-order (single-term) admittance and (B) the first-order (two-terms) admittance. Figure 2 shows the normalized incident-wave Green's function as a function of normalized time [cf. Eq. (6)].

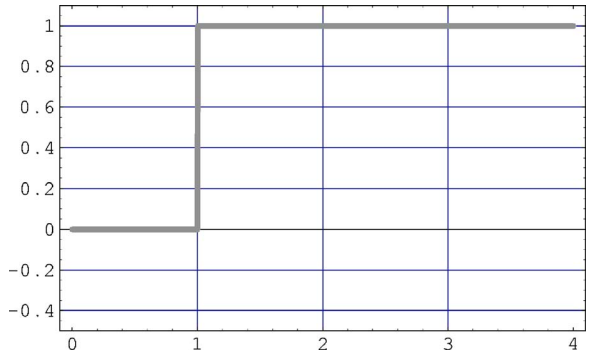


FIG. 2. Normalized incident-wave Green's function $4\pi D_0 G^i$ as a function of normalized time t/T_0.

A. Zero-order boundary admittance

For the zero-order boundary admittance we have

$$\hat{Y}(s) = Y^{\infty}, \tag{51}$$

which corresponds to the time-domain acoustic admittance

$$Y(t) = Y^{\infty}\delta(t) \tag{52}$$

and the time-domain boundary condition [cf. Eq.(3)]

$$v_z(x,y,0,t) = -(\rho_0 c_0)^{-1} Y^{\infty} p(x,y,0,t). \tag{53}$$

This section mainly serves to illustrate the influence of Y^{∞} on the reflection problem. Figure 3 shows the normalized reflected-wave Green's function as a function of normalized time [cf. Eqs. (27) and (28)] at (a) $r=10$ m, $z=0$ (propagation parallel to the boundary) and (b) $r=0$, $z=1$ m (propagation normal to the boundary), for three different values of Y^{∞}. Note that for propagation parallel to the boundary the normalized Green's function always starts at the value -1, irrespective of the value of Y^{∞}, while for propagation normal to the boundary the starting value is positive for $Y^{\infty}>1$, zero for $Y^{\infty}=1$ (admittance matched to the plane-wave value at normal incidence), and negative for $Y^{\infty}<1$.

B. First-order boundary admittance

For the first-order boundary admittance we have [cf. Eqs. (36)–(38)]

$$\hat{Y}(s) = Y^{\infty} + \frac{A_1}{s+\alpha_1}, \tag{54}$$

which we rewrite as

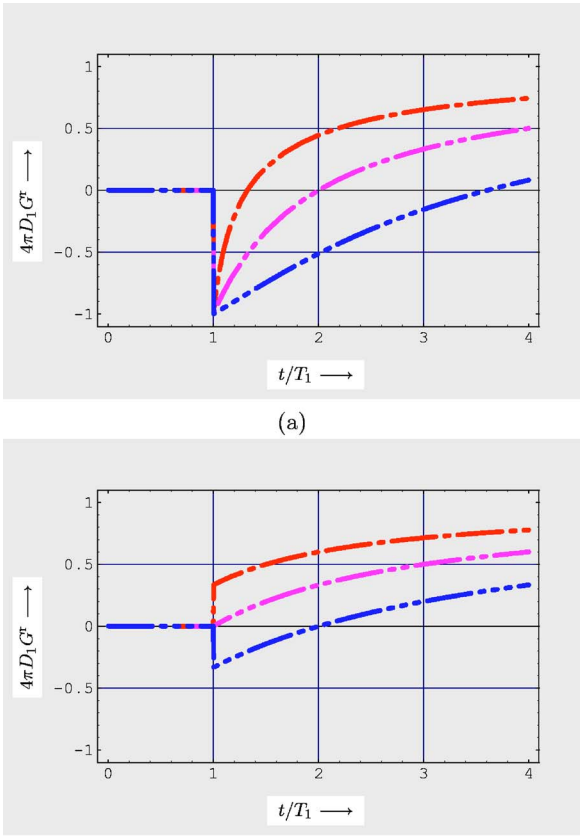


FIG. 3. Normalized reflected-wave Green's function $4\pi D_1 G^r$ as a function of normalized time t/T_1. Zero-order acoustic boundary admittance $Y=Y^\infty$. Source at boundary ($h=0$); $c_0=330$ m/s. (a) Propagation parallel to boundary ($r=10$ m, $z=0$), (b) propagation normal to boundary ($r=0$, $z=1$ m). Curves: (- . -) $Y^\infty=2.0$, (- .. -) $Y^\infty=1.0$ (matched to plane-wave value at normal incidence), (- ... -) $Y^\infty=0.5$.

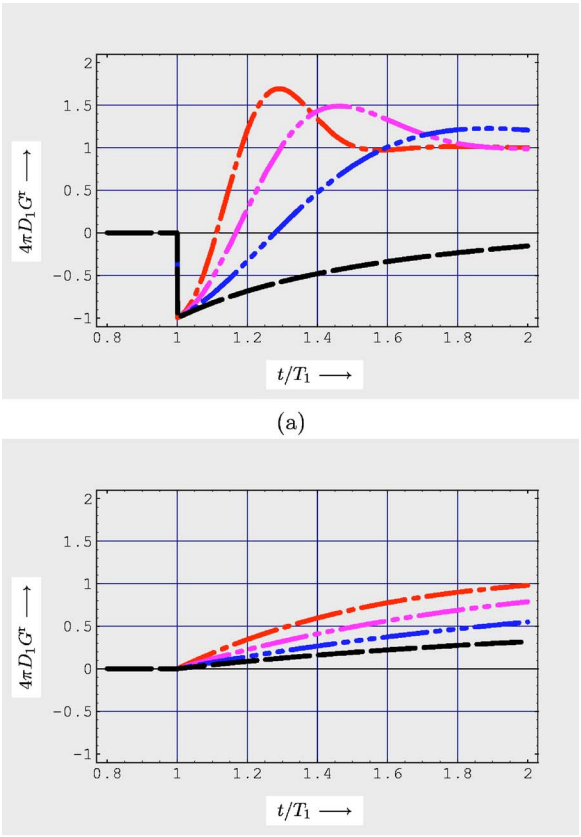


FIG. 4. Normalized reflected-wave Green's function $4\pi D_1 G^r$ as a function of normalized time t/T_1. First-order acoustic boundary admittance: $(\partial_t + 1/\tau_v)v = -(\rho_0 c_0)^{-1} Y^\infty (\partial_t + 1/\tau_p)p$ at boundary. Source at boundary ($h=0$); $Y^\infty=1.0$ (matched to plane-wave value at normal incidence), $c_0=330$ m/s. (a) Propagation parallel to boundary ($r=10$ m, $z=0$), (b) propagation normal to boundary ($r=0$, $z=1$ m). Curves: (- . -) $\tau_v=1.0\times10^{-3}$ s, $\tau_p=5.0\times10^{-2}$ s, $A_1=-9.8\times10^2$ s^{-1}; (- .. -) $\tau_v=2.0\times10^{-3}$ s, $\tau_p=5.0\times10^{-2}$ s, $A_1=-4.8\times10^2$ s^{-1}; (- ... -) $\tau_v=5.0\times10^{-3}$ s, $\tau_p=5.0\times10^{-2}$ s, $A_1=-1.8\times10^2$ s^{-1}; (- - -) $\tau_v=1.0\times10^{-1}$ s, $\tau_p=5.0\times10^{-2}$ s, $A_1=1.0\times10^1$ s^{-1}.

$$\hat{Y}(s) = Y^\infty \frac{s+z_1}{s+p_1}, \tag{55}$$

where $-p_1=-\alpha_1$ is the pole of $\hat{Y}(s)$ and $-z_1$ is the zero of $\hat{Y}(s)$, both located in the left half of the complex s plane, and

$$A_1 = Y^\infty(z_1 - p_1) \tag{56}$$

is the residue at the pole. Equations (54) and (55) correspond to the time-domain acoustic admittance

$$Y(t) = Y^\infty \delta(t) + A_1 \exp(-\alpha_1 t)H(t) \tag{57}$$

and the time-domain boundary condition [cf. Eq.(3)]

$$(\partial_t + 1/\tau_v)v_z(x,y,0,t) = -(\rho_0 c_0)^{-1} Y^\infty (\partial_t + 1/\tau_p)p(x,y,0,t), \tag{58}$$

where $\tau_v=1/p_1$ is the velocity relaxation time and $\tau_p=1/z_1$ is the pressure relaxation time (Christensen, 2003, pp. 17–19; Meinardi, 2002, p. 105). This section mainly serves to illustrate the influence of τ_v and τ_p on the reflection problem. Therefore, we take $Y^\infty=1$, which implies matching to the plane-wave admittance at normal incidence.

Figure 4 shows the normalized reflected-wave Green's function as a function of normalized time [cf. Eqs. (28) and (33)] at (a) $r=10$ m, $z=0$ (propagation parallel to the boundary) and (b) $r=0$, $z=1$ m (propagation normal to the boundary), for four different values of τ_v, with τ_p fixed. As Fig. 4(a) shows, strong oscillations occur at propagation parallel to the boundary, which phenomenon has been referred to in Sec. I. No such oscillations show up in the propagation normal to the boundary, as Fig. 4(b) shows. It can be argued that this behavior can be inferred from Eq. (31), where Γ_1 is related to the offset normal to the boundary and occurs in the damping exponential function, while Γ_2 is related to the offset parallel to the boundary and occurs in the oscillating Bessel functions. Apparently, such an easy interpretation does not apply to Eq. (43), where for $A_n>0$ the Bessel functions are oscillatory, while for $A_n<0$ they change into modified Bessel functions of the first kind that show a monotonic behavior.

VIII. DISCUSSION OF THE RESULTS

Via the combined applications of the modified Cagniard method and the Schouten–Van der Pol theorem of the unilateral Laplace transformation the time-domain acoustic pressure of the monopole (volume injection) point-source excited wave reflected against a locally reacting, absorptive and dis-

persive boundary has been expressed as a multiple sequence of operations acting on the source signature. Each of the kernel functions in the expression contains separately the configurational parameters of the measurement setup (location of source, receiver and boundary, and propagation through the fluid) and the parameters by which the absorptive and dispersive properties of the boundary can be characterized. Two parametrizations of the boundary's complex frequency domain acoustic admittance have been discussed in detail: the partial-fraction parametrization and the plane-wave admittance parametrization. The explicit attribution of a sequence of parameters to their corresponding kernel functions is conjectured to play an illuminating role in the use of the reflection measurement setup to characterize (via an appropriate inversion algorithm applied to the measured values of the acoustic pressure) the absorption and dispersion properties of the boundary, while the obtained expression itself is directly amenable to carry out the relevant parameter sensitivity analysis. It is noted that the multiple time convolutions that occur in the final expression for the acoustic pressure can numerically most profitably be evaluated through the use of the FFT algorithm.

ACKNOWLEDGMENT

The authors want to thank the (anonymous) reviewers for their constructive criticism and their suggestions for improving the paper.

APPENDIX: EVALUATION OF THE INTEGRAL IN EQ. (22)

In this Appendix the integral occurring in Eq. (22)

$$I = \frac{2}{\pi}\mathrm{Re}\left[\int_{\psi=0}^{\pi/2} \frac{1}{c_0\bar{\gamma}_0 + \hat{Y}(s)}d\psi\right]$$
$$= \frac{2}{\pi}\mathrm{Re}\left[\int_{\psi=0}^{\pi/2} \frac{1}{\Gamma_1 - i\Gamma_2\cos(\psi) + \hat{Y}(s)}d\psi\right]$$
$$= \frac{2}{\pi}\int_{\psi=0}^{\pi/2} \frac{\Gamma_1 + \hat{Y}(s)}{[\Gamma_1 + \hat{Y}(s)]^2 + \Gamma_2^2\cos^2(\psi)}d\psi, \tag{A1}$$

with [cf. Eqs. (24) and (25)]

$$\Gamma_1 = c_0\tau(z+h)/D_1^2, \tag{A2}$$

$$\Gamma_2 = c_0(\tau^2 - T_1^2)^{1/2}r/D_1^2, \tag{A3}$$

is evaluated. Using the standard integral

$$\frac{2}{\pi}\int_{\psi=0}^{\pi/2} \frac{A}{A^2 + B^2\cos^2(\psi)}d\psi = \frac{1}{(A^2+B^2)^{1/2}}, \tag{A4}$$

we obtain

$$I = \frac{1}{\{[\Gamma_1 + \hat{Y}(s)]^2 + \Gamma_2^2\}^{1/2}}. \tag{A5}$$

This result is used in the main text.

Abramowitz, M. and Stegun, I. A. (**1968**). *Handbook of Mathematical Functions* (Dover, Mineola, NY).

Cagniard, L. (**1962**). *Reflection and Refraction of Progressive Seismic Waves* (McGraw-Hill, New York), pp. 47–50 and p. 244. [Translation and revision of Cagniard, L. (**1939**). *Réflexion et Réfraction des Ondes Séismiques Progressives* edited by E. A. Flinn and C. H. Dix (Gauthier-Villars, Paris)].

Christensen, R. M. (**2003**). *Theory of Viscoelasticity*, 2nd ed. (Dover, Mineola, NY).

Daigle, G. A., Stinson, M. R., and Havelock, D. I. (**1996**). "Experiments on surface waves over a model impedance plane using acoustical pulses," J. Acoust. Soc. Am. **99**, 1993–2005.

De Hoop, A. T. (**1960**). "A modification of Cagniard's method for solving seismic pulse problems," Appl. Sci. Res., Sect. B **8**, 349–356.

De Hoop, A. T. (**1995**). *Handbook of Radiation and Scattering of Waves* (Academic, London).

De Hoop, A. T. (**2002**). "Reflection and transmission of a transient, elastic line-source excited SH-wave by a planar, elastic bounding surface in a solid," Int. J. Solids Struct. **39**, 5379–5391.

De Hoop, A. T. and Van der Hijden, J. H. M. T. (**1984**). "Generation of acoustic waves by an impulsive point source in a fluid/solid configuration with a plane boundary," J. Acoust. Soc. Am. **75**, 1709–1715.

Donato, R. J. (**1976a**). "Propagation of a spherical wave near a plane boundary with a complex impedance," J. Acoust. Soc. Am. **60**, 34–39.

Donato, R. J. (**1976b**). "Spherical-wave reflection from a boundary of reactive impedance using a modification of Cagniard's method," J. Acoust. Soc. Am. **60**, 999–1002.

Ingard, K. U. (**1951**). "On the reflection of a spherical sound wave from an infinite plane," J. Acoust. Soc. Am. **23**, 329–335.

Ju, H. B. and Fung, K. Y. (**2002**). "Time-domain simulation of acoustic sources over an impedance plane," J. Comput. Acoust. **10**, 311–329.

Kwakernaak, H. and Sivan, R. (**1991**). *Modern Signals and Systems* (Prentice–Hall, Englewood Cliffs, NJ), pp. 463–466.

Lam, C. H., Kooij, B. J., and De Hoop, A. T. (**2004**). "Impulsive sound reflection from an absorptive and dispersive planar boundary," J. Acoust. Soc. Am. **118**, 677–685.

Meinardi, F. (**2002**). "Linear viscoelasticity", in *Acoustic Interactions with Submerged Elastic Structures*, edited by A. Uran, A. Boström, O. Leroy, and G. Maze, (World Scientific, Englewood Cliffs, NJ), pp. 97–126.

Schouten, J. P. (**1934**). "A new theorem in operational calculus together with an application of it," Physica (Amsterdam) **1**, 75–80.

Schouten, J. P. (**1961**). *Operatorenrechnung* (Springer, Berlin), pp. 124–126.

Thomasson, S. I. (**1976**). "Reflection of waves from a point source by an impedance boundary," J. Acoust. Soc. Am. **59**, 780–785.

Van den Nieuwenhof, B. and Coyette, J. P. (**2001**). "Treatment of frequency-dependent admittance boundary conditions in transient acoustic finite/infinite-element models," J. Acoust. Soc. Am. **110**, 1743–1751.

Van der Pol, B. (**1934**). "A theorem on electrical networks with an application to filters," Physica (Amsterdam) **1**, 521–530.

Van der Pol, B. and Bremmer, H. (**1950**). *Operational Calculus Based on the Two-sided Laplace Transform* (Cambridge University Press, Cambridge, UK), pp. 232–236.

Wenzel, S. R. (**1974**). "Propagation of waves along an impedance boundary," J. Acoust. Soc. Am. **55**, 956–963.

Widder, D. V. (**1946**). *The Laplace Transform* (Princeton University Press, Princeton, NJ), pp. 63–65.

A time-domain model of transient acoustic wave propagation in double-layered porous media

Z. E. A. Fellah and A. Wirgin
Laboratoire de Mécanique et d'Acoustique, CNRS-UPR 7051, 31 chemin Joseph Aiguier, Marseille, 13009, France

M. Fellah
Laboratoire de Physique Théorique, Institut de Physique, USTHB, BP 32 El Alia, Bab Ezzouar 16111, Algeria

N. Sebaa and C. Depollier
Laboratoire d'Acoustique de l'Université du Maine, UMR-CNRS 6613, Université du Maine, Avenue Olivier Messiaen, 72085 Le Mans Cedex 09, France

W. Lauriks
Laboratorium voor Akoestiek en Thermische Fysica, Katholieke Universiteit Leuven, Celestijnenlaan 200 D, B-3001 Heverlee, Belgium

(Received 1 March 2005; revised 10 May 2005; accepted 23 May 2005)

This paper concerns a time-domain model of transient wave propagation in double-layered porous materials. An analytical derivation of reflection and transmission scattering operators is given in the time domain. These scattering kernels are the medium's responses to an incident acoustic pulse. The expressions obtained take into account the multiple reflections occurring at the interfaces of the double-layered material. The double-layered porous media consist of two slabs of homogeneous isotropic porous materials with a rigid frame. Each porous slab is described by a temporal equivalent fluid model, in which the acoustic wave propagates only in the fluid saturating the material. In this model, the inertial effects are described by the tortuosity; the viscous and thermal losses of the medium are described by two susceptibility kernels which depend on the viscous and thermal characteristic lengths. Experimental and numerical results are given for waves transmitted and reflected by double-layered porous media formed by air-saturated plastic foam samples. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1953247]

PACS number(s): 43.20.Bi, 43.20.Hq, 43.20.Gp, 43.20.El [ANN] Pages: 661–670

I. INTRODUCTION

The ultrasonic characterization of porous materials saturated by air[1,2] is of great interest for a large class of industrial applications. These materials are frequently used in the automotive and aeronautics industries and in the building trade.

Ultrasonic characterization of materials is often achieved by measuring the attenuation coefficient and phase velocity in the frequency domain,[3,4] or by solving the direct and inverse problems directly in the time domain.[5–13] In the frequency domain, measurements of the attenuation coefficient may be more robust than measurements of phase velocity. In these situations, the application of the Kramers–Kronig[11–13] dispersion relations may allow the determination of the phase velocity from the measured attenuation coefficient.

Many applications, such as medical imaging or inverse scattering,[14] require a study of the behavior of pulses traveling into porous media.[3,4,8–11] When a broadband ultrasound pulse passes through a layer of a medium, the pulse waveform changes as a result of attenuation and dispersion of the medium. The classic method for predicting a change in the waveform of a signal passing through a medium relies on the system's impulse response. According to the theory of linear systems,[15] the output signal is a convolution of the input signal and the system impulse response. Many media, including porous materials and soft tissues, have been observed to have an attenuation function that increases with frequency.[16] As a result, higher frequency components of the pulse are attenuated more than lower frequency components. After passing through the layer, the transmitted pulse is not just a scaled-down version of the incident pulse, but has a different shape. Dispersion refers to the phenomenon observed when the phase velocity of a propagating wave changes with frequency.[17] Dispersion causes the propagating pulse waveform to change because wave components with different frequencies travel at different speeds. An understanding of the interaction of ultrasound with a porous medium in both the time and frequency domains, and the ability to determine the change of waveform when propagating ultrasound pulses, should be useful in designing array transducers and in quantitative ultrasound tissue characterization.[18,19]

This time-domain model is an alternative to the classical frequency-domain approach.[1,3,4] It is an advantage of the time domain that the results are immediate and direct.[5–13] The attractive feature of a time-domain-based approach is that the analysis is naturally bounded by the finite duration of ultrasonic pressures, and is consequently the most appropri-

ate approach for the transient signal. However, for wave propagation generated by time-harmonic incident waves and sources (monochromatic waves), the frequency analysis is more appropriate.[1] A time-domain approach differs from frequency analysis in that the susceptibility functions describing viscous and thermal effects are convolution operators acting on velocity and pressure, and therefore a different algebraic formalism must be applied to solve the wave equation. The time-domain response of the material is described by an instantaneous response and a "susceptibility" kernel responsible for memory effects.

In the past, many authors have used fractional calculus[20] as an empirical method to describe the properties of viscoelastic materials, e.g., see Caputo[21] and Bagley and Torvik.[22] The observation that asymptotic expressions of stiffness and damping in porous materials are proportional to the fractional powers of frequency[23] suggests that time derivatives of a fractional order might describe the behavior of sound waves in this kind of material, including relaxation and frequency dependence. In this work, fractional calculus is used to describe viscous and thermal interaction between the fluid and the structure in double-layered porous media consisting of two slabs of homogeneous porous materials. Given the medium's response to an incident pulse, reflection and transmission scattering operators are calculated for double-layered porous media. Experimental results are compared with theoretical predictions, giving good correlation.

The outline of this paper is as follows. Section II shows a temporal equivalent fluid model, the connection between the fractional derivatives and the wave propagation in rigid porous media in high-frequency range is established, and the basic equations are written in the time domain. Sections III and IV are devoted to formulating the problem and analytical derivation of the reflection and transmission scattering kernels for double-layered porous media consisting of two slabs of homogeneous porous materials. The scattering responses of the media take into account the multiple reflections at the double-layered porous media interfaces. Finally, in Sec. V, experimental validation using ultrasonic measurement in transmission and reflection is discussed for air-saturated industrial plastic foams.

II. TEMPORAL EQUIVALENT FLUID MODEL

The quantities involved in sound propagation in porous materials can be defined locally, on a microscopic scale. However, this study is generally difficult because of the complicated frame geometries. Only the mean values of the quantities involved are of practical interest. Averaging must be performed on a macroscopic scale, using volumes with large enough dimensions for the average to be significant. At the same time, these dimensions must be much smaller than the wavelength. Even on a macroscopic scale, describing sound propagation in porous material can be very complicated, since sound also propagates in the frame of the material. If the frame is motionless, the porous material can be replaced on a macroscopic scale by an equivalent fluid.

In porous material acoustics, a distinction can be made between two situations depending on whether the frame is moving or not. In the first case, the wave dynamics due to coupling between the solid frame and the fluid is clearly described by the Biot theory.[24,25] In air-saturated porous media, the structure is generally motionless and the waves propagate only in the fluid. This case is described by the equivalent fluid model which is a particular case in the Biot model, in which fluid–structure interactions are taken into account by the viscous susceptibility kernel, χ_v, and the thermal susceptibility kernel, χ_{th}, as follows:[8,26]

$$\rho_f \alpha_\infty \partial_t \mathbf{v}(\mathbf{r},t) + \int_0^t \chi_v(t-t')\partial_t \mathbf{v}(\mathbf{r},t')dt' = -\nabla p(\mathbf{r},t), \quad (1)$$

$$\frac{1}{K_a}\partial_t p(\mathbf{r},t) + \int_0^t \chi_{th}(t-t')\partial_t p(\mathbf{r},t')dt' = -\nabla \mathbf{v}(\mathbf{r},t). \quad (2)$$

Constitutive relations in the time domain result from arguments based on invariance under time translation and causality.[11–13] In these equations, p is the acoustic pressure, $\mathbf{v}$ is the particle velocity, ρ_f and K_a are the density and compressibility modulus of the fluid, respectively. The parameter α_∞ reflects the instantaneous response of the porous medium and describes the inertial coupling between fluid and structure. Instantaneous therefore means that the response time is much shorter than the typical time scale for acoustic field variation. The susceptibility kernels χ_v and χ_{th} are memory functions which determine the dispersion of the medium.

The medium varies with depth x only, and the incident wave is planar and normally incident. With no lack of generality, the pressure acoustic field can be assumed to have only one component, denoted $p(x,t)$. It is assumed that the pressure field in the medium is zero prior to $t=0$. The wave equation for the pressure acoustic field of a porous dispersive medium with a rigid frame is obtained from the constitutive equations (1) and (2), and is in the form

$$\partial_x^2 p(x,t) - \frac{1}{c_0^2}\Bigg[\alpha_\infty \partial_t^2 p(x,t) + \Bigg(\alpha_\infty K_a \chi_{th} + \frac{\chi_v}{\rho_f} + c_0^2 \chi_{th} * \chi_v\Bigg) * \partial_t^2 p(x,t)\Bigg] = 0, \quad (3)$$

where $c_0 = (K_a/\rho_f)^{1/2}$ is the speed of free fluid. The following notation is used for the convolution integral:

$$[f * g](x,t) = \int_0^t f(x,t-t')g(x,t')dt'. \quad (4)$$

Viscous and thermal exchanges between a fluid-saturated porous medium and its structure are responsible for acoustic field damping. In the asymptotic regime, corresponding to high-frequency limit,[8,26] viscous and thermal interactions are modeled by the memory relaxation operators $\chi_v(t)$ and $\chi_{th}(t)$ given by[26]

$$\chi_v(t) = \frac{2\rho_f \alpha_\infty}{\Lambda}\sqrt{\frac{\eta}{\pi \rho_f}}t^{-1/2}, \quad (5)$$

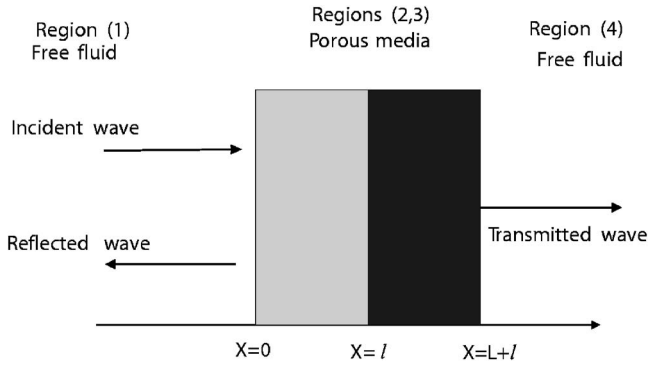


FIG. 1. Problem geometry.

$$\chi_{th}(t) = \frac{2(\gamma-1)}{K_a\Lambda'}\sqrt{\frac{\eta}{\pi P_r \rho_f}}t^{-1/2}, \tag{6}$$

where P_r is the Prandtl number, η and γ are the fluid viscosity and adiabatic constant, respectively. The model's relevant physical parameters are tortuosity α_∞, and viscous and thermal characteristic lengths, Λ and Λ'. In this model the time convolution of $t^{-1/2}$ with a function is interpreted as a semiderivative operator according to the definition of the fractional derivative of order ν given[27] by

$$D^\nu[x(t)] = \frac{1}{\Gamma(-\nu)}\int_0^t (t-u)^{-\nu-1}x(u)du, \tag{7}$$

where $\Gamma(x)$ is the gamma function.

III. FORMULATING THE PROBLEM

Consider a double-layered porous medium consisting of two homogeneous slabs with different acoustic parameters. The geometry of the problem is shown in Fig. 1. The first porous slab occupies the region $0 \leqslant x \leqslant \ell$ and the second one occupies the region $\ell \leqslant x \leqslant L$. Each porous slab is assumed to be isotropic and to have a rigid frame. A short sound pulse impinges normally on the medium from the left [free fluid—region (1)]. It gives rise to an acoustic pressure field $p_i(x,t)$, $i=2,3$ and an acoustic velocity field $v_i(x,t)$, $i=2,3$ within each layer [the symbol $i=2,3$ denotes regions (2) and (3), respectively] which satisfy the propagation equation (3).

It is assumed that the pressure field is continuous at the boundary of each layer

$$p(0^+,t) = p(0^-,t), \quad p(\ell^-,t) = p(\ell^+,t),$$

$$p[(\ell+L)^-,t] = p[(\ell+L)^+,t] \tag{8}$$

(where ± superscript denotes the limit from left and right, respectively) and the initial conditions

$$p(x,t)|_{t=0} = 0 \qquad \left.\frac{\partial p}{\partial t}\right|_{t=0} = 0, \tag{9}$$

which mean that the medium is idle for $t=0$.

If the incident sound wave is launched in region $x \leqslant 0$ [region (1)], then the pressure field in the region on the left of the double-layered material is expressed as the sum of the incident and reflected fields

$$p_1(x,t) = p^i\left(t-\frac{x}{c_0}\right) + p^r\left(t+\frac{x}{c_0}\right), \quad x<0, \tag{10}$$

where $p_1(x,t)$ is the field in region $x<0$, p^i and p^r denote the incident and reflected fields, respectively. A transmitted field is also produced in the region on the right of the double-layered material. This takes the form

$$p_4(x,t) = p^t\left(t-\frac{\ell}{c_2}-\frac{L}{c_3}-\frac{x-\ell-L}{c_0}\right), \quad x>\ell+L. \tag{11}$$

[$p_4(x,t)$ is the field in region (4): $x>\ell+L$ and p^t the transmitted field]. c_2 and c_3 represent the acoustic velocities in regions (2) and (3), respectively, defined by the relation $c_i = c_0/(\alpha_i)^{1/2}$, $i=2,3$, where α_i represents the tortuosity of each porous layer.

The incident and scattered fields are related by scattering operators (i.e., reflection and transmission operators) for the material. These are integral operators represented by

$$p^r(x,t) = \int_0^t \tilde{R}(\tau)p^i\left(t-\tau+\frac{x}{c_0}\right)d\tau$$

$$= \tilde{R}(t) * p^i(t) * \delta\left(t+\frac{x}{c_0}\right), \tag{12}$$

$$p^t(x,t) = \int_0^t \tilde{T}(\tau)p^i\left(t-\frac{\ell}{c_2}-\frac{L}{c_3}-\frac{x-\ell-L}{c_0}\right)d\tau$$

$$= \tilde{T}(t) * p^i(t) * \delta\left(t-\frac{\ell}{c_2}-\frac{L}{c_3}-\frac{x-\ell-L}{c_0}\right), \tag{13}$$

where $\delta(t)$ is the Dirac distribution. In Eqs. (12) and (13) the functions $\tilde{R}$ and $\tilde{T}$ are the respective reflection and transmission kernels for incidence from the left. Note that the lower limit of integration in (12) and (13) is set to 0, which is equivalent to assuming that the incident wavefront first impinges on the material at $t=0$. The operators $\tilde{R}$ and $\tilde{T}$ are independent of the incident field used in the scattering experiment and depend only on the properties of the materials.

In region $x \leqslant 0$, the field $p_1(x,t)$ is given by

$$p_1(x,t) = \left[\delta\left(t-\frac{x}{c_0}\right) + \tilde{R}(t) * \delta\left(t+\frac{x}{c_0}\right)\right] * p^i(t). \tag{14}$$

To simplify the analysis, we will use the Laplace transform which is appropriate for our problem. We note $\tilde{P}_i(x,z)$, $i=1,2,3,4$, the Laplace transform of $p_i(x,t)$, $i=1,2,3,4$ defined by

$$\tilde{P}_i(x,z) = \mathcal{L}[p_i(x,t)] = \int_0^\infty \exp(-zt)p_i(x,t)dt. \tag{15}$$

The Laplace transform of the field outside the double-layered medium is given by

$$\tilde{P}_1(x,z) = \left[\exp\left(-z\frac{x}{c_0}\right) + R(z)\exp\left(z\frac{x}{c_0}\right)\right]\varphi(z), \quad x \leqslant 0, \tag{16}$$

$$\tilde{P}_4(x,z) = T(z)\exp\left[-\left(\frac{\ell}{c_2}+\frac{L}{c_3}+\frac{x-\ell-L}{c_0}\right)z\right]\varphi(z),$$
$$x \geqslant \ell + L, \tag{17}$$

Here, $\tilde{P}_1(x,z)$ and $\tilde{P}_4(x,z)$ are the Laplace transform of the field on the left and right of the double-layered porous media, respectively, $\varphi(z)$ denotes the Laplace transform of the incident field $p^i(t)$, and finally $R(z)$ and $T(z)$ are the Laplace transforms of the reflection and transmission kernels, respectively.

The acoustic pressure fields $p_i(x,t)$, $i=2,3$ inside each layer of the porous media [regions (2) and (3)] satisfy the propagation equation (3), which can be written in the Laplace domain as

$$\frac{\partial^2 \tilde{P}_i(x,z)}{\partial x^2} - \frac{f_i(z)}{c_i^2}\tilde{P}_i(x,z) = 0, \quad i=2,3, \quad 0 \leqslant x \leqslant \ell + L. \tag{18}$$

The function $f_i(z)$ is given by the following expression:

$$f_i(z) = z^2 c_i^2[\rho_f \alpha_i + \tilde{\chi}_{vi}(z)] \cdot [1/Ka + \tilde{\chi}_{thi}(z)], \quad i=2,3, \tag{19}$$

where $\tilde{\chi}_{vi}(z)$ and $\tilde{\chi}_{thi}(z)$ represent the Laplace transform of $\chi_{vi}(t)$ and $\chi_{thi}(t)$, respectively, and their expressions in the time domain are given by

$$\chi_{vi}(t) = \frac{2\rho_f \alpha_i}{\Lambda_i}\sqrt{\frac{\eta}{\pi\rho_f}}t^{-1/2}, \quad i=2,3. \tag{20}$$

$$\chi_{thi}(t) = \frac{2(\gamma-1)}{K_a \Lambda_i'}\sqrt{\frac{\eta}{\pi P_r \rho_f}}t^{-1/2}, \quad i=2,3. \tag{21}$$

Λ_i and Λ_i', $i=2,3$ are the viscous and thermal characteristic lengths of each porous layer.

By developing expression (19), we obtain the following relation for $f_i(z)$:

$$f_i(z) = z^2 + 2\sqrt{\frac{\eta}{\rho}}\left(\frac{1}{\Lambda_i}+\frac{\gamma-1}{\sqrt{P_r}\Lambda_i'}\right)z\sqrt{z} + \frac{4(\gamma-1)\eta}{\rho_f \Lambda_i \Lambda_i' \sqrt{P_r}}z, \quad i=2,3. \tag{22}$$

The solution of Eq. (18) can be expressed by

$$\tilde{P}_i(x,z) = \left[A_i(z)\exp\left(-\frac{\sqrt{f_i(z)}}{c_i}x\right) + B_i(z)\exp\left(\frac{\sqrt{f_i(z)}}{c_i}x\right)\right]\varphi(z), \quad i=2,3, \tag{23}$$

where coefficients $A_i(z)$ and $B_i(z)$ can be determined by the physical conditions at the boundary of each layer. This is given in the next section.

IV. REFLECTION AND TRANSMISSION SCATTERING OPERATORS

To derive reflection and transmission coefficients, we use the continuity relations of the acoustic pressure fields [Eqs. (8)] given in the Laplace domain by

$$\tilde{P}_1(0^-,z) = \tilde{P}_2(0^+,z), \quad \tilde{P}_2(\ell^-,z) = \tilde{P}_3(\ell^+,z),$$
$$\tilde{P}_3[(\ell+L)^-,z] = \tilde{P}_4((\ell+L)^+,z). \tag{24}$$

Using the expressions of the pressure fields in each layer [Eqs. (16), (17), and (23)] and the conditions (24), we obtain the followings relations for the coefficients $A_i(z)$ and $B_i(z)$, $i=2,3$, and the coefficients $R(z)$ and $T(z)$:

$$A_2(z) + B_2(z) = \tilde{P}_2(0,z) = 1 + R(z), \tag{25}$$

$$A_2(z)\exp\left(-\frac{\sqrt{f_2(z)}}{c_2}\ell\right) + B_2(z)\exp\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right) = A_3(z)\exp\left(-\frac{\sqrt{f_3(z)}}{c_3}\ell\right) + B_3(z)\exp\left(\frac{\sqrt{f_3(z)}}{c_3}\ell\right), \tag{26}$$

$$A_3(z)\exp\left(-\frac{\sqrt{f_3(z)}}{c_3}(\ell+L)\right) + B_3(z)\exp\left(\frac{\sqrt{f_3(z)}}{c_3}(\ell+L)\right) = T(z)\exp\left[-\left(\frac{\ell}{c_2}+\frac{L}{c_3}\right)z\right]. \tag{27}$$

The Euler equation in each region is written as

$$\rho_f \alpha_i \partial_t v_i(x,t) + \chi_{vi}(t) * \partial_t v_i(x,t) = -\partial_x p_i(x,t), \quad i=1,\ldots,4. \tag{28}$$

In these equations $v_i(x,t)$ $i=1,\ldots,4$ is the acoustic velocity field in regions (1),..., (4). Note that in regions (1) and (4), corresponding to the free fluid, porosity and tortuosity values are 1 ($\alpha_1=\alpha_4=1$ and $\phi_1=\phi_4=1$), and the viscous susceptibility kernel vanishes ($\chi_{vi}=0, i=1,4$) outside the double-layered porous media.

The equation for flow continuity between each interface ($x=0$, $x=\ell$, and $x=\ell+L$) is given by

$$\phi_i v_i(x,t) = \phi_{i+1} v_{i+1}(x,t), \quad i=1,2,3, \tag{29}$$

where ϕ_i, $i=1,\ldots,4$ is the porosity of each layer.

Using the relations (28) and (29), we obtain the following relations between the acoustic pressure $p_i(x,t)$ and physical properties of each layer:

$$\phi_{i+1}[\rho_f \alpha_i \partial_x p_{i+1}(x,t) + \chi_{vi}(t) * \partial_x p_{i+1}(x,t)] = \phi_i[\rho_f \alpha_{i+1}\partial_x p_i(x,t) + \chi_{v(i+1)}(t) * \partial_x p_i(x,t)], \quad i=1,2,3. \tag{30}$$

Using the Laplace transform of Eq. (30) and the pressure field expressions for each layer [Eqs. (16), (17), and (23)], we obtain the following relations at the interface of each layer:

$$B_2(z) - A_2(z) = K_1(R(z)-1), \tag{31}$$

$$B_3(z)\exp\left(\frac{\sqrt{f_3(z)}}{c_3}\ell\right) - A_3(z)\exp\left(-\frac{\sqrt{f_3(z)}}{c_3}\ell\right) = K_2\left[B_2(z)\exp\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right) - A_2(z)\exp\left(-\frac{\sqrt{f_2(z)}}{c_2}\ell\right)\right], \tag{32}$$

$$B_3(z)\exp\left(\frac{\sqrt{f_3(z)}}{c_3}(\ell+L)\right)-A_3(z)\exp\left(-\frac{\sqrt{f_3(z)}}{c_3}(\ell+L)\right)$$
$$=K_3T(z)\exp\left[-\left(\frac{\ell}{c_2}+\frac{L}{c_3}\right)z\right], \tag{33}$$

with

$$K_1=\frac{\sqrt{\alpha_2}}{\phi_2},\quad K_2=\frac{\phi_2\sqrt{\alpha_3}}{\phi_3\sqrt{\alpha_2}},\quad K_3=\frac{\sqrt{\alpha_3}}{\phi_3},\quad K_1K_2=K_3. \tag{34}$$

Using the relations (25)–(27) and (31)–(33) (see Appendix A), we obtain the following expressions of the reflection and transmission coefficients:

$$R(z)=d_1\frac{\Gamma(z)}{\Xi(z)}, \tag{35}$$

$$T(z)=h_1\exp\left[\left(\frac{\ell}{c_2}+\frac{L}{c_3}\right)z\right]\frac{\exp\left(-2\frac{\sqrt{f_2(z)}}{c_2}\ell-\frac{\sqrt{f_3(z)}}{c_3}L\right)}{\Xi(z)}, \tag{36}$$

with

$$\Gamma(z)=1+d_2\exp\left(-2\frac{\sqrt{f_2(z)}}{c_2}\ell\right)+d_3\exp\left(-2\frac{\sqrt{f_3(z)}}{c_3}L\right)$$
$$-d_4\exp\left(-2\frac{\sqrt{f_2(z)}}{c_2}\ell-2\frac{\sqrt{f_3(z)}}{c_3}L\right),$$

$$\Xi(z)=1+h_2\exp\left(-2\frac{\sqrt{f_2(z)}}{c_2}\ell\right)+h_3\exp\left(-2\frac{\sqrt{f_3(z)}}{c_3}L\right)$$
$$+h_4\exp\left(-2\frac{\sqrt{f_2(z)}}{c_2}\ell-2\frac{\sqrt{f_3(z)}}{c_3}L\right).$$

$$d_1=\frac{K_1-1}{K_1+1},\quad d_2=\frac{(K_1+1)(K_2-1)}{(K_1-1)(K_2+1)},$$

$$d_3=\frac{(K_3-1)(K_2-1)}{(K_3+1)(K_2+1)},\quad d_4=-\frac{(K_3-1)(K_1+1)}{(K_3+1)(K_1-1)},$$

$$h_1=\frac{4K_1K_2}{(1+K_3)(1+K_1+K_2+K_3)},$$

$$h_2=\frac{(1-K_2)(1-K_1)}{(1+K_1)(1+K_2)},\quad h_3=\frac{(K_3-1)(1-K_2)}{(K_3+1)(1+K_2)},$$

$$h_4=\frac{(1-K_3)(K_1-1)}{(1+K_3)(K_1+1)}.$$

To express n-multiple reflections in porous layers, we shall write the reflection and transmission coefficients as follows:

$$R(z)=d_1\Gamma(z)\sum_{n\geqslant 0}(-1)^n(\Xi(z)-1)^n, \tag{37}$$

$$T(z)=h_1\exp\left[\left(\frac{\ell}{c_2}+\frac{L}{c_3}\right)z\right]\exp\left(-2\frac{\sqrt{f_2(z)}}{c_2}\ell-\frac{\sqrt{f_3(z)}}{c_3}(L)\right)\sum_{n\geqslant 0}(-1)^n(\Xi(z)-1)^n. \tag{38}$$

Using the identity

$$(x+y+z)^n=\sum_{n_1+n_2+n_3=n}\frac{n!}{n_1!n_2!n_3!}x^{n_1}y^{n_2}z^{n_3}, \tag{39}$$

where $n!=\Gamma(n+1)$, the reflection and transmission expressions become

$$R(z)=d_1\Gamma(z)\sum_{n\geqslant 0}(-1)^n n!\sum_{n_1+n_2+n_3=n}\frac{h_2^{n_1}h_3^{n_2}h_4^{n_3}}{n_1!n_2!n_3!}$$
$$\times\exp\left(-2\frac{\sqrt{f_2(z)}}{c_2}(n_1+n_3)\ell\right)$$
$$\times\exp\left(-2\frac{\sqrt{f_3(z)}}{c_3}(n_2+n_3)L\right), \tag{40}$$

and

$$T(z)=h_1\exp\left[\left(\frac{\ell}{c_2}+\frac{L}{c_3}\right)z\right]$$
$$\times\sum_{n\geqslant 0}(-1)^n n!\sum_{n_1+n_2+n_3=n}\frac{h_2^{n_1}h_3^{n_2}h_4^{n_3}}{n_1!n_2!n_3!}$$
$$\times\exp\left(-\frac{\sqrt{f_2(z)}}{c_2}(2n_1+2n_3+1)\ell\right)$$
$$\times\exp\left(-\frac{\sqrt{f_3(z)}}{c_3}(2n_2+2n_3+1)L\right). \tag{41}$$

By setting $z=j\omega$, where $j^2=-1$ and ω is the angular frequency, we can easily deduce the expressions of the reflection and transmission coefficients in the frequency domain.

Recall that the inverse Laplace transforms of $\exp[-(\ell/c_2)\sqrt{f_2(z)}]$ and $\exp[-(L/c_3)\sqrt{f_3(z)}]$ are the Green function[28] of the first and second porous slab, respectively.

In the time domain, the transmission scattering operator is expressed as

$$\tilde{T}(t)=h_1\sum_{n\geqslant 0}(-1)^n n!\sum_{n_1+n_2+n_3=n}\frac{h_2^{n_1}h_3^{n_2}h_4^{n_3}}{n_1!n_2!n_3!}F_2\left[t+\frac{\ell}{c_2},(2n_1+2n_3+1)\frac{\ell}{c_2}\right]*F_3\left[t+\frac{L}{c_3},(2n_2+2n_3+1)\frac{L}{c_3}\right], \tag{42}$$

where F_i, $i=2,3$ is the Green function[28] of porous layers (2) and (3), respectively (see Appendix B).

The reflection scattering operator is expressed by the relation

$$\tilde{R}(t) = d_1 \sum_{n\geqslant 0} (-1)^n n! \sum_{n_1+n_2+n_3=n} \frac{h_2^{n_1} h_3^{n_2} h_4^{n_3}}{n_1! n_2! n_3!} \times \left[F_2\left[t, 2(n_1+n_3)\frac{\ell}{c_2} \right] * F_3\left[t, 2(n_2+n_3)\frac{L}{c_3} \right] + d_2 F_2\left[t, 2(n_1+n_3+1)\frac{\ell}{c_2} \right] * F_3\left[t, 2(n_2+n_3)\frac{L}{c_3} \right] + d_3 F_2\left[t, 2(n_1+n_3)\frac{\ell}{c_2} \right] * F_3\left[t, 2(n_2+n_3+1)\frac{L}{c_3} \right] + d_4 F_2\left[t, 2(n_1+n_3+1)\frac{\ell}{c_2} \right] * F_3\left[t, 2(n_2+n_3+1)\frac{L}{c_3} \right] \right]. \quad (43)$$

If only the first reflections at interfaces $x=0$, $x=\ell$ and $x=L$ are taken into account, the reflection scattering kernel expression becomes

$$\tilde{R}(t) = d_1\delta(t) + d_1(d_2-h_2) F_2\left[t, 2\frac{\ell}{c_2} \right] + d_1(d_4 - h_4 - d_2 h_3 - d_3 h_2 + 2h_2 h_3) F_2\left[t, 2\frac{\ell}{c_2} \right] * F_3\left[t, 2\frac{L}{c_3} \right]. \quad (44)$$

The transmission scattering kernel describing the direct wave transmitted through two layers of porous materials without internal reflection is expressed by

$$\tilde{T}(t) = h_1 F_2\left[t + \frac{\ell}{c_2}, \frac{\ell}{c_2} \right] * F_3\left[t + \frac{L}{c_3}, \frac{L}{c_3} \right] = h_1 \int_0^t F_2\left[\tau + \frac{\ell}{c_2}, \frac{\ell}{c_2} \right] F_3\left[t - \tau + \frac{L}{c_3}, \frac{L}{c_3} \right] d\tau. \quad (45)$$

The first term on the right-hand side of Eq. (44), $d_1\delta(t) = [(\sqrt{\alpha_2}-\phi_2)/(\sqrt{\alpha_2}+\phi_2)]\delta(t)$, is equivalent to the instantaneous reflected response of the first layer (region 2). The part of the wave which is equivalent to this term corresponds to the wave reflected by the first interface $x=0$ of the first porous layer. It depends only on the porosity and tortuosity of the first porous slab. The wave reflected by the first interface has the advantage of not being dispersive, but simply attenuated by factor d_1. This result is in agreement with the conclusions obtained in other works for wave reflected by a slab of porous material.[9,29,30] This shows that it is possible to measure the porosity and tortuosity of the first porous layer just by measuring its first reflected wave.

The second term on the right-hand side of Eq. (44), $d_1(d_2-h_2)F_2[t, 2(\ell/c_2)]$, corresponds to the second interface reflection contribution, $x=\ell$. This term depends on the porosity and tortuosity of the two porous layers. The Green's function F_2 describes the propagation and dispersion of an acoustic wave making one round trip inside the first porous slab. This Green's function depends on the viscous and thermal characteristics lengths Λ and Λ' of the first layer (region 2). This result means that it can be possible to get information of all the acoustical properties (porosity, tortuosity, viscous and thermal characteristics lengths) of the first porous layer (region 2), and also, the porosity and tortuosity of the second porous layer (region 3), but the viscous and thermal characteristics lengths of the second porous layer (region 3) do not intervene.

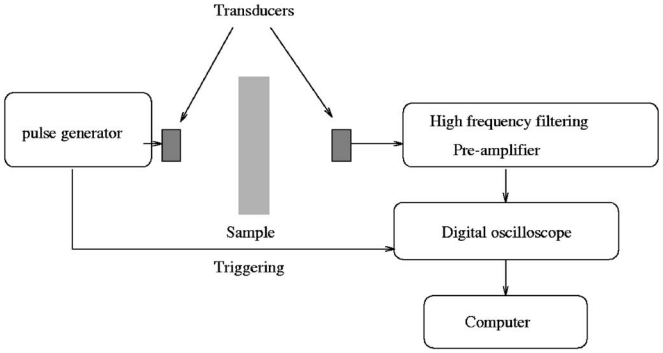


FIG. 2. Experimental setup of the ultrasonic measurements in transmitted mode.

Finally, the term $d_1(d_4-h_4-d_2h_3-d_3h_2+2h_2h_3)F_2[t,2(\ell/c_2)]*F_3[t,2(L/c_3)]$ represents the reflection contribution of the third interface, $x=L$. The corresponding wave makes one round trip inside the two porous layers. Evidently this wave contribution depends on all acoustical parameters of each porous layer.

The advantage of the obtained time-domain expression of the reflection and transmission scattering operators [Eqs. (44) and (45)] is to show analytically the effect of the acoustical parameters (porosity, tortuosity, viscous and thermal characteristic lengths) of each porous layer on the reflection contributions by the interfaces of the double-layered media.

V. EXPERIMENTAL VALIDATION

In application of this model, several numerical simulations for reflected and transmitted acoustic waves by two layered porous materials are compared to experimental data. Experiments are performed in the air using two broadband Ultran NCT202 transducers with a central frequency of 190 kHz in air and a bandwidth of 6 dB from 150 to 230 kHz. Pulses of 400 V are provided by a 5052PR Panametrics pulser/receiver. The signals received are amplified to 90 dB and filtered above 1 MHz to avoid high-

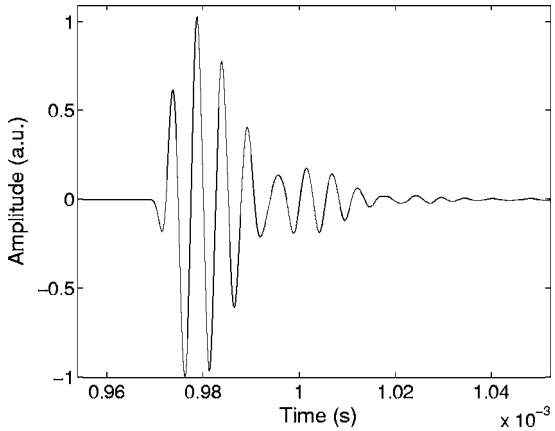


FIG. 3. Incident signal given out by the transducer in transmitted mode.

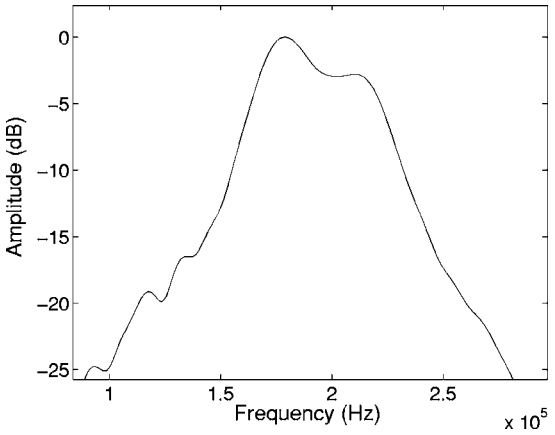


FIG. 4. Spectrum of the incident signal generated by the transducer in transmitted mode.

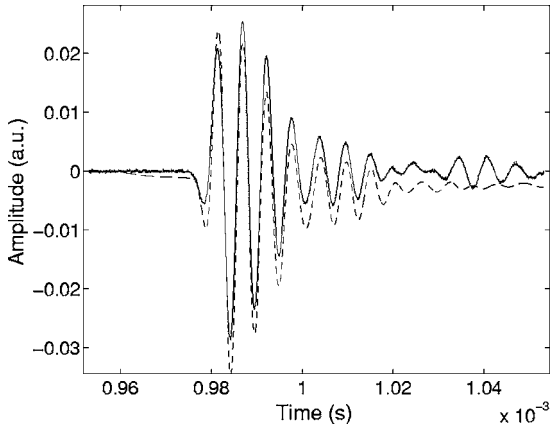


FIG. 5. Comparison between experimental transmitted signal (solid line) and simulated transmitted signal (dashed line) for a double-layered medium consisting of 0.86 cm of M1 and 0.81 cm of M2.

frequency noise (energy is totally filtered by the sample in this upper-frequency domain). Electronic interference is eliminated by 1000 acquisition averages. The experimental setup is shown in Fig. 2. The experimental incident signal generated by the transducer is given in Fig. 3. The amplitude is represented by an arbitrary unit (a.u.) and the point number represented in the abscissa is proportional to time. Signal duration is important as its spectrum must verify the condition of high-frequency approximation[8,26] referred to in Sec. II. The spectrum of the incident signal is given in Fig. 4.

Measurements were made on plastic foam samples M1–M4. Their acoustic characteristics were determined independently using classical methods[8] (which were developed for a slab of porous material). The acoustical parameters of the plastic foam samples are given in Table I.

Three samples of double-layered porous materials were considered, the first consists of 0.86 cm of M1 and 0.81 cm of M2, the second of 4.13 cm of M1 and 1.99 cm of M3, and finally the third of 2.98 cm of M1 and 1.99 cm of M3. Numerical simulation and experimental results (transmitted signal) for the three samples of double-layered materials are presented in Figs. 5–7, respectively. The numerical results are obtained from convolution of the transmission operator [Eq. (45)] with the signal generated by the transducer shown in Fig. 3. The reader can see, from Figs. 5–7, the good correlation obtained between the experimental transmitted signal (solid line) and simulated signal (dashed line). This result validates the expression of the transmission scattering operator [Eq. (45)].

Reflected waves were processed by another experimental setup shown in Fig. 8. One transducer was used alternatively as a transmitter and receiver to detect the reflected wave. The experimental incident signal used in reflected mode is given in Fig. 9 and its spectrum in Fig. 10.

A double-layered porous medium consisting of 1.11 cm of M4 and 0.87 cm of M1 was considered. Figure 11 shows a comparison between a simulated reflected signal (dashed line) and an experimental reflected signal (solid line). The simulated signal was obtained through convolution of the reflection scattering kernel given in Eq. (44) with the incident signal given in Fig. 9. Three reflected signals can be seen in Fig. 11. The first corresponds to the reflection of the first porous layer M4 ($x=0$) by the first interface, this reflected wave corresponds to the first term on the right-hand side of Eq. (44): $d_1\delta(t)=[(\sqrt{\alpha_2}-\phi_2)/(\sqrt{\alpha_2}+\phi_2)]\ \delta(t)$. The second reflected wave given in Fig. 11 corresponds to the reflection between the second M4 interface and the first M1 interface ($x=\ell$); this wave corresponds to the second term on the right-hand side of Eq. (44): $d_1(d_2-h_2)F_2[t,2(\ell/c_2)]$. Finally, the third signal corresponds to reflection by the second M1 interface ($x=L$), which is given by the term $d_1(d_4-h_4-d_2h_3-d_3h_2+2h_2h_3)F_2[t,2(\ell/c_2)]*F_3[t,2(L/c_3)]$. Generally, it is not possible to see the other reflection contributions experimentally because of the high damping of ultrasonic waves in air-saturated plastic foams. However, the third reflection contribution at $x=L$ is not always seen experimen-

TABLE I. Acoustical characteristics of the plastic foams samples.

Material	M1	M2	M3	M4
Viscous characteristic length [$\Lambda(\mu m)$]	200	30	330	230
Thermal characteristic length [$\Lambda'(\mu m)$]	600	90	990	690
Tortuosity (α)	1.07	1.4	1.02	1.05
Porosity (ϕ)	0.97	0.85	0.90	0.98

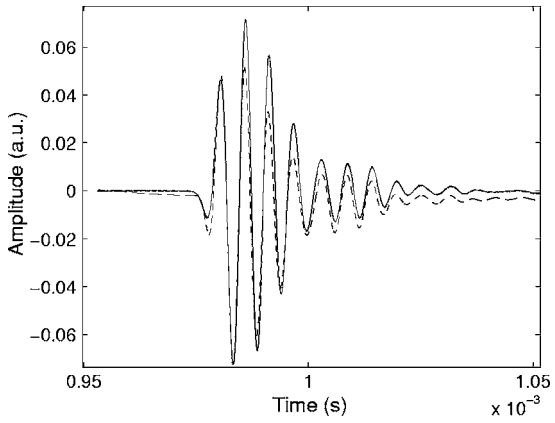


FIG. 6. Comparison between experimental transmitted signal (solid line) and simulated transmitted signal (dashed line) for a double-layered medium consisting of 4.13 cm of M1 and 1.99 cm of M3.

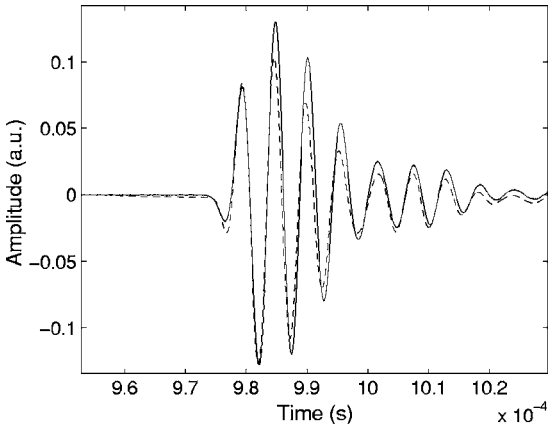


FIG. 7. Comparison between experimental transmitted signal (solid line) and simulated transmitted signal (dashed line) for a double-layered medium consisting of 2.98 cm of M1 and 1.99 cm of M3.

tally; for example, Fig. 12 shows a comparison between theoretical predictions (dashed line) and experimental results (solid line) for a double-layered medium consisting of 0.88 cm of M1 and 0.83 cm of M2. Acoustic attenuation in plastic foam sample M2 is higher than in the other samples. Sample M2 has high tortuosity and low characteristic lengths compared to those of the other plastic foam samples, which indicates high acoustic damping. In Fig. 12, we can only see the two reflected waves corresponding to reflection by the first M1 (first layer) interface ($x=0$) and reflection between the second M1 interface and the first M2 (second layer) interface ($x=\ell$), respectively. The reflected wave of the second M2 interface ($x=L$) is fully absorbed by the two layers, M1 and M2. We can also see in Fig. 12 that the amplitude of the second reflected wave is greater than that of the first reflected wave. This is due to the high resistivity of sample M2 near sample M1 and the thickness of M1, which also plays an important part in attenuating the wave reflected at the second interface, $x=\ell$.

VI. CONCLUSION

In this paper the analytical expressions of reflection and transmission scattering operators are derived for double-layered porous media consisting of two homogeneous isotropic materials. Simple relationships are given between these operators and the acoustic parameters of the medium. It is shown that the scattering operators are equal to the sum of the contribution of each interface to the double-layered porous medium. The advantages of the analytical expressions of reflection and transmission scattering operators in the time domain is to show easily the effect of the acoustical parameters on the multiple reflections at the interfaces of the double-layered medium.

Ultrasonic measurements in the transmission and reflection mode were processed using different experimental setups. A slight difference was observed between theoretical predictions and experimental data in the two modes (reflection and transmission). This leads to the conclusion that the expressions of scattering operators obtained are correct. Future studies will concentrate on the inverse problem, and methods and inversion algorithms will be developed to optimize the acoustic properties of double-layered air-saturated porous media for reflected and transmitted ultrasonic waves.

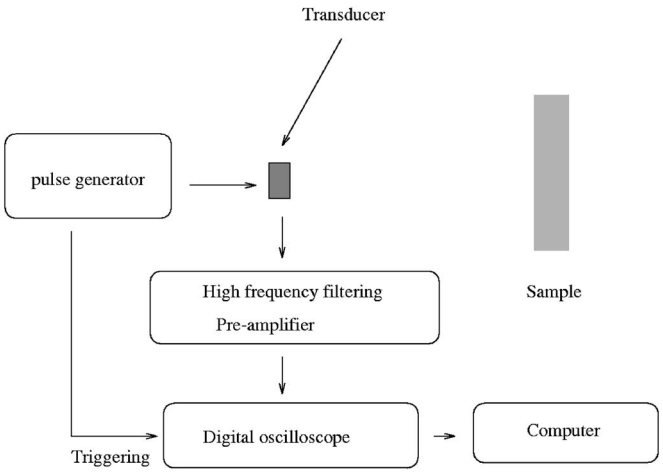


FIG. 8. Experimental setup of the ultrasonic measurements in reflected mode.

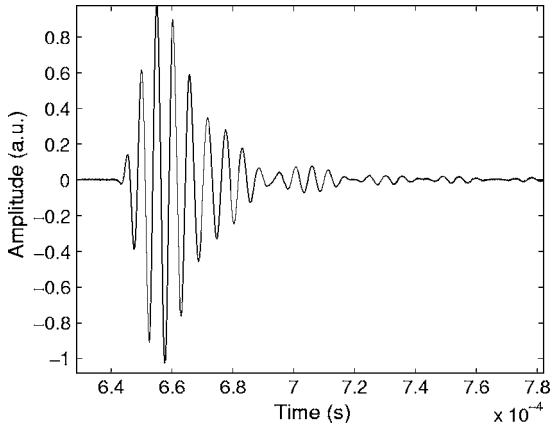


FIG. 9. Incident signal given out by the transducer in reflected mode.

APPENDIX A: EXPRESSION OF THE REFLECTION AND TRANSMISSION OPERATORS

Using Eqs. (25) and (31), we can write the following equation system given the relations between $A_2(z)$, $B_2(z)$, and $R(z)$:

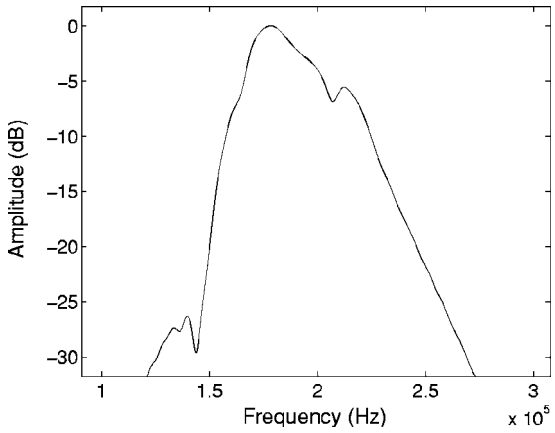


FIG. 10. Spectrum of the incident signal generated by the transducer in reflected mode.

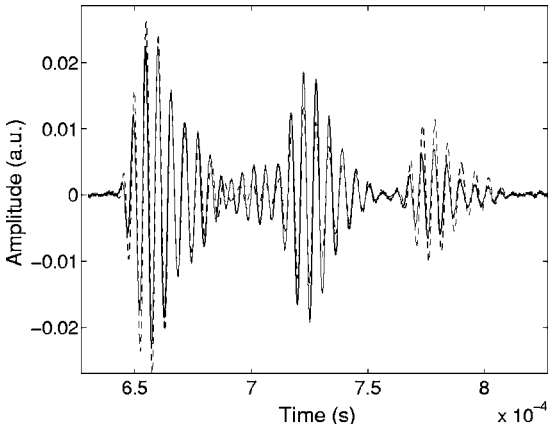


FIG. 11. Comparison between experimental reflected signal (solid line) and simulated reflected signal (dashed line) for a double-layered medium consisting of 1.11 cm of M4 and 0.87 cm of M1.

$$\begin{pmatrix} A_2(z) \\ B_2(z) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1+R(z) \\ K_1(R(z)-1) \end{pmatrix}. \tag{A1}$$

From Eqs. (27) and (33), one has

$$\begin{pmatrix} A_3(z) \\ B_3(z) \end{pmatrix} = \frac{T'(z)}{2}\begin{pmatrix} (1+K_3)\exp\left(\frac{\sqrt{f_3(z)}}{c_3}(\ell+L)\right) \\ (1-K_3)\exp\left(-\frac{\sqrt{f_3(z)}}{c_3}(\ell+L)\right) \end{pmatrix}. \tag{A2}$$

where $T'(z)=T(z)\exp\{-[(\ell/c_2)+(L/c_3)]z\}$.

Using Eqs. (46),(47), and Eqs. (26) and (32), we obtain the following linear system given the reflection and transmission coefficients $R(z)$ and $T(z)$:

$$\begin{aligned} &R\left[\cosh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right)+K_1\sinh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right)\right]-T'(z) \\ &\quad\times\left[\cosh\left(\frac{\sqrt{f_3(z)}}{c_3}L\right)+K_3\sinh\left(\frac{\sqrt{f_3(z)}}{c_3}L\right)\right] \\ &\quad=K_1\sinh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right)-\cosh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right), \end{aligned}$$

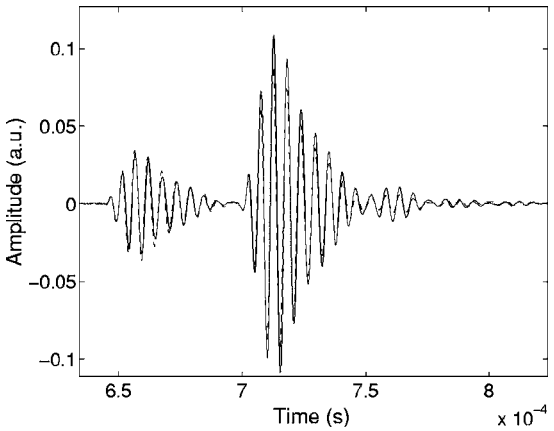


FIG. 12. Comparison between experimental reflected signal (solid line) and simulated reflected signal (dashed line) for a double-layered medium consisting of 0.88 cm of M1 and 0.83 cm of M2.

$$\begin{aligned} &R\left[K_2\sinh\left(\frac{\sqrt{f_2(z)}}{c_3}\ell\right)+K_3\cosh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right)\right]+T'(z) \\ &\quad\times\left[\sinh\left(\frac{\sqrt{f_3(z)}}{c_3}L\right)+K_3\cosh\left(\frac{\sqrt{f_3(z)}}{c_3}L\right)\right] \\ &\quad=K_3\cosh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right)-K_2\sinh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right). \end{aligned}$$

By setting

$$\begin{aligned} D(z) &= (1+K_3^2)\cosh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right)\sinh\left(\frac{\sqrt{f_3(z)}}{c_3}L\right) \\ &\quad+(K_1+K_2K_3)\sinh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right)\sinh\left(\frac{\sqrt{f_3(z)}}{c_3}L\right) \\ &\quad+2K_3\cosh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right)\cosh\left(\frac{\sqrt{f_3(z)}}{c_3}L\right) \\ &\quad+(K_2+K_1K_3)\sinh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right)\cosh\left(\frac{\sqrt{f_3(z)}}{c_3}L\right), \end{aligned}$$

and

$$\begin{aligned} D_1(z) &= (K_1-K_2K_3)\sinh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right)\sinh\left(\frac{\sqrt{f_3(z)}}{c_3}L\right) \\ &\quad+(K_3^2-1)\cosh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right)\sinh\left(\frac{\sqrt{f_3(z)}}{c_3}L\right) \\ &\quad+(K_1K_3-K_2)\sinh\left(\frac{\sqrt{f_2(z)}}{c_2}\ell\right)\cosh\left(\frac{\sqrt{f_3(z)}}{c_3}L\right), \end{aligned}$$

one has the following expressions for $R(z)$ and $T(z)$:

$$R=\frac{D_1(z)}{D(z)} \quad \text{and} \quad T'(z)=\frac{2K_1K_2}{D(z)},$$

which are equivalent to the expressions (35) and (36) given in Sec. IV.

APPENDIX B: GREEN FUNCTION OF THE MEDIUM

The Green function[28] F_i, $i=2,3$ of the porous layer (2) and (3), respectively is given by

$$F_i(t,k) = \begin{cases} 0 & \text{if } 0 \leqslant t \leqslant k \\ \Im_i(t)+\Delta_i\int_0^{t-k}\varnothing_i(t,\xi)d\xi, & \text{if } t \geqslant k, \quad i=2,3 \end{cases}$$

with

$$\Im_i(t)=\frac{\psi_i}{4\sqrt{\pi}}\frac{k}{(t-k)^{3/2}}\exp\left(-\frac{\psi_i^2k^2}{16(t-k)}\right), \quad i=2,3,$$

where $\varnothing_i(\tau,\xi)$, $i=2,3$ has the following form:

$$\varnothing_i(\xi,\tau) = -\frac{1}{4\pi^{3/2}}\frac{k}{\sqrt{(\tau-\xi)^2-k^2}}\frac{1}{\xi^{3/2}} \times \int_{-1}^{1} \exp\left(-\frac{\mathfrak{N}_i(\mu,\tau,\xi)}{2}\right)(\mathfrak{N}_i(\mu,\tau,\xi)-1) \times \frac{\mu\, d\mu}{\sqrt{1-\mu^2}}, \quad i=2,3,$$

and where

$$\mathfrak{N}_i(\mu,\tau,\xi) = (\Delta_i\mu\sqrt{(\tau-\xi)^2-k^2}+\psi_i(\tau-\xi))^2/8\xi,$$

$$\psi_i = 2\alpha_i\sqrt{\frac{\eta}{\pi}}\left(\frac{1}{\Lambda_i}+\frac{\gamma-1}{\sqrt{P_r}\Lambda_i'}\right), \quad \varphi_i = \frac{4\alpha_i(\gamma-1)\eta}{\Lambda_i\Lambda_i'\sqrt{P_r}\rho_f},$$

$$\Delta_i^2 = \psi_i^2 - 4\varphi_i, \quad i=2,3.$$

when $k\to\infty$, $\mathfrak{J}_i$ and $\varnothing_i(\xi,\tau)$ tends to zero, then the Green function $F_i(t,k)$ also tends to zero.

[1] J. F. Allard, *Propagation of Sound in Porous Media: Modeling Sound Absorbing Materials* (Chapman and Hall, London, 1993).
[2] K. Attenbourough, "On the acoustic slow wave in air-filled granular media," J. Acoust. Soc. Am. **81**, 93–102 (1986).
[3] P. Leclaire, L. Kelders, W. Lauriks, N. R. Brown, M. Melon, and B. Castagnède, "Determination of the viscous and thermal characteristics lengths of plastic foams by ultrasonic measurements in helium and air," J. Appl. Phys. **80**, 2009–2012 (1996).
[4] P. Leclaire, L. Kelders, W. Lauriks, C. Glorieux, and J. Thoen, "Determination of the viscous characteristic length in air-filled porous materials by ultrasonic attenuation measurements," J. Acoust. Soc. Am. **99**, 1944–1948 (1996).
[5] G. Caviglia and A. Morro, "A closed-form solution for reflection and transmission of transient waves in multilayers," J. Acoust. Soc. Am. **116**, 643–654 (2004).
[6] G. V. Norton and J. C. Novarini, "Including dispersion and attenuation directly in time domain for wave propagation in isotropic media," J. Acoust. Soc. Am. **113**, 3024–3031 (2003).
[7] W. Chen and S. Holm, "Modified Szabo's wave equation models for lossy media obeying frequency power law," J. Acoust. Soc. Am. **113**, 3024–3031 (2003).
[8] Z. E. A. Fellah, M. Fellah, W. Lauriks, and C. Depollier, "Direct and inverse scattering of transient acoustic waves by a slab of rigid porous material," J. Acoust. Soc. Am. **114**, 2570–2574 (2003).
[9] Z. E. A. Fellah, C. Depollier, S. Berger, W. Lauriks, P. Trompette, and J. Y. Chapelon, "Determination of transport parameters in air-saturated porous materials via reflected ultrasonic waves," J. Acoust. Soc. Am. **114**(5), 2561–2569 (2003).
[10] Z. E. A. Fellah, F. G. Mitri, C. Depollier, S. Berger, W. Lauriks, and J. Y. Chapelon, "Characterization of porous materials with a rigid frame via reflected waves," J. Appl. Phys. **94**, 7914–7922 (2003).
[11] Z. E. A. Fellah, S. Berger, W. Lauriks, and C. Depoller, "Verification of Kramers–Kronig relationships in porous materials having a rigid frame," J. Sound Vib. **270**, 865–885 (2004).
[12] T. L. Szabo, "Time domain wave equations for lossy media obeying a frequency power law," J. Acoust. Soc. Am. **96**, 491–500 (1994).
[13] T. L. Szabo, "Causal theories and data for acoustic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **97**, 14–24 (1995).
[14] *Inverse Problems in Mathematical Physics*, edited by L. Päivärinta and E. Somersalo (Springer, Berlin, 1993).
[15] F. Mainardi, "Transient waves in linear viscoelasticity," in *Vibration and Control of Structures*, edited by A. Gurran (World Scientific, Singapore, 1997).
[16] D. L. Johnson, J. Koplik, and R. Dashen, "Theory of dynamic permeability and tortuosity in fluid-saturated porous media," J. Fluid Mech. **176**, 379–402 (1987).
[17] W. Sachse and Y. H. Pao, "On the determination of phase and group velocities of dispersive waves in solids," J. Appl. Phys. **49**, 4320–4327 (1978).
[18] K. V. Gurumurthy and R. M. Arthur, "A dispersive model for the propagation of ultrasound in soft tissue," Ultrason. Imaging **49**, 355–377 (1982).
[19] R. Kuc, "Modeling acoustic attenuation of soft tissue with a minimum-phase filter," Ultrason. Imaging **6**, 24–36 (1984).
[20] F. Yu, A. Rossikhin, and M. V. Shitikova, "Application of fractional calculus to dynamic problems of linear hereditary mechanics of solids," Appl. Mech. Rev. **50**, 15–67 (1997).
[21] M. Caputo, "Vibration of an infinite plate with a frequency dependent Q," J. Acoust. Soc. Am. **60**, 634–639 (1976).
[22] R. L. Bagley and P. J. Torvik, "On the fractional calculus model of viscoelastic behavior," J. Rheol. **30**, 133–155 (1986).
[23] A. Hanyga and V. E. Rok, "Wave propagation in micro-heterogeneous porous media: A model based on an integro-differential wave equation," J. Acoust. Soc. Am. **107**, 2965–2972 (2000).
[24] M. A. Biot, "The theory of propagation of elastic waves in fluid-saturated porous solid. I. Low frequency range," J. Acoust. Soc. Am. **28**, 168–178 (1956).
[25] M. A. Biot, "The theory of propagation of elastic waves in fluid-saturated porous solid. II. Higher frequency range," J. Acoust. Soc. Am. **28**, 179–191 (1956).
[26] Z. E. A. Fellah and C. Depollier, "Transient wave propagation in rigid porous media: A time domain approach," J. Acoust. Soc. Am. **107**, 683–688 (2000).
[27] S. G. Samko, A. A. Kilbas, and O. I. Marichev, *Fractional Integrals and Derivatives: Theory and Applications* (Gordon and Breach Science, Amsterdam, 1993).
[28] Z. E. A. Fellah, M. Fellah, W. Lauriks, C. Depollier, J. Y. Chapelon, and Y. C. Angel, "Solution in time domain of ultrasonic propagation equation in a porous material," Wave Motion **38**, 151–163 (2003).
[29] Z. E. A. Fellah, S. Berger, W. Lauriks, C. Depollier, C. Aristegui, and J. Y. Chapelon, "Measuring the porosity and tortuosity of porous materials via reflected waves at oblique incidence," J. Acoust. Soc. Am. **113**(5), 2424–2433 (2003).
[30] Z. E. A. Fellah, S. Berger, W. Lauriks, C. Depollier, and M. Fellah, "Measuring the porosity of porous material having rigid frame via reflected waves: A time domain analysis with fractional derivatives," J. Appl. Phys. **93**, 296–303 (2003).

Method of superposition applied to patch near-field acoustic holography

Angie Sarkissian
Naval Research Laboratory, Washington, DC 20375-5350

(Received 16 November 2004; revised 13 April 2005; accepted 11 May 2005)

The method of superposition may be applied to reconstruct the field on a partial surface on a radiating structure from measurements made on a nearby limited surface. Unlike conformal near-field holography, where the measurement surface surrounds the entire structure, in patch holography the measurement surface need only be approximately as large as the patch on the structure surface where the reconstruction is required. Using the method of superposition, the field on and near the measurement surface may be approximated by the field produced by a source distribution placed on a surface inside the structure. The source strengths are evaluated by applying boundary conditions on the measurement surface. The algorithm requires the inversion of the Green's function matrix which may be ill-conditioned. Truncated singular value decomposition is used to invert it. The field on the structure surface is then approximated by the field produced by the source distribution. The algorithm is easier to implement than the boundary elements method because it does not require integrations over singular integrands and may be applied to flat or curved surfaces. [DOI: 10.1121/1.1945470]

PACS number(s): 43.20.Tb, 43.40.Rj [EGW] Pages: 671–678

I. INTRODUCTION

Near-field acoustic holography allows a high resolution reconstruction of the field on the surface of a radiating structure from measurements made in the near-field.[1,2] The problem is ill-posed because of the presence of evanescent waves that decay rapidly from the structure surface to the measurement surface. Thus measurements are typically made very close to the structure in order to obtain the high resolution reconstruction. To treat the ill-posed nature of the problem, for planar surfaces Fourier decomposition is used with the series truncated to remove the very short wavelength components of the field that decay very rapidly.[1,2] For an arbitrarily shaped structure, boundary elements method may be applied to numerically compute the Neumann Greens function that relates the field on the measurement surface to the surface normal velocity in matrix form.[3–6] This ill-conditioned matrix is then inverted using either singular value decomposition[3–5] or eigenfunction decomposition[6] to remove the components of the field that have very short wavelengths on the structure surface and thus decay very rapidly away from the surface. Stepanishen[7] has developed an algorithm based on the method of superposition and singular value decomposition for the inverse holography problem in the case where the measurement surface surrounds the entire structure.

Conformal near-field holography requires measurement of the field on a surface that surrounds the entire structure. For large structures, if the reconstruction of the field is required only on a partial structure surface, patch holography may be applied where measurements need to be made only on a surface that is approximately as large as the patch on the structure surface where the reconstruction is required.[8–12] For planar surfaces, Fourier decomposition may be applied again.[9,10] A regularized least squares solution has also been applied to planar patch holography.[11,12] For surfaces having arbitrary shapes, the boundary elements method may be applied with singular value decomposition.[8,9] This requires the evaluation of the Neumann Green's function numerically which is tedious involving integrations that contain singular integrands. The method of superposition[7,13–16] is applied here which is faster and much easier to implement. It may be applied to flat or curved surfaces. This algorithm also uses singular value decomposition to treat the ill-posed nature of the problem. A similar algorithm has previously been applied to extend the measurement surface tangentially outward.[17] Bobrovnitskii[18] has also applied singular value decomposition algorithm to extrapolate the field on the structure surface from measurements made on a partial surface on the structure. Here we apply method of superposition with singular value decomposition to reconstruct the surface field in patch holography.

II. ALGORITHM

Figure 1 shows the geometry. The structure shown radiates a harmonic field. The $e^{-i\omega t}$ time dependence has been suppressed later. The surface ∂S is the patch on the structure surface where the reconstruction is required. The measurement surface lies at a constant distance δ away from ∂S. We define surface σ to lie in the interior of the structure at a constant distance δ' from surface ∂S. Using the method of superposition, the field produced by the structure on and near the measurement surface may be approximated by the field produced by a source distribution placed on surface σ. There was no reason to use a nonuniform distribution of sources on this surface. We thus chose an approximately uniform distribution, where source i has coordinates $\vec{r_i^{\sigma}}$ and source strength q_i. The field produced by the source distribution at a measurement location i having coordinates $\vec{r_i^{m}}$ is

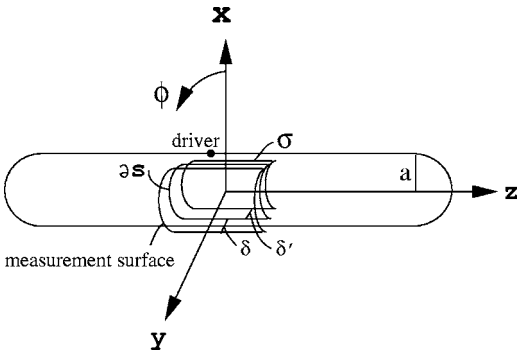


FIG. 1. Geometry.

$$p(\vec{r_i^m}) = \sum_{j=1}^{N'} G_{ij} q_j, \tag{1}$$

where N' is the number of sources,

$$G_{ij} = \frac{e^{ik|\vec{r_i^m} - \vec{r_j^\sigma}|}}{|\vec{r_i^m} - \vec{r_j^\sigma}|}; \tag{2}$$

$k = \omega/c$ is the wave number and c is the speed of sound in the fluid outside the structure. Equation (2) may be rewritten in matrix form as

$$\mathbf{p^m} = \mathbf{Gq}. \tag{3}$$

We define N to be the number of measurement locations. Matrix $\mathbf{G}$ then has size $N \times N'$. If matrix $\mathbf{G}$ can be inverted, the source strengths can be evaluated from the measured field values p_i^m using Eq. (3). The field at a point $\vec{r_i^s}$ on the partial structure surface ∂S can then be evaluated simply by summing the contributions of the fields produced by the sources at that location

$$p_j^s = \sum_{j=1}^{N'} G_{ij}^s q_j, \tag{4}$$

where

$$G_{ij}^s = \frac{e^{ik|\vec{r_i^s} - \vec{r_j^\sigma}|}}{|\vec{r_i^s} - \vec{r_j^\sigma}|}. \tag{5}$$

Similarly the normal velocity may be evaluated

$$v_j = \sum_{j=1}^{N'} G_{ij}^{sv} q_j, \tag{6}$$

where

$$G_{ij}^{sv} = \frac{1}{i\omega\rho_o} \hat{n} \cdot \nabla \left(\frac{e^{ik|\vec{r} - \vec{r_j^\sigma}|}}{|\vec{r} - \vec{r_j^\sigma}|} \right) \Bigg|_{\vec{r} = \vec{r_i^s}}; \tag{7}$$

ρ_o is the density of the fluid medium outside the structure.

If N' is large, matrix $\mathbf{G}$ in Eq. (3) will be ill-conditioned because the inverse problem is of ill-posed nature. We apply singular value decomposition to invert it. Matrix $\mathbf{G}$ may be written as

$$\mathbf{G} = \mathbf{USV}^\dagger, \tag{8}$$

where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{S}$ is a diagonal matrix containing real, positive singular values in order of decreasing magnitude

$$S_{ij} = \delta_{ij}\sigma_j; \tag{9}$$

$$\sigma_j \geqslant \sigma_{j+1}. \tag{10}$$

Since $\mathbf{U}$ and $\mathbf{V}$ are unitary, using Eq. (8) the inverse of $\mathbf{G}$ may simply be written as

$$\mathbf{G}^{-1} = \mathbf{VS}^{-1}\mathbf{U}^\dagger. \tag{11}$$

Using the above with Eqs. (3) and (9) we may approximate q_i,

$$q_i \simeq \sum_{j=1}^{N_t} \sum_{k=1}^{N} V_{ij} \sigma_j^{-1} U_{kj}^* p_k^m. \tag{12}$$

The sum over j above is truncated to keep only N_t singular values. The higher order singular values are removed because they represent waves that decay very rapidly. We later discuss how to determine where to truncate the sum. Using an approximation for the source strengths q_i above, the surface pressure or normal velocity can be approximated using Eqs. (4) or (6).

III. IMPLEMENTATION

The algorithm is applied to numerically simulated data. Finite elements/infinite elements method[19] is applied to numerically compute the radiated pressure and normal velocity on the surface of a framed cylindrical shell. Field values are also computed on the measurement surface. We normalize the field values such that the surface normal velocity has maximum magnitude of unity.

The cylinder is hemispherically end-capped. It has radius a, total length, including end-caps, of $14a$ and thickness of 0.0074a. It has straight frames that are uniformly spaced with frame spacing of $0.14a$. The frames have two different lengths. The shorter frames have length $0.078a$ and thickness 0.0065a. Every 13th frame is longer having length $0.14a$ and thickness 0.0074a. The properties of nickel are used for the cylinder and the frames, with density of 8800 kg/m^3, Young's modulus of 2.1×10^{11} Pa, Poisson's ratio of 0.30 and loss factor of 0.001. The sound speed in the fluid outside the structure is 1500 m/s and the density is 1000 kg/m^3.

A. Example 1

Using cylindrical coordinates with the origin at the center of the cylinder, the driver is located at $(a, 0, -a/3)$, as shown in Fig. 1. We first simulate the field at a $ka=3$. The measurement surface contains a 33×33 grid of points on the curved surface ($\rho = a + \delta$, $\pi/4 \leqslant \phi \leqslant 3\pi/4$, $-0.75a \leqslant z \leqslant 0.75a$), where the value of δ is varied. The reconstruction is performed at a 33×33 grid of points again, on the surface ($\rho = a$, $\pi/4 \leqslant \phi \leqslant 3\pi/4$, $-0.75a \leqslant z \leqslant 0.75a$). The surface σ also contains a 33×33 grid of points on the surface ($\rho = a - \delta'$, $\pi/4 \leqslant \phi \leqslant 3\pi/4$, $-0.75a \leqslant z \leqslant 0.75a$). We thus have $N = N'$ for our example with a square matrix $\mathbf{G}$. Bobrovnitskii

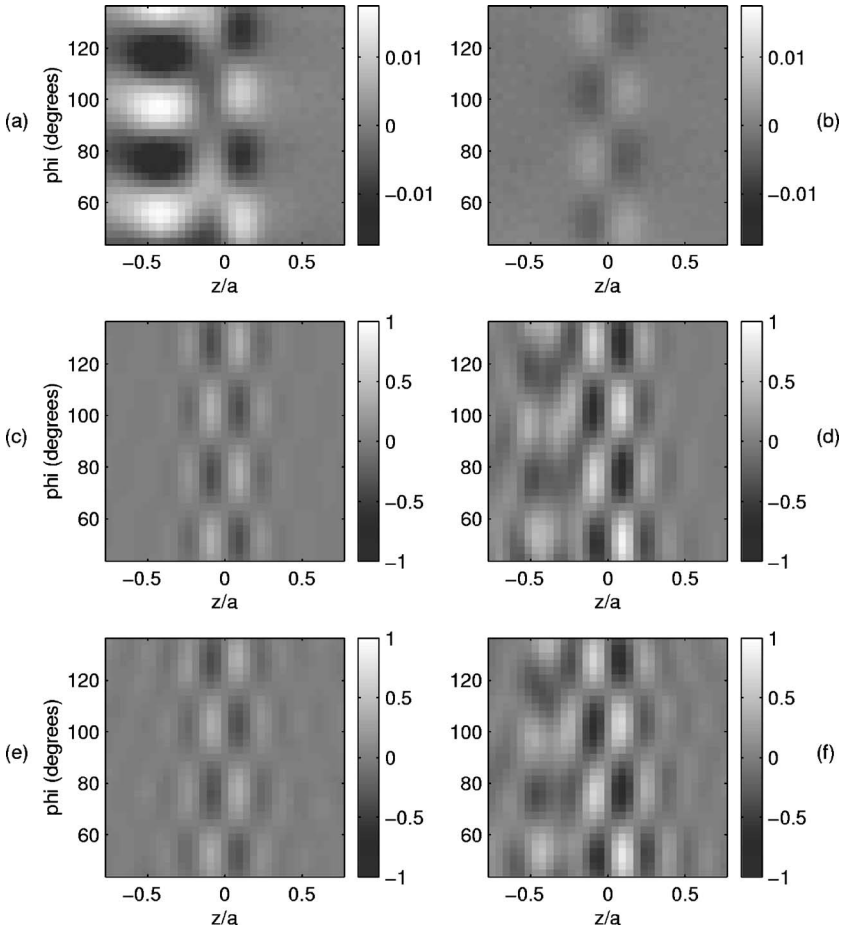


FIG. 2. (a) Real and (b) imaginary parts of the field on the measurement surface for $ka=3$, $k\delta=0.5$; (c) real and (d) imaginary parts of the normal velocity on the structure surface; (e) real and (f) imaginary parts of the reconstructed normal velocity.

and Tomilina[16] discuss the ill-posedness of the inverse problem that requires the inversion of $\mathbf{G}$ as the number of sources N' becomes too large. A large number of sources N' does not cause difficulties in our case because we apply singular value decomposition to invert matrix $\mathbf{G}$ and the smaller singular values are not included in the reconstruction. Similarly we found no difficulties in choosing $N=N'$ because we use the truncated singular value decomposition algorithm.

The algorithm is first applied to a measurement surface having an offset distance $\delta=a/6$. Thus $k\delta=0.5$. We set the offset distance for surface σ at $\delta'=a/3$, making $k\delta'=1$. To the computed field values p_j^m on the measurement surface we add 5% random noise to obtain

$$\vec{p_j^m} = p_j^m + n_j, \tag{13}$$

where n_j is the noise and the error is defined as

$$E = \left(\frac{\sum_{j=1}^{N} |\vec{p_j^m} - p_j^m|^2}{\sum_{j=1}^{N} |p_j^m|^2} \right)^{1/2} \times 100. \tag{14}$$

Figures 2(a) and 2(b) show the real and imaginary parts of the field plus noise values on the measurement surface. Pressure values are divide by $\rho_o c$ throughout the article and we recall that all field values are normalized such that the surface normal velocity has maximum magnitude of unity. Figures 2(c) and 2(d) below show the real and imaginary parts of the normal velocity on the partial structure surface ∂S computed using the finite elements algorithm. We observe shorter wavelengths on the structure surface. Since the shorter wavelength components decay more rapidly from the structure surface to the measurement surface, longer wavelength components dominate on the measurement surface.

The singular value decomposition algorithm decomposes the field into states. Each states is related to a singular value. To determine the contribution of the surface normal velocity to each state at location $\vec{r_i^s}$ we compute

$$f_n(\vec{r_i^s}) = \sum_{j=1}^{N'} \sum_{k=1}^{N} G_{ij}^{s\nu} V_{jn} \sigma_n^{-1} U_{kn}^* \vec{p_k^m}. \tag{15}$$

We next define F_n as

$$F_n = \sum_i [|f_n(\vec{r_i^s})|^2]. \tag{16}$$

Similarly the contribution of the measured pressure to each state at location $\vec{r_i^m}$ is given by

$$g_n(\vec{r_i^m}) = \sum_{k=1}^{N} U_{in} U_{kn}^* \vec{p_k^m}. \tag{17}$$

We similarly define G_n as

$$G_n = \sum_i [|g_n(\vec{r_i^m})|^2]. \tag{18}$$

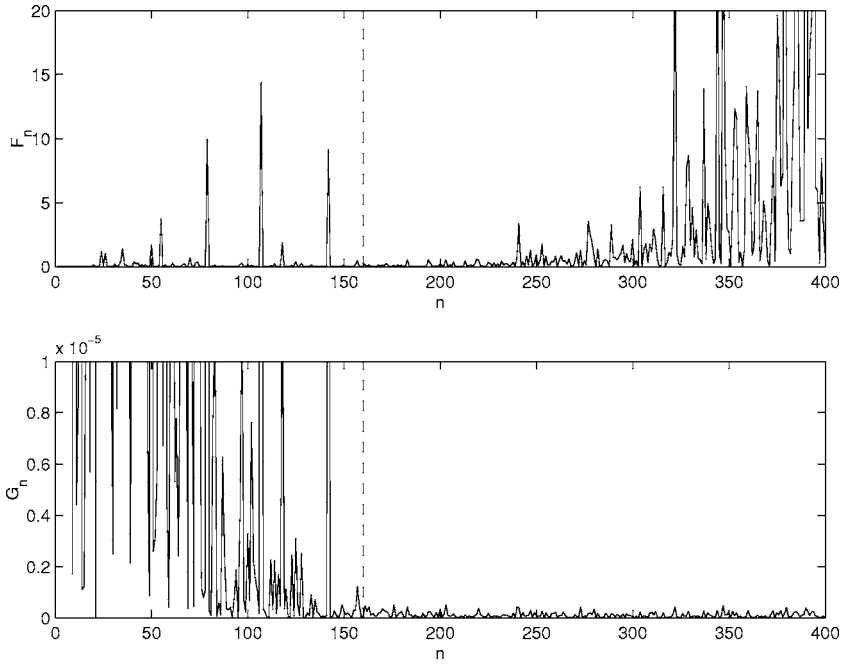


FIG. 3. F_n and G_n as a function of index n for $ka=3$, $k\delta=0.5$.

Summing the function f_n over the index n produces the surface normal velocity; f_n represents the decomposition of the field into velocity states. Similarly, summation of function g_n over index n produces the measured pressure; g_n represents the decomposition of the measured pressure into pressure states. These pressure and velocity states are produced by the singular value decomposition algorithm.

Functions F_n and G_n are plotted in Fig. 3. We observe that for small index n, relatively large values of G_n (which represent states of the field on the measured surface) is produced by relatively small values of F_n (which represents the velocity states on the structure surface). These states with small indices n are the propagating states. For $n>160$, the lower plot of G_n is dominated by random noise. We thus truncate the sum at $N_t=160$, as shown by the dashed lines. In the upper plot, large peaks can be observed for index n below 160. States $160<n<230$ do not contain high peaks. For n above 230, the plot of G_n below shows the presence of mainly random noise yet these noise values produce large velocity components F_n above. These are the evanescent states with short wavelengths. If these states are included in the sum for the reconstruction, they would produce very large errors.

The real and imaginary parts of the reconstructed total surface normal velocity with $N_t=160$ are shown in Figs. 2(e) and 2(f). They are compared to the actual values (computed using finite elements algorithm) shown in Figs. 2(c) and 2(d). The error of the reconstructed field is 20%. The error is computed using an equation similar to Eq. (14). We expect the error to be higher around the edges. Removing three reconstruction grid points from each of the sides thus keeping the central 27×27 reconstruction grid reduces the error to 18%

In the earlier example, the measurement surface is sufficiently close to the structure to accomplish the reconstruction of the short wavelength components present on the surface of the structure. Increasing the distance δ between the structure surface and measurement surface should result in a degraded reconstruction. Next, that distance is doubled to $k\delta=1$. We keep the offset distance to surface σ at $k\delta'=1$. Again, we add 5% random error to the simulated field values on the measurement surface. Figures 4(a) and 4(b) show the real and imaginary parts of the field including noise on the measurement surface. Since this is more distant than the example shown in Figs. 2(a) and 2(b), longer wavelengths dominate in Figs. 4(a) and 4(b). Figure 5 shows F_n and G_n as a function of index n. In this case, we truncate the sum at $N_t=115$ because, for $n>115$, G_n appears to contain mostly noise. Any pressure value, for these high order states is too small to be distinguished from noise and, for large n, the noise in G_n produces large normal velocity values in F_n.

The real and imaginary parts of the reconstructed velocity for $k\delta=1$ and $N_t=115$ is shown in Figs. 4(c) and 4(d). Comparison to the actual values shown in Figs. 2(c) and 2(d), shows the reconstruction is not as good as the earlier case. The error, in this case, is 54%. Again, the error is, on the average, higher around the edges. Removing three grid points from each of the sides thus keeping a 27×27 reconstruction grid reduces the error to 49%.

To understand why the error is excessively high, we recompute functions F_n and G_n in the absence of the random noise. They are plotted in Fig. 6. We see that above the $n=115$ where the truncation was applied earlier, F_n has four high peaks that are mixed with the noise in Fig. 4. The exclusion of these evanescent peaks in the previous reconstruction contributed to the large, 54% error. The reconstruction algorithm is next applied to the noise free example. We choose to truncate at $N_t=200$ in this case, to include the four additional peaks. Results of the real and imaginary parts of the reconstructed normal velocity are shown in Figs. 7(a) and 7(b). We compare this reconstruction to the actual field shown in Figs. 2(c) and 2(d)). We find an improved recon-

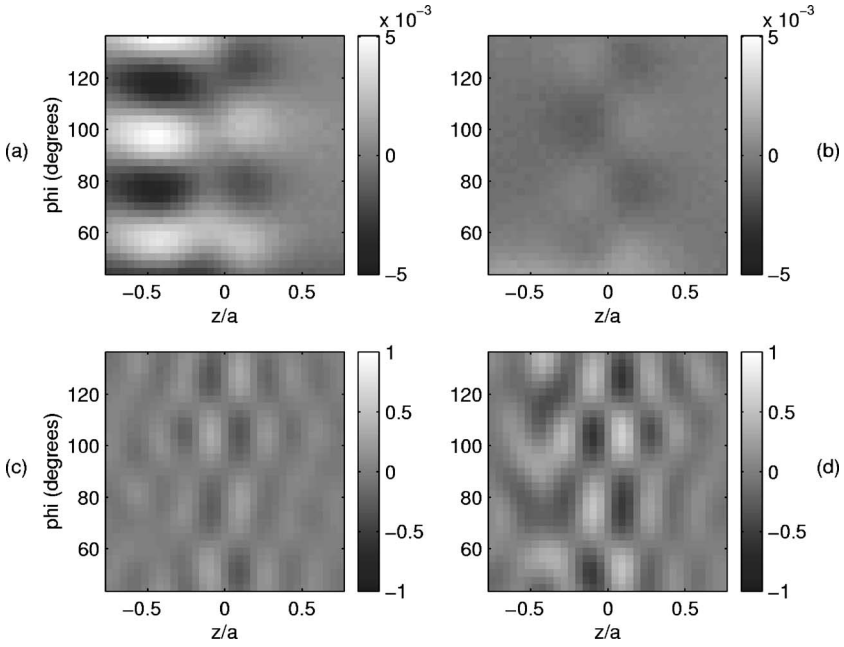


FIG. 4. (a) Real and (b) imaginary parts of the field on the measurement surface for $ka=3$, $k\delta=1$; (c) real and (d) imaginary parts the reconstructed normal velocity.

struction. The error in this case is 19%. Excluding three grid points from each side of the reconstruction reduces the error to 18%.

Truncating the sum at $N_t=115$ in the noise free case produces reconstruction error of 52% which is comparable to the noisy case. These high errors are produced because the reconstruction does not include short wavelength components of the field that decays fast away from the surface. When the measurement surface is distant from the structure, these short wavelength components will decay to field values on the measurement surface that are so small that they can not be distinguished from the noise that is present in the data.

B. Example 2

Keeping the driver at the same location, we now increase the frequency to $ka=7$. Since shorter wavelengths are excited for this higher frequency case, we reduce the size of the partial surfaces, keeping a 33×33 grid of points on each surface. The measurement surface becomes ($\rho=a+\delta$, $\pi/4\leqslant\phi\leqslant3\pi/4$, $-0.5a\leqslant z\leqslant0.5a$), ∂S becomes ($\rho=a$, $\pi/4\leqslant\phi\leqslant3\pi/4$, $-0.5a\leqslant z\leqslant0.5a$) and σ becomes ($\rho=a-\delta'$, $\pi/4\leqslant\phi\leqslant3\pi/4$, $-0.5a\leqslant z\leqslant0.5a$). We set $k\delta=0.5$ and $k\delta'=1.5$. Thus the distance between surface σ and the structure surface is three times the distance between the structure and the measurement surface in this case.

Figures 8(a) and 8(b) show the real and imaginary parts of the field, including noise, on the measurement surface; 5% noise has been added for this case also. Figures 8(c) and 8(d) show the real and imaginary parts of the computed surface normal velocity. The reconstructed normal velocity is shown below in Figs. 8(e) and 8(f). To determine where to truncate the sum for the reconstruction, functions F_n and G_n are plot-

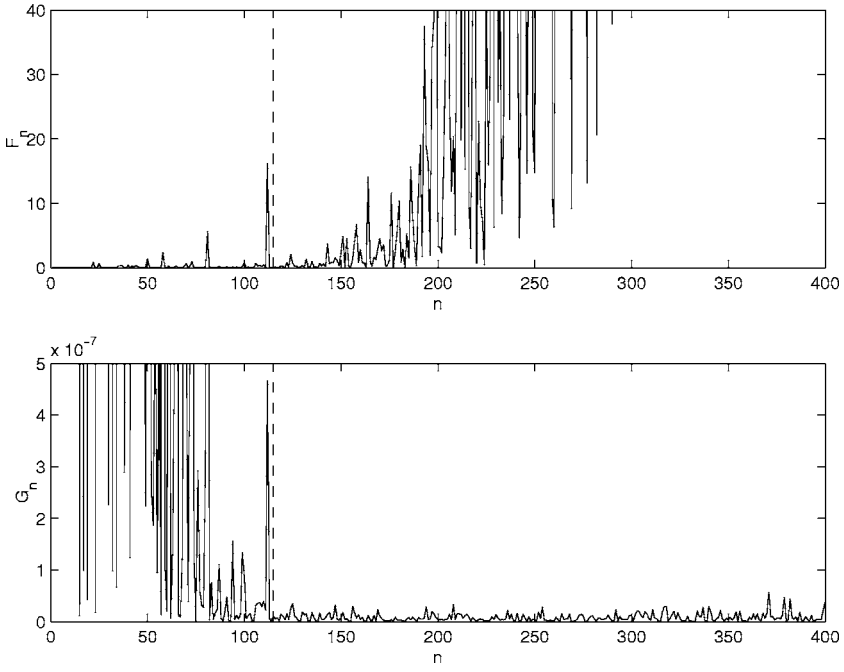


FIG. 5. F_n and G_n as a function of index n for $ka=3$, $k\delta=1$.

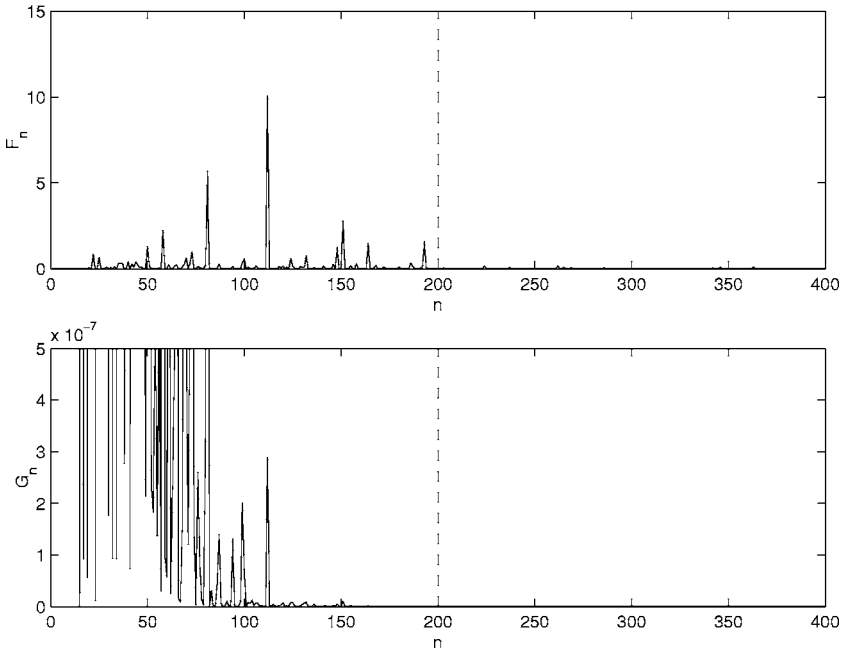


FIG. 6. F_n and G_n as a function of index n for $ka=3$, $k\delta=1$ in the noise free case.

ted in Fig. 9. In the lower plot, we observe that noise dominates for index $n>280$. We thus truncate the sum at $N_t=280$. The reconstruction error for this case is 28%.

Examination of the velocity plots show that the errors are highest around the edges. Removing three grid points from each of the sides of the measurement surface thus keeping the central 27×27 grid reduces the error to 24%. Edge effects have previously been observed in the nearfield acoustical holography when using Fourier decomposition.[20] Williams[9] has also observed that, with the boundary elements method based singular value decomposition approach, where singular value decomposition is applied to the Neumann Greens function relating the field on the measurement surface to the normal velocity on the structure surface, the reconstructed field at the edges is not as accurate as the reconstructed field in the interior. The same is observed in the examples here but the errors are not excessively high at the edges in our examples.

IV. REGULARIZATION

Next, Tikhonov regularization is applied to the examples in order to determine if it produces improved reconstructions. Instead of the truncated singular value decomposition approximation of q_i given by Eq. (12), we approximate $\mathbf{q}$ by[20]

$$q_i \simeq \sum_{j=1}^{N'} \sum_{k=1}^{N} V_{ij} \frac{\sigma_j}{\sigma_j^2 + \alpha} U_{kj}^* p_k^m, \tag{19}$$

where α is the regularization parameter. Using Eq. (19) with Eq. (3), an approximation for the pressure on the measurement surface is given by

$$\hat{p}(\vec{r_n^m}) \simeq \sum_{i=1}^{N'} \sum_{j=1}^{N'} \sum_{k=1}^{N} G_{ni} V_{ij} \frac{\sigma_j}{\sigma_j^2 + \alpha} U_{kj}^* p_k^m. \tag{20}$$

Combining the above with Eq. (8) and observing that matrix $\mathbf{V}$ is unitary we obtain

$$\hat{p}(\vec{r_n^m}) \simeq \sum_{j=1}^{N} \sum_{k=1}^{N} U_{nj} \frac{\sigma_j^2}{\sigma_j^2 + \alpha} U_{kj}^* p_k^m. \tag{21}$$

To determine the regularization parameter α, we use Morozov discrepancy principle,[20] where α is varied until the approximation to the pressure on the measurement surface, computed using Eqs. (20) or (21) differs from the measured pressure by the amount of noise present. The noise in our examples is known to be 5%.

Tikhonov regularization is first applied to example 1 with an offset distance of $\delta=a/6$. The reconstructed field on the measurement surface is shown in Fig. 10. They may be compared to the actual values shown in Figs. 2(c) and 2(d).

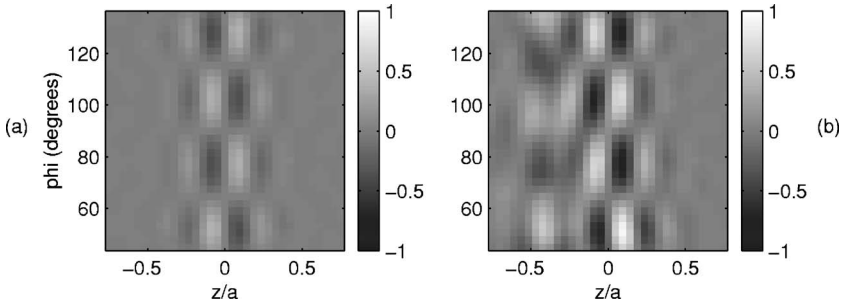


FIG. 7. (a) Real and (b) imaginary parts of the reconstructed normal velocity for $ka=3$, $k\delta=1$ in the noise free case.

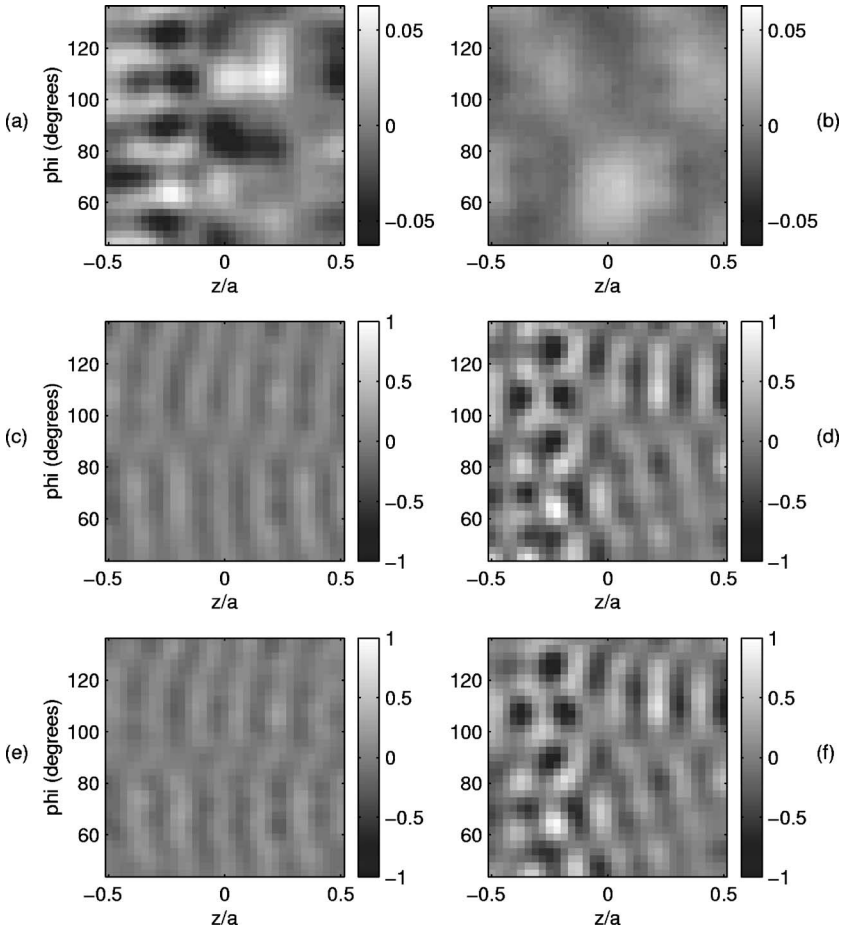


FIG. 8. (a) Real and (b) imaginary parts of the field on the measurement surface for $ka=7$, $k\delta=0.5$; (c) real and (d) imaginary parts of the normal velocity on the structure surface; (e) real and (f) imaginary parts of the reconstructed normal velocity.

The error in this case is 25% which is higher than the 20% error obtained with the truncated singular value decomposition algorithm.

Increasing the offset distance to $k\delta=1$, Tikhonov regularization algorithm produces an error of 71% for the case where 5% random noise is present in the pressure on the measurement surface. The error in this case is again higher than the 54% error obtained with the truncated singular value decomposition algorithm.

Finally applying the algorithm to example 2 where $ka=7$, the reconstruction error obtained using Tikhonov regu-

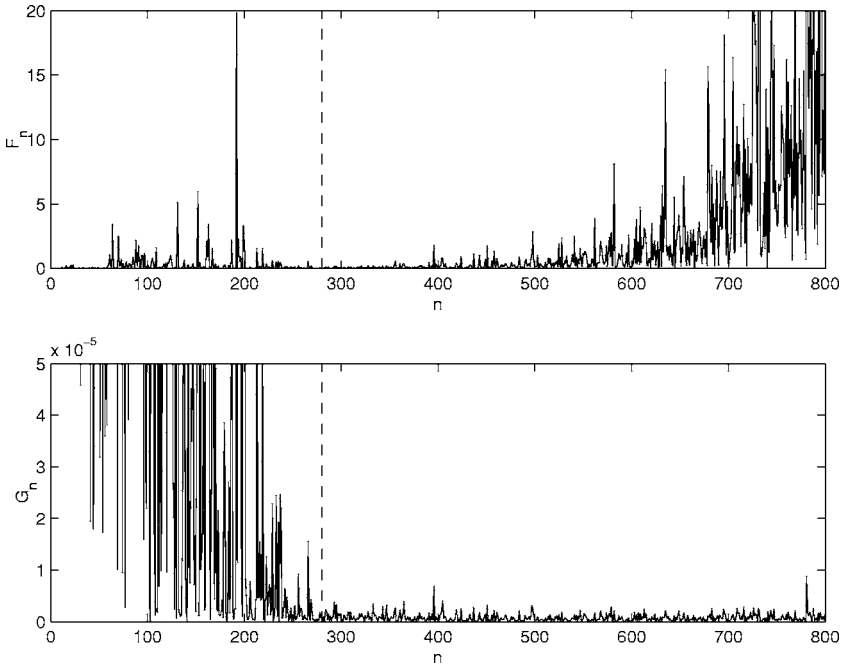


FIG. 9. F_n and G_n as a function of index n for $ka=7$, $k\delta=0.5$.

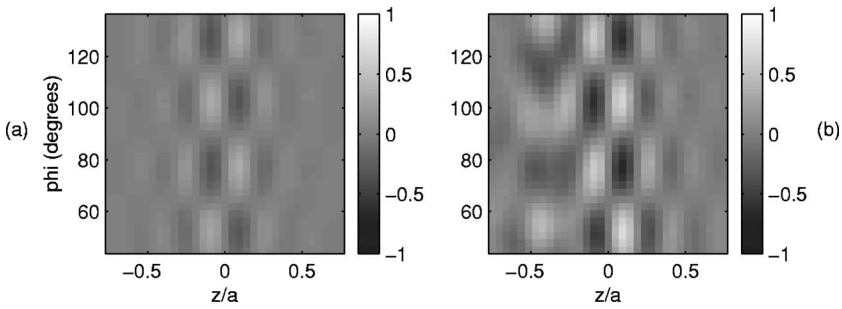


FIG. 10. (a) Real and (b) imaginary parts of the reconstructed normal velocity for $ka=3$, $k\delta=0.5$ using Tikhonov regularization.

larization is 30% which is comparable to the 28% error obtained using the truncated singular value decomposition algorithm.

V. DISCUSSION

The field on a patch on a structure surface may be reconstructed from measurements made on a nearby partial surface. The algorithm employed here approximates the field produced by the structure with the field produced by a source distribution placed at a distance δ' inside the structure. There are no concrete guidelines for choosing surface σ. We found using a surface the same length and having the same angular extent as the measurement surface and placed at a constant distance away from the measurement surface produced reasonable results. For the two cases presented at $ka=3$ and $ka=7$, an offset distance from measurement surface to surface σ, in the range $3\lambda/4\pi \leqslant \delta+\delta' \leqslant \lambda/\pi$, where $\lambda=2\pi/k$ is the acoustic wavelength, produced reasonable results.

The length and the width of the reconstruction surface were approximately 3/4 of an acoustic wavelength λ in example 1. They were approximately $1.7\lambda \times 1.1\lambda$ in example 2. The measurement surfaces were only slightly larger. We observe that a reconstruction region having a length that is as small as a wavelength or less still produced reasonable results.

Because of the ill-posed nature of this inverse problem an ill-conditioned matrix must be inverted. Singular value decomposition is applied to invert the matrix. We observe that larger number of states are kept in the reconstruction of the higher frequency case, i.e., N_t is larger at the higher frequency, because of the presence of the shorter wavelength components at the higher frequencies.

The example calculations show the importance of making the measurement very close to the structure. The presence of the frames in the cylinder produces surface fields with short wavelengths that necessitate having a measurement surface very close to the structure. In the example calculation $k\delta=0.5$ produced surface normal velocity fields with errors of 20%–28% with the truncated singular value decomposition algorithm when edges are included in the reconstruction.

The application of Tikhonov regularization with Morozov discrepancy principle did not improve the quality of the reconstruction. It produced errors that were comparable to or higher than the truncated singular value decomposition algorithm.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research.

[1]E. G. Williams and J. D. Maynard, "Holographic imaging without the wavelength resolution limit," Phys. Rev. Lett. **45**, 554–557 (1980).
[2]J. D. Maynard, E. G. Williams, and Y. Lee, "Nearfield acoustic holography: I. Theory of generalized holography and the development of NAH," J. Acoust. Soc. Am. **78**, 1395–1413 (1985).
[3]W. A. Veronesi and J. D. Maynard, "Digital holographic reconstruction of sources with arbitrarily shaped surfaces," J. Acoust. Soc. Am., **85**, 588–598 (1989).
[4]G. T. Kim and B. H. Lee, "3-D sound source reproduction using the Helmholtz integral equation," J. Sound Vib. **136**, 245–261 (1990).
[5]G. V. Borgiotti, A. Sarkissian, E. G. Williams, and L. Schuetz, "Conformal generalized near-field acoustic holography for axisymmetric geometries," J. Acoust. Soc. Am. **88**, 199–209 (1990).
[6]A. Sarkissian, "Near-field acoustic holography for axisymmetric geometries: A new formulation," J. Acoust. Soc. Am. **88**, 961–966 (1990).
[7]P. Stepanishen, "A generalized internal source density method for the forward and backward projection of harmonic pressure fields from complex bodies," J. Acoust. Soc. Am. **101**, 3270–377 (1997).
[8]A. Sarkissian, C. F. Gaumond, E. G. Williams and B. H. Houston, "Reconstruction of the acoustic field over a limited surface area on a vibrating cylinder," J. Acoust. Soc. Am. **93**, 48–54 (1993).
[9]E. G. Williams and B. H. Houston, "Fast Fourier transform and singular value decomposition formulation for patch nearfield acoustical holography," J. Acoust. Soc. Am. **114**, 1322–1333 (2003).
[10]K. Saijyou and S. Yoshikawa, "Reduction methods of the reconstruction error for large-scale implementation of near-field acoustical holography," J. Acoust. Soc. Am. **110**, 2007–2023 (2001).
[11]J. Hald, "Patch near-field acoustical holography using a new statistically optimal method," *Proceedings of Inter-noise 2003*, p. 2203.
[12]R. Steiner and J. Hald, "Near-field acoustical holography without the errors and limitations caused by the use of spatial DFT," Int. J. Acoust. Vib. **6**, 83–89 (2001).
[13]G. H. Koopmann, L. Song, and J. B. Fahnline, "A method for computing acoustic fields based on the principle of wave superposition," J. Acoust. Soc. Am. **86**, 2433–2438 (1989).
[14]R. D. Miller, E. T. Moyer, Jr., H. Huang, and H. Uberall, "A comparison between the boundary element method and the wave superposition approach for the analysis of the scattered fields from rigid bodies and elastic shells," J. Acoust. Soc. Am. **89**, 2185–2196 (1991).
[15]A. Sarkissian, "Method of superposition applied to scattering from a target in shallow water," J. Acoust. Soc. Am. **95**, 2340–2345 (1994).
[16]Y. I. Bobrovnitskii and T. M. Tomilina, "General properties and fundamental errors of the method of equivalent sources," Acoust. Phys. **41**, 649–660 (1995).
[17]A. Sarkissian "Extension of measurement surface in near-field acoustic holography," J. Acoust. Soc. Am. **115**, 1593–1596 (2004).
[18]Y. I. Bobrovnitskii, "The problem of vibration field reconstruction: Statement and general properties," J. Sound Vib. **247**, 145–163 (2001).
[19]H. Allik, R. Dees, S. Moore, and D. Pan, *SARA-2D User's Manual*, BBN Systems and Technologies, New London, CT, 1995.
[20]E. Williams, "Continuation of acoustic near-fields," J. Acoust. Soc. Am. **113**, 1273–1281 (2003).

A space–time filtered gradient method for detecting directions of echoes and transient sounds[a)]

Terry L. Henderson[b)] and Terry J. Brudner[c)]
Applied Research Laboratories, The University of Texas at Austin, Austin, Texas 78713-8029

(Received 19 November 2003; revised 15 April 2005; accepted 2 May 2005)

The gradient vector (e.g., of the acoustic pressure) indicates the direction to the source of a wave, but it is easily corrupted by interference from other directions. However the gradient concept, even for higher orders, can be applied rigorously to a beamforming aperture that shields against interference, thereby allowing precise determination of the direction of sound echoes or emissions, especially for very brief, broadband transient sounds. In this treatment there is no gradient sensor *per se*; the aperture weighting supplants that function. Various geometric shapes can be used as apertures, but simple plates are often best, and the required weightings can be realized by patterned electrodes. The method is shown to be a natural extension of earlier techniques and inventions, and useful interpretations and generalizations are provided, such as compound and steered apertures, instantantly re-steerable nulls, and an equivalence to tracking acoustic particle motion after acoustical shielding from interference. There are two stages: aperture signal extraction, and ratio processing based upon Watson–Watt concepts, for which statistically based formulas are useful. In-water test results are provided.
 [DOI: 10.1121/1.1940427]

PACS number(s): 43.20.Ye, 43.30.Yj, 43.60.Qv, 43.60.Fg [WMC] Pages: 679–695

I. INTRODUCTION

An acoustic sensor is the preferred tool to investigate underwater scenery from a distance, because sound propagates well and is usually reflected by, and sometimes emitted by, conspicuous features such as shipwrecks, seamounts, fish, sea mines, wellheads, and submarines. Much about their geometry can be discerned from the angles of acoustic rays intercepted by the sensor, but its angular resolution is only about λ/d_Q, where d_Q is its aperture diameter. This limit applies whenever rays arrive simultaneously from a distribution of angles. ("Superdirective" arrays overcome the limit in theory, but succeed only marginally in practice.[1])

A pulse echo or transient emission from a distinct feature *can* have a single, discrete ray angle—if only momentarily, until corrupted by multipath. In theory, the direction of a single wave is revealed instantly by its gradient, in a tiny aperture, with perfect accuracy. In practice, sensor noise, distortion, or interfering waves limit accuracy; nevertheless, it can be orders of magnitude better than λ/d_Q if interfering waves are controlled. Active sonars use directional transmissions to avoid extraneous echoes, but a beam can be formed by applying a space–time filter to the *receiving* aperture of either an active or passive system. If this beam, despite being no narrower than λ/d_Q, can block enough interference then the space–time filtered wave field's gradient will indicate the sound angle accurately.[2]

Notation and assumptions: t, c, f, λ, and $\mathbf{x}$ denote time, sound speed, frequency, wavelength, and position vector (with Cartesian coordinates x_i). In a slight break with tradition, $\mathbf{n}$ denotes the unit vector pointing to the wave's *source*, not its propagating direction.[3] Its Cartesian components n_i are the direction cosines of the source; moreover, $n_i = \sin\theta_i$, where θ_i is the source bearing from the x_i axis's normal. The wave field $p(\mathbf{x},t)$ is acoustic pressure or any other scalar field (e.g., eastward component of acoustic displacement). To emphasize frequency independence and application to broadband echoes and transients, we stay in the time domain where possible, but $p(\mathbf{x},t)$ can have an imaginary part via Hilbert transform. Source distances are assumed quite larger than $d_Q^2/\lambda_{\min}$ so their emanations look like plane waves within the receiving aperture Q, which is assumed to be a volume aperture[4] at the origin.

II. FIRST ORDER SPACE–TIME FILTERED GRADIENT METHOD

A. Discerning direction from the gradient vector: Filtered or unfiltered

A plane wave obeys $p(\mathbf{x},t) = p(0, t + c^{-1}\mathbf{n}\cdot\mathbf{x})$, so $\nabla p = c^{-1}\dot{p}\mathbf{n}$. Thus, ∇p always points at the source[5] or opposite it, alternating with the sign of $\dot{p}$. If both ∇p and $\dot{p}$ are measured this alternation can be rectified in two ways: $c\dot{p}^{-1}\nabla p = \mathbf{n}$, or $\dot{p}^*\nabla p = c^{-1}|\dot{p}|^2\mathbf{n}$ (* denotes conjugate). If the incident field p_{main} from a main source at direction $\mathbf{n}_{\text{main}}$ is corrupted by an interference p_{int} at direction $\mathbf{n}_{\text{int}}$ separated by angle $\Delta\Theta$, then the new gradient, $c^{-1}(\dot{p}_{\text{main}}\mathbf{n}_{\text{main}} + \dot{p}_{\text{int}}\mathbf{n}_{\text{int}})$, is "pulled" or "pushed," depending on the sign of $\dot{p}_{\text{int}}/\dot{p}_{\text{main}}$ (see Fig. 1), by

[a)]This paper expands upon material presented at the Fall 1994 and Spring 1995 meetings of the Acoustical Society of America.
[b)]Electronic mail: henderson@arlut.utexas.edu
[c)]Electronic mail: brudner@arlut.utexas.edu

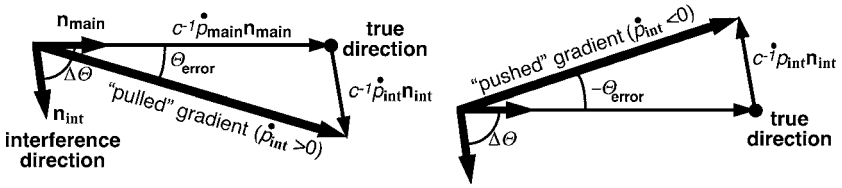


FIG. 1. Perturbation of the gradient due to interference p_{int} at angle $\Delta\Theta$ from the main wave field p_{main} (depicted for $\dot{p}_{\text{main}} > 0$). This also applies with the space–time filtered wave fields $(\dot{p}_{\text{main}})_{\text{sh}}$ and $(\dot{p}_{\text{int}})_{\text{sh}}$ substituted for $\dot{p}_{\text{main}}$ and $\dot{p}_{\text{int}}$.

$$\Theta_{\text{error}} = \arctan\left[\frac{\dot{p}_{\text{int}} \sin\Delta\Theta}{\dot{p}_{\text{main}} + \dot{p}_{\text{int}} \cos\Delta\Theta}\right] \approx \frac{\dot{p}_{\text{int}}}{\dot{p}_{\text{main}}} \sin\Delta\Theta, \tag{1}$$

with the approximation being valid when $|\dot{p}_{\text{int}}/\dot{p}_{\text{main}}| \ll 1$. A 20 dB weaker interference 90° away can cause a ±9° error. This is usually unacceptable.

A selective filter, expressed as a space–time convolution by some kernel γ,

$$p_{\text{sh}}(\mathbf{x},t) = \iiint\int_{-\infty}^{\infty} \gamma(\mathbf{y},\tau) p(\mathbf{x}-\mathbf{y}, t-\tau) dy_1 dy_2 dy_3 d\tau, \tag{2}$$

can *shield* against such interferences: For a plane wave

$$p(\mathbf{x}-\mathbf{y}, t-\tau) = p(0, t + c^{-1}\mathbf{n}\cdot\mathbf{x} - \tau - c^{-1}\mathbf{n}\cdot\mathbf{y}), \tag{3}$$

so p_{sh} depends only on $(t+c^{-1}\mathbf{n}\cdot\mathbf{x})$, ensuring that p_{sh} is still a plane wave; i.e., the filter preserves directions and speeds of plane waves. Therefore $\nabla p_{\text{sh}} = c^{-1}\dot{p}_{\text{sh}}\mathbf{n}$ as well. If one (i) chooses an arbitrary aperture weighting function $W(\mathbf{x})$ that is zero outside some aperture volume Q, (ii) chooses an arbitrary time-domain filter impulse response $h(t)$, (iii) sets $\gamma(\mathbf{x},t) = h(t)W(-\mathbf{x})$, and then (iv) samples $c^{-1}\dot{p}_{\text{sh}}$ and ∇p_{sh} at the origin, the results will be as follows [denoted as $\alpha(t)$ and $\mathbf{a}(t)$]:

$$\begin{aligned}\alpha(t) &\triangleq (c^{-1}\dot{p}_{\text{sh}})_{\mathbf{x}=0} \\ &= c^{-1}\frac{d}{dt}\left[h(t) * \iiint\int_{-\infty}^{\infty} W(-\mathbf{y}) \times p(0-\mathbf{y}, t) dy_1 dy_2 dy_3\right],\end{aligned} \tag{4}$$

i.e.,

$$\alpha(t) = c^{-1}\dot{h}(t) * \iiint_{Q} W(\mathbf{x}) p(\mathbf{x},t) dV, \tag{5}$$

$$\begin{aligned}\mathbf{a}(t) &\triangleq (\nabla p_{\text{sh}})_{\mathbf{x}=0} = h(t) * \iiint\int_{-\infty}^{\infty} W(-\mathbf{y}) \\ &\quad \times [\nabla_{\mathbf{x}} p(\mathbf{x}-\mathbf{y}, t)]_{\mathbf{x}=0} dy_1 dy_2 dy_3\end{aligned} \tag{6}$$

$$\begin{aligned}&= h(t) * \iiint\int_{-\infty}^{\infty} W \nabla p dV \\ &= h(t) * \iiint\int_{-\infty}^{\infty} \{\nabla[Wp] - [\nabla W]p\} dV,\end{aligned} \tag{7}$$

i.e.,

$$\mathbf{a}(t) = h(t) * \left\{\iint_{\partial Q} W(\mathbf{x}) p(\mathbf{x},t) \hat{\mathbf{n}}(\mathbf{x}) dS - \iiint_{Q} [\nabla W(\mathbf{x})] p(\mathbf{x},t) dV\right\}, \tag{8}$$

where $*$ denotes time-convolution. ∂Q is the aperture's boundary surface and $\hat{\mathbf{n}}(\mathbf{x})$ is its outward unit-normal (as contrasted with $\mathbf{n}$, the unit vector pointing toward the source). The surface integral in Eq. (8) arises due to a Divergence Theorem corollary.[6]

Since $\nabla p_{\text{sh}} = c^{-1}\dot{p}_{\text{sh}}\mathbf{n}$ for a plane wave from direction $\mathbf{n}$, the results of Eqs. (5) and (8) must satisfy

$$\mathbf{a}(t) = \alpha(t)\mathbf{n}. \tag{9}$$

To produce $\alpha(t)$ and $\mathbf{a}(t)$ via Eqs. (5) and (8) one applies weightings W and ∇W and integrates the wave field p over Q and its surface ∂Q, filtering the results with impulse responses $h(t)$ and $c^{-1}\dot{h}(t)$. By Eq. (9), the vector $\mathbf{a}(t)$ points at or directly opposite the source, alternating with the sign of $\alpha(t)$. The alternation can be rectified in two ways: $\mathbf{a}(t)/\alpha(t) = \mathbf{n}$, or $\alpha^*(t)\mathbf{a}(t) = |\alpha(t)|^2\mathbf{n}$. The source's direction cosines are

$$n_i = \frac{a_i(t)}{\alpha(t)} \quad \text{for } i = 1,2,3. \tag{10}$$

(Division by a zero-crossing signal is problematic, but Sec. VI discusses remedies.)

Equations (5) and (8) employ no gradient sensor of $p(\mathbf{x},t)$ *per se*. Instead, the gradient has already operated on the weighting $W(\mathbf{x})$. The designer chooses $W(\mathbf{x})$ and the aperture Q to make the gradient simple and to shield interference from other sources, especially those at large angular separations—the most damaging ones. Figure 1 and Eq. (1) still apply with p replaced by p_{sh}, but the aperture's directivity can make $(\dot{p}_{\text{int}})_{\text{sh}}/(\dot{p}_{\text{main}})_{\text{sh}}$ much smaller than $\dot{p}_{\text{int}}/\dot{p}_{\text{main}}$, reducing the error proportionately. Due to linearity in Eqs. (4), (6), and (9), interferences at multiple directions $\mathbf{n}_{[m]}$ will deflect the space–time filtered gradient to

$$(\nabla p_{\text{sh}})_{\mathbf{x}=0} = \mathbf{a}(t) = \alpha_{\text{main}}(t)\mathbf{n}_{\text{main}} + \sum_m \alpha_{[m]}(t)\mathbf{n}_{[m]}. \tag{11}$$

B. Implementation issues

In a complete implementation Eqs. (5) and (8) must physically produce $\alpha(t)$ and the three Cartesian components of the vector $\mathbf{a}(t)$ to determine the source direction vector $\mathbf{n}$. The time-domain filters $c^{-1}\dot{h}(t)$ and $h(t)$ can be realized exactly in analog form as single-pole filters with identical RC time constants—the first being high-pass, the second low-

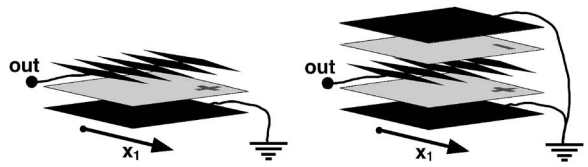


FIG. 2. Exploded views of a sheet sensor with an electrode pattern for nonuniform sensitivity, specifically, linearly tapered sensitivity with respect to x_1. The (gray) sheet-transducers, e.g., of voided PVDF, have their "+" sides against the patterned electrode. The (black) triangular "spears" are a thin metallic coating that forms the patterned-electrode surface. The full-rectangle (black) metallic coatings are the ground electrodes. The construction on the left is simpler, but the ground-encapsulated version on the right avoids conductive or capacitative short-circuits due to stacking with other sheet sensors, and minimizes noise pickup from power sources or radio stations.

pass: $h(t)=(RC)^{-1}\exp(-t/RC)$. Realization in software is also easy, especially in echoranging systems that use a transmitted wave form replica for matched filtering. The replica used for the $\alpha(t)$ processing channel can be differentiated and scaled by $-c^{-1}$, while the $\mathbf{a}(t)$ channels' replica is unmodified. Also, Eq. (9) is preserved by such operations as time-varied gains, filtering, transforms (Hilbert or Fourier), and frequency shifting, if strictly matched in the $\alpha(t)$ and $\mathbf{a}(t)$ channels.

The spatial integrals in Eqs. (5) and (8) are harder. One approach is to sum weighted outputs from point sensors of $p(\mathbf{x},t)$ distributed throughout the aperture Q and its surface. The sensors must be small enough to avoid perturbing the plane waves, yet closely spaced to form the integrals accurately.[7] This is feasible at low frequencies (e.g., 100 Hz), but at high frequencies (e.g., 100 kHz) it is difficult to make tiny transducers with negligible electrical noise. Thus our examples include an optional alternative for the surface integral: an acoustically transparent transducer sheet, between negligibly thin surface electrodes, integrating $p(\mathbf{x},t)$ into voltage. A good realization is polyvinylidene flouride (PVDF) sheet[8] whose acoustic impedance has been matched to water by "working" it to create microscopic voids to reduce its density. Attenuation and refraction are negligible for thin sheets. In principle, the volume integral can be formed by sheets that are trimmed and stacked to fill Q, but absorption and refraction can build up unless the aperture Q is a thin plate or shell.

If $W(\mathbf{x})=1$ everywhere in Q then the volume integral in Eq. (8) vanishes, leaving only a surface integral of a simple integrand. If, instead, W tapers to zero on the boundary, the surface integral vanishes. (Examples of both types are included in Sec. III, and in-water test results are presented in Sec. VI B.) Nonuniform sensitivity is required wherever W, ∇W, or $\hat{\mathbf{n}}(\mathbf{x})$ are not constant under the integrals. Graduated poling or annealing of a piezoelectric sheet can give a continuously varying sensitivity, but a more obvious method is to etch or cut out patterns in its electrode, leaving a partial coverage with a density that approximates a continuous weighting (with zones of negative weighting connected to a separate output). For example, spear-shaped "tiles" can approximate a tapered weighting for a thin plate aperture Q (see Fig. 2).

C. Channel-count reduction

Independent production of $\alpha(t)$ and all three Cartesian components $a_i(t)$ of $\mathbf{a}(t)$ requires four processing channels, but fewer will usually suffice.

(1) *If only one direction component n_J is sought* then $\alpha(t)$ and $a_J(t)$ are enough (and only n_J has to be distinct). For example, suppose a short-pulse bathymetric sidescan sonar scans a patch of seafloor to measure its elevation. If the $+x_3$ axis points up, then $n_3=\sin\theta_3$, where θ_3 is the elevation angle of the patch. Since the slant range is $r=ct/2$, where t is the time delay of the echo, it follows that the elevation of the patch relative to the receiving aperture is

$$\text{elevation} = r\sin\theta_3 = \frac{ct}{2}n_3(t) = \frac{ct}{2}\frac{a_3(t)}{\alpha(t)} \quad \text{(typically negative)}. \tag{12}$$

(2) *If the goal is just the azimuth*, $\phi=\arctan(n_2/n_1)$, then $a_1(t)$ and $a_2(t)$ are enough since $n_2/n_1=a_2(t)/a_1(t)$. The 180° ambiguity must be resolved independently. Time-domain filtering is not needed. [Let $h(t)=\delta(t)$.]

(3) *If the source direction is confined to a half-space* then one member of the set $\{\alpha,a_1,a_2,a_3\}$ can be omitted. For example, if α is omitted then the direction is indicated by $\pm\mathbf{a}(t)$, giving a polar-opposite ambiguity that can be resolved by the constraint. If one of the a_i's is omitted instead, e.g., a_3, then $n_i=a_i(t)/\alpha(t)$ for $i=1,2$ and $n_3=\pm\sqrt{1-n_1^2-n_2^2}$, giving a mirror ambiguity that can be resolved by the constraint if the axes are properly oriented. If Q is near a large, reflective, barrier plane (e.g., a mounting plate), orthogonal to the x_3 axis, then the source's reflection image has already introduced this mirror ambiguity; yet, it also resolves it since the barrier rules out sources behind it.

D. The directional shielding pattern $D_Q(\mathbf{n}/\lambda)$

Even if $\alpha(t)$ is not measured, its directional response is needed to assess performance via Eq. (11). Denoting the Fourier transforms of $h(t)$ and $p(\mathbf{x},t)_{\mathbf{x}=0}$ as $H(f)$ and $P_0(f)$, and transforming Eq. (5) with aid of the plane wave assumption, one gets

$$\mathcal{FT}\{\alpha(t)\} \triangleq \int_{-\infty}^{\infty}\alpha(t)e^{-j2\pi ft}dt = [j2\pi fH(f)c^{-1}D_Q(\mathbf{n}/\lambda)]P_0(f), \tag{13}$$

where

$$D_Q(\mathbf{n}/\lambda) \triangleq \iiint_Q W(\mathbf{x})\exp(j2\pi\mathbf{x}\cdot\mathbf{n}/\lambda)dV \tag{14}$$

is the directional response of the weighted aperture before time-domain filtering, which is the three-dimensional (3D) Fourier transform of $W(\mathbf{x})$ with spatial frequency variables replaced by $-n_i/\lambda$. It has conjugate symmetry with respect to direction reversal, $D_Q(-\mathbf{n}/\lambda)=D_Q^*(\mathbf{n}/\lambda)$, so if $|D_Q|$ or $\angle D_Q$ is depicted on the $\mathbf{n}$ sphere it suffices to show one side.

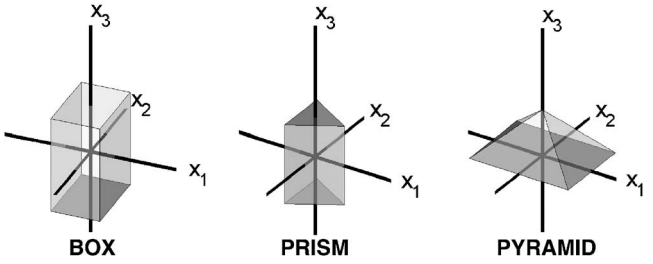


FIG. 3. Polyhedron aperture shapes used as examples: rectangular box, prism with equilateral triangle bases, pyramid with rectangular base. See Table I for more details.

Reduction to a surface integral is possible if $W=1$ on Q. In that case the volume integral in Eq. (8) vanishes, and, since $\alpha(t)=(\mathbf{n}\cdot\mathbf{n})\alpha(t)=\mathbf{n}\cdot\mathbf{a}(t)$, it follows that

$$\alpha(t)=\mathbf{n}\cdot h(t)*\iint_{\partial Q} p(\mathbf{x},t)\hat{\mathbf{n}}(\mathbf{x})dS, \tag{15}$$

which, via the plane wave assumption, results in an alternate formula [if $W(\mathbf{x})=1$]:

$$D_Q(\mathbf{n}/\lambda)=(j2\pi/\lambda)^{-1}\mathbf{n}\cdot\iint_{\partial Q}\exp(j2\pi\mathbf{x}\cdot\mathbf{n}/\lambda)\hat{\mathbf{n}}(\mathbf{x})dS. \tag{16}$$

III. SIMPLE APERTURE EXAMPLES

A. Polyhedron with uniform weighting

If Q is a polyhedron of M planar polygon facets $F_1,F_2,\ldots,F_M$, where $W(\mathbf{x})=1$ in Q and on its surface ∂Q, then the second integral in Eq. (8) vanishes, so that $\mathbf{a}(t)$ is just the time-domain-filtered sum of the facet integrals:

$$\mathbf{a}(t)=h(t)*\sum_{m=1}^{M}\bar{F}_m(t)\hat{\mathbf{n}}_{F_m}, \tag{17}$$

where $\hat{\mathbf{n}}_{F_m}$ is the outward unit normal vector on the mth facet, and

$$\bar{F}_m(t)\triangleq\iint_{F_m}p(\mathbf{x},t)dS. \tag{18}$$

These equations can be implemented by thin, flat, sheet-sensors of uniform sensitivitity. The volume integral for $\alpha(t)$ is unnecessary if independent knowledge is available to resolve the sign of $\pm\mathbf{n}$ after it is inferred from observations of $\mathbf{a}(t)$.[9] Three polyhedron examples are depicted in Fig. 3 and summarized in Table I. Directional patterns of $|D_Q(\mathbf{n}/\lambda)|$ are depicted on the unit sphere in Fig. 4 for selected sizes, calculated from Eq. (16). Predictably, prominent directional maxima appear that are normal to the facets, but they have deep frequency response notches when the spacing of opposing parallel facets is $m\lambda$, due to cancellation. In particular, for an $L_1\times L_2\times L_3$ rectangular box the $\hat{\mathbf{n}}_{F_m}$'s point along the axes, so that if $h(t)=\delta(t)$ then

$$a_i(t)=\bar{F}_i(t)-\bar{F}_{i+3}(t) \tag{19}$$

[assuming the $+x_i$ and $-x_i$ axes impale the ith and $(i+3)$th facets]. The frequency response notches vanish if, for example, $L_3\leqslant\lambda/2$ so that Q is a thin plate with a strong lobe along the x_3 axis. But for very small L_3 the narrow edge-facets used to get $a_1(t)$ and $a_2(t)$ give weak, noisy outputs, and the differencing of large, closely spaced, identical, sheet sensors to get $a_3(t)$ is problematic. The prism also has opposing facets (its bases) with frequency response notches, but the response there is weaker and less relevant than in the three directions normal to its long-panel facets

TABLE I. The aperture Q as a box, prism, or pyramid with uniform weighting, $W(\mathbf{x})=1$. The matrix of face-normal (column) vectors and directional response formulas are listed.

	Rectangular box ($M=6$)	Prism w/equilateral triangle bases ($M=5$)	Pyramid w/rectangular base ($M=5$)
Uniformly weighted aperture: $Q\Rightarrow$	An $L_1\times L_2\times L_3$ rectangular box centered at the origin, its six faces F_i skewered on the coordinate axes: F_i at $x_i=L_i/2$, F_{i+3} at $x_i=-L_i/2$.	A prism of length L_3; Bases F_4, F_5 are equilateral triangle of edge L_e skewered on x_3 axis at $x_3=L_3/2$; F_1 skewered on the x_1 axis at $x_1=-L_e/2\sqrt{3}$. F_2, F_3 face 60° right and left of the $+x_1$ axis.	A pyramid of altitude L_3, its $L_1\times L_2$ base F_5 on the x_1x_2 plane, centered, with sides parallel to the axes. Sides F_1, F_2 foot on the $+x_1$, $+x_2$ axes; F_3, F_4 foot on the $-x_1$, $-x_2$ axes.
Unit-normals to faces: (as columns)	$\begin{bmatrix}1&0&0&-1&0&0\\0&1&0&0&-1&0\\0&0&1&0&0&-1\end{bmatrix}$	$\begin{bmatrix}-1&1/2&1/2&0&0\\0&\sqrt{3}/2&-\sqrt{3}/2&0&0\\0&0&0&-1&1\end{bmatrix}$	$\begin{bmatrix}\check{\chi}_1&0&-\check{\chi}_1&0&0\\0&\check{\chi}_2&0&-\check{\chi}_2&0\\\chi_1&\chi_2&\chi_1&\chi_2&-1\end{bmatrix}$ $\chi_n=1/\sqrt{1+4L_3^2/L_n^2}$; $\check{\chi}_n=2L_3\chi_n/L_n$.
Possible scenario	Sign of $\pm\mathbf{n}$ resolvable by independent means, or via a crude approximation $\tilde{\alpha}(t)$. No requirement for $\alpha(t)$.	Only the source's azimuth angle in the x_1x_2 plane is to be determined, i.e., $\phi=\arctan(n_2/n_1)=\arctan[a_2(t)/a_2(t)]$. No requirement for $\alpha(t)$ or $a_3(t)$.	Base mounted on absorptive wall. Sign of n_3 is therefore positive. No requirement for $\alpha(t)$.
Direc tional response: $D(\mathbf{n}/\lambda)=$	$L_1L_2L_3\,\mathrm{sinc}[(L_1/\lambda)n_1]\times\mathrm{sinc}[(L_2/\lambda)n_2]\mathrm{sinc}[(L_3/\lambda)n_3]$	$\frac{\sqrt{3}L_3L_e^2}{8j\pi\nu_2}\,\mathrm{sinc}\,\nu_3 e^{j\pi\nu_1/\sqrt{3}}\times[e^{j\pi\nu_2}\,\mathrm{sinc}(\nu_2-\sqrt{3}\nu_1)-e^{-j\pi\nu_2}\,\mathrm{sinc}(\nu_2+\sqrt{3}\nu_1)]$ $\nu_3=n_3L_3/\lambda$; $\nu_n=\frac{1}{2}n_nL_e/\lambda(n=1,2)$..	$L_1L_2L_3/j\nu_1\nu_2\sum_{m=1}^{2}\sum_{n=1}^{2}e^{j\nu_{m,n}}-e^{j\nu_3}/(\nu_3-\nu_{m,n})(-1)^{m+n}$ $\nu_m=2\pi n_mL_m/\lambda$; $\nu_{m,n}=[(-1)^m\nu_1+(-1)^n\nu_2]/2$.

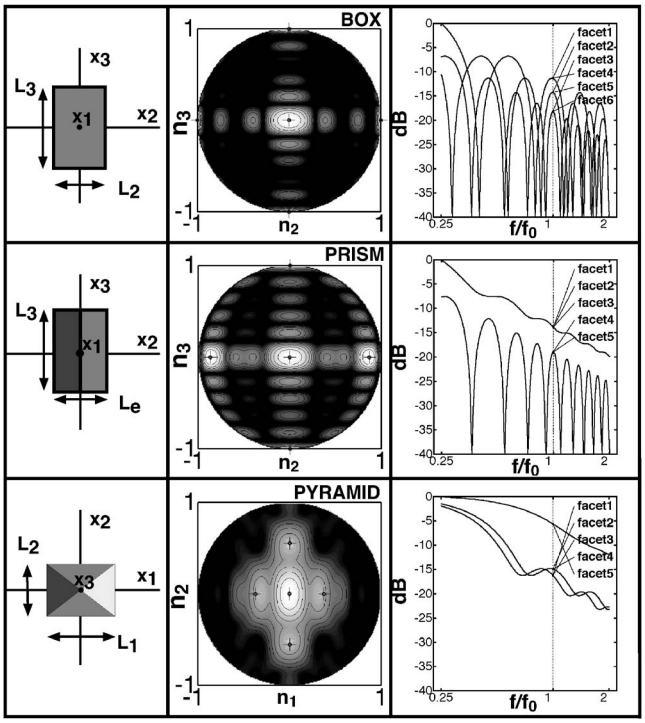


FIG. 4. Direction/frequency response, $20\log_{10}|D_Q(\mathbf{n}/\lambda)|$, for box (upper), prism (middle), and pyramid (lower). (See Table I). Directional response at nominal frequency f_0 ($=c/\lambda_0$) is plotted in 3 dB contours and grayscale on the unit sphere as seen from a relevant direction. Facet-normal directions marked with cross-stars. Also shown: Frequency response at face-normal directions for $0.25f_0<f<2f_0$. *Dimensions*: box—$L_1=2.5\lambda_0$, $L_2=3.5\lambda_0$, $L_3=5.5\lambda_0$; prism—triangular bases of edge-length $L_e=3.5\lambda_0$; altitude $L_3=5.5\lambda_0$; pyramid—Rectangular base, $L_1=5\lambda_0$ by $L_2=3\lambda_0$; altitude $L_3=1\lambda_0$.

(six if you count the backlobes), where the frequency response is relatively smooth.

Frequency response notches are avoided by a polyhedron without parallel facets, e.g., a many-faceted hemisphere joined at the equator to one big nearly circular polygon. But a simple pyramid of low altitude also largely avoids these problems while still clustering its directivity around the x_3 axis, as Fig. 4 shows. (Its altitude can go to zero if the azimuth angle, $\arctan[a_2(t)/a_1(t)]$, is the sole objective.)

B. Thin plate of uniform thickness, with $W=W(x_1,x_2)$

For a flat frequency-response beam along the x_3 axis, Q can be a thin plate of x_3 thickness $\epsilon<\lambda_{\min}/4$, its boundary ∂Q consisting of faces F_+ and F_- that sandwich the x_1x_2 plane and an orthogonal edge-strip F_e around the arbitrary perimeter, with $W=W(x_1,x_2)$. From Eq. (14) the directional response $D_Q(\mathbf{n}/\lambda)$ reduces to the two-dimensional Fourier transform of $W(x_1,x_2)$, its spatial frequencies replaced by $-n_i/\lambda$, $(i=1,2)$, with a trailing factor $\epsilon\,\mathrm{sinc}(n_3\epsilon/\lambda)$. In Cartesian coordinates Eq. (8) simplifies to

$$a_i(t)=h(t)*\left\{\iint_{F_e}W(x_1,x_2)p(\mathbf{x},t)\hat{n}_i(\mathbf{x})dS-\iiint_Q\frac{\partial W}{\partial x_i}p(\mathbf{x},t)dV\right\}\quad(i=1,2),\tag{20}$$

$$a_3(t)=h(t)*\left\{\iint_{F_+}W(x_1,x_2)p(\mathbf{x},t)dS-\iint_{F_-}W(x_1,x_2)p(\mathbf{x},t)dS\right\},\tag{21}$$

where $\hat{n}_i(\mathbf{x})$ denotes the Cartesian components of $\hat{\mathbf{n}}(\mathbf{x})$. For thin plates Eq. (21) is ill-conditioned, but Eq. (5) can be implemented to produce $\alpha(t)$ to get the direction cosines $n_i=a_i(t)/\alpha(t)$, $(i=1,2)$; then $n_3=\pm\sqrt{1-n_1^2-n_2^2}$ ("+" if Q is backed by a barrier plane[10] as described in Sec. II C). Here are four categories of weighting.

1. Uniform weighting: $W(x_1,x_2)=1$

To get $\alpha(t)$, $p(\mathbf{x},t)$ is integrated over the plate. To get $a_i(t)$ from Eq. (20) for $i=1,2$, $p(\mathbf{x},t)$ is weighted by $\hat{n}_i(\mathbf{x})$ and integrated around the perimeter. For so thin a plate the perimeter integrals are, in effect, line integrals, and can be approximated by narrow strips of PVDF whose width is proportional to the required weighting. For a rectangular plate $\hat{n}_i(\mathbf{x})=1$ or 0 on the perimeter, so strip-width variation is not needed, and as an approximation the strips can be laid down on the plane of the plate (like four mitered pieces of a flat picture frame). Implementation is only slightly more complicated for a circular plate since $\hat{n}_i(\mathbf{x})=x_i/R$ on the perimeter, for $i=1,2$, where R is its radius. All integrals can be accomplished by a planar array of hydrophones if the PVDF implementation is not appropriate.

2. Disk with tapered, radius-dependent weighting: $W=W(r)$

If Q is circular, with radius R, thickness ϵ, and weighting $W=W(r)$, where (r,ϕ) represent (x_1,x_2) in polar coordinates, then in the D_Q formula of Eq. (14) the x_3 integration gives a factor $\epsilon\,\mathrm{sinc}(n_3\epsilon/\lambda)$, leaving a two-dimensional Fourier transform of W (spatial frequencies f_i replaced by $-n_i/\lambda$, for $i=1,2$), which reduces[11] to a (truncated) one-dimensional Hankel transform,

$$D_Q(\mathbf{n}/\lambda)=\varepsilon\,\mathrm{sinc}\left(\frac{\varepsilon n_3}{\lambda}\right)2\pi\int_0^R rW(r)J_0\times\left(2\pi\frac{r}{\lambda}\sqrt{n_1^2+n_2^2}\right)dr.\tag{22}$$

If $W\to 0$ at the edge then the edge-strip integrals vanish, so $a_1(t)$, $a_2(t)$, and $\alpha(t)$, by Eqs. (5) and (20), are formed by integrating $p(\mathbf{x},t)$ over the plate with three weightings $\partial W/\partial x_1=(dW/dr)\cos\phi$, $\partial W/\partial x_2=(dW/dr)\sin\phi$, and $W(r)$, respectively. For example, a plate of radius 2 with weightings $W(0\leqslant r\leqslant 1)=1$ and $W(1\leqslant r\leqslant 2)=2-r$ gives $\partial W/\partial x_1=\cos\phi$ and $\partial W/\partial x_2=\sin\phi$. The directional response for $\lambda=R/2$ is shown in Fig. 5, accompanied by electrode patterns to do the weightings on three thin double-layers of PVDF, using the ground-encapsulated type of Fig. 2 to avoid capacitative short circuits. Or the weights can be applied to outputs of an array.

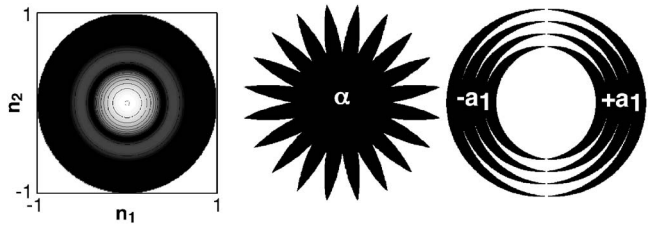


FIG. 5. Left panel: ideal directional response, $20\log_{10}|D_Q(\mathbf{n}/\lambda)|$, for the example of a circular plate with radial weighting, at nominal frequency f_0 ($=c/\lambda_0$), plotted in 3 dB contours and grayscale on the unit sphere (as seen from $+x_3$ direction) for a plate diameter of $4\lambda_0$. Middle and right panels: electrode patterns to produce $\alpha(t)$ and $a_1(t)$ using three laminated sheet sensors (the electrode pattern for a_2 is just a 90°-rotated version of that shown for a_1).

3. Rectangle with tapered, symmetric, uniaxial weighting: $W=W(|x_1|)>0$

If Q is an $L_1\times L_2\times\epsilon$ plate, $W=W(|x_1|)\geqslant 0$, and $W(\pm L_1/2)=0$, then only n_1 is measurable, and Eqs. (5) and (8) give $\alpha(t)$ and $a_1(t)$ as weighted integrals of $p(\mathbf{x},t)$,

$$\alpha(t)=c^{-1}\dot{h}(t)*\iiint_Q Wp(\mathbf{x},t)\,dV \tag{23a}$$

$$a_1(t)=h(t)*\iiint_Q \frac{dW}{dx_1}p(\mathbf{x},t)\,dV, \tag{23b}$$

to be done simultaneously over the plate. This is simple for a grid of array elements, but cost or other factors may favor PVDF, especially if both α and a_1 can be produced by a single sheet: Suppose electrode patterns E_α,E_+,E_- are designed to approximate $W(x_1)$ and the "+" and "−" portions of dW/dx_1 (we ignore scaling factors here). Then each of the possibly overlapping plane subsets E_α,E_+,E_- can be expressed as a disjunctive canonical form on the Boolean subalgebra they generate.[12] This means that each's output can be expressed as a sum of outputs of elemental, nonoverlapping subsets $\mathcal{E}_i$ lying in the same plane. Normally there would be $2^3=8$ such subsets, but since $E_-\cap E_+=\emptyset$ only four are needed:

$$\mathcal{E}_1\triangleq E_\alpha^c\cap E_-,\quad \mathcal{E}_2\triangleq E_\alpha\cap E_-,\quad \mathcal{E}_3\triangleq E_\alpha\cap E_+,\quad \times\mathcal{E}_4\triangleq E_\alpha^c\cap E_+. \tag{24}$$

The required electrodes areas are then formed on a single sheet as

$$E_\alpha=\mathcal{E}_2+\mathcal{E}_3,\quad E_-=\mathcal{E}_1+\mathcal{E}_2,\quad E_+=\mathcal{E}_3+\mathcal{E}_4, \tag{25}$$

where "+" designates set union as well as implying addition of output voltages.[13] For example, consider a linearly tapered weighting of the form $W(|x_1|)=1-|x_1|$ on a plate of length $L_1=2$, so that $dW/dx_1=-\text{sign}(x_1)$. The $\mathcal{E}_i$'s of Eqs. (24) and (25)are shown in Fig. 6(a) for a single narrow "tile" which can be repeated to fill out the entire width L_2. All of the PVDF area gets used, without multiple layers.

Alternatively, one can design a weighting W with partial-coverage electrodes E_α,E_+,E_- that use all of the PVDF with no overlap ($\mathcal{E}_i$'s are not needed):

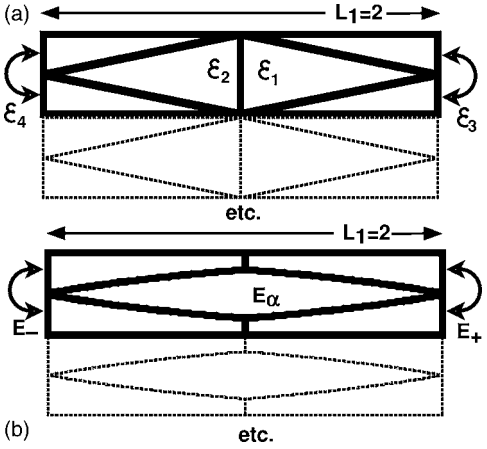


FIG. 6. Two example tile patterns for partial-coverage electrodes (the vertical scale is exagerrated) on a thin $L_1\times L_2$ rectangular transducer plate. (See Sec. III B 3.) Prototype tiles are shown in heavy black outline, and are repeated to fill out the L_2 dimension of the plate. Matching sections must be connected electrically, either externally or by small conductive bridges. A linear-taper weighting is depicted in (a), with $E_\alpha=(\mathcal{E}_1)_{\text{voltage}}+(\mathcal{E}_2)_{\text{voltage}}$, $E_{1+}=(\mathcal{E}_1)_{\text{voltage}}+(\mathcal{E}_3)_{\text{voltage}}$, and $E_{1-}=(\mathcal{E}_2)_{\text{voltage}}+(\mathcal{E}_4)_{\text{voltage}}$. The weighting of Eq. (26) for $A=0.875$ is depicted in (b), showing only E_α, E_+, and E_- (no $\mathcal{E}$'s are needed).

$$W(x_1)=1-\exp(A|x_1|-1)\quad \text{for } |x_1|\leqslant 1, \tag{26}$$

with $A>0$. Since $A^{-1}|dW/dx_1|+W=1$, a tile of unit width with a portion of its area for W and the remainder for $A^{-1}dW/dx_1$ can accomplish the required weightings (as A^{-1} is merely a scale factor). The pattern is shown in Fig. 6(b).

Fully tiled rectangular-plate electrode patterns can be formed in this way. Several PVDF apertures were fabricated at our laboratory using chemically etched and knife-scribed electrode patterns, backed by 1/4 in brass barrier plate and potted in polyurethane (see Fig. 7). We showed these examples at the IEEE Oceans '90 Conference, and our laboratory collaborated with Draper Laboratory on a related project. Similar configurations were further studied at Draper,[14] and more recently at MIT by Paradiso and his research group.[15]

To simulate a continuously varied sensitivity the tiles need to be spaced much more closely than the waves sliding across the face of the electrode, i.e., than the *trace wavelength* $\lambda_{\text{tr}}=c_{\text{tr}}/f$, where c_{tr} is the wave's apparent speed along the surface, $c_{\text{tr}}=c/\sin\psi$, and ψ is the angle of incidence from $\hat{\mathbf{n}}$, the unit normal to the plate:

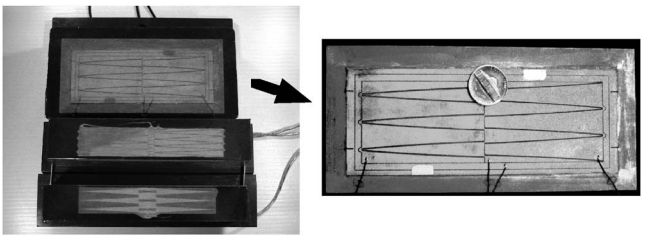


FIG. 7. Some of the $W(x_1)$-pattern PVDF plate transducers that were fabricated at our laboratory in the late 1980s. After the front electrode was patterned, the PVDF was bonded to brass plate approximately 1/4 in thick, then potted in polyurethane. The one with a 5-cent coin on it is the "no overlap" pattern with staggered spears to discourage grating lobes and ease connectivity (photographed before potting).

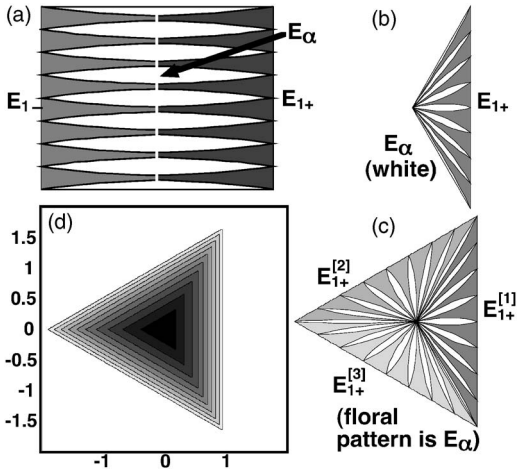


FIG. 8. Constructing a single-layer, M-sided polygon, plate electrode pattern for $\alpha(t)$, $a_1(t)$, and $a_2(t)$, shown for $M=3$. (See Sec. III B 4.) (a) A fully-tiled pattern for electrodes E_α, E_{1+}, and E_{1-}, for uniaxial weighting per Eq. (26), for $A=1$; (b) the right-half of (a) after it has been vertically morphed to form a wedge of angle $2\pi/M$; (c) the aggregation of the M rotated versions of (b), with the E_α areas combined to form a single E_α (the floral pattern) whose electrical output is used to extract $\alpha(t)$, but the rotated E_{1+} patterns, labeled as $E_{1+}^{[m]}$, are wired separately to give electrical outputs $a_1^{[m]}$ for $m=1,2,\ldots,M$, for extraction of $a_1(t)$ and $a_2(t)$ by the formula in the text following Eq. (28); and (d) a contour plot of the resulting weighting function $W_{\text{poly}}(x_1,x_2)$ that is expressed in Eq. (28).

$$\lambda_{\text{tr}} = \frac{c/f}{\sin\psi} = \frac{\lambda}{\sqrt{1-(\hat{\mathbf{n}}\cdot\mathbf{n})^2}}. \tag{27}$$

[*How many tiles*? The beamwidth is about λ/L_1 rad if $W(x_1)$ is fairly uniform, so $|\sin\psi| \leqslant |\psi| \leqslant 0.5\lambda/L_1$ for within-the-beam operation, implying $\lambda_{\text{tr}} \geqslant 2L_1$. A spacing under $\lambda_{\text{tr}}/20$ then would require 10 tiles, more if $W(x_1)$ varies rapidly, or if outer sidelobe behavior needs tighter control. Use of too many will raise interelectrode edge-capacitance and fringing E-field issues.]

4. Regular polygon on single-layer PVDF, with radially symmetric weighting

For a non-negative weighting that depends on both x_1 and x_2, five electrode areas $E_\alpha, E_{1+}, E_{1-}, E_{2+}, E_{2-}$ are needed. To put them in one layer requires, in general, 17 $\mathcal{E}_i$'s, but here is a way to make an M-sided polygon, with only $M+1$ nonoverlapping electrodes (see Fig. 8, for $M=3$): First select the *right half* of a fully tiled pattern for Eq. (26) for $A=1.6$ (for example); then shrink *its* left edge to zero height to make a triangle while scaling the right side's height so it subtends an angle of $2\pi/M$. The electrodes retain their partial-coverage densities, still correct for $\alpha(t)$ and $\mathbf{a}(t)$, for the $W(x_1)$ of Eq. (26), for $x_1 \geqslant 0$. (Perimeter integrals would also be needed for this triangle, but not when adjoined with others in the next step.) Next, replicate the triangle $M-1$ times, rotating in steps of $2\pi/M$ rad, each triangle carrying its weighting with it. This forms an M-sided polygon with weighting $W_{\text{poly}}(x_1,x_2)$ whose contours are also polygons [see Fig. 8(d)], defined via the $W(x_1)$ of Eq. (26):

$$W_{\text{poly}}(x_1,x_2) = W(\max_m [x_1\cos(2\pi m/M) + x_2\sin(2\pi m/M)]). \tag{28}$$

The inner electrode areas merge to extract the correct $\alpha(t)$ output. The M outer electrode areas must not be merged; instead, their separate outputs $a_1^{[m]}(t)$ are weighted and added: $a_1(t) = \Sigma_m \cos(2\pi m/M) a_1^{[m]}(t)$ and $a_2(t) = \Sigma_m \sin(2\pi m/M) a_1^{[m]}(t)$.

IV. SIMPLIFICATIONS AND EXTENSIONS OF THE THEORY

A. Plane and line apertures

Any thin plate, as in Sec. III B, effectively collapses to a planar aperture S if

$$W(\mathbf{x}) = W_{12}(x_1,x_2)\,\delta(x_3), \tag{29}$$

and this reduces Eqs. (5) and (8) to surface and line integrals in the x_1x_2 plane:

$$\alpha(t) = \left[\frac{1}{c}\frac{dh(t)}{dt}\right] * \left[\iint_S W_{12}(x_1,x_2)p(\mathbf{x},t)dS\right], \tag{30}$$

$$a_i(t) = h(t) * \left\{\oint_{\partial S} W_{12}(x_1,x_2)p(\mathbf{x}_{12},t)\hat{n}_i(\mathbf{x})d\ell - \iint_S \frac{\partial W_{12}(x_1,x_2)}{\partial x_i}p(\mathbf{x},t)dS\right\} \tag{31}$$

for $i=1,2$, where the cross section of Q in the x_1x_2 plane is S. Its perimeter is ∂S (ℓ is distance around), requiring a perimeter line receiver unless $W_{12}(x_1,x_2) \to 0$ there.

The aperture can be further collapsed to a *line* of length L_1 along the x_1 axis:

$$W(\mathbf{x}) = W_1(x_1)\,\delta(x_2)\,\delta(x_3), \tag{32}$$

$$\alpha(t) = \left[\frac{1}{c}\frac{dh(t)}{dt}\right] * \left[\int_{-L_1/2}^{L_1/2} W_1(x_1)p_1(x_1,t)dx_1\right], \tag{33}$$

$$a_1(t) = W_1(L_1/2)h(t) * p_1(L_1/2,t) - W_1(-L_1/2)h(t) * p_1(-L_1/2,t) - h(t) * \int_{-L_1/2}^{L_1/2} \frac{dW_1(x_1)}{dx_1}p_1(x_1,t)dx_1, \tag{34}$$

where $p_1(x_1,t)$ is $p(\mathbf{x},t)$ evaluated on the x_1 axis. The first two terms in Eq. (34) imply point-receivers at the ends, but they disappear if W_1 tapers to zero. Then Eqs. (33) and (34) describe two line receivers collocated on the x_1 axis (i.e., one line receiver with two applied weightings), whose weighting patterns are related by a derivative. In the monopulse radar literature this is the *Kerr–Murdock condition*, as noted in Sec. V D. If θ_1 is the bearing angle relative to the x_2x_3 plane then $n_1=a_1(t)/\alpha(t)=\sin\theta_1$, independent of frequency, confirming a previous result.[16]

B. Reshaping, reorienting, and/or resizing the aperture

The size, shape, and/or orientation of an aperture Q can be changed by a linear transformation: $Q_{\text{new}}=\{\mathbf{Ax}:\mathbf{x}\in Q\}$, where $\det\mathbf{A}\neq 0$. (Here vectors are regarded as 3×1 matrices of Cartesian components, with $\mathbf{A}$ as a 3×3 matrix.) If the weighting goes with the aperture then $W_{\text{new}}(\mathbf{x})=W(\mathbf{A}^{-1}\mathbf{x})$ and $\nabla W_{\text{new}}=(\mathbf{A}^{-1})^{\mathrm{T}}\nabla W(\mathbf{x})$, where $(\cdot)^{\mathrm{T}}$ denotes transpose, in Eqs. (5) and (8). Putting $W_{\text{new}}(\mathbf{x})$ into Eq. (14) gives, after a change of variables, $D_{Q\text{new}}(\mathbf{n}/\lambda)=D_Q(\mathbf{An}/\lambda)|\det\mathbf{A}|$.

C. Compound aperture

Suppose there are M apertures $Q^{[m]}$ with weightings $W^{[m]}(\mathbf{x})$, producing M outputs $\alpha^{[m]}(t)$ and $\mathbf{a}^{[m]}(t)$, produced in accordance with the integral formulas of Eqs. (5) and (8). Since $\mathbf{a}^{[m]}(t)=\mathbf{n}\alpha^{[m]}(t)$ and $\mathbf{n}$ is the same for all m (no parallax), it follows that $\Sigma_1^M\mathbf{a}^{[m]}(t)=\mathbf{n}\Sigma_1^M\alpha^{[m]}(t)$, i.e., the direction-finding nature is preserved. Moreover,

$$\mathbf{a}^{\Sigma}(t)=\alpha^{\Sigma}(t)\mathbf{n}, \tag{35}$$

where α^{Σ} and $\mathbf{a}^{\Sigma}$ are delayed-and-weighted sums:

$$\mathbf{a}^{\Sigma}(t)\triangleq\sum_1^M C^{[m]}\mathbf{a}^{[m]}(t-\tau^{[m]}), \tag{36}$$

$$\alpha^{\Sigma}(t)\triangleq\sum_1^M C^{[m]}\alpha^{[m]}(t-\tau^{[m]}). \tag{37}$$

These sums can form a beam to shield out interference. For example, suppose the $Q^{[m]}$'s represent vertical line apertures that form the staves of a vertical-axis cylindrical array with identical vertical aperture weightings $W^{[m]}(\mathbf{x})$. For beamforming, the $\tau^{[m]}$'s can be chosen to synchronize the stave outputs for a plane wave incident from the "look" azimuth, with tapered weighting coefficients $C^{[m]}$'s chosen to limit the horizontal sidelobes of the resulting beam pattern. [Optionally, the temporal filters $h(t)$ and $dh(t)/dt$ of Eqs. (5) and (8) can be applied *after* beamforming.] If the goal is to measure the elevation of sonar echoes while shielding out interference from other azimuths, then the vertical direction cosine can be calculated as $n_3=a_3^{\Sigma}(t)/\alpha^{\Sigma}(t)$.

D. Steering the entire aperture

In forming $\alpha(t)$ and $\mathbf{a}(t)$ one can apply a position-dependent delay $\tau=c^{-1}\check{\mathbf{n}}\cdot\mathbf{x}$ to the sensed wave field, to "steer" the aperture Q in the direction of a chosen unit vector $\check{\mathbf{n}}$, so that a plane wave from the $\check{\mathbf{n}}$ direction seems to have arrived simultaneously at every point in Q, maximizing sensitivity in that direction. This makes the aperture act like a plate that can reorient merely by changing delays.[17] In effect, it substitutes

$$\check{p}(\mathbf{x},t)\triangleq p(\mathbf{x},t-c^{-1}\check{\mathbf{n}}\cdot\mathbf{x}) \tag{38}$$

for the wave field $p(\mathbf{x},t)$ in the space–time filtered gradient integrals of Eqs. (5) and (8). From the plane wave assumption $p(\mathbf{x},t-c^{-1}\check{\mathbf{n}}\cdot\mathbf{x})=p(0,t+c^{-1}[\mathbf{n}-\check{\mathbf{n}}]\cdot\mathbf{x})$, which is *exactly* the same as if the wave arrived at speed $c/|\mathbf{n}-\check{\mathbf{n}}|$ from the direction of the unit vector $(\mathbf{n}-\check{\mathbf{n}})/|\mathbf{n}-\check{\mathbf{n}}|$. But sound speed affects our direction finder only in the scaling $\alpha(t)$ in Eq. (5). Thus the direction indicator formula of Eq. (9) is modified to

$$\mathbf{a}(t)=|\mathbf{n}-\check{\mathbf{n}}|\alpha(t)(\mathbf{n}-\check{\mathbf{n}})/|\mathbf{n}-\check{\mathbf{n}}|=\alpha(t)(\mathbf{n}-\check{\mathbf{n}}), \tag{39}$$

and the directional response becomes $|\mathbf{n}-\check{\mathbf{n}}|^{-1}D_Q([\mathbf{n}-\check{\mathbf{n}}]/\lambda)$. All of our direction finding methods still apply, with $(\mathbf{n}-\check{\mathbf{n}})$ substituted for $\mathbf{n}$, except that $(\mathbf{n}-\check{\mathbf{n}})$ is not a unit vector. That presents little difficulty since $\check{\mathbf{n}}$ was selected *a priori*. For example, the decoupled formula of Eq. (10) becomes $n_i=\check{n}_i+a_i(t)/\alpha(t)$. In particular, if $\check{\mathbf{n}}$ is aimed along the $+x_3$ axis, then the formulas for n_1 and n_2 are unaltered. Another strategy, if $\alpha(t)$ is available, is to convert $\mathbf{a}(t)$ to a corrected version $\check{\mathbf{a}}(t)\triangleq\mathbf{a}(t)+\alpha(t)\check{\mathbf{n}}$, so that Eq. (39) assumes the form of the unsteered case:

$$\check{\mathbf{a}}(t)=\alpha(t)\mathbf{n}. \tag{40}$$

To apply continuously spatially varying delays to steer Q may be infeasible, but discrete delays are adequate for a discrete array of omnidirectional sensors.

E. Higher-order gradients (treating only one dimension, for simplicity)

Suppose Q is an extrusion of length L along the x_1 axis, with $W=W(x_1)$, where W and its first $M-1$ derivatives go to zero at the ends, $x_1=\pm L_1/2$. Then Eq. (23a) and (23b) applies, even if the extrusion is not a plate. But in Eq. (23a) and (23b) one could have put the mth-derivative $W^{(m)}$ in place of W (thereby requiring $W^{(m+1)}$ in place of W'), and put the mth-derivative $h^{(m)}$ in place of h. Thus, if one defines

$$s_m(t)\triangleq -(-c)^{m-M}h^{(M-m)}\times(t)*\left[\iiint_Q W^{(m)}(x_1)p(\mathbf{x},t)dV\right] \tag{41}$$

for $m=0,1,2,\ldots,M$, then consecutive $s_n(t)$'s behave like $\alpha(t)$ and $a_1(t)$, in the sense that $s_{m+1}(t)=n_1s_m(t)$. If these $s_m(t)$'s are produced as outputs from Q, they can be regarded as space–time filtered, higher-order gradients. As in the theory of higher-order gradient receivers without filtering,[18] one can use an order-reducing transformation to put broadband nulls into the directional response at selected angles, and/or discern multiple source directions. For example, J *broadband nulls* are introduced if the set $\{s_m(t):m=0,\ldots,M\}$ is transformed into a smaller set $\{\bar{s}_m(t):m=0,\ldots,M-J\}$ by

$$\begin{bmatrix} \bar{s}_0(t) \\ \bar{s}_1(t) \\ \bar{s}_2(t) \\ \dots \\ \bar{s}_{M-J}(t) \end{bmatrix} = \begin{bmatrix} B_0 & B_1 & B_2 & \dots & B_J & 0 & 0 & \dots & 0 \\ 0 & B_0 & B_1 & B_2 & \dots & B_J & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & B_0 & B_1 & B_2 & \dots & B_J \end{bmatrix} \begin{bmatrix} s_0(t) \\ s_1(t) \\ s_2(t) \\ \dots \\ s_M(t) \end{bmatrix} = \begin{bmatrix} 1 \\ n_1 \\ n_1^2 \\ \dots \\ n_1^{M-J} \end{bmatrix} B(n_1)s_0(t); \tag{42}$$

the latter equality applies to a single wave from arbitrary direction cosine n_1, with

$$B(n_1) = B_0 + B_1 n_1 + B_2 n_1^2 + \dots + B_J n_1^J = B_J \prod_{i=1}^{J} (n_1 - \beta_i), \tag{43}$$

i.e., $B(n_1)$ is a frequency-independent polynomial whose roots are the direction cosines β_i (measured from the $+x_1$ axis) of the intended null angles, producing frequency-independent notches in the beam pattern to shield out broadband interferences. Except for sharing these nulls, the new $\bar{s}_m(t)$'s behave like the original $s_m(t)$'s; i.e., $\bar{s}_{m+1}(t) = n_1 \bar{s}_m(t)$. For example, $\bar{s}_0(t)$ and $\bar{s}_1(t)$ behave like $\alpha(t)$ and $a_1(t)$.

On the other hand, if instead of a single source there are $I(\leqslant M)$ wave sources with distinct direction cosines $n_{1[1]}, n_{1[2]}, \dots, n_{1[I]}$, and if one forms an $(M+1) \times K$ data matrix from the measured $s_m(t)$'s, sampled at times $t_1, t_2, \dots, t_K$, it must obey

$$\begin{bmatrix} s_0(t_1) & s_0(t_2) & \dots & s_0(t_K) \\ s_1(t_1) & s_1(t_2) & \dots & s_1(t_K) \\ s_2(t_1) & s_2(t_2) & \dots & s_2(t_K) \\ \dots & \dots & \dots & \dots \\ s_M(t_1) & s_M(t_2) & \dots & s_M(t_K) \end{bmatrix} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ n_{1[1]} & n_{1[2]} & \dots & n_{1[I]} \\ n_{1[1]}^2 & n_{1[2]}^2 & \dots & n_{1[I]}^2 \\ \dots & \dots & \dots & \dots \\ n_{1[1]}^M & n_{1[2]}^M & \dots & n_{1[I]}^M \end{bmatrix} \begin{bmatrix} s_{0[1]}(t_1) & s_{0[1]}(t_2) & \dots & s_{0[1]}(t_K) \\ s_{0[2]}(t_1) & s_{0[2]}(t_2) & \dots & s_{0[2]}(t_K) \\ s_{0[3]}(t_1) & s_{0[3]}(t_2) & \dots & s_{0[3]}(t_K) \\ \dots & \dots & \dots & \dots \\ s_{0[I]}(t_1) & s_{0[I]}(t_2) & \dots & s_{0[I]}(t_K) \end{bmatrix}, \tag{44}$$

where $s_{0[i]}(t)$ denotes the contribution of the ith source to $s_0(t)$. The $(M+1) \times I$ matrix of progressive powers of the $n_{1[i]}$'s, call it **N**, is a Vandermonde-type matrix of rank I. The data matrix (on the left) thus cannot have a rank exceeding I, and if it is wide enough to have rank I then its column-space coincides with that of **N**, uniquely determining the set of $n_{1[i]}$'s, i.e., the sources' direction cosines with respect to the x_1 axis. Decomposition algorithms used in single-frequency, signal-subspace, direction finding[19,20] can extract the Vandermonde factor **N**. (A MATLAB algorithm is provided in Appendix A.)

Higher-order gradients offer a frequency-independent alternative to wave-number space,[21] but problems arise for M above 3 or 4, reminiscent of superdirectivity failure. Indeed, $J=M$ in Eq. (42) gives $\bar{s}_0(t) = B(n_1)s_0(t)$ as output, where $B(n_1)$ is an *arbitrary* Mth-order polynomial superdirective beam shaping factor that is independent of frequency and aperture! Sadly, $W^{(m)}(x_1)$ is not accurately realizable for large m.

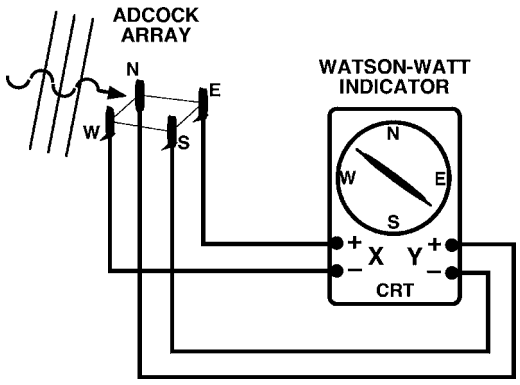


FIG. 9. Adcock's square, horizontal array (with edge-dimension $\ll\lambda$) that measures horizontal components of wave gradient vector, shown with a simple Watson–Watt indicator (a CRT) that displays azimuth instantaneously.

F. Discrete line arrays (which may be parts of a compound aperture)

Suppose Q is a line array of L points, uniformly spaced at interval d_1 along the x_1 axis, at $x_1 = \ell d_1$ for $\ell = 0, \dots, L-1$. The discrete analog of Eq. (41) is

$$s_m(t) \triangleq -(-c)^{m-M} h^{(M-m)}(t) * \sum_{\ell=0}^{L-1} \mathcal{W}_\ell^{(m)} [p(\mathbf{x},t)]_{\substack{x_1=\ell d_1 \\ x_2=x_3=0}} \tag{45}$$

for $m=0,1,2,\ldots,M$, where $\mathcal{W}_0^{(m)},\ldots,\mathcal{W}_{L-1}^{(m)}$ are the array (aperture) weights for $s_m(t)$ at the L point-sensors. Since in Eq. (41) $W^{(m)}(x_1)=d[W^{(m-1)}(x_1)]/dx_1$, it is only a modest leap of inspiration to require that the discrete weights obey

$$\begin{aligned}&\text{conv}([\mathcal{W}_0^{(m)},\ldots,\mathcal{W}_{L-1}^{(m)}],[1,1])/2 \\ &\quad = \text{conv}([\mathcal{W}_0^{(m-1)},\ldots,\mathcal{W}_{L-1}^{(m-1)}],[1,-1])/d_1,\end{aligned} \tag{46}$$

i.e., to specify that the following relation be satisfied without remainder:

$$\begin{aligned}[\mathcal{W}_0^{(m)},\ldots,\mathcal{W}_{L-1}^{(m)}] = (2/d_1)\text{deconv} \\ \times(\text{conv}([\mathcal{W}_0^{(m-1)},\ldots,\mathcal{W}_{L-1}^{(m-1)}],[1, \\ -1]),[1,1]),\end{aligned} \tag{47}$$

where conv(·,·) and deconv(·,·) denote discrete convolution and deconvolution (as in MATLAB). Indeed, this condition is enough to ensure (see proof in Appendix B) that consecutive $s_m(t)$'s satisfy $s_{m+1}(t)=\tilde{n}_1 s_m(t)$ with an *inflected* direction cosine

$$\tilde{n}_1 \triangleq \kappa^{-1}\tan(\kappa n_1) = n_1 + \kappa^2 n_1^3/3 + 2\kappa^4 n_1^5/15 + \cdots, \tag{48}$$

where $\kappa=\pi d_1/\lambda=\pi f d_1/c$. If $d_1^2 \ll \lambda^2/(3\pi^2)$ then $\tilde{n}_1 \approx n_1 = \sin\theta_1$ regardless of direction. Otherwise, if the bearing is small enough that $|\sin\theta_1|^2 \ll \lambda^2/(3\pi^2 d_1^2)$ then $\tilde{n}_1 \approx n_1$. At other angles $\tilde{n}_1$ departs from n_1 at high frequencies.[22] If the array is steered to $\breve{n}_1$ as in Sec. IV D, then $(n_1-\breve{n}_1)$ is substituted for n_1 in Eq. (48), so the zero-inflection direction shifts from $n_1=0$ to the steered direction cosine, $n_1=\breve{n}_1$.

Excellent $\mathcal{W}_\ell^{(m)}$'s are obtained by designing a "base" sequence of $L-M$ weights in a customary way to minimize sidelobes, then convolving with the two-element sequence [1,1] repeatedly, M times, to get $[\mathcal{W}_0^{(0)},\ldots,\mathcal{W}_{L-1}^{(0)}]$, and applying Eq. (47) recursively. An equivalent, if obfuscated, form of Eqs. (46)–(48) was published in 1985.[23]

V. RELATION TO OTHER DIRECTION FINDING METHODS

The space–time filtered gradient method is best understood in the context of other direction finding techniques that tolerate brief or instantaneous observations.

A. The Adcock array and the Watson–Watt indicator

Figure 9 shows the square, horizontal, four-element array patented by Adcock[24,25] in 1919 for determining the azimuth of a terrestrial radio wave. It directly senses the two horizontal components of the incident wave's gradient vector by differencing diagonally opposite elements, usually spaced at well under $\lambda/2$. These differences operate like our $a_1(t)$ and $a_2(t)$ of Sec. II C, in the sense that $\arctan(Y/X)=\arctan(n_2/n_1)=$azimuth. The Adcock array's small size makes it omnidirectional and, therefore, susceptible to interference. A Watson–Watt[26,27] indicator is shown in the figure, i.e., a cathode ray tube (CRT) with the difference-signals applied to the vertical and horizontal plates, so azimuth is indicated instantly by the tilt of the displayed "blip." The 180° azimuth ambiguity can be resolved by summing the four array outputs to the cathode through a 90°-lead filter [much like our $\alpha(t)$], to brighten the correct end of the trace.[28] An acoustic adaptation, called DIFAR,[29] is often used in sonobuoys. 3D versions of the Adcock array have also been reported.[30]

B. Bearing deviation indicator and sector scan indicator

The *bearing deviation indicator* (BDI) was a narrowband echoranging sonar of modest aperture used in World War II to chase submarines.[31] The first model compared output amplitudes of two slightly skewed beams, but advanced designs embodied a rudimentary form of gradient measurement. The aperture was divided into left and right halves with centers spaced by d_p along the x_1 axis; their output phasors A and B had a measured phase difference $\angle(AB^*)=\Phi=(2\pi d_p/\lambda)n_1$. A *sum-and-difference* scheme was used, best understood through the identity

$$\angle(AB^*) = 2\,\text{Re}\left[\arctan\left(\frac{A-B}{j(A+B)}\right)\right]. \tag{49}$$

If $\Sigma \triangleq A+B$ and $\Delta \triangleq (A-B)/j$, and division by j denotes a 90° phase-lag, then

$$\begin{aligned}n_1 = \frac{\lambda}{2\pi d_p}\Phi = \frac{\lambda}{2\pi d_p}\angle(AB^*) = \frac{\lambda}{\pi d_p}\,\text{Re}\left[\arctan\left(\frac{\Delta}{\Sigma}\right)\right] \\ \approx \frac{\lambda}{\pi d_p}\,\text{Re}\left(\frac{\Delta}{\Sigma}\right),\end{aligned} \tag{50}$$

where the final step only applies if $n_1 d_p \ll \lambda/2$. The ratio Δ/Σ has the same role as a_1/α in our Eq. (10), namely, $n_1=a_1(t)/\alpha(t)$. However, directivity can be improved only by making $d_p \gg \lambda/2$, so that Δ becomes the output of a widely spaced interferometer. That exacerbates the frequency dependence, nonlinearity, and multivaluedness lurking in Eq. (50), when used to calculate n_1 from $\angle(AB^*)$ or Δ/Σ.

The BDI was manually aimed for a strong output, hoping to put the target in the main lobe so that $-\pi<\Phi<\pi$, making the principal value of arctan(·) correct. Horton[32] described a BDI of 5λ aperture, with Δ and Σ applied to the deflection plates of a Watson–Watt CRT indicator marked with slopes for specific relative bearings $\theta_1=\arcsin n_1$ calibrated per Eq. (50), using the principal value of arctan(·).

The *sector scan indicator* (SSI), developed later in the war, measured split-aperture phase shift $\angle(AB^*)$ directly and indicated it on a raster-scanned CRT plotting bearing versus range. In its 90°-canted configuration it was touted as a seafloor profile plotter,[33] anticipating bathymetric sidescan sonars by 40 years.

The acronyms BDI and SSI are rarely seen today, but split-aperture methods based implicitly upon Eq. (50) abound, despite its narrowband assumption, even for broadband transient sounds, echoes of frequency modulated trans-

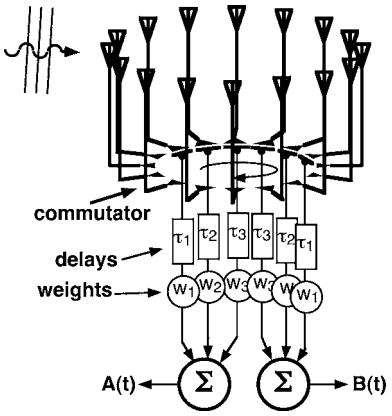


FIG. 10. Wullenweber array consisting of a ring of antennas and a continuous-scanning commutator beamformer with a selectable null beam for fine azimuth determination. Later sonar versions have used multiple fixed beams, sometimes measuring fine azimuth by phase-comparison between left/right aperture halves.

missions, random processes such as seafloor reverberation, and other large-bandwidth waves. Although ideally one should set $\lambda = c/f_t$ in Eq. (50), with f_t being the *instantaneous frequency* of the arriving wave,[34] it is customary to use the center frequency f_0, thereby making the calculated direction cosine n_1 erroneous by an erratically fluctuating factor f_t/f_0, even without noise or interference! The usual accommodation for this fluctuation is to use a smoothed phase shift $\angle\langle AB^*\rangle_T$, where $\langle\cdot\rangle_T$ denotes time average over interval T. But T needs to be relatively long and the energy spectrum of the wave must be well-balanced around f_0.[35] Regrettably, this smoothing degrades the range-resolving performance in echoranging systems, and compromises accuracy in sensing directions of extremely brief pulses in passive applications.

C. Wullenweber arrays

A *Wullenweber array*,[36] as used in Germany during World War II, is a ring of identical, fixed, radio-intercept antenna masts connected to a beamformer that uses delay lines to correct for the wave's nonsimultaneous arrival at the masts.[37] The masts feed the beamformer through a 120°-arc commutator that sweeps the beam in azimuth (see Fig. 10). Capacitive rotor/stator coupling lets it sweep continuously rather than in jumps,[38] with the detected amplitude presented on a polar CRT display. To give a sharp central null for precise aiming, it can use the amplitude of the *split-aperture difference-beam* (the electrical difference between the aperture halves A and B).[27]

This scheme had already been patented for sonar[39] and implemented on the German submarine U570 (captured in 1941 and renamed HMS Graph).[40] In the postwar era commutator beamformers for scanning sonars eventually evolved to simultaneous digital beamforming at many fixed azimuths, sometimes using phase comparison of the half-aperture outputs A and B to provide interpolation between beams (reliable only for prominent, discrete-angle targets). Indeed, $\angle(AB^*)$ is a better-behaved function of azimuth for a circular-arc aperture than for a straight-line aperture: Its inverse function is single-valued over a larger sector, mitigating angle ambiguities.[41]

D. Monopulse radar

Monopulse radar, the obvious electromagnetic-wave analog of BDI, was studied by wartime laboratory groups, some of which had been associated with BDI or SSI. Its development eclipsed that of its sonar kin in postwar years, leading to hundreds of technical papers and reports, numerous chapters in reference books, and a few entire books.[42,43] In due course it was shown that the phase-comparison, amplitude-comparison, and sum-and-difference schemes were mathematically equivalent, and could be transformed into each other at the signal processing stage.[44] The now-standard "Δ/Σ" (or "D/S") notation emerged as the canonical form, and it was realized that Σ and Δ need not actually be formed from the sum and difference of a pair of half-aperture phasors A and B as in Eq. (50). Briefly stated, Σ can be produced by applying even-symmetric weighting along the aperture, and Δ by odd-symmetric weighting with a 90° phase lag.[45] For a complex-sinusoidal wave the *monopulse ratio*, Δ/Σ, is a real-valued, odd-symmetric, wavelength-dependent function of n_1, denoted as $M_\lambda(n_1)$ for our purposes. After the function $M_\lambda(\cdot)$ is computed or measured during calibration tests, the direction of an incident wave is given by $n_1 = M_\lambda^{-1}[\Delta/\Sigma]$, where $M_\lambda^{-1}[\cdot]$ is the inverse function. For example, a split-aperture governed by Eq. (50) gives $M_\lambda(n_1) = \tan(\pi d_p n_1/\lambda)$. Its inverse, $n_1 = (\lambda/\pi d_p)\arctan(\Delta/\Sigma)$, is multivalued if $d_p > \lambda$, but the principal value is valid near boresight.

For a line aperture, Kerr and Murdock[46] found the monopulse ratio to be linear in n_1, i.e., $M_\lambda(n_1) = K_\lambda n_1$, only if the weighting functions used to produce Σ and Δ are related by a derivative. This "Kerr–Murdock condition" was proposed to simplify the feedback characteristics of a servo-controlled tracking radar,[47] but it also makes the inverse function single-valued.[45] $n_1 = M_\lambda^{-1}[\Delta/\Sigma] = (\Delta/\Sigma)K_\lambda^{-1}$.

In the sonar context it was later realized that compensatory filtering could make $K_\lambda = 1$ identically, so that $n_1 = \Delta/\Sigma$, independent of frequency, removing narrow-band restrictions.[48] This result is expressed in Eqs. (33) and (34) with α and a_1 in the roles of Σ and Δ, so that $n_1 = a_1(t)/\alpha(t) = \Delta/\Sigma$. This concept led to the space–time filtered gradient method, termed "wideband monopulse sonar" during its evolution.

VI. RATIO PROCESSING

A. Considerations arising from the Watson–Watt concept

The simple Watson–Watt indicator of Fig. 9 shows the source's azimuth if its deflection plates are driven by $a_1(t)$ and $a_2(t)$ as defined in Sec. II C. But if $\alpha(t)$ and $a_i(t)$ are used, i.e., $X = \alpha(t)$ and $Y = a_i(t) = n_i\alpha(t)$, then the trace's tilt indicates the direction cosine rather than the angle itself, in the sense that a perfectly straight line of slope n_i will be displayed when noise is absent [see Fig. 11(a)].

If transient waves come from *two* directions n_i and $\tilde{n}_i$ their contributions to X and Y just add. The result looks like Fig. 11(b) if they arrive at different times. But simultaneous sinusoid arrivals create a Lissajous-type pattern bounded by a parallelogram whose sides parallel the individual traces, as

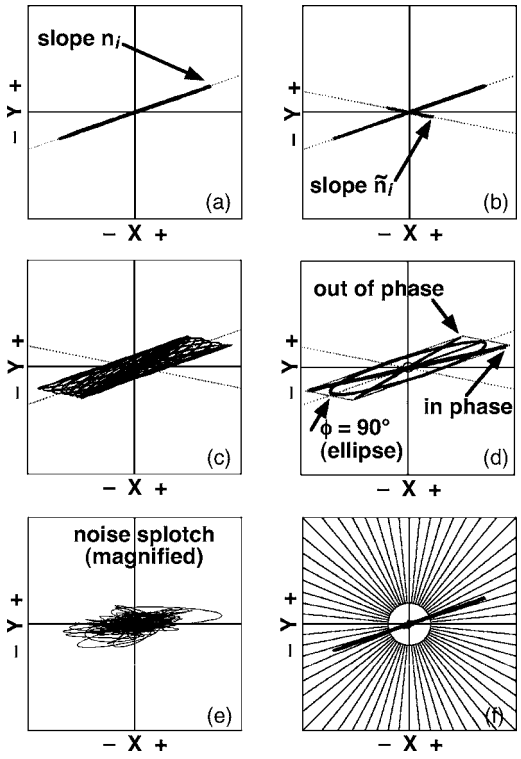


FIG. 11. Important characteristics of the Watson–Watt indicator, plotting the locus of $a_i(t)$ vs $\alpha(t)$ (for example). (a) A single incident wave. (b) Two nonsimultaneous waves from two directions. (c) Two waves as before, but now simultaneous and sinusoidal, with different frequencies. (d) Two sinusoidal waves of the same frequency, in phase, out of phase (by 180°), and with 90° phase (all shown superimposed). (e) An enlarged view of the noise "splotch." (f) Angle bins for slope detection, with a threshold circle of radius ν to squelch unreliable data

shown in Fig. 11(c) for arrivals at slightly different frequencies. If the waves have the same frequency and constant amplitude but wandering phase, the displayed ellipse undulates within the bounding parallelogram. Figure 11(d) shows that the ellipse is fat for a 90° phase difference, but collapses to a diagonal if they are in phase, like a solitary wave from somewhere between the actual sources. If they are exactly out of phase (by 180°) the trace collapses to the other diagonal, mimicking a wave from a bogus direction not bounded by the two real directions! This anomaly is not a quirk of the Watson–Watt indicator—it is a consequence of destructive interference misleadingly termed "glint" in the radar literature, for which Howard[49,50] gave this explanation (paraphrased): The radiation pattern of two (or more) coherent but spatially separated sources is characterized by lobes and nulls. The abrupt phase jump across each null makes a kink in the emanating wave front along the null direction, where the waves destructively interfere. Near this kink the gradient, always normal to the wave front, can point to a grossly deviant direction rather than toward the sources. *Any* small-aperture method, whether Adcock, monopulse, split-beam, or space-time filtered gradient method, will indicate a spurious angle if it lies in the path of this wave front kink (unless it employs enough directivity to shield out the interference, thereby removing the kink).

A keen observer of the wobbly ellipse on a Watson–Watt display might discern the actual source directions—from the angles of the sides of its bounding parallelogram. In fact, for radio direction finding Bailey and McClurg proposed that idea to discern N sources that produce a $2N$-sided polygon pattern.[38,51,52] This observation applies to any Watson–Watt display, whichever pair it employs of the space–time filtered gradient outputs $\{\alpha(t), a_1(t), a_2(t), a_3(t)\}$.

Broadband spectra soften the polygon borders (and smooth out the wave front kinks). We have observed that when no prominent directional wave is present the *noise splotch* [see Fig. 11(e)] often displays an elliptical elongation, sometimes tilted obliquely, depending upon noise directionality. In such cases the display coordinates can be rotated and scaled by a 2×2 matrix $\mathbf{\Gamma}$ to "circularize" the noise splotch,

$$\begin{bmatrix} X \\ Y \end{bmatrix} = \mathbf{\Gamma} \begin{bmatrix} \alpha(t) \\ a_i(t) \end{bmatrix}. \tag{51}$$

For example, if the noise contributions to α and a_i have a measured 2×2 covariance matrix $\mathbf{A}_i$, then $\mathbf{A}_i^{-1/2}$ can be used as $\mathbf{\Gamma}$, so that $\mathbf{\Gamma A}_i\mathbf{\Gamma}^{\mathrm{T}}$, the covariance matrix of the coordinates X and Y, becomes the identity matrix. Whatever its rationale, a coordinate transformation matrix $\mathbf{\Gamma}$ changes the slope of the displayed trace from n_i to $(\Gamma_{21}+\Gamma_{22}n_i)/(\Gamma_{11}+\Gamma_{12}n_i)$, so the direction cosine of the source should be computed from the measured slope μ in the XY display plane as

$$n_i = (\Gamma_{11}\mu - \Gamma_{21})/(\Gamma_{22} - \Gamma_{12}\mu). \tag{52}$$

Even with a normalizing transformation, a large bandwidth can prevent the noise splotch from appearing truly rotationally symmetric. The directional beam is narrower at high frequencies, so the only high frequency components that survive its shielding effect are those well-aligned to the main lobe direction. The result is a subtle grain direction in the noise splotch. A formal interpretation is that the noise power spectra in $\alpha(t)$ and $a_i(t)$ are different. If the main lobe is orthogonal to the x_i axis, for example, then $\alpha(t)$ displays stronger high frequency content than $a_i(t)$.

In a Watson–Watt display of $\sqrt{3}a_i$ vs α, a spherically isotropic broadband noise field (each n_i being uniformly distributed between ±1 with $\mathrm{E}[\mathrm{n}_\mathrm{i}^2]=1/3$) *will* produce a rotationally symmetric noise splotch if the aperture has no directivity. But if, for example, it has its directivity concentrated orthogonal to the x_i axis, then the normalizing factor on a_i has to be larger than $\sqrt{3}$, and a grain parallel to the α axis can become visible for the reasons just given.

B. Angle bins: Waterfall display

Even when not implemented directly, the Watson–Watt indicator inspires various direction finding algorithms and graphical aids. For example, one can specify K slope boundaries $-\infty \leqslant \cdots < \mu_k < \mu_{k+1} < \cdots \leqslant \infty$ and construct a series of detectors to test for slopes lying between them, with detector outputs

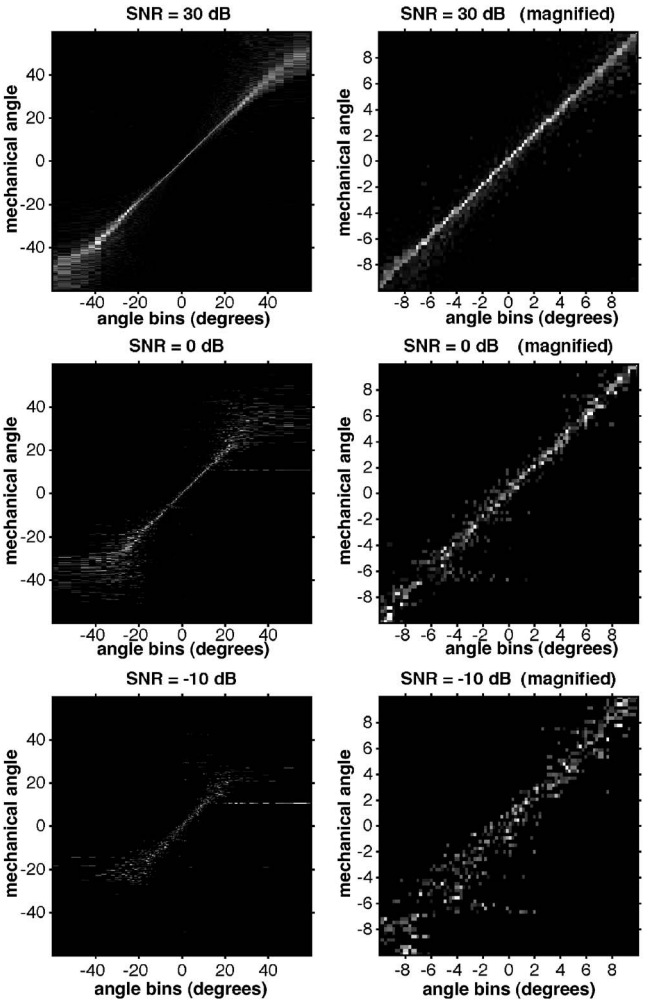


FIG. 12. Waterfall display of acoustic data recorded at the Lake Travis Test Station, for SNRs of 30, 0, and −10 dB, using a source of 3-octave bandwidth. The history of angle-bin occupancy is shown as the source moved from left to right, with each display scrolling downward. The known source angle (mechanically measured) is labeled on the vertical time axis.

$$\zeta_k(t) \triangleq \begin{cases} \sqrt{X^2+Y^2} & \text{if } \mu_k < Y/X \leqslant \mu_{k+1}, \quad \sqrt{X^2+Y^2} > \nu, \\ 0 & \text{otherwise}, \end{cases} \tag{53}$$

wherein the threshold ν is set high enough to reject poor data or limit false alarms. [For complex data $\mathrm{Re}(Y/X)$ replaces Y/X; see Eqs. (56) and (60) *ff.*] Figure 11(f) shows a set of slope boundaries $\{\mu_k = \tan[\pi(k/K - 1/2)] : k = 0, \ldots, K\}$ for $K = 32$. A nonzero value of $\zeta_k(t)$ offers evidence of an incident wave whose direction cosine lies between the n_i values computed from μ_k and μ_{k+1} via Eq. (52), assuming X and Y are as defined in Eq. (51).

The detector outputs $\zeta_k(t)$ can be presented in a "waterfall" display, as illustrated in Fig. 12. (Imagine it scrolling downward, the newest data at the top as an intensity-encoded histogram of bearing.) These data were recorded at our Lake Travis Test Station, for an underwater acoustic source whose bearing angle $\theta_i = \arcsin n_i$ was moved slowly from −85° to +85° as it repeatedly transmitted linear-frequency-modulated "pings" of about three octaves bandwidth—each with a duration × bandwidth product of 200 and a geometric-mean wavelength of $\lambda_0 = \sqrt{\lambda_{\text{lower}} \times \lambda_{\text{upper}}} \approx 12$ cm. Not surprisingly, the angle indications got more reliable as the received signal power level was artificially increased relative to the ambient lake noise. [Signal-to-noise ratio (SNR) was calculated for an in-water point receiver, band-limited to the signal bandwidth, prior to replica-correlation filtering.]

More details behind Fig. 12: The acoustic aperture was a $2.5\lambda_0 \times 2.5\lambda_0$ surface of hard alumina, subdivided into contiguous cells and backed by piezoelectric transducers, with a polyurethane facing. The linearly tapered weighting of Sec. III B 3 was applied electronically, but in a ten-step staircase approximation due to the cell configuration. The measured half-power beamwidth was 19° with no sidelobes (thanks to the three-octave bandwidth). The time-domain filters (see Sec. II B) included identical signal-conditioning filters in the form of replica correlators, acting to deconvolve and compress the duration-bandwidth products to unity at the angle detector inputs. The noise splotch was circularized, complex data were extracted with $\mathrm{Re}(Y/X)$ used in Eq. (53), the μ_k slopes were set to give 1° spacing in the normalized XY plane, and the time samples of $\zeta_k(t)$ used for the display were integrate-and-dump samples with a capture period of 80× the compressed ping duration; however, the waterfalls of Fig. 12 include only the time-bins that actually captured pings, the intervening intervals being almost devoid of false-alarm events because the ν threshold was set at four times the root-mean-square radius of the noise splotch. Because of the aperture's discrete cell structure, the "inflected direction cosine" formula of Eq. (48) was applied, with κ calculated at the center of the frequency band.

C. Solid-angle bins: $\mathbf{a}(t)$ as shielded acoustic particle motion

Instead of operating in an XY plane one could use angle bins in 3D to analyze $\mathbf{a}(t)$. The kth bin's boundaries would prescribe inequality tests of a_i vs α, setting $\zeta_k(t) = |\mathbf{a}(t)| = |\alpha(t)|$ if the bin were occupied, zero otherwise. By Eq. (7), with $W(-\mathbf{x})$ as specified following Eq. (4), $\mathbf{a}(t)$ is the gradient (at the origin) of the filtered wave field $p_{\text{sh}} = h(t) * W(-\mathbf{x}) \star p$, where $\star$ denotes 3D spatial convolution. If p is acoustic pressure then for linear acoustics $-\nabla p = \rho_0 \ddot{\xi}$ where ξ is particle displacement and ρ_0 is the medium's density. Derivatives and convolutions commute, so

$$\begin{aligned} \mathbf{a}(t) &= (\nabla p_{\text{sh}})_{\mathbf{x}=0} = (\nabla [h(t) * W(-\mathbf{x}) \star p])_{\mathbf{x}=0} \\ &= [h(t) * W(-\mathbf{x}) \star \nabla p]_{\mathbf{x}=0} \end{aligned} \tag{54}$$

$$\begin{aligned} &= [h(t) * W(-\mathbf{x}) \star (-\rho_0 \ddot{\xi})]_{\mathbf{x}=0} \\ &= -\rho_0 \ddot{h}(t) * [W(-\mathbf{x}) \star \xi]_{\mathbf{x}=0}. \end{aligned} \tag{55}$$

If $h(t)$ is chosen so $-\rho_0 \ddot{h}(t) = \delta(t)$, then $\mathbf{a}(t)$ is the value, at the origin, of $W(-\mathbf{x}) \star \xi$, i.e., the acoustic particle displacement *as shielded* by the spatial filtering. The $\mathbf{a}(t)$ vector tracks this motion and reveals directions of incident waves.

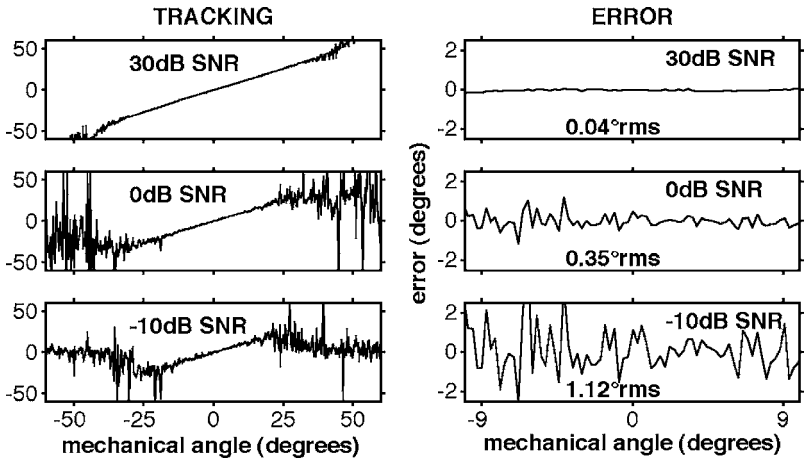


FIG. 13. Performance of the real-part-of-complex-ratio estimate for the Lake Travis acoustic data. The right-hand plots show the tracking errors within the half-power beamwidth of the receiver's beam pattern.

D. Conditioning of denominator when complex data are available

The formula $n_i = a_i(t)/\alpha(t)$ of Eq. (10) is poorly conditioned due to frequent zero-crossings of its denominator. But if complex-valued $\alpha(t)$ and $a_i(t)$ are used (imaginary parts reconstructed by Hilbert transforms, quadrature demodulation, etc.) then

$$n_i = \mathrm{Re}\left[\frac{a_i(t)}{\alpha(t)}\right] = \frac{\mathrm{Re}[a_i(t)\alpha^*(t)]}{|\alpha(t)|^2}, \tag{56}$$

which is well behaved as long as the modulus of $\alpha(t)$ is nonzero; indeed, any slight phase shift φ between the α and a_i electrical paths now just induces a mildly distortive factor of $\cos\varphi \approx 1$, instead of wild gyrations. For the Lake Travis acoustic data, Fig. 13 plots the bearing estimate $\hat{\theta}_1 = \arcsin \hat{n}_1$, with $\hat{n}_1$ calculated using Eq. (56) at the peaks of the pulse-compressed pings. For 30 dB SNR, the tracking error had a root-mean-square (rms) value of 0.04° (0.70 mrad) when the source was within the 19° half-power beamwidth, or almost three orders of magnitude smaller than λ/d_Q ($\approx$0.4 rad at the geometric-mean frequency).

The $\mathrm{Re}[\cdot]$ operation is superfluous in the ideal case, but necessary if noise or interference is present, in which case a more robust formula is

$$n_i = \frac{\langle \mathrm{Re}[a_i(t)\alpha^*(t)]\rangle_T}{\langle |\alpha(t)|^2\rangle_T}, \tag{57}$$

where $\langle\cdot\rangle_T$ denotes a sliding-window average of small duration T,

$$\langle x(t)\rangle_T \triangleq \frac{1}{T}\int_{-T/2}^{T/2} x(t+\tau)\,d\tau. \tag{58}$$

Formulas like Eq. (57) have been exploited in radars to calculate "monopulse ratio," with Δ and Σ in place of $a_i(t)$ and $\alpha(t)$, but they average over an ensemble of repeated radar transmissions instead of by a sliding window. (Ensemble averaging over repeated sonar "pings" is feasible, but usually ill-advised due to target or sensor movement during the long echo interval for the 200 000× slower sound propagation.) Note that T, really a detector time constant, can either be made so short that the incident wave angle is sensed almost instantly, or long enough to smooth out noise or interference effects—albeit with a loss of timing resolution. The monopulse radar literature gives much attention to formulas like Eq. (56) and its imaginary counterpart,

$$n_i^{\mathrm{I}} = \frac{\mathrm{Im}[a_i(t)\alpha^*(t)]}{|\alpha(t)|^2}, \tag{59}$$

which becomes nonzero only in the presence of something more complicated than a solitary plane wave. In fact, the imaginary part of the monopulse ratio has been used in radars to detect multipath distortions, interference, or the presence of multiple scattering centers within the time-resolved range "slice" across the target.[53]

Finally, although the coordinate normalizing transformation of Eq. (51) was designed for real-valued data, it can be applied to complex valued data (with real $\mathbf{\Gamma}$) after appropriate modifications to Eqs. (56) and (57). In particular, Eq. (56) becomes

$$\mu = \mathrm{Re}\left[\frac{Y(t)}{X(t)}\right] = \frac{\mathrm{Re}[Y(t)X^*(t)]}{|X(t)|^2}. \tag{60}$$

This μ can be put in Eq. (52) to estimate the direction cosine n_i; or, one can use $\mu = \mathrm{Re}[M_T(t)]$, where

$$M_T(t) \triangleq \frac{\langle Y(t)X^*(t)\rangle_T}{\langle |X(t)|^2\rangle_T}. \tag{61}$$

If $\Upsilon \triangleq |\mu|/\sqrt{\langle |Y(t)|^2\rangle_T/\langle |X(t)|^2\rangle_T}$ then $\Upsilon \leqslant 1$ by the Schwarz inequality, with equality if there is no noise or interference; thus, small Υ implies inconclusive data.

E. Other formulas

If $[X(t), Y(t)]^T = \mathbf{\Gamma}[\alpha(t), a_i(t)]^T$ its correlation matrix has a real-part

$$\mathbf{R}(t) \triangleq \mathrm{Re}\begin{bmatrix} \langle |X|^2(t)\rangle_T & \langle X(t)Y^*(t)\rangle_T \\ \langle Y(t)X^*(t)\rangle_T & \langle |Y|^2(t)\rangle_T \end{bmatrix}. \tag{62}$$

A solitary incident wave gives $a_i(t) = n_i\alpha(t)$ in the absence of noise or interference, so $Y(t) = \mu X(t)$, with a slope $\mu = (\Gamma_{21} + \Gamma_{22}n_i)/(\Gamma_{11} + \Gamma_{12}n_i)$, and

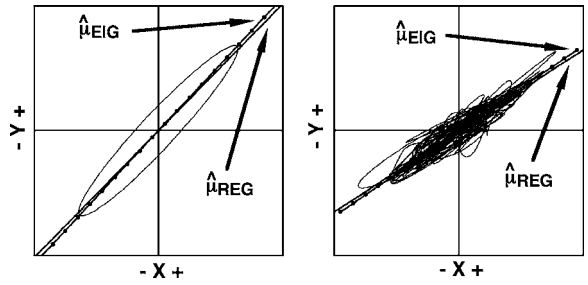


FIG. 14. Geometric comparison of the different slope estimates, μ_{reg} and μ_{eig}. Left panel: Two incident sinusoidal waves of the same frequency but moderate phase difference will produce an ellipse. (Any parallelogram that exactly circumscribes the ellipse defines a pair of possible source directions and their respective amplitudes.) The ellipse's major axis aligns with μ_{eig}, but μ_{reg} intersects the point where the curve is exactly vertical. Right panel: A broadband, angularly diffuse wave makes a splotchy figure. If the figure is rotated to bring the μ_{eig} line to the horizontal, the figure will appear level, and both μ_{eig} and μ_{reg} will be zero upon recalculation.

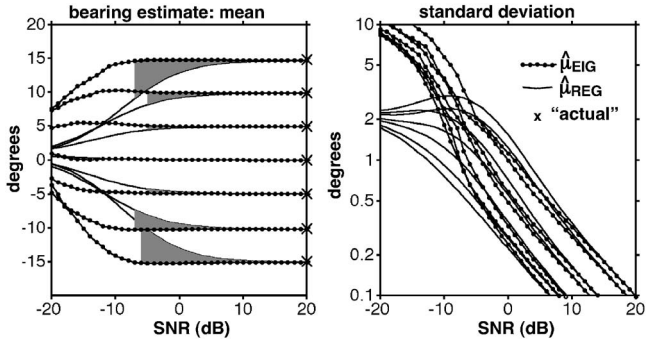


FIG. 15. Comparison of statistics of μ_{reg} and μ_{eig} as a function of SNR, for sources located at specific known angles, marked by "×" in the left-hand plot. The left-hand plot is also marked with gray webbing to indicate where (as determined from the right-hand plot) μ_{eig} has less standard deviation than μ_{reg}.

$$\mathbf{R}(t)=\begin{bmatrix} R_{11}(t) & R_{12}(t) \\ R_{21}(t) & R_{22}(t) \end{bmatrix} \overset{\text{ideally}}{=} \langle |X|^2(t)\rangle_T \begin{bmatrix} 1 & \mu \\ \mu & \mu^2 \end{bmatrix} = \langle |X|^2(t)\rangle_T \begin{bmatrix} 1 \\ \mu \end{bmatrix} [1\ \mu]. \tag{63}$$

The fact that $\mathbf{R}$ takes this form ideally suggests ways to extract a slope estimate for calculating n_i via Eq. (52) under nonideal conditions. The most obvious is

$$\mu_{\text{reg}}=\frac{R_{21}(t)}{R_{11}(t)}=\frac{\langle \text{Re}[Y(t)X^*(t)]\rangle_T}{\langle |X|^2(t)\rangle_T}=\text{Re}[M_T(t)], \tag{64}$$

which is the slope of the (regression) line through the origin that fits the XY data (real pairs and imaginary pairs) with least mean square *vertical* error, in the sliding-window spanned by $t\pm T/2$. Figure 14 illustrates this line for an elliptical XY pattern produced by a strong/weak pair of pure tones coming from different directions, and for an angularly diffuse broadband wave. Clearly Eq. (64) is the XY version of Eq. (57), and if $\mathbf{\Gamma}$ is a diagonal matrix Eqs. (64) and (52) reduce to

$$n_i=\frac{\langle \text{Re}[a_i(t)\alpha^*(t)]\rangle_T}{\langle |\alpha|^2(t)\rangle_T} \text{ (via } \mu_{\text{reg}}\text{, for diagonal } \mathbf{\Gamma}). \tag{65}$$

Unfortunately, noise inflates the denominator of Eq. (64) and biases μ_{reg} toward zero. A different estimate μ_{eig} that resists this tendency can be derived from signal subspace ideas: From Eq. (63) it is evident that the 2×2 matrix $\mathbf{R}(t)$ is an outer product, with unit rank, until that rank is inflated by perturbations due to noise or interference. For small perturbations the eigen-expansion, $\mathbf{R}(t)=V_s\mathbf{v}_s\mathbf{v}_s^T+V_n\mathbf{v}_n\mathbf{v}_n^T$, will reveal a weak *noise* eigenvalue V_n and a strong *signal* eigenvalue V_s, and the eigenvector $\mathbf{v}_s$ associated with the strong eigenvalue V_s will approximate $[1\ \mu]^T$ robustly, to within a scalar multiple. So we set $\mu_{\text{eig}}\triangleq(\mathbf{v}_s)_2/(\mathbf{v}_s)_1$. The eigen-expansion of $\mathbf{R}(t)$ yields to some tedious but straightforward algebra, giving

$$\mu_{\text{eig}}=\frac{2R_{21}(t)}{R_{11}(t)-R_{22}(t)+\sqrt{[R_{11}(t)-R_{22}(t)]^2+4R_{21}^2(t)}}. \tag{66}$$

As it turns out, this is the slope of the line through the origin that fits the data with least mean square error of the *perpendicular* distances from the line to the XY data[54,55] (real pairs and imaginary pairs) in the interval spanned by $t\pm T/2$. Another interpretation is that μ_{eig} corresponds to the angle of rotation necessary to make the displayed pattern outline appear untilted. It is the slope that would likely be estimated by visual inspection, and is most appropriate when the noises in X and Y are of similar power, as when the coordinate transformation $\mathbf{\Gamma}$ is chosen to circularize the noise splotch. (However, this slope migrates slowly through the pattern if one axis is rescaled, because μ_{eig} gives a bearing estimate that depends on relative scaling of the X and Y axes, whereas μ_{reg} does not.)

To compare the performance of μ_{eig} and μ_{reg} we used recorded pings from the Lake Travis acoustic data of Fig. 12, for seven specific source bearing angles spread across the main lobe of the beam pattern, ranging from $-15°$ to $+15°$, and added an entire ensemble of contemporaneously recorded lake noise at various levels to compare the statistics of the bearing estimates (after being limited to ±19°, a span twice the half-power beamwidth). The results (mean and standard deviation) are in Fig. 15. The bearing estimates computed from μ_{eig} and μ_{reg} give virtually identical results at high SNR, but μ_{reg} gives a strong bias toward $\theta=0$ as the SNR worsens, especially for directions far from the center of the beam. The estimate using μ_{eig} appears to hold off the effects of bias for at least 10 dB more loss in SNR than μ_{reg} does, and even has a somewhat lower standard deviation, but only down to the point where the standard deviation for μ_{eig} finally blossoms disturbingly.

If an aggregation of sources with direction cosines $n_i^{[1]},\ldots,n_i^{[m]},\ldots$, contribute waves that are uncorrelated over $t\pm T/2$, i.e., $\langle X_{[\ell]}X_{[m]}^*\rangle_T=\langle X_{[\ell]}Y_{[m]}^*\rangle_T=\langle Y_{[\ell]}Y_{[m]}^*\rangle_T=0$ for $\ell\neq m$, then the $\mathbf{R}$ matrix of Eq. (62) can be expressed as

$$\begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix} = 2 \times \begin{bmatrix} \sum_m \mathcal{P}^X_{[m]} & \sum_m \mathcal{P}^X_{[n]} \mu^{[m]} \\ \sum_n \mathcal{P}^X_{[m]} \mu_{[m]} & \sum_m \mathcal{P}^X_{[m]} (\mu^{[m]})^2 \end{bmatrix}, \tag{67}$$

where $\mathcal{P}^X_{[m]} = \frac{1}{2}\langle |X_{[m]}|^2(t)\rangle_T$ (average signal power of mth source as seen in X) and $\mu^{[m]} = (\Gamma_{21} + \Gamma_{22} n_i^{[m]})/(\Gamma_{11} + \Gamma_{12} n_i^{[m]})$ (mapped direction-slope of mth source). The power-weighted mean of the $\mu^{[m]}$'s is $\Sigma_m \mathcal{P}^X_{[m]} \mu^{[m]} / \Sigma_m \mathcal{P}^X_{[m]} = R_{21}/R_{11} = \mu_{\text{reg}}$; power-weighted variance is $\Sigma_m \mathcal{P}^X_{[m]} (\mu^{[m]} - \mu_{\text{reg}})^2 / \Sigma_m \mathcal{P}^X_{[m]} = R_{22}/R_{11} - \mu^2_{\text{reg}}$.

VII. CONCLUSIONS

The Adcock, Wullenweber, and Δ/Σ (sum and difference, BDI, monopulse, etc.) methods follow a natural progression. The space–time filtered gradient method extends it by doing space–time filtering so that the ratio a_i/α indicates the direction cosines n_i directly, independent of frequency and bandwidth, using apertures of (up to) three dimensions. It is immune to fluctuations of instantaneous frequency in broadband waves, and has direction-indicating characteristics like an Adcock array, while offering directionality to shield out interference. Over the past two decades we have employed it in many practical applications, through various phases of its evolution. The mathematical "toolbox" that we have presented here is intended to aid others in its use. Application of higher-order gradient methods, using the discrete version of the gradient for point-sensor arrays, is left to a separate paper.[56]

ACKNOWLEDGMENTS

This paper was written with the support of the Office of Naval Research under Contract No. N00014-99-1-0362 and under ARL:UT Internal Research and Development Contract No. FEE-800. The work described therein was aided by the considerable contributions of many participants during the years since 1982, including Garland Barnard, George Coble, John Maxwell, Steve Blackstock, John Huckabay, George Koehler, John Brady, Ted Stanford, Sudha Reese, Steve Lacker, Rob Stewart, Charles Loeffler, Min Chang, Dave Murray, Steve Morrissette, Danny Dickens, Danny Shrode, Scott Duff, Mike DeSimone, Jim Baughman, Greg Allen, and Colin Bown. T. L. H. owes much to the teaching and inspiration of Professor F. V. Hunt. Special thanks go to James Fung for checking the equations and derivations, and to the reviewers, whose criticisms led to an improved manuscript.

APPENDIX A: FACTORING A VANDERMONDE PRODUCT (SEE SEC. IV E)

This MATLAB function extracts a Vandermonde factor by ESPRIT methods:

```
function [V,S0]=vanderESPRIT(Smk,I)
%. Find I-width Vandermonde V so Smk=V*S0
[V,Lam]=eig(-Smk*Smk'); mS=size(Smk,1);
Lam=eig(V(1:(end-1),1:I)\V(2:end,1:I)).';
V=Lam(ones(mS,1),:).^[[0:(mS-1)]'*ones(1,I)];
S0=V\Smk;[v,n]=sort(-sum(real(S0.*conj(S0)),2));
V=V(:,n); S0=S0(n,:); % strongest first
```

APPENDIX B: PROOF THAT EQS. (45) and (46) IMPLY $s_{m+1} = \tilde{n}_1 s_m$, i.e., $s_m = \kappa^{-1} \tan(\kappa n_1) s_{m-1}$

A (sinusoidal) component of wavelength $\lambda = c/f$ and direction cosine n_1 makes $[p(\mathbf{x},t)]\ldots = A_0 \exp[\mathrm{j}2\pi(ft + \ell n_1 d_1/\lambda)]$ in Eq. (45). So, with $z \triangleq \exp(-\mathrm{j}2\pi n_1 d_1/\lambda)$,

$$s_m(t) = -\left(\frac{-c}{\mathrm{j}2\pi f}\right)^{m-M} [h(t) * A_0 \exp(\mathrm{j}2\pi f t)] \sum_{\ell=0}^{L-1} \mathcal{W}_\ell^{(m)} z^{-\ell}. \tag{B1}$$

The summation is the z transform of the weights, i.e., $\mathcal{ZT}\{[\mathcal{W}_0^{(m)}, \ldots, \mathcal{W}_{L-1}^{(m)}]\}$, so

$$\frac{s_m(t)}{s_{m-1}(t)} = \left(\frac{-\lambda}{\mathrm{j}2\pi}\right) \frac{\mathcal{ZT}\{[\mathcal{W}_0^{(m)}, \ldots, \mathcal{W}_{L-1}^{(m)}]\}}{\mathcal{ZT}\{[\mathcal{W}_0^{(m-1)}, \ldots, \mathcal{W}_{L-1}^{(m-1)}]\}}, \tag{B2}$$

using $\lambda = c/f$. The z transform turns convolutions into products, so Eq. (46) implies

$$\frac{\mathcal{ZT}\{[\mathcal{W}_0^{(m)}, \ldots, \mathcal{W}_{L-1}^{(m)}]\}}{\mathcal{ZT}\{[\mathcal{W}_0^{(m-1)}, \ldots, \mathcal{W}_{L-1}^{(m-1)}]\}} = \left[\frac{2}{d_1} \frac{\mathcal{ZT}\{[1,-1]\}}{\mathcal{ZT}\{[1,1]\}}\right]. \tag{B3}$$

Since $\mathcal{ZT}\{[1,\pm 1]\} = 1 \pm z^{-1} = 1 \pm \exp(\mathrm{j}2\pi n_1 d_1/\lambda)$, the right side simplifies to $(-\mathrm{j}2/d_1)\tan(n_1 \pi d_1/\lambda)$, so Eqs. (B2) and (B3) imply $s_m = (\pi d_1/\lambda)^{-1} \tan(n_1 \pi d_1/\lambda) s_{m-1}$ Q.E.D.

[1] C. A. Balanis, *Antenna Theory-Analysis and Design* (Harper & Row, New York, 1982), pp. 257–260.

[2] The space–time filtered wave field will exist only in the signal processing realm; nevertheless, its gradient can be measured.

[3] Allan D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981), Chap. 1.

[4] L. J. Ziomek, *Acoustics: Fundamentals of Acoustic Field Theory and Space-time Signal Processing* (CRC Press, Ann Arbor, 1995), p. 402.

[5] Statements about the direction of ∇p [or, later in this section, of $\mathbf{a}(t)$] assume a real p, since the direction of a 3D vector defined over the field of complex scalars is troublesome to contemplate. However, $\mathbf{n}$ is always real and can be calculated even if p is complex, as discussed in Sec. VI.

[6] K. F. Riley, M. P. Hobson, and S. J. Bence, *Mathematical Methods for Physics And Engineering* (Cambridge University Press, New York, 198), p. 317 (Eq. 9.21).

[7] *How close*? The answer depends on the likelihood of interferors in grating lobes, and upon one's ability to tolerate the frequency-dependent *inflection* of the direction cosine(s) that is described in Sec. IV F. For irregularly spaced arrays the formula of Eq. (48) does not hold, and tedious simulation may be required, although a spacing $<\lambda_{\min}/4$ probably makes inflection inconsequential. But this may be impractical or even unnecessary if the array is steered as in Sec. IV D.

[8] M. B. Moffet, J. M. Powers, and J. C. McGrath, "A ρc hydrophone," J. Acoust. Soc. Am. **80**, 375–381 (1986).

[9] If a few interior sheets can, without harm, generate an approximation $\tilde{\alpha}(t)$ to $\alpha(t)$, then $\tilde{\alpha}(t)\mathbf{a}(t)$ will point correctly except between the imperfectly synchronized zero-crossings of $\tilde{\alpha}(t)$ and $\alpha(t)$, when it points to the opposite direction.

[10] If the thin-plate aperture is mounted on a hard barrier plate it probably does not need to match the acoustic impedance of water. This means it could be nonvoided PVDF or some other material.

[11] R. N. Bracewell, *The Fourier Transform and its Applications* (McGraw-Hill, New York, 1978), pp. 244–250.

[12] G. Birkhoff and S. MacLane, *A Survey of Modern Algebra*, 4th ed. (Macmillan, New York, 1977), pp. 368–371.

[13] Electrode preamplifier outputs may need calibration factors before the

addition, depending upon preamplifier input impedances, cable loading, and whether the preamplifiers are of the charge-amplifying or voltage-amplifying variety.
[14]S. Burke and P. Rosenstrach, "High resolution monopulse piezopolymer sonar sensor," Proceedings of the 1992 Symposium on Autonomous Underwater Vehicle Technology, pp. 1, 209–213.
[15]P. J. Rowe, "Characteristics of a wideband monopulse direction finder," Masters thesis in Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1999.
[16]T. L. Henderson, "Wide-band monopulse sonar: Processor performance in the remote profiling application," IEEE J. Ocean. Eng. **OE-12**, 182–197 (1987).
[17]The reoriented plate's effective size and shape also change, to maintain the same silhouette as the original aperture (as viewed from the source).
[18]H. F. Olson, "Gradient microphones," J. Acoust. Soc. Am. **17**, 192–198 (1946).
[19]R. Kumeresan, "Spectral Analysis," in *Handbook for Digital Signal Processing*, edited by Sanjit K. Mitra and James F. Kaiser (Wiley, New York, 1993), pp. 1203–1237.
[20]T. L. Henderson, "Rank reduction for broadband waves incident on a linear receiving aperture," *Proceedings of the Nineteenth Asilomar Conference on Circuits, Systems, and Computers*, 6–8 November 1985 (IEEE Computer Society Press, Washington, DC, 1986), pp. 462–466.
[21]D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques* (Prentice-Hall, Englewood Cliffs, NJ, 1993) pp. 40–45 and Chap. 7.
[22]Admittedly, Eq. (48) presupposes a sinusoid, but Fourier analysis decomposes any nonsinusoidal incident wave into sinusoids, each with its own inflected direction cosine $\bar{n}_1$, with the aggregate behaving rather like multiple incident waves from somewhat different directions, amenable to the treatment of Sec. VI A.
[23]T. L. Henderson, "Matched beam theory for unambiguous broadband direction finding," J. Acoust. Soc. Am. **78**, 563–574 (1985).
[24]F. Adcock, "Improvements in means for determining the direction of a distant source of electro-magnetic radiation," British Patent 130,490 (1919).
[25]L. A. deRosa, "Direction-finding," in *Electronic Countermeasures*, edited by J. A. Boyd, D. B. Harris, D. D. King, and H. W. Welch, Jr. (Peninsula, Los Altos, CA, 1978), pp. 10-38–10-95.
[26]R. A. Watson-Watt and J. F. Herd, "An instantaneous direct-reading goniometer," Journal of the Institution of Electrical Engineers (London) **64**, 611–622 (1926) [read before the Wireless Section on 3 March, 1926].
[27]H. D. Kennedy and W. Wharton, "Direction-finding antennas and systems," in *Antenna Engineering Handbook*, 2nd ed., edited by R. C. Johnson and H. Jasik (McGraw-Hill, New York, 1984), pp. 39-18–39-32.
[28]W. G. Bruner, R. C. Bauer, S. N. Broady, R. B. Hughes, and J. C. Kirk, "Radar and navigational systems," in *Electronic Designers' Handbook*, 2nd ed., edited by L. J. Giacoletto (McGraw-Hill, New York, 1977), pp. 26-72–26-75.
[29]J. A. Hildebrand and W. S. Hodgkiss, "Large-aperture arrays for VLF ambient noise and signal propagation studies," Proc. IEEE Oceans '90 Conference, pp. 24–29.
[30]F. Adcock, "Radio direction finding in three dimensions," Proc. Inst. Radio Eng. Australia **20**, 7–11 (1959).
[31]O. H. Schock, C. K. Stedman, J. L. Hathaway, and A. N. Butz, "Bearing deviation indicator for sonar," AIIE Trans. **66**, 1285–1295 (1947).
[32]J. W. Horton, *Fundamentals of Sonar* (United States Naval Institute, Annapolis, 1959), pp. 274–286.
[33]H. L. Saxton, "Sector scan indicator QXA," NRL Rep. S-2631, Naval Research Laboratory, Washington, DC, 30 August 1945.
[34]Leon Cohen, *Time-Frequency Analysis* (Prentice Hall, Englewood Cliffs, NJ, 1995), Chap. 2.
[35]G. R. Gapper and T. Hollis, "The accuracy of an inteferometric sidescan sonar," Proc. Inst. Acoust. **7**, 126–134 (1985).
[36]H. Rindfleisch, "The Wullenweber wide aperture direction finder," Nachrichtenech. Z. **9**, 119–123 (1956) [The German title of the paper is: "Die Grossbasis-Peilanlage 'Wullenwever' "].
[37]R. E. Franks, "Direction-finding antennas," in *Antenna Handbook*, edited by Y. T. Lo and S. W. Lee (Van Nostrand Reinhold, New York, 1988), pp. 25-9–25-26.
[38]P. J. D. Gething, *Radio Direction-Finding* (Peregrinus, Southgate House, Stevenage, Herts, UK, 1978), pp. 1–9, 87–97, 123–137, 161–168.
[39]W. R. Kiel, "Apparatus for the directional transmission or reception of wave energy," U.S. Patent 1,977,974 (1934). [German Patent 544,267 (1931)].
[40]F. V. Hunt *et al.* (of the Harvard Underwater Sound Laboratory), "Underwater Sound Equipment. III. Scanning Sonar Systems," NDRC Summary Technical Report, Div. 6, (Washington, DC, 1946) Vol. **16**, p. 5.
[41]M. A. Chramiec and J. T. Kroenert, "A sonar for smaller ships," Proceedings of the IEEE EASCON '78 Conference, pp. 747–754.
[42]A. I. Leonov and K. I. Fomichev, *Monopulse Radar* (Artech House, Norwood, MA, 1986).
[43]S. M. Sherman, *Monopulse Principles and Techniques* (Artech House, Norwood, MA, 1984).
[44]D. R. Rhodes, *Introduction to Monopulse* (McGraw-Hill, New York, 1959).
[45]G. M. Kirkpatrick, "Final engineering report on angular accuracy improvement," a General Electric Company Report dated 1 Aug. 1952, reprinted in *Monopulse Radar*, Radars Vol. **1**, edited by D. K. Barton (Artech House, Norwood MA, 1974).
[46]W. L. Murdock and J. S. Kerr, "Relations between the far field and the illumination of antenna apertures," General Electric Company, Electronics Laboratory, TIS 51E234, 1 November, 1951. Pertinent results are abstracted in *Monopulse Radar*, Radars Vol. **1**, edited by D. K. Barton (Artech House, Norwood, MA, 1974), pp. 26, 46, 89, and also in *Introduction To Monopulse*, by D. R. Rhodes, (McGraw-Hill, New York, 1959), pp. 92–104.
[47]S. M. Sherman (see Ref. 43, pp. 330–331).
[48]T. L. Henderson, "Fractional beamwidth resolution with a paired beamformer," Proceedings of the IEEE EASCON '82 Conference, pp. 251–252.
[49]J. H. Dunn and D. D. Howard, "Target Noise," in *Radar Handbook*, edited by M. I. Skolnik (McGraw-Hill, New York, 1970), pp. 28-8–28-15.
[50]B. Edde, *Radar: Principles, Technology, Applications* (Prentice Hall, Englewood Cliffs, NJ, 1993), pp. 203–208.
[51]W. M. Sherrill, "A survey of HF interferometry for ionospheric propagation research," Radio Sci. **6**, 549–566 (especially Fig. 9). (1985).
[52]A. D. Bailey and W. C. McClurg, "A sum-and-difference interferometer system of HF radio direction finding," IEEE Trans. Aerosp. Navig. Electron. **ANE-10**, 65–72 (1963).
[53]S. J. Asseo, "Detection of target multiplicity using quadrature monopulse ratio," IEEE Trans. Aerosp. Electron. Syst. **AES-17**, 271–280 (1981).
[54]R. Burlington and D. May, *Handbook of Probability And Statistics with Tables*, 2nd ed. (McGraw-Hill, New York, 1970), p. 152.
[55]E. Gose, R. Johnsonbaugh, and S. Jost, *Pattern Recognition And Image Analysis* (Prentice-Hall, Upper Saddle River, NJ, 1996), pp. 358–362.
[56]T. J. Brudner and T. L. Henderson, "Wideband direction finding via shielded gradient beamspace techniques," J. Acoust. Soc. Am. **114**, 2427–2428 (Paper 4aUWa12 presented at the 146th Meeting of the Acoustical Society of America, 14 November 2003).

Variability of focused sonic booms from accelerating supersonic aircraft in consideration of meteorological effects

Reinhard Blumrich[a)]
Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, 82234 Weßling, Germany

François Coulouvrat
Laboratoire de Modélisation en Mécanique, Université Pierre et Marie Curie and CNRS (UMR 7607), 4 Place Jussieu, 75252 Paris Cedex 05, France

Dietrich Heimann
Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, 82234 Weßling, Germany

(Received 11 February 2005; revised 2 May 2005; accepted 2 May 2005)

Statistics of the meteorologically induced variability of focused sonic boom characteristics due to unsteady, accelerated supersonic flights were derived. The simulations were performed with an advanced sonic boom software including a numerical solver of the Nonlinear Tricomi Equation modeling the focused sound pressure field around caustics. A one-year set of daily meteorological data along the Northern Atlantic flight corridor was used as input. The statistics comprise the location, strength and geometrical extension of caustics at ground level for assumed flights of a supersonic commercial transport from Paris to New York and back. Caustics only occurred during the acceleration phase above the English Channel for flights from Paris to New York, never during the deceleration phase of return flights. The mean value of the maximum focused overpressure is found to be about 4 times the maximum overpressure for cruise flight at Mach 2 in similar conditions, but shows a large variability. The caustics intersection with the ground varies between 3 and 8 km in width and between 30 and 180 km in length. A linear correlation analysis showed the relationship between meteorological profile shape parameters and the focused sonic boom characteristics. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1938547]

PACS number(s): 43.25.Cb, 43.28.Mw, 43.28.Lv, 43.28.Fp [MFH] Pages: 696–706

I. INTRODUCTION

The annoyance of sonic boom is a key environmental barrier to the development of civil supersonic aircraft. The Concorde was operated commercially at supersonic speed mostly overseas, but future high speed aircraft, either commercial transports or business jets, are likely to be profitable only if they are allowed to fly supersonically overland. Much effort is devoted nowadays to the design of a low boom aircraft, with an aerodynamic shape optimized to reduce the sonic boom at cruising speed below an acceptable level. However, the sonic boom of aircraft at cruising speed is not the one with the highest peak overpressure. During the acceleration phase from subsonic to supersonic speeds, the sonic boom is focused because of the convergence of the acoustical rays launched normally to the narrowing Mach cone (Fig. 1). The envelope of rays forms a caustic surface around which the sonic boom is amplified (a focused sonic boom is sometimes referred to as superboom). Superbooms can also be observed in the case of turn maneuvers, or for cruise at low supersonic Mach numbers, but these cases can be avoided, contrarily to the boom focusing due to acceleration.

The phenomenon of sonic boom focusing is known since the pioneering work of Esclangon (1925) during World War I on the sonic boom of supersonic gun shells. Sonic boom focusing was modeled by Guiraud (1965), who showed that the superboom satisfies the so-called nonlinear Tricomi equation. Approximate analytical solutions based on the hodograph transform and Guiraud's similitude were obtained by Seebass (1971) and Gill and Seebass (1973) and integrated within the PCBoom3 software by Plotkin (2002). A numerical algorithm to solve the nonlinear Tricomi equation was developed by Auger and Coulouvrat (2002). It has been improved by Marchiano, Coulouvrat, and Grenon (2003) and coupled to CFD numerical simulations of the aircraft nearfield at low supersonic Mach 1.2.

Experimental studies on sonic boom focusing include either laboratory scale experimental simulations (Sanai *et al.*, 1976) or superboom recordings during test flights (Wanner *et al.*, 1972). Marchiano, Thomas, and Coulouvrat (2003) showed good agreement between numerical solution of the nonlinear Tricomi equation, and weak shock experiments in a water tank scaled to sonic boom. After extensive flight tests, Downing *et al.* (1998) found a reasonable agreement between measured data and superboom simulations.

In a previous paper Blumrich, Coulouvrat, and Heimann (2005) (referred to as BCH'04 in the following) showed that, for cruising flights, the area affected by sonic boom from

[a)]Present address: Forschungsinstitut für Kraftfahrwesen und Fahrzeugmotoren (FKFS), Pfaffenwaldring 12, 70569 Stuttgart, Germany.

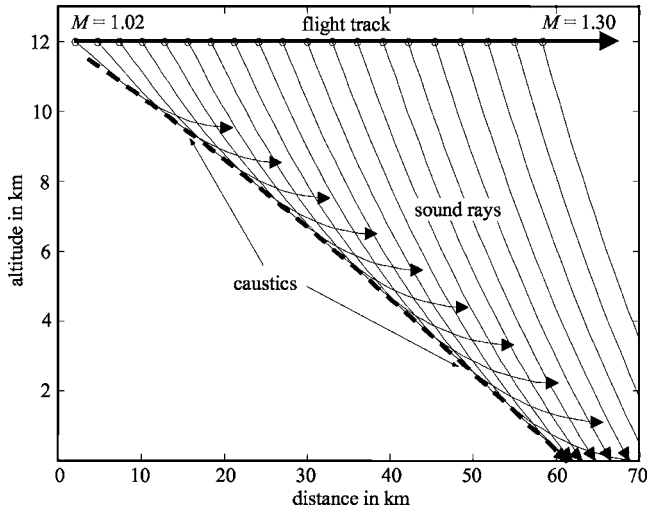


FIG. 1. Schematic view of sonic-boom focusing caused by an aircraft accelerating from Mach=1.02 to Mach=1.30 at 12 km altitude (thick solid arrow). The thin arrows indicate sound rays which emerge from the aircraft during acceleration. The thick dashed line shows the position where the sound rays form caustics.

cruising flight largely depends on the actual meteorological condition. Hence, it is expected that the onset location and amplification of booms heard at the ground level due to an aircraft in accelerated flight also depends on the meteorological condition. A statistical analysis of the focused boom characteristics is the purpose of the present study. The investigation is restricted to the area of the English Channel where assumed future supersonic aircraft from Paris to New York and back accelerate and decelerate, respectively. Various real meteorological situations that appeared during 1 year in this area are taken into account. Aircraft trajectories are determined according to exploitation procedures similar to those applied by Air France for the former Paris to New York Concorde flights. Simulations are performed for an Airbus mock-up of the planned European Supersonic Commercial Transport (ESCT) vehicle. It is a mock-up with conventional supersonic design and the following dimensions: an 89 m long aircraft of 340 tons maximum take-off weight, with a wing span of 42 m (wing surface 836 m^2) and an almost cylindrical fuselage (4 m diameter), designed for carrying 250 passengers at a Mach 2 cruise.

The meteorological and trajectory data base is presented in Sec. II. Section III briefly describes the sonic boom propagation algorithm which was used to calculate the focusing effects and the resulting sonic boom characteristics. The output of the focused sonic boom calculations and statistical evaluations are discussed in Sec. IV, whereas the dependency on the meteorology is presented in Sec. V. Eventually, conclusions are presented in Sec. VI.

II. METEOROLOGICAL DATA BASE AND FLIGHT TRAJECTORIES

The investigation is based on real meteorological data. Detailed aircraft configuration and flight trajectory data depend on the meteorological conditions. They were determined in an iterative way in cooperation with Airbus France SAS. In a first step, two mean ground tracks (Paris to New York and return) were determined, close to those used for Concorde to avoid nuisance due to sonic boom (Fig. 2). Onboard aircraft flight control system will keep lateral deviations of actual ground tracks due to lateral winds within ranges much smaller than the horizontal grid spacing of the meteorological data. Therefore precision is sufficient to extract the real meteorological data that will then be used to compute actual tracks, at the vertical of points along the mean ground tracks. It is to be noticed that subsonic stages at Mach 0.95 during the climb and the descent are required over land and that acceleration beyond Mach 0.95 is allowed only outside sensitive areas to avoid sonic boom impact at ground. The point of passing Mach 0.95 is calculated on the base of Concorde flight procedures.

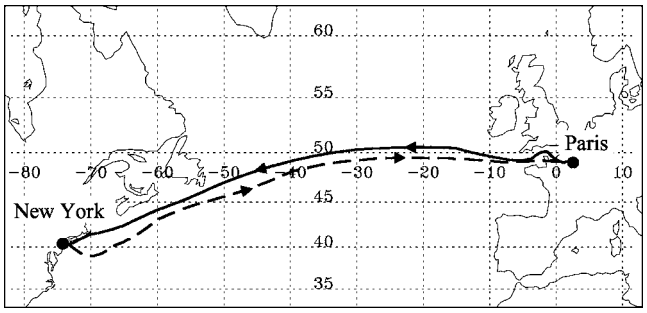


FIG. 2. Assumed flight tracks from Paris to New York (solid) and from New York to Paris (dashed).

The meteorological data were extracted from the ERA15 reanalysis database of the European Centre for Medium Range Weather Forecast (ECMWF; Gibson *et al.*, 1997). They were interpolated from the original grid (horizontal spacing: 1.125°, 31 levels between surface and 30 km altitude) to way points of the flight tracks Paris–New York and New York–Paris as they are shown in Fig. 2.

For computational reasons the investigation was restricted to 1 year. The year 1993 was selected out of the 15-year data set because it exhibits a near-average behavior with respect to the height of the thermal tropopause h_{TP} and the NAO-Index. The thermal tropopause as defined in WMO (1957) was determined following Zängl and Hoinka (2000). The NAO index (North Atlantic Oscillation Index; Barnston and Livezey, 1987) is defined as the difference in sea surface air pressure between Iceland and the Azores Islands. High values of the NAO index stand for mainly zonal flow, while low values indicate mainly meridional flow.

Meteorological data above sea surfaces do not show much diurnal variation because the sea surface temperature is nearly constant during the course of a day. Therefore only the 00 UTC data (UTC=Universal Time Coordinated) of each day are considered. In total, 365 sets of meteorological data and corresponding aircraft data were used. The meteorological data comprise height, pressure, temperature, specific humidity, wind speed, and wind direction at all way points between Paris and New York and back.

From these meteorological data, the detailed 730 flight trajectories were calculated for each meteorological situation. The tool used to calculate trajectories includes models for engine and airframe geometry, weight, aerodynamic, and fuel consumption. This tool has been adapted to calculate

TABLE I. Meteorological profile shape parameters (set 1).

m	Symbol	Description
11	h_{TP}	height of the thermal tropopause
12	$\bar{\gamma}_B$	mean vertical temperature gradient near the ground
13	$\bar{\gamma}_T$	mean vertical temperature gradient in the troposphere
14	$\bar{\gamma}_S$	mean vertical temperature gradient in the stratosphere
15	q_B	specific humidity near the ground
16	h_{JS}	height of the jet stream wind speed maximum
17	u_{JS}	wind speed maximum
18	$\bar{\eta}_T$	mean vertical wind speed gradient below h_{JS}
19	$\bar{\eta}_S$	mean vertical wind speed gradient above h_{JS}

flight profiles in a real atmosphere (instead of standard atmosphere), accounting for actual pressure, temperature and wind at each altitude along the flight path. Trajectory integration is made by quadratic approximation using adaptive steps for each flight phase. More specifically, output parameters per selected point on trajectories are (1) time and distance, (2) true altitude and pressure altitude, (3) longitude and latitude, (4) true air speed, (5) acceleration, (6) angle of attack, (7) Mach number (and derivative vs time), (8) heading angle (and derivative vs time), (9) climb angle (and derivative vs time). Finally, these trajectory data were used as an input for the sonic boom propagation code detailed in the next section.

In the following, the acoustically relevant vertical profiles of the meteorological parameters are expressed by shape parameters as they were defined in BCH'04. A first set of shape parameters refers to meteorological parameters which are independent of the flight direction. They characterize the temperature profile, the humidity and the wind profile and comprise the height of the thermal tropopause h_{TP} (WMO, 1957; Zängle and Hoinka, 2000) as a major discontinuity in the vertical temperature profile separating troposphere and stratosphere, the mean vertical temperature gradients near the ground $\bar{\gamma}_B$ (i.e., lowest 300 m), in the troposphere $\bar{\gamma}_T$, and the lower stratosphere $\bar{\gamma}_S$, the specific humidity near the ground q_B, the height of the jet stream wind speed maximum h_{JS}, the maximum wind speed of the jet stream u_{JS}, and the mean vertical wind speed gradients in the troposphere $\bar{\eta}_T$ and the lower stratosphere $\bar{\eta}_S$. A second set of shape parameters also involves the angle $\alpha=\varphi_{\mathrm{flight}}-\varphi_{\mathrm{wind}}$ between the direction of the flight $\varphi_{\mathrm{flight}}$ and the wind direction φ_{wind}. These parameters refer to the mean vertical gradients near the ground (index B), in the troposphere (index T) and stratosphere (index S) of the effective speed of sound for propagation into the four main directions relative to the direction of flight: left-wing (port) direction $(\bar{\chi}_{-x,B},\bar{\chi}_{-x,T},\bar{\chi}_{-x,S})$, right wing (starboard) direction $(\bar{\chi}_{+x,B},\bar{\chi}_{+x,T},\bar{\chi}_{+x,S})$, backward (tail) direction $(\bar{\chi}_{-y,B},\bar{\chi}_{-y,T},\bar{\chi}_{-y,S})$, and forward (head) direction $(\bar{\chi}_{+y,B},\bar{\chi}_{+y,T},\bar{\chi}_{+y,S})$. The effective speed of sound controls the refraction of sound waves and is composed of the adiabatic speed of sound in resting air and the wind component in the respective direction of sound propagation,

$$c_{\mathrm{eff},-x}=\sqrt{\kappa R_d T}+V\sin(\alpha)$$

sound propagation into port direction,

$$c_{\mathrm{eff},+x}=\sqrt{\kappa R_d T}-V\sin(\alpha)$$

sound propagation into starboard direction,

$$c_{\mathrm{eff},-y}=\sqrt{\kappa R_d T}+V\cos(\alpha)$$

sound propagation into tail direction,

$$c_{\mathrm{eff},+y}=\sqrt{\kappa R_d T}-V\cos(\alpha)$$

sound propagation into head direction,

with the ratio of the specific heat capacities for constant pressure and constant volume $\kappa=c_p/c_v=1.4$, the gas constant of dry air $R_d=287\ \mathrm{J\,kg^{-1}\,K^{-1}}$, and the wind speed V. The wind direction φ_{wind} is defined following meteorological convention ($\varphi_{\mathrm{wind}}=0°$: wind blowing from north to south, $\varphi_{\mathrm{wind}}=90°$: wind blowing from east to west, etc.), while the flight direction $\varphi_{\mathrm{flight}}$ is defined according to aeronautical conventions ($\varphi_{\mathrm{flight}}=0°$: aircraft flies from south to north, $\varphi_{\mathrm{flight}}=90°$: wind aircraft flies from west to east, etc.). The meteorological shape parameters are summarized in Tables I and II. For further details we refer to BCH'04.

III. SUPERBOOM MODELING

The theoretical model for sonic boom focusing was established by Guiraud (1965). It includes two physical effects: diffraction and nonlinearity. The dominant one is diffraction, which is neglected in the geometrical approximation. Diffraction catastrophe theory is known in linear wave physics (Berry, 1976) to provide a universal description of focusing effects depending only on the caustic topology. It predicts the transformation of the shape of the sonic boom from an "N" to a "U" wave. However, diffraction alone is unable to provide the amplitude of the boom overpressure, as shock waves lead to singular (unbounded) peaks of the "U" wave. Nonlinearities must be taken into account as an additional "limiting" effect to remove these singularities. The resulting equation is the so-called Nonlinear Tricomi Equation (NTE). It is a scalar, nonlinear, partial differential equation for the pressure field in two variables (time and distance from the caustics). One peculiarity of the NTE is its mixed-type character, as it may be either hyperbolic or elliptic, depending on which side ("illuminated" or "shadow zone") of the caustic the observer is. This explains the sharp transition in the wave

TABLE II. Meteorological profile shape parameters (set 2): effective sound speed profile shape parameters relative to flight direction. (A) Port direction, (B) starboard direction, (C) tail direction, (D) head direction.

	Direction				
m	A	B	C	D	Description
21	$\bar{X}_{-x,B}$	$\bar{X}_{+x,B}$	$\bar{X}_{-y,B}$	$\bar{X}_{+y,B}$	mean vertical gradient of the effective sound speed near the ground
22	$\bar{X}_{-x,T}$	$\bar{X}_{+x,T}$	$\bar{X}_{-y,T}$	$\bar{X}_{+y,T}$	mean vertical gradient of the effective sound speed in the troposphere
23	$\bar{X}_{-x,S}$	$\bar{X}_{+x,S}$	$\bar{X}_{-y,S}$	$\bar{X}_{+y,S}$	mean vertical gradient of the effective sound speed in the lower stratosphere

shape, but makes the numerical resolution especially tricky. Several authors have derived Guiraud's theory in a more tractable way, the most recent one being the work of Auger (2001) for the general case of a 3D, heterogeneous atmosphere with slow winds. The reader is referred to this reference for details of the theory and an extensive bibliography.

The numerical modeling of sonic boom focusing is carried out in four successive steps: ray tracing, determination of the ground position of the geometrical caustics (the envelope of rays where the ray tube algebraic area vanishes), determination of the caustic local geometry, and numerical resolution of the Nonlinear Tricomi Equation. The three first steps are presented in the Appendix. The numerical solver for the NTE is described in Auger and Coulouvrat (2002) and Marchiano, Coulouvrat, and Grenon (2003).

This numerical solution provides the sound pressure time waveforms only in the direction normal to the geometrical caustic. To get the pressure field at the ground level, a projection in the direction tangent to the caustic has to be done (Fig. 3). A reflection coefficient equal to 1 (pressure doubling) is assumed, thus amounting to a perfectly rigid and flat surface, an excellent approximation for the (flat) sea surface. A separate study (Rendón Garrido and Coulouvrat, 2005) showed that, at the ground surface, even in the nonlinear case, a linear reflection coefficient could be used. Numerical simulations based on an impedance model for a rough sea surface (Boulanger and Attenborough, 2004) indicate the influence of the sea surface roughness is not sufficient to strongly modify the superboom.

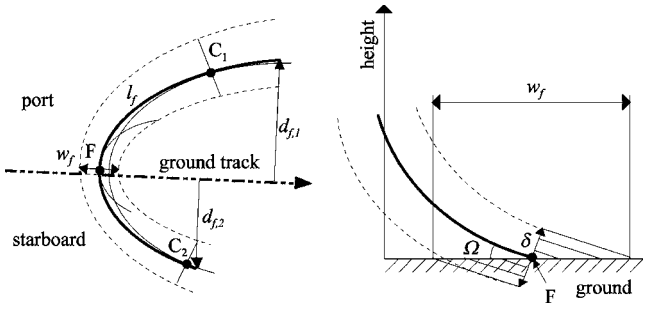


FIG. 3. Sketched view of caustic geometry. Left: horizontal view. An accelerated flight (the dashed–dotted arrow is the ground projection of the flight track) leads to a geometrical caustic, with a crescent-like ground intersection (thick line). Two isoemission lines (thin lines) tangential to the caustic are drawn, one at the "focal" point F (zero azimuthal emission angle), one laterally on port side at point C_1 and on starboard side at point C_2. The curvilinear length of the ground caustic is l_f, its port-side extent is $d_{f,1}$ and its starboard extent is $d_{f,2}$. The width of the physical caustic (within the two dotted lines) at the focal point F is w_f. Right: vertical view. The caustic (thick line) is incident on the ground at the focal point F with an angle Ω. The NTE solver provides the pressure field in the normal direction up to distances equal to one caustic thickness δ. The ground pressure field is then obtained over a segment of width w_f by projection in the direction tangential to the caustic and doubling (rigid surface).

For each point along the geometrical caustic, the NTE numerical solver provides the pressure field in the normal direction at a finite distance from the geometrical caustic. In practice, in the numerical solver, that distance is chosen equal to one diffraction boundary layer thickness on each side of the geometrical caustic (Auger and Coulouvrat, 2002). After ground projection, this defines the physical caustic, a narrow strip of width w_f around the geometrical caustic were diffraction effects cannot be neglected (Fig. 3). That thickness depends on the duration of the boom and on the caustic local geometry (Appendix). Finally, it has also to be recalled that the point of maximum overpressure is not located exactly on the geometrical caustic, but is slightly shifted towards the boom carpet. The shift is equal to a fraction of the caustic thickness (typically 10%–30% of w_f). It is a nonlinear effect, the shift increasing with the boom amplitude (Marchiano, Coulouvrat, and Grenon, 2003).

The geometrical parameters of caustics (if any) have been determined for all computed cases (365). To reduce computation time however, focused pressure waveforms were computed only at the caustic "focal" point F, i.e., the point where a ray emitted in forward direction (azimuth angle $\Phi=0$) is tangential to the caustic (see Appendix for more details). For an acceleration caustic with a typical crescent ground shape (Fig. 3), that point is located close to the apex of the crescent and is almost the nearest to the point of breaking the sound-barrier. As expected, numerical simulations in a few cases have shown that the focal point F is also the point along the caustic where the focused boom reaches the maximum overpressure. Therefore, the simulation result at the "focal" point provides a measure of the maximum impact during a supersonic flight. It has also to be noted that the physics of the focusing at the extremities of the caustic crescent when sound is grazing over the ground is not really understood.

The ray tracing algorithm cannot deal with horizontal variations in the atmosphere. Therefore, only vertical meteo-

TABLE III. Flight data during the time of sonic boom focusing. The average and standard deviation refer to the 348 cases of focusing over the English Channel due to accelerating flights from Paris to New York.

	Mean value	Standard deviation
Flight altitude	12 090 m	347 m
Heading angle	319°	2.3°
Climb angle	1.68°	0.21°
Mach number	1.34	0.076
Time derivative of Mach number	0.0025 s^{-1}	0.00038 s^{-1}

rological profiles were used in the simulations. They always refer to the aircraft position at which the ray associated to the "focal" point was emitted.

The pressure wavefield used as input data for the acoustical propagation model is the Whitham' function (Whitham, 1952), generalized for a nonaxisymmetric body with wings (Walkden, 1958) and engines. The use of Whitham's functions is nowadays often superseded by direct CFD simulations (Plotkin and Page, 2002), also for focused booms (Marchiano, Coulouvrat and Grenon, 2003). However, as the present study focuses on the influence of meteorology rather than on the detailed influence of the aircraft design, it is sufficient for our purpose. Whitham's functions were computed for azimuth angles of emission normal to the Mach cone between $\Phi=0°$ and $\Phi=\pm 70°$ in 5° steps. As acceleration-induced focusing takes place shortly after takeoff, the angle of attack was chosen equal to 4°, corresponding to a heavy fuel-loaded aircraft. Whitham's functions were computed for 5 Mach numbers (1.2, 1.4, 1.6, 1.8, and 2.0). For a given Mach number along the trajectory, the computed Whitham's function with the closest Mach number was selected and rescaled using Whitham's factor (describing the dependency on Mach number for an axisymmetric body).

IV. SUPERBOOM STATISTICS

Simple ray tracing calculations were performed for all way points along the flight trajectory in order to identify those parts of the flight track that give rise to focusing in the area of the English Channel. Only for these parts comprehensive calculations were accomplished to determine all parameters of focused sonic booms. In the cases of decelerating flights from New York to Paris focusing was not found at all. For accelerating flights from Paris to New York, focusing occurred in 348 of all 365 cases. Sixteen cases had to be disregarded because of numerical problems, and in one case the software did not explicitly detect focusing. Although only 1 year was investigated, it can be concluded that sonic boom focusing during deceleration is unlikely under meteorological conditions which are encountered near the French coast. On the other hand, as expected, focusing in the area of the French coast is seemingly the normal case for accelerating supersonic flights.

The average flight data (altitude, heading and climb angles, speed, and acceleration) at the time of focusing (e.g., at the time of emission of the ray reaching the "focal" point F) are listed in Table III. Note that the mean value of the smallest Mach number leading to focusing is relatively large (1.34) compared to the critical value (Mach 1.15) for a boomless flight in the ICAO standard atmosphere (Maglieri and Plotkin, 1995). This is explained by the dominant head winds, which amplify upward refraction effects. However, the standard deviation is also relatively large. The detailed output of the sonic boom model calculations is aggregated to a set of 8 shape parameters. They are listed in Table IV. The strength of the focused sonic boom at the "focal" point F is expressed by the peak sound pressure $p_{f,\max}$ and the maximum A-weighted and C-weighted Sound Exposure Levels $L_{A,f,\max}$ and $L_{C,f,\max}$. A further parameter which is useful to assess the psychoacoustical effect of a sonic boom is the pressure risetime $t_{r,f,\min}$, i.e., the minimum time which is needed to jump from 10% to 90% of the peak overpressure. In addition, geometrical parameters of the caustic intersection with the ground are stored: the width w_f of the caustic at the "focal" point F (see Sec. III), the total curvilinear length l_f, and the lateral extension of the geometrical caustic in left-wing (port) direction $d_{f,1}$ and right-wing (starboard) direction $d_{f,2}$. For a better understanding the geometrical parameters are sketched in Fig. 3.

In order to illustrate the results of the focused boom calculations, four characteristic examples are given for flights from Paris to New York. These examples were chosen according to their meteorological characteristics: two situations with average wind condition and two situations with extreme wind conditions:

11 January 1993,00 UTC: Extremely strong head wind;

28 March 1993,00 UTC: Near-average situation;

19 October 1993,00 UTC: Near average situation;

22 October 1993,00 UTC: Extremely strong cross-wind.

Figure 4 shows the part of the aircraft ground track during which a caustic appeared, a sequence of impact lines of synchronously emitted sound rays at ground (iso-emission lines) responsible for the caustics, the intersection of the caustic at ground, and the location of the maximum overpressure at the "focal" point F. In addition, the value of maximum sound pressure $p_{f,\max}$ at this point of the caustic is indicated.

The four examples show that different meteorological conditions lead to largely different focused sonic boom characteristics. In the case of 11 January 1993 the caustic intersection with the ground is rather long ($l_f=80$ km) while the maximum sound pressure is relatively small ($p_{f,\max}=267$ Pa). The shift of the focal point away from the ground track due to strong winds is also clearly visible. On 22 Oct 2003 the intersection is relatively short ($l_f=31$ km), while the maximum pressure is high ($p_{f,\max}=878$ Pa).

All 348 cases of sonic boom focusing due to accelerating aircraft on the route Paris–New York were evaluated with respect to the location of caustic impact at ground level. The aircraft leave the French air space between Caen and Le Havre with an average heading of 319° towards the northwest. While crossing the coast they accelerate to supersonic

TABLE IV. Focused sonic boom parameters.

Number	Symbol	Description
01	$p_{f,\max}$	maximum sound pressure of focused sonic boom at "focal" point F
02	$L_{A,f,\max}$	maximum A-weighted sound exposure level of focused sonic boom at "focal" point F
03	$L_{C,f,\max}$	maximum C-weighted sound exposure level of focused sonic boom at "focal" point F
04	$t_{r,f,\min}$	minimum risetime of focused sonic boom at "focal" point F
05	w_f	width of the caustic ground intersection at "focal" point F
06	l_f	length of the caustic ground intersection
07	$d_{f,1}$	maximum port-side extension of the caustic ground intersection
08	$d_{f,2}$	maximum starboard extension of the caustic ground intersection

speed. During further acceleration, caustics occur and focused sonic booms hit the ground. The spatial distribution of the frequency of occurrence of focused sonic booms is shown in Fig. 5. The assumed initial ground track of the flights is indicated. The actual flight tracks deviate from it by up to 10 km since they depend on the respective meteorological situation. In total, an area of 7750 km^2 is affected by focused sonic booms. In most cases the ground impact of caustics is simulated over water. In 26 of the 348 cases caustics were simulated over land (near Cherbourg). The probability that a certain land area is affected by a focused sonic boom hardly exceeds 1%.

The variability of the selected superboom parameters (Table IV, Fig. 3) are displayed in Fig. 6. The maximum sound pressure $p_{f,\max}$ varies mainly between 0.2 and 0.7 kPa with a mean value of 0.41 kPa and a standard deviation of 0.14 kPa. However, in a few cases peak values of up to 1.5 kPa were calculated. Note that the $p_{f,\max}$ is typically 4 times larger than the mean value (120 Pa) of the ground track sonic boom of a westbound Mach 2 flight cruising over Western Europe with the same mock-up and the same angle of attack [BCH'04, Fig. 8]. Also noticeable is the great variability of the peak pressure, while for Mach 2 cruise the peak pressure never left the narrow interval between 108 Pa and 126 Pa. This is due both to the great variability of the caustics geometrical parameters (the radius of curvature defining the caustic thickness, Appendix, can vary between a few tens of kilometers to a few thousands of kilometers) and to the low Mach number which causes grazing propagation near the ground being highly sensitive to sharp vertical gradients of the lowest atmosphere. The spread of values of the maxi-

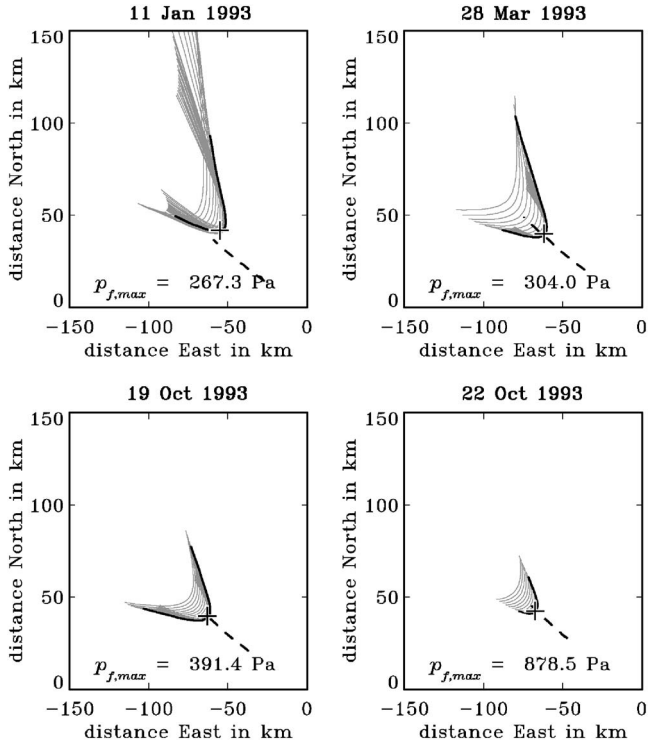


FIG. 4. Four cases of sonic boom focusing caused by acceleration during flights from Paris to New York. The grey lines indicate the synchronous impact of sound rays at the ground while the aircraft moves along the dashed line towards the northwest. The solid black line shows the position of the geometrical caustic. The cross marks the position of maximum sound pressure ($p_{f,\max}$) at "focal" point F.

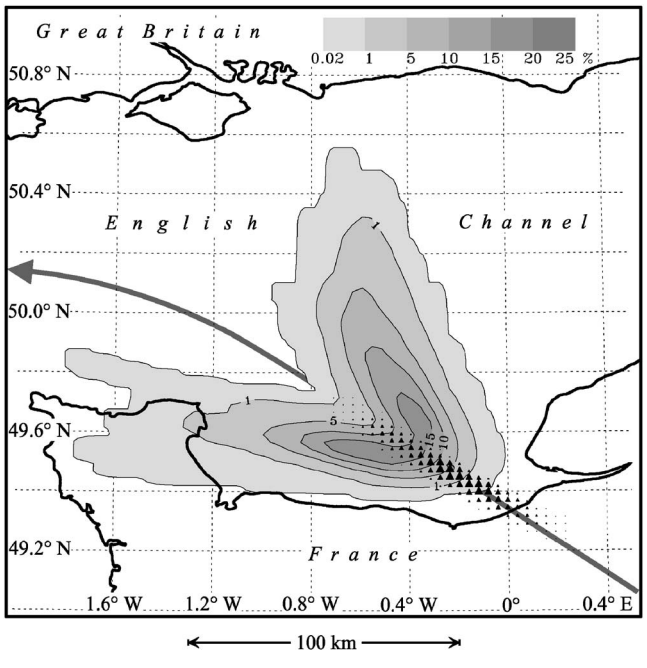


FIG. 5. Frequency distribution and location of simulated caustics (in grey levels) for accelerating flights from Paris to New York in the area of the English Channel. The thick arrow shows the assumed flight track. The triangles mark the probability (proportional to the size of the triangle) of aircraft ground position at the time of emission of the focused boom nearest to the coast (focal point). The map shows the coast lines of Great Britain and France.

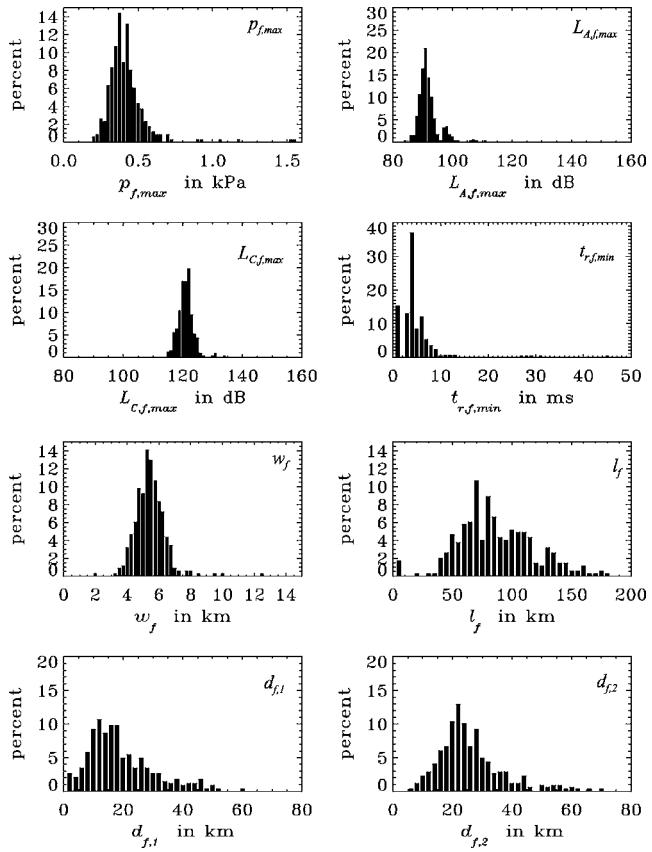


FIG. 6. Frequency distribution of simulated focused sonic-boom parameters (Table I) due to accelerating flights from Paris to New York.

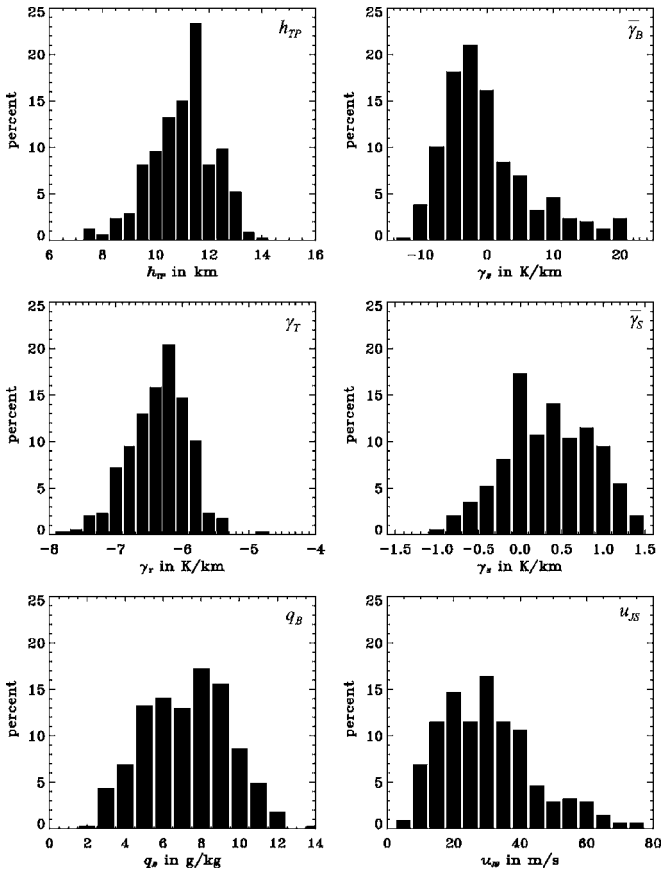


FIG. 7. Frequency distribution of flight-independent meteorological parameters (Table II) during focusing events due to accelerating flights from Paris to New York (348 cases in the year 1993).

mum A-weighted and C-weighted Sound Exposure Levels ($L_{A,f,\max}, L_{C,f,\max}$) amounts to about 10 dB. Due to the high portion of low frequencies in the focused sonic boom events, the A-weighted levels (in average 91.4 dB) are lower than the C-weighted levels (in average 120.5 dB). With few exceptions the minimum rise times $t_{r,f,\min}$ are below 10 ms with an average value of 4.3 ms, more than ten times the mean value at Mach 2 cruise. This increase is related to the "U" shape of the focused boom, which implies a longer rise time than an "N" wave. It is noted that the computed rise time does not take into account the influence of the atmospheric turbulence, which is the main mechanism responsible for the finite rise times of unfocused booms. However, the influence of turbulence on focused booms remains largely unexplored. As for the unfocused boom, a few recordings (Downing *et al.*, 1998) indicate a mean tendency to reduction, but not systematically. The width w_f of the ground caustics at "focal'" point F has a mean value of 5.3 km and a standard deviation of 1.0 km. The ground caustics are in average 84.7 km long with a standard deviation of 31.8 km. The average value of the lateral extension of the caustics in port direction $d_{f,1}$ amounts to 17.8 km (standard deviation 11.0 km). For the extension of the caustics in starboard direction $d_{f,2}$ the average value reads 24.8 km (standard deviation 10.6 km). That strong asymmetry is again due to dominant west winds, which tend to push the sound towards starboard.

V. DETERMINATION OF SPECIFIC METEOROLOGICAL INFLUENCES

In this chapter it is investigated if and to what extent the meteorological condition exerts an influence on the focused sonic boom characteristics.

Figure 7 shows how the values of the meteorological profile shape parameters (Table I) are distributed if sonic boom focusing occurs for the assumed accelerating flights from Paris to New York in 1993. The distribution of the flight-independent meteorological parameters in 1993 corresponds quite well to their long-term climatology in the region of the British Channel as a comparison with the 10-year distribution of these parameters at the nearby St. George's Channel (Fig. 2 in BCH'04) shows. Therefore it can be concluded that the year 1993 is representative of a longer climate period also in the area of the focusing.

To elaborate the dependency of the characteristics focused sonic booms on meteorological parameters, the linear correlation coefficient r_{ma} was determined for all 168 paired combinations of the 21 meteorological parameters (Tables I and II) and the 8 acoustical parameters (Table IV). The result of this correlation analysis is visualized in Fig. 8. The correlation coefficient is shown only for those combinations for which it significantly deviates from zero. The statistical significance was tested with a *t*-test according to the significance level of 0.01. In general, the flight-dependent meteorological parameters (Table II) show a stronger correlation

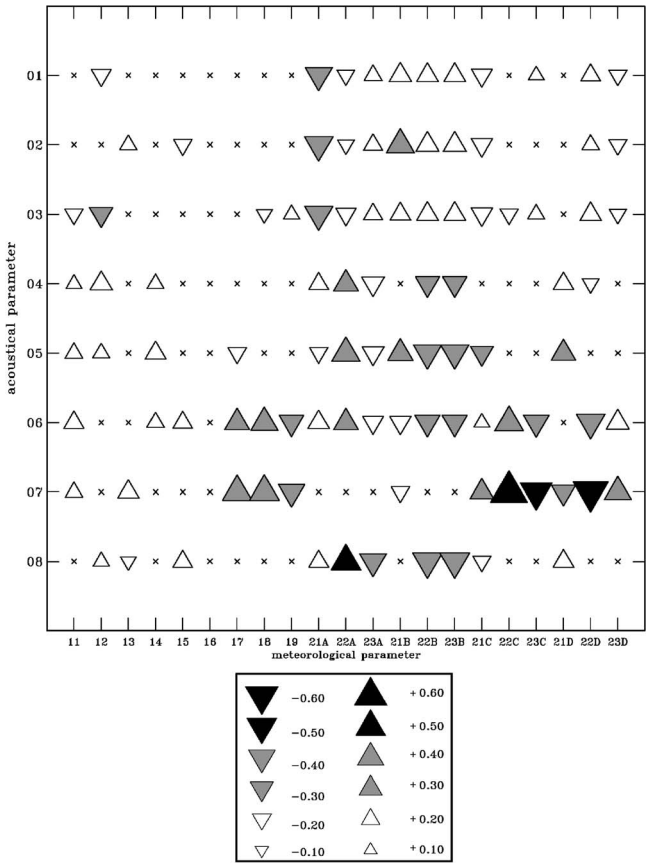


FIG. 8. Correlation coefficients r_{ma} of all combinations of meteorological and acoustical shape parameters (see Tables I, II, and IV). Crosses indicate correlation coefficients which are not significantly different from zero at a significance level of 0.01. Read text for further explanations.

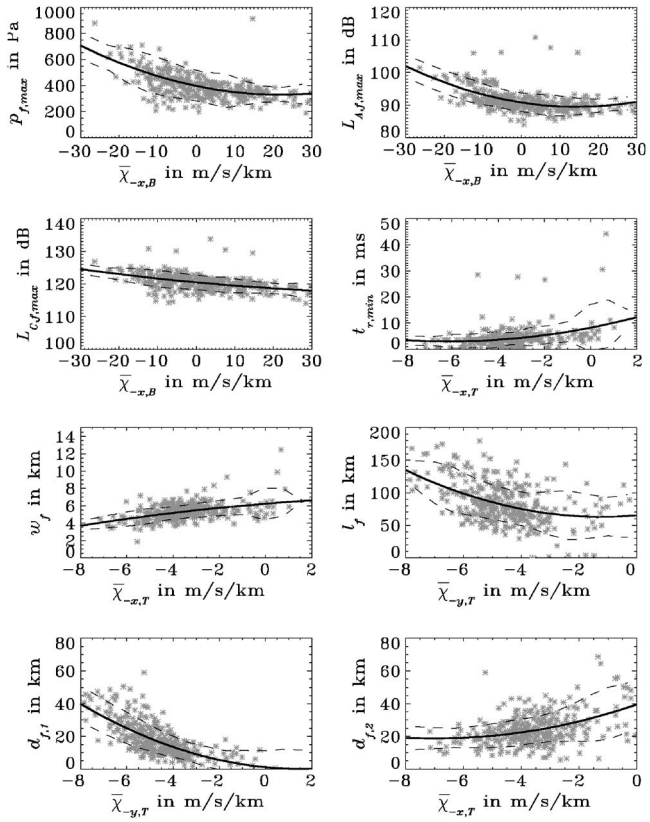


FIG. 9. Scatter diagrams of selected combinations of meteorological and acoustical shape parameters (Tables II and IV). The solid line shows the result of a second order polynomial fit. The dashed curves show the fitted values +/− the standard deviations of the simulated parameter values relative to the fitted values.

than the flight-independent parameters (Table I). This was expected because the effective speed of sound is linked more closely to the refraction of propagating sound waves. The peak sound pressure $p_{f,\max}$ correlates most of all with the vertical gradient of the effective speed of sound near the ground with respect to sound propagation into port ($\bar{\chi}_{-x,B}$; $r_{ma}=-0.41$) and starboard ($\bar{\chi}_{+x,B}$; $r_{ma}=+0.29$) directions. The maximum A-weighted and C-weighted Sound Exposure Levels depend in a similar way on the flight-independent meteorological parameters. The minimum rise time $t_{r,f,\min}$ depends more on the mean vertical gradients of the effective speed of sound in the troposphere and lower stratosphere (mainly $\bar{\chi}_{-x,T}$, $\bar{\chi}_{+x,T}$, $\bar{\chi}_{-x,S}$, and $\bar{\chi}_{+x,S}$). The geometrical parameters of the caustic intersection with the ground are generally stronger correlated with the meteorological parameters than the strength parameters do. The gradients near the ground however, play a minor role. The length of the caustic ground intersection l_f and the extension in starboard direction $d_{f,2}$ are also positively correlated with the height of the jet stream speed maximum h_{JS} and the mean vertical wind speed gradient in the troposphere $\bar{\eta}_T$. They are negatively correlated with the mean vertical wind speed gradient in the lower stratosphere $\bar{\eta}_S$. The strongest correlation between an acoustical parameter and meteorological parameters is found for the extension of the caustic ground intersection in port direction $d_{f,1}$ and the mean vertical gradient of the effective speed of sound in the troposphere and the lower stratosphere in backward and forward direction $\bar{\chi}_{-y,T}(r_{ma}=+0.76)$, $\bar{\chi}_{+y,T}(r_{ma}=-0.69)$, $\bar{\chi}_{-y,S}(r_{ma}=-0.57)$, and $\bar{\chi}_{+y,S}(r_{ma}=+0.42)$.

In order to parametrize sonic boom characteristics by meteorological quantities, a second order polynomial fit was performed for pairs of acoustical parameters and mean vertical gradients of the effective speed of sound. These pairs of parameters were selected because of their high linear correlation. The fitted curves (Fig. 9) show that some of the focused sonic boom characteristics can be estimated fairly well from a single flight-dependent meteorological shape parameter. The coefficients are listed in Table V which also includes the absolute and relative root mean square errors of the fitted curves. The maximum A- and C-weighted Sound Exposure Levels $L_{A,f,\max}$ and $L_{C,f,\max}$ can be determined from the mean vertical gradient of the effective speed of sound near the ground $\chi_{-x,B}$ with an average error of less than 3 dB. Reasonable results are also obtained for the width of the caustic at ground w_f. The absolute root mean square error amounts to less than 1 km. The minimum rise time $t_{r,f,\min}$, the length of the caustic l_f and the lateral extensions of the caustic $d_{f,1}$ and $d_{f,2}$ can be estimated only with some uncertainty using second order polynomial fits. For these sonic boom characteristics the mean relative errors range between 30% and 80%.

TABLE V. Results of least-square fit to a second-order polynomial $a=a_0+a_1m+a_2m^2$ with the meteorological parameter m as the independent variable and the acoustical parameter a as the dependent variable.

					rms error	
a	m	a_0	a_1	a_2	Absolute	Relative
$p_{f,\max}$/Pa	$\chi_{-x,B}$/ks^{-1}	369.93	−6.15	+0.14	125.8 Pa	0.307
$L_{A,f,\max}$/dB	$\chi_{-x,B}$/ks^{-1}	90.80	−0.18	+0.0063	2.93 dB	0.032
$L_{C,f,\max}$/dB	$\chi_{-x,B}$/ks^{-1}	120.50	−0.11	+0.00087	2.26 dB	0.019
$t_{r,f,\min}$/ms	$\chi_{-x,T}$/ks^{-1}	8.26	+1.72	+0.14	3.71 ms	0.788
w_f/km	$\chi_{-x,T}$/ks^{-1}	6.24	+0.22	−0.013	0.86 km	0.159
l_f/km	$\chi_{-y,T}$/ks^{-1}	65.19	+3.81	+1.56	27.7 km	0.337
$d_{f,1}$/km	$\chi_{-x,T}$/ks^{-1}	39.59	−6.02	+0.43	8.99 km	0.351
$d_{f,2}$/km	$\chi_{-y,T}$/ks^{-1}	1.04	−1.35	+0.44	7.80 km	0.676

VI. CONCLUSION

An advanced sonic boom propagation model was run in the case of unsteady supersonic flights to study the variability of focused sonic booms at the ground level. The flights were assumed to serve the route from Paris to New York and back. Calculations were performed for 365 real meteorological conditions that appeared during one year and for realistic flight configurations and trajectories based on former Concorde procedures and estimated performances of a future supersonic commercial aircraft. Under the selected meteorological conditions, focusing occurred in the area of the English Channel during all cases of accelerated flights leaving the European continent towards New York. In the opposite, focusing never occurred for decelerating flights inbound to Paris. Therefore, the detailed focused sonic boom calculations were limited to the cases of accelerated westbound flights.

The results of the simulation revealed a high variability of the superboom characteristics due to the varying weather conditions that were encountered during 1 year. The mean Mach number associated with ground track focusing is 1.34. An average maximum sound pressure of 0.41 kPa and a standard deviation of 0.14 kPa are computed. This is about 4 times more than for cruising flight at Mach 2. In few cases peak sound pressure values of even more than 1 kPa and up to 1.5 kPa are simulated. The variability of the focused sonic boom turns out to be dramatically larger than that of a Mach 2 cruising-flight boom. The respective average Sound Exposure Levels amount to 91.4 dB (A-weighted) and 120.5 dB (C-weighted), both showing a variability of about 10 dB. The pressure rise time is normally below 10 milliseconds. The average width of the caustics comes up to 5.3 km with only a small standard deviation of 1.0 km. The length of the caustic ground intersection appears to vary over a rather large range (approx. 40–180 km) with an average value of 85 km. In general, because of dominant cross winds, the caustics extend farther in the starboard (right wing) direction (average 24.8 km, standard deviation 10.6 km) than in the port (left wing) direction (average 17.8 km, standard deviation 11.0 km).

For the given flight trajectories, the area for which at least one focused sonic boom impact at ground level is calculated covers 7750 km^2. Inside an area of 660 km^2 focused sonic boom events occurs in more than 15% of the cases. In 7.5% of the cases the caustics touches land areas. These landfall cases could be avoided by shifting the flight trajectories by about 30 km to the northeast.

The study shows that sonic boom focusing in acceleration phase remains a key barrier to be overcome for a future supersonic fleet. In the case of a commercial aircraft flying supersonically only overseas, future routes will have to be optimized precisely to avoid overland focusing when the aircraft leaves the coast. Concorde exploitation and the present study shows that this would be possible, but a specific analysis depending on local geography and climatology should be carried out for each operated route. Cases of complex coast lines with numerous islands near main transcontinental lines (Baltic Sea, English Channel, Mediterranean Sea in Europe, Caribbean Sea in North America, all East Asia Coasts from Singapore to Japan, etc.) should be addressed with high priority. For supersonic business jets intended for overland supersonic flight, the situation is more critical. Low boom designs may enable cruise boom amplitudes below an "acceptable" level, if such a level can be defined. However, given the boom amplification (here by a factor 4 in the mean, but 15 at maximum) and its great sensitivity to meteorology, it remains unlikely that also superbooms can be kept below such an "acceptable" level. Therefore, supersonic acceleration may have to be restricted to "superboom corridors" over very low populated areas, and carefully defined procedures to maneuver in these corridors may have to be prescribed.

Finally, we point out that the present study has been restricted to a commercial supersonic transport with conventional design, giving rise to a relatively large boom. Of course, for a quantitative investigation, it should be repeated for other aircraft configurations. For instance, for a supersonic business jet with unconventional low-boom design, we may expect pressure levels to be largely reduced (assuming that unconventional low-boom design for cruising speed is efficient also at lower acceleration Mach numbers!). Nevertheless, it is very unlikely that *meteorologically-induced* variability of focused boom may be significantly reduced by changes of the aircraft design. Therefore, whatever the aircraft shape, we expect that some meteorological situations may lead to much more intense superbooms than under standard-atmosphere conditions.

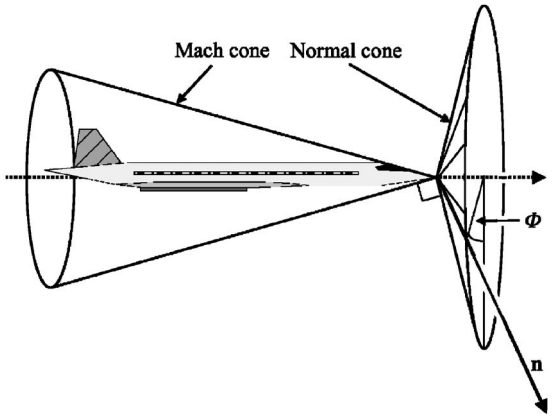


FIG. 10. The azimuthal angle Φ determines the orientation of an acoustical ray on the cone normal to the Mach cone of half apex angle $\sin\theta = 1/M$ with M the Mach number at the time of emission.

ACKNOWLEDGMENTS

The study was carried out as part of the project "Sonic Boom European Research Programme: Numerical and Laboratory-Scale Experimental Simulation" (SOBER) which was cofunded by the European Commission under Contract No. G4RD-CT-2000-00398. The authors are obliged to Dr. Klaus-Peter Hoinka (DLR) who extracted the ERA-15 model level data from ECMWF within the Special Project "The Climatology of the Global Tropopause." Dr. Günther Zängl (University of Munich) is thanked for providing his Fortran code to determine the thermal tropopause height. Roland Etchevest (Airbus France S.A.S., Toulouse) and Stephane Illa (Transiciel Technologies, Toulouse) patiently helped us to implement the sonic boom propagation code at DLR. The contributions of Joseph Carla and Benjamin Vandenbroucke (Airbus France S.A.S.) providing input flight trajectories and Whitham functions were also highly appreciated. The progresses made by Régis Marchiano (Université Pierre et Marie Curie, Paris) to improve the performance of the numerical solver of the Nonlinear Tricomi Equation were also very useful in minimizing the computation costs of the study.

APPENDIX

Ray tracing in a stratified atmosphere is performed by standard numerical procedures for solving differential systems of equations. An individual ray is classically parameterized by the emission time t determining the position of the aircraft along its trajectory and the associated flight parameters, and the azimuthal angle Φ determining the launch direction of the ray on the cone normal to the Mach cone, Fig. 10 (Hayes *et al.*, 1969). The two parameters (t,Φ) are the ray coordinates. If parameterized by the eikonal function Ψ (the propagation time along a given ray), rays $\mathbf{x}(t,\Phi,\Psi)$, wave-front normal $\mathbf{n}(t,\Phi,\Psi)$ and infinitesimal algebraic ray-tube areas

$$A(t,\Phi,\Psi) = \left(\frac{\partial \mathbf{x}(t,\Phi,\Psi)}{\partial t} \times \frac{\partial \mathbf{x}(t,\Phi,\Psi)}{\partial \Phi} \right) \cdot \mathbf{n} \tag{A1}$$

are computed until the rays touch the ground at $z=0$ for the value $\Psi^G(t,\Phi)$. The ground ray tube algebraic area $A^G(t,\Phi)=A(t,\Phi,\Psi^G(t,\Phi))$ is therefore a function of the two ray coordinates (t,Φ).

Caustics are defined as the locus of points where the ray tube algebraic area vanishes, i.e., $A(t,\Phi,\Psi)=0$, so that the geometrical approximation breaks down there. This enables to define the eikonal function $\Psi_C(t,\Phi)$ at points on the caustic surface as a function of the ray coordinates, such that

$$A(t,\Phi,\Psi_C(t,\Phi)) = 0. \tag{A2}$$

Caustic points are given by

$$\mathbf{x}_C(t,\Phi) = \mathbf{x}(t,\Phi,\Psi_C(t,\Phi)). \tag{A3}$$

The ground intersection with the caustics is the locus of points where $A^G(t,\Phi)=0$. For a fixed emission time, the ground position of caustics is determined by searching for discrete intervals $[\Phi_i,\Phi_{i+1}]$ where the sign of the ray tube algebraic area changes. Then, a quadratic interpolation of the ray tube area is performed to approximate the azimuthal angle $\Phi_C^G(t)$ of the ray launched at emission time t and that is tangential to the caustic at the ground level. The corresponding coordinates $\mathbf{x}_C(t,\Phi_C^G(t))$ provide the ground position of the caustic. For an accelerating aircraft and for each emission time, there can be 0 (no caustic), 1 (one on one side of the caustic) or 2 (one on each side of the trajectory) emission angles $\Phi_C^G(t)$ depending on flight parameters and local meteorology (Fig. 3). The same procedure can be performed by searching, for a fixed emission angle, the emission time $t_C^G(\Phi)$ of the ray that will be tangent to the caustic at the ground level. This procedure is implemented only for the zero azimuthal angle $\Phi=0$, and determines the "focal" point $F(x_F,y_F)=\mathbf{x}_C(t_C^G(0),0)$. For an acceleration-caused caustic with a typical crescent shape, that point is practically the closest to the sound-barrier breaking point. The highest boom overpressure all along the flight path will be recorded in the close vicinity of that point.

To determine the input parameters of the nonlinear Tricomi equation, it is necessary to compute elements of the caustic local geometry, e.g., the caustic unit normal vector $\mathbf{N}^C$ and the radius of curvature R^C of the intersection of the caustic with the plane $(\mathbf{N}^C,\mathbf{t}^C)$ where $\mathbf{t}^C$ is the unit tangent vector to the ray that is tangential to the caustic at the point of interest. The unit tangent vector is defined as

$$\mathbf{N}^C = \pm \frac{\dfrac{\partial \mathbf{x}_C}{\partial t} \times \dfrac{\partial \mathbf{x}_C}{\partial \Phi}}{\left| \dfrac{\partial \mathbf{x}_C}{\partial t} \times \dfrac{\partial \mathbf{x}_C}{\partial \Phi} \right|}. \tag{A4}$$

Differentiating Eqs. (A1)–(A3) with respect to t and Φ shows that the unit normal vector $\mathbf{N}^C$ can be expressed as a function of the first and second derivatives of the position of a ray point $\mathbf{x}(t,\Phi,\Psi)$ with respect to t and Φ. Generalizing the method of Candel (1977), these derivatives can be obtained by differentiating the 6 differential ray equations $(\partial\mathbf{x}/\partial\Psi, \partial\mathbf{n}/\partial\Psi)$ governing position and the wave-front normal vector along a single ray (Pierce 1989, Auger 2001). This leads to a system of 36 coupled differential equations. Finally, the set of unit orthogonal vectors $(\mathbf{N}^C,\mathbf{t}^C)$ is com-

pleted by the unit vector $\mathbf{e}^C$ to form a local orthonormal basis. This allows us to introduce a local parametrization $\mathbf{x}^C(\sigma,\lambda)$ of the caustic surface by the two orthogonal curvilinear coordinates (σ,λ) (Babič and Buldyrev, 1991) such that

$$\frac{\partial \mathbf{x}_C}{\partial \sigma} = \mathbf{t}_C, \qquad \frac{\partial \mathbf{x}_C}{\partial \lambda} = \mathbf{e}_C. \tag{A5}$$

The radius of curvature R^C can be shown to be (Auger, 2001):

$$R^C = -\left(\mathbf{N}^C \cdot \frac{\partial \mathbf{t}^C}{\partial \sigma}\right)^{-1} \tag{A6}$$

and the derivative $\partial \mathbf{t}^C / \partial \sigma$ can be expressed also as a function of the first and second derivatives of the position of a ray point $\mathbf{x}(t,\Phi,\Psi)$ with respect to t and Φ. That method turns out to be more efficient than using the geometrical definition of R^C, which would imply computing the third derivatives of a ray point with respect to the two ray coordinates (Plotkin, 1995), and thus to solve a differential system of 60 equations.

Finally, once the geometrical parameters are known, the aerodynamic pressure field is propagated in the atmosphere along the ray tangent to the caustic at a selected point, including nonlinear distortion. The computation is stopped at some distance from the caustic and the input parameters for the NTE are deduced accordingly. In particular, the thickness of the diffraction boundary layer around the geometrical caustic is deduced according to its definition

$$\delta = (2c_0^2 T^2 R)^{-1/3}; \tag{A7}$$

where $R = (1/R^C + 1/R^{\text{ray}})^{-1}$, R^{ray} is the radius of curvature of the ray at the point where it is tangential to the caustic and T is the duration of the incoming sonic boom. The width of the ground caustic is therefore equal to $w_f = 2\delta/\sin\Omega$ with Ω the angle of the caustic with the horizontal ground (Fig. 3).

Auger, Th. (2001). "Modélisation et simulation numérique de la focalisation d'ondes de choc acoustiques en milieu en mouvement. Application à la focalisation du bang sonique en accélération," thèse de doctorat de l'Université Pierre et Marie Curie (Paris 6) (in French).

Auger, Th. and Coulouvrat, F. (**2002**). "Numerical simulation of sonic boom focusing," AIAA J. **40**, 1726–1734.

Babič, V. M. and Buldyrev, V. S. (**1991**). *Short-Wavelength Diffraction Theory. Asymptotic Methods* (Springer-Verlag, Berlin).

Barnston, A. G. and Livezey, R. E. (**1987**). "Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns," Mon. Weather Rev. **115**, 1083–1126.

Berry, M. V. (**1976**). "Waves and Thom's theorem," Adv. Phys. **25**, 1–26.

Blumrich, R., Coulouvrat, F., and Heimann, D. (**2005**). "Meteorologically induced variability of sonic boom characteristics of supersonic aircraft in cruising flight," J. Acoust. Soc. Am. **118**, 687–702.

Boulanger, P. and Attenborough, K. (**2005**). "Effective impedance spectra for predicting rough dea effects on atmospheric impulsive sounds," J. Acoust. Soc. Am. **117**, 751–762.

Candel, S. (**1977**). "Numerical solution of conservation equation arising in linear wave theory: application to aeroacoustics," J. Fluid Mech. **83**, 465–493.

Downing, M., Zamot, N., Moss, C., Morin, D., Wolski, E., Chung, S., Plotkin, K. J., and Maglieri, D. (**1998**). "Controlled focused sonic booms from manoeuvreing aircraft," J. Acoust. Soc. Am. **104**, 112–121.

Esclangon, E. (**1925**). *L'acoustique des Canons et des Projectiles* (Imprimerie Nationale, Paris) (in French).

Gibson, R., Kallberg, P., Uppala, S., Hernandez, A., Nomura, A., and Serrano, E. (1997). "ERA description," ECMWF ReAnalysis Project Report Series 1, 72 pp. Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, U.K.

Gill, P. M. and Seebass, A. R. (1973). "Nonlinear acoustic behavior at a caustic: An approximate analytical solution," AIAA Paper No. 73-1037.

Guiraud, J.-P. (**1965**). "Acoustique géométrique, bruit balistique des avions supersoniques et focalisation (Geometrical acoustics, balistic noise of supersonic aircraft and focusing)," J. Mec. **4**, 215–267 (in French).

Hayes, W. D., Haefeli, R. C., and Kulsrud, H. E. (1969). "Sonic boom propagation in a stratified atmosphere with computer programme," NASA CR-1299.

Maglieri, D. J. and Plotkin, K. J. (**1995**). "Sonic boom," in *Aeroacoustics of Flight Vehicles*, edited by H. H. Hubbard (Acoustical Society of America, Woodbury), Vol. **1**, pp. 519–561.

Marchiano, R., Coulouvrat, F., and Grenon, R. (**2003**). "Numerical simulation of shock waves focusing at fold caustics, with application to sonic boom," J. Acoust. Soc. Am. **114**, 1758–1771.

Marchiano, R., Thomas, J.-L., and Coulouvrat, F. (**2003**). "Experimental simulation of supersonic superboom in a water tank: Nonlinear focusing of weak shock waves at a fold caustic," Phys. Rev. Lett. **91**(18), 184301 (1–4).

Pierce, A. D. (**1989**). *Acoustics, An Introduction to its Physical Principles and Applications*, 1st ed. in 1981 (Acoustical Society of America, New York).

Plotkin, K. J. (**1995**). "The theoretical and computational basis of focused sonic booms," J. Acoust. Soc. Am. **97**, 3257–3258 (129th Meeting of the Acoustical Society of America, 1pPA6).

Plotkin, K. J. (**2002**). "State of the art of sonic boom modelling," J. Acoust. Soc. Am. **111**, 530–536.

Plotkin, K. J. and Page, J. A. (2002). "Extrapolation of sonic boom signatures from CFD solutions," AIAA Paper No. 2002-0922.

Rendón Garrido, P. L. and Coulouvrat, F. (2005). "Nonlinear ground reflection of caustics and focused sonic booms," Wave Motion (in press).

Sanai, M., Toong, T. Y., and Pierce, A. D. (**1976**). "Ballistic range experiments on superboom generated at increasing flight Mach numbers," J. Acoust. Soc. Am. **59**, 520–524.

Seebass, A. R. (1971). "Nonlinear behavior at a caustic," Boeing Scientific Research Laboratories Document 01-82-1039.

Walkden, F. (**1958**). "The shock pattern of a wing-body combination, far from the flight path," Aeronaut. Q. **IX**, 164–194.

Wanner, J.-C., Vallée, J., Vivier, C., and Théry, C. (**1972**). "Theoretical and experimental studies of the focus of sonic booms," J. Acoust. Soc. Am. **52**, 1–32.

Whitham, G. B., (**1952**). "The flow pattern of a supersonic projectile," Commun. Pure Appl. Math. **5**, 301–348.

WMO (**1957**). "Meteorology—A three-dimensional science: Second session of the commission for aerology," WMO Bull. **IV**(4), 134–138.

Zängl, G. and Hoinka, K. P. (**2000**). "The tropopause in the polar regions," J. Clim. **14**, 3117–3139.

Meteorologically induced variability of sonic-boom characteristics of supersonic aircraft in cruising flight

Reinhard Blumrich[a)]
Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, 82234 Weßling, Germany

François Coulouvrat
Laboratoire de Modélisation en Mécanique, Université Pierre et Marie Curie & CNRS (UMR 7607), 4 place Jussieu, 75252 Paris cedex 05, France

Dietrich Heimann[b)]
Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, 82234 Weßling, Germany

(Received 10 November 2004; revised 18 May 2005; accepted 23 May 2005)

The influence of the meteorological variability on the characteristics of the primary sonic boom emerging from an aircraft in cruising flight is investigated. The sonic-boom propagation is calculated by means of an advanced ray-tracing algorithm which takes meteorological influences into account. Real meteorological situations are considered based on a full 10-year data set in 12- and/or 24-h resolution. Three different climate regions are studied: a mid-latitude coastal sea region, a tropical coastal sea area, and a subpolar land region. Frequency distributions of sonic-boom characteristics such as wave amplitude, rise time, and carpet width are shown for each area, all seasons, and opposing flight directions. It turns out that while variability is low at the ground track, it is high laterally for carpet width or boom amplitude at the outer carpet edges. A correlation analysis is applied which shows specific relationships between meteorological profile parameters and acoustical response. In addition, a meteorological classification is introduced and tested.
 [DOI: 10.1121/1.1953208]

PACS number(s): 43.28.Mw, 43.28.Lv, 43.28.Fp, 43.28.Js [LCS] Pages: 707–722

I. INTRODUCTION

Supersonic air transport is accompanied by several environmental nuisances such as noise or air pollution. A major problem, and the only one specific to supersonic speed, is the sonic boom associated with long-range evolution of the shock waves generated by the aerodynamic flow around the aircraft (Maglieri and Plotkin, 1995; Plotkin, 2002). The geometrical and physical characteristics of the sonic booms (such as the position and width of the sonic-boom carpet at ground, the peak sound pressure, and the rise time) depend on both the source and the atmospheric propagation. While the source is controlled by the aircraft shape and the flight parameters (for cruising flight: Mach number, altitude, angle of attack, and heading), the propagation is strongly influenced by refraction, nonlinear distortion, and sound absorption which depend on atmospheric parameters (temperature, density, wind, and humidity) along the propagation path.

The key role of meteorological parameters on sonic boom refraction was investigated as early as World War I for sonic booms produced by gun shells (Esclangon, 1925). The progressive nonlinear distortion of the pressure waveform until the ultimate "N" wave shape has been recognized as a key effect by Whitham (1956) in a uniform atmosphere, then generalized by Guiraud (1965) to a moving, nonuniform medium and numerically implemented by Hayes *et al.* (1969). Atmospheric absorption has been outlined as a main effect on sonic-boom rise time by Hodgson (1973) and implemented in different numerical procedures (Cleveland *et al.*, 1996). It is caused primarily by the vibrational relaxation of diatomic nitrogen and oxygen molecules and strongly influenced by humidity (Bass *et al.*, 1984).

Because atmospheric parameters vary in time and space, sonic-boom characteristics also vary widely. Moreover, the local long-term statistics of sonic boom characteristics depend on the climate, i.e., the frequency distribution of the relevant meteorological parameters and their vertical gradients. For an environmentally acceptable supersonic air transport it is necessary to estimate the size of the area affected by sonic booms and their strength beneath potential flight tracks as a function of weather conditions. As a benefit flight tracks can be defined such that unwanted impact on sensitive ground areas can be avoided depending on the actual state of the atmosphere.

The long-term variability of sonic-boom characteristics in a certain region would be best determined with the aid of long-term measurements. Though thousands of sonic-boom measurements are available (for a review see Maglieri and Plotkin, 1995), including recordings of sonic-boom variability near the cutoff (Hubbard *et al.*, 1971; Haglund and Kane, 1974), no long-term monitoring of primary booms from regularly occurring supersonic cruise operations has ever

[a)]Present address: Forschungsinstitut für Kraftfahrwesen und Fahrzeugmotoren (FKFS), Pfaffenwaldring 12, 70569 Stuttgart, Germany.
[b)]Author to whom correspondence should be addressed.

been performed, as this would have required a permanent offshore measuring system beneath Concorde flight path. Such long-term data records exist only for ground recorded secondary booms (Le Pichon *et al.*, 2002). Therefore, long-term variability has to be estimated by numerical simulations of sonic-boom propagation. Lundberg (1994) made a first step in this direction by studying seasonal variability of an F111 sonic boom (Mach 1.3), but using only (five) seasonal averages of atmospheric data near Edwards Air Force Base. The present study aims at systematizing this approach to investigate the propagation of primary sonic booms emerging from high-flying aircraft in cruising flight conditions. It focuses especially on a statistical analysis of the meteorologically induced variability of the sonic-boom characteristics on the ground. Here the study targets primarily application to commercial aircraft, as military aircraft in the U.S.A. or in Europe normally fly supersonic training missions only over restricted areas. Nevertheless, the present results would also be applicable to military fighters with a view to enlarge these areas.

A previous study (Heimann, 2001) applied a simple ray-tracing model to daily meteorological profiles over a 12-year period above the British-Atlantic coast to simulate the sonic boom of an aircraft flying 15 km high at infinite Mach number, i.e., (2D) cylindrical propagation from the straight flight path. The total width of the sonic-boom carpet due to the variability of the atmosphere ranged between 70 and more than 200 km. The present study generalizes this preliminary approach to realistic configurations (3D-propagation at finite cruising Mach number in a vertically stratified atmosphere) and extends the statistical analysis beyond geometrical parameters (carpet width) to physical sonic-boom characteristics such as peak sound pressure, rise time (defined as the time interval necessary for the head shock pressure to jump from 10% to 90% of the maximum overpressure), and carpet width above a prescribed sound level. The numerical sonic-boom propagation model is applied to meteorological analysis data (temperature, wind, humidity; see Sec. II). Because meteorological parameters are subject to daily and seasonal variability, a long time-series with high temporal resolution is required to obtain reliable statistics of sonic-boom behavior at a particular location.

We wish to investigate various geographically diverse locations for which meteorological statistics are required. In practice, the high computational demand of repeated simulations confines the number of meteorological profiles for which the sonic-boom behavior can be determined. Nevertheless, it was possible to determine sonic-boom statistics for a 10-year period and for three different climates based upon simulations at 12- and 24-h intervals. The choice of the three target areas along potential future supersonic routes is presented in Sec. II. In total, 29 200 calculations of sonic-boom ground distributions were performed, each with a mean number of 36 ground "impact" points, thus providing more than one million sonic-boom pressure waveforms to be numerically evaluated. To achieve this, an efficient algorithm was required. This is briefly described in Sec. III. Section IV presents the statistical distribution of the main acoustical parameters for the different headings and target areas. Section V investigates their correlation to specific meteorological parameters. For future applications, e.g., the determination of commercial routes and the real-time adjustment of supersonic aircraft operations, the high computational costs require a further reduction of the simulation effort. Therefore, a new classification scheme of meteorological situations was tested in Sec. VI with the aim of restricting propagation simulations to a relatively small number of vertical meteorological profiles which are representative of meteorological categories. Different climates are then characterized by different frequency distributions of the categories.

TABLE I. Selected target areas including a brief topographical description.

Target area	Location	Longitude/ latitude	Climate zone	Surface
TA1	St. George's Channel	6.750° W 50.625° N	mid-latitudes	sea
TA2	Coast of Vietnam	106.875° E 9.000° N	tropics	sea
TA3	Mackenzie/ Canada	115.875° W 60.750° N	subpolar	land

II. TARGET AREAS AND METEOROLOGICAL DATA BASE

The present investigation focuses on three target areas which have been selected in view of their relevance regarding supersonic transport, along routes of expected commercial importance, and in order to cover different climate situations. Two target areas are sea areas close to densely populated coasts (TA1: St.George's Channel between England and Ireland, TA2: Vietnamese coastal sea) while one is a land area (TA3: Mackenzie/Canada). They represent mid-latitude (TA1), tropical (TA2), and subpolar (TA3) way points along flight routes between major population agglomerations. The target areas are specified in Table I. The location of the target areas are shown in Fig. 1. For all target areas, two opposing flight directions, corresponding to the great circle between relevant destinations, and appropriate cruising altitudes and angles of attack were considered. They are specified in Table II. The flight directions are also indicated in Fig. 1. The air speed was assumed to be Mach 2 for all target areas and flight directions.

Meteorological data were taken from the ERA15 reanalysis database of the European Centre for Medium Range Weather Forecast (ECMWF; Gibson *et al.*, 1997). The ERA15 analyis data are based on observations but also fulfill

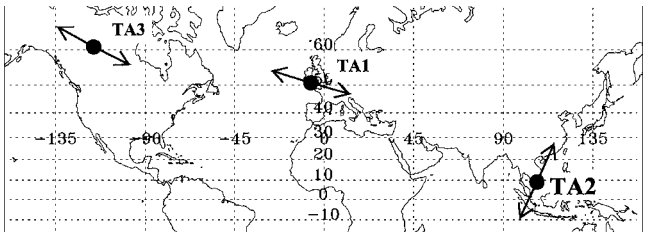


FIG. 1. Location of the target areas (see Table I). The arrows indicate the two opposing flight directions (Table II) at each target area. The numbers indicate longitude and latitude in degree.

TABLE II. Assumed flight parameters for each target area.

Target area	Nearest great circle route	Flight direction	Cruising altitude (m)	Angle of attack (deg)
TA1	New York—Paris	106.5° (eastbound)	19 812	2.5
	Paris—New York	286.5° (westbound)	18 318	4
TA2	Hong Kong—Singapore	205.0° (southbound)	19 812	2.5
	Singapore—Hong Kong	25.0° (northbound)	19 644	4
TA3	Tokyo—Detroit	117.7° (eastbound)	18 980	2.5
	Detroit—Tokyo	297.7° (westbound)	17 511	2.5

the mass, momentum, and energy conservation equations of the atmosphere through a four-dimensional spatio-temporal data assimilation procedure. The global data set covers a 15-year period (1979–1993) in time intervals of 6 h with a horizontal resolution of 1.125°. Vertically, the data refer to 30 levels between the ground and an altitude of 31 km. The vertical resolution varies between approximately 100 m near the ground and approximately 4000 m near the top level. The vertical profiles at the grid points nearest to the target areas were extracted from the database.

The data comprise air pressure p_{air}, density ρ, horizontal wind speed V, and direction δ (vertical wind is not taken into account), temperature T, and specific humidity q (that can be converted into mole ratio or, more commonly, relative humidity quite easily). These parameters are sufficient to determine refraction, nonlinear distortion, and absorption of sound waves. Air absorption is controlled by temperature T, (specific) humidity q, and atmospheric pressure for higher frequencies and higher altitudes.

This study refers to a 10-year subset (1984–1993) of 24-h intervals (00 UTC; TA1 and TA2) or 12-h intervals (00 and 12 UTC; TA3) analyses. The evaluation either refers to seasons (winter: December, January, February; spring: March, April, May; summer: June, July, August; autumn: September, October, November) or it refers to the full year.

In order to reduce the dimensionality of the problem and for classification purposes a limited number of meteorological profile shape parameters m were defined. In the following M denotes the entirety of meteorological shape parameters with $m \in M$. The shape parameters characterize the atmospheric stratification in an acoustically relevant sense, because they refer to the primary gradients and the depth of the layer in which they are effective for refraction.

A first set of meteorological shape parameters (Table III) characterizes the vertical structure of temperature and wind speed in the layer relevant for the propagation of sonic-boom waveforms. In contrast to the second set, this set is independent of flight parameters. The temperature profile was divided into three layers and is represented by the mean vertical temperature gradient within these layers: "boundary layer" $\bar{\gamma}_B$, "troposphere" $\bar{\gamma}_T$, and "lower stratosphere" $\bar{\gamma}_S$. The first one, $\bar{\gamma}_B$, always refers to the lowest layer of the ERA15 grid (roughly the lowest 300 m), independent of the actual height of the atmospheric boundary layer. This simplification was used because the vertical resolution of the ERA15 data does not admit the specification of the actual boundary-layer height. The tropopause height h_{TP}, which separates troposphere and stratosphere, is defined in the sense of the "thermal tropopause" (significant change of the vertical temperature gradient) following WMO (1957). The algorithm used to determine h_{TP} is described in Zängl and Hoinka (2000). The mean vertical gradients $\bar{\gamma}_T$ and $\bar{\gamma}_S$ refer to the layers $[300\ \mathrm{m}, h_{TP}]$ and $[h_{TP}, 25\ \mathrm{km}]$, respectively. The vertical wind profile was divided into two layers separated by the height h_{JS} of the "jet stream" wind speed maximum u_{JS}. The mean gradients of temperature and wind speed in the respective layers were determined by linear regression. An additional parameter is the specific humidity near the ground (q_B) as it influences molecular air absorption. We preferred specific humidity because relative humidity is strongly correlated with the temperature near the ground and thus the vertical temperature gradient $\bar{\gamma}_B$.

Figures 2–5 show frequency distributions of the meteorological shape parameters (Table III) at all three target areas. For TA1 and TA2 the evaluation was based on the 00 UTC analyses (Figs. 2 and 3). For the land area TA3, where diurnal variations play a role, frequency distributions are shown separately for 00 UTC=16 Local Time (Fig. 4) and 12 UTC=04 Local Time (Fig. 5). The corresponding values of the International Civil Aviation Organization (ICAO) standard atmosphere (ICAO, 1993) are also marked in these figures. The ICAO standard atmosphere defines the temperature profile, but it does not consider wind.

According to these data the actual height of the tropopause h_{TP} varies between 7 and 14 km at the mid and high latitude targets TA1 and TA3 ($h_{TP,ICAO}$=11 km). In the tropics (TA2) the tropopause is significantly higher and varies within a rather small range. Generally, the tropical troposphere (TA2) deviates significantly from the standard atmosphere. The mean vertical temperature gradient in the bound-

TABLE III. Meteorological profile shape parameters (set 1).

m	Symbol	Description
11	h_{TP}	height of the thermal tropopause in km
12	$\bar{\gamma}_B$	mean vertical temperature gradient near the ground in K/km
13	$\bar{\gamma}_T$	mean vertical temperature gradient in the troposphere in K/km
14	$\bar{\gamma}_S$	mean vertical temperature gradient in the stratosphere in K/km
15	q_B	specific humidity near the ground in g/kg
16	h_{JS}	height of the jet stream wind speed maximum km
17	u_{JS}	wind speed maximum in m/s
18	$\bar{\eta}_1$	mean vertical wind speed gradient below h_{JS} in 1/s
19	$\bar{\eta}_2$	mean vertical wind speed gradient above h_{JS} in 1/s

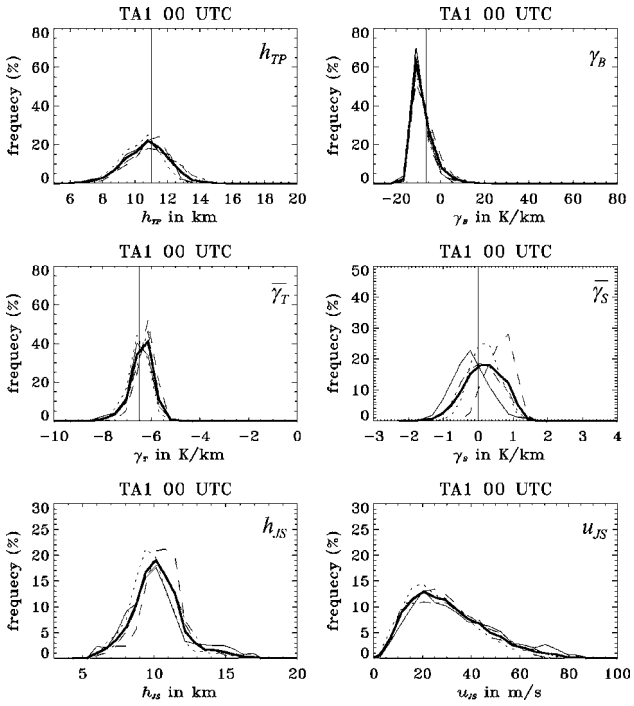


FIG. 2. Frequency distributions at target area TA1 (00 UTC) of the tropopause height (h_{TP}) in km, the mean vertical temperature gradients near the ground ($\overline{\gamma}_B$), in the troposphere ($\overline{\gamma}_T$), and in the stratosphere ($\overline{\gamma}_S$) in K/km, the height of the jet stream wind maximum (h_{JS}) in km, and the maximum jet stream wind speed (u_{JS}) in m/s (see also Table III). The thick solid line refers to the full year. The thin lines refer to single seasons (winter: solid; spring: dots; summer: dashes; autumn: dash-dots). The vertical line indicates the ICAO standard atmosphere (only for the tropopause height and temperature gradients).

ary layer $\overline{\gamma}_B$ varies most over land (TA3) because of the pronounced diurnal cycle of surface heating (day) and cooling (night). Over sea (TA1 and TA2) the heat capacity of the ocean water permits only small diurnal temperature variations such that the typical adiabatic temperature gradient ($\overline{\gamma}_B \approx -10$ K/km) of a neutrally stratified, well-mixed boundary layer occurs most frequently. Within the troposphere the mean vertical temperature gradient $\overline{\gamma}_T$ is generally negative. It varies most at TA3 and least at TA2. For TA1 and TA3 the median is close to the corresponding value of the ICAO standard atmosphere ($\gamma_{T,ICAO} = -6.5$ K/km). The mean vertical temperature gradient of the lower stratosphere $\overline{\gamma}_S$ varies between −1 and +1 K/km (ICAO standard atmo-

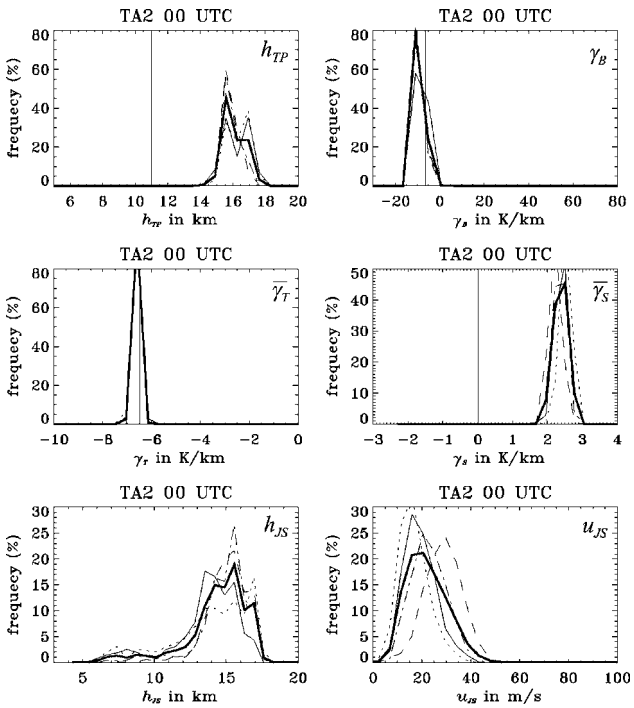


FIG. 3. Same as Fig. 2 but for TA2 (00 UTC).

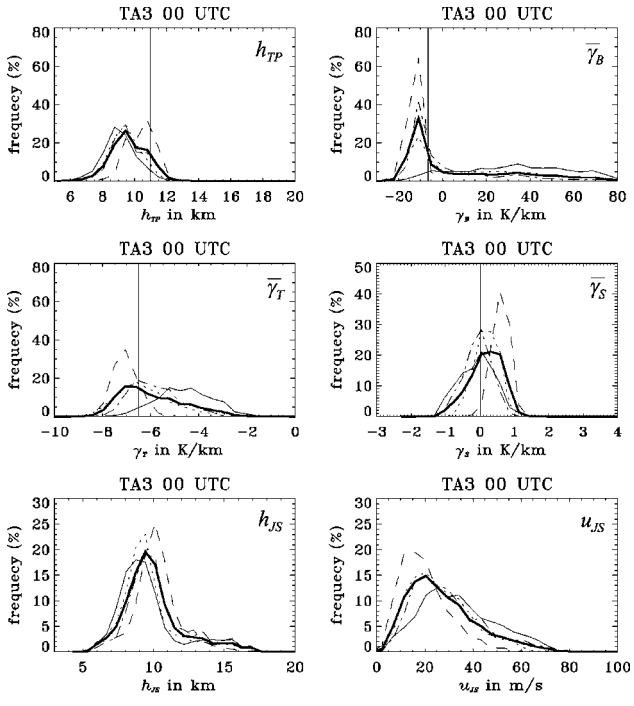


FIG. 4. Same as Fig. 2 but for TA3 (00 UTC).

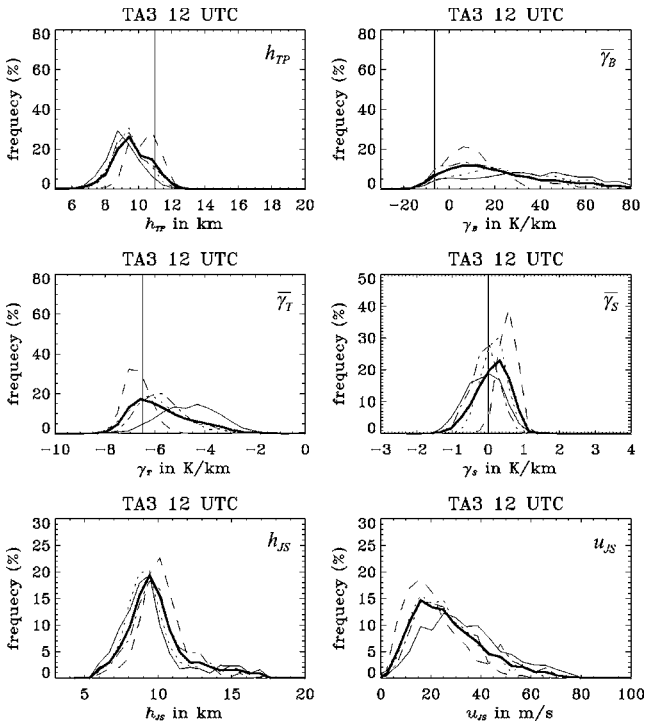


FIG. 5. Same as Fig. 2 but for TA3 (12 UTC).

TABLE IV. Meteorological profile shape parameters (set 2): effective sound speed profile shape parameters relative to flight direction. A: port direction, B: starboard direction, C: tail direction, D: head direction.

	Direction				
m	A	B	C	D	Description
21	$\bar{\chi}_{-x,B}$	$\bar{\chi}_{+x,B}$	$\bar{\chi}_{-y,B}$	$\bar{\chi}_{+y,B}$	mean vertical gradient of the effective sound speed near the ground
22	$\bar{\chi}_{-x,T}$	$\bar{\chi}_{+x,T}$	$\bar{\chi}_{-y,T}$	$\bar{\chi}_{+y,T}$	mean vertical gradient of the effective sound speed in the troposphere
23	$\bar{\chi}_{-x,S}$	$\bar{\chi}_{+x,S}$	$\bar{\chi}_{-y,S}$	$\bar{\chi}_{+y,S}$	mean vertical gradient of the effective sound speed in the lower stratosphere

sphere: $\gamma_{S,ICAO}=0$ K/km) at TA1 and TA3. In the tropical target area TA2 $\bar{\gamma}_S$ is always positive $(+2,\ldots,+3$ K/km). The variability range of the height of maximum wind speed (jet stream level) h_{JS} does not differ much between TA1 and TA3. In the tropics (TA2) the jet stream is normally situated around 5 km higher than in the mid and high latitudes. The maximum wind speed u_{JS} of the jet stream shows similar distributions for TA1 and TA3. The high-speed tail is shorter at TA2. In general, the atmospheric variability is highest in mid latitudes (TA1) and smallest in the tropics (TA2).

The second set of meteorological shape parameters (Table IV) is used to characterize the vertical profiles of the "effective" speed of sound $c_{eff}(z)$:

$$c_{eff}(z) = c(z) + u(z) = \sqrt{\kappa p_{air}(z)/\rho(z)} + V(z)\cos(\alpha(z)),$$

where α is the angle between wind direction and the direction of propagation. κ is the ratio of specific heats at constant volume and constant pressure. These profiles do not only depend on the local meteorological state, but also on the angle α between wind direction and the direction of sonic-boom propagation. In practice, direction of propagation is not known *a priori*, as it varies along a ray and between rays. So, for practical reasons it is convenient to define the direction of propagation relative to the flight direction. Four directions of propagation were considered: left wing or port direction: index $-x$; right wing or starboard direction: index $+x$; backward or tail direction, index $-y$; forward or head direction: index $+y$. The mean vertical gradients $\bar{\chi}_B$, $\bar{\chi}_T$, and $\bar{\chi}_S$ of the effective speed of sound again refer to the lowest 300 m, the troposphere [300 m, h_{TP}], and the lower stratosphere [h_{TP}, 25 km], respectively.

III. SONIC-BOOM MODELING

The sonic-boom simulation code used in the present paper consists of a source model (the aerodynamic flow around the aircraft) and an atmospheric propagation model. A brief summary is presented here.

The proper matching between near-field aerodynamics and far-field acoustical propagation was formulated by Whitham (1952) for a body of revolution. This method was later extended by Walkden (1958) for a non-axisymmetric body with lift. It is nowadays often superseded by direct CFD simulations (e.g., Plotkin and Page, 2002). However, as the present study focuses on the influence of meteorology rather than on the detailed influence of the aircraft design, it is sufficient to model the source term with Whitham's F-functions.

The source term is adapted to the Airbus mock-up for a planned European Supersonic Commercial Transport (ESCT), an 89-m-long aircraft with a wing span of 42 m (wing surface 836 m^2) and an almost cylindrical fuselage (4 m width), designed for carrying 250 passengers at a Mach 2 cruise. Whitham's functions were computed for azimuth angles of emission normal to the Mach cone between 0° and 70° in steps of 5°. The angle of attack was selected depending on the aircraft position along its route (4° for a heavy aircraft near take-off, 2.5° for a lighter aircraft near landing; see Table II). The cruising altitude given in Table II corresponds to flight level with respect to the ICAO standard atmosphere. The true altitude may deviate from the nominal flight level depending on the actual meteorological condition.

Like other sonic-boom codes (Hayes *et al.*, 1969; Thomas, 1972; Cleveland *et al.*, 1996; Plotkin, 2002) the propagation model is based on full ray tracing in a stratified, moving atmosphere (Blokhintsev, 1946; Pierce, 1989, Chapt. 8). To avoid singularities near cutoff when rays turn up, the position of a point along a given ray is parametrized by the eikonal function rather than by its altitude. Along each ray, the acoustical pressure p satisfies a nonlinear, generalized Burgers' equation. Nonlinear effects are deemed essential in the long-range propagation of finite-amplitude waves such as sonic booms (Whitham, 1956, 1974). They are responsible for the slow evolution of the waveform until the typical "N" shape is achieved which is frequently recorded at ground level. To calculate the pressure p along the rays in a stratified medium, our implementation follows the method of Candel (1977) and requires the numerical integration of 13 differential equations [4 for ray tracing, 8 for determination of the ray-tube cross-section (4 for the derivative of ray equations with respect to each of the two ray coordinates parameterizing a single ray, e.g., time of emission and azimuthal angle), and 1 for the nonlinear age variable] by using standard quadrature algorithms (such as Runge-Kutta).

The acoustical pressure is determined only for rays that touch the ground. The ground impact points define the so-called geometrical carpet. Because of atmospheric refraction, not all emitted rays do touch the ground. A special procedure has been developed to determine the limiting rays, e.g., the two (one starboard, one port side) extreme rays delineating the carpet. They can be of two types whether the atmosphere near the ground is upward or downward refracting. In the upward case, the limiting ray grazes over the ground and separates the geometrical carpet from the shadow zone. In the second case, the limiting ray remains horizontal above the ground so that the carpet is not confined. In this case, the lateral extent of the carpet is limited geometrically by the impact point at the ground level of the ray emitted at an angle slightly below the limiting angle. It is to be noticed that this situation is very frequent, occurring at least on one side of the carpet in about 50% of all investigated cases.

For an upward refracting atmosphere, geometrical acoustics does not predict any signal inside the shadow zone. However, diffraction theory allows extending the signal there (Coulouvrat, 2002). Sound propagation near the cutoff and inside the shadow zone is dominated by diffraction effects, but is mostly linear. In the frequency domain, an expression for the pressure field in the shadow zone can be given in terms of the Fock integral (Pierce, 1989, Sec. 9.5, Coulouvrat, 2002) which matches to geometrical acoustics. It simplifies into a creeping-wave series expansion inside the shadow zone, where boom amplitude decreases exponentially and rise time increases almost linearly. Near the cutoff, the grazing boom is sensitive to the ground nature (surface roughness or porosity) with ground absorption significantly increasing the rise time. Since the present study focuses on meteorological effects, a perfectly flat and rigid surface is nevertheless assumed for simplicity.

Atmospheric absorption by classical and rotational losses and by molecular relaxation of diatomic nitrogen and oxygen relaxation is a key factor for sonic boom rise times. A full integration of the generalized Burgers' equation (nonlinear distortion+classical and rotational losses+relaxations) along each ray would have led to a dramatic increase of the overall computation time (Cleveland *et al.*, 1996, by a factor at least 10 according to our own estimations) and would have made the present statistical study impossible. To keep the computational effort within manageable limits, a steady-state approximation (Hodgson, 1973; Pierce and Kang, 1990; Coulouvrat and Auger, 1996) was made. It assumes that the boom wave has propagated over a sufficiently long distance to reach an almost steady time signal. Then the rise time emerges simply as a local balance *at the ground level* between nonlinear effects that tend to steepen the signal and absorption that tends to smear it out. In an atmosphere with constant humidity and temperature, the steady-state approximation overestimates the rise time. It is more accurate [see Cleveland (1995) for a detailed comparison] the longer the signal (compared with the rise time), the longer the propagation, and the more humid the air. In this study the approximation is well justified because the aircraft is rather long (producing a long signal), flies at high altitudes (ensuring a long distance down to the ground), and, at least at target areas TA1 and TA2, the sonic boom propagates into humid maritime air. Also note that near the ground, relaxation absorption is dominant over classical and rotational ones, which can be neglected except for very strong shocks (Coulouvrat and Auger, 1996). That would not be true if we had considered absorption losses all along the ray path at high altitudes (Sutherland and Bass, 2004) without the steady-state approximation, but again the computational cost would then have made the present study impossible. Finally, it is to be noted that inside the shadow zone near the carpet edge, the finite rise time is due mostly not to absorption but to diffraction effects, fully taken into account by the model.

Let us also note that the model does not take into account the influence of atmospheric turbulence, though it is known to affect booms near the ground. Despite the model capabilities, the meteorological data set resolves neither turbulent atmospheric motions nor mesoscale circulation patterns. Therefore, the variability of sonic boom characteristics, which is investigated in this paper, is exclusively due to the large-scale variability of the atmosphere caused by planetary waves and transient synoptic pressure systems.

The results of the sonic boom propagation simulations are given as time evolutions of the sound pressure $p(x,y,t)$ along the intersection of the sonic-boom normal cone (isoemission line) with the ground, with x and y being the horizontal coordinates relative to the aircraft ground position and t being the time. Typically, the time traces of a sonic-boom pressure wave have an N-like shape with a rapid increase from background pressure to maximum overpressure. With respect to the noise impact and annoyance of the affected population, the maximum sound pressure, the rise time, the boom duration (controlling the low-frequency content responsible for indoor building vibrations and rattling noise), and the carpet width are the most relevant parameters to describe the characteristics of sonic booms. Therefore the maximum sound pressure $p_{\max}$, the rise time t_r, and the A-weighted sound exposure level (A-SEL) L_A over the wave event were derived from the simulated time evolutions of the sound pressure as a function of the lateral distance x from the ground track (x is positive in the starboard direction and negative in the port direction). The method of calculation to obtain the metric A-SEL from the simulated ground pressure signatures is described by Shepherd and Sullivan (1991). Note that the C-weighting metric, which emphasizes the contribution of low frequencies, is generally used in environmental situations dealing with loud impulse sounds in an attempt to incorporate the effects of structural vibrations and rattling noise. It has been adopted by the American National Standards Institute (ANSI S12.9-1996, Part 4, Annex B) for assessing the impact of sonic booms. However, A-weighting, which is commonly used as a standard metric for community noise from (subsonic) aircraft, emphasizes more the audible frequencies and is much more sensitive to the rise time which is known for a long time as a key parameter for the impact of sonic boom on people (Dancer, 2004). For instance, for the present study, the C-SEL metric compared to the A-SEL one would imply adding about +13 dB right below the aircraft, but up to +30 dB at distant lateral points, where rise time is much larger because of diffraction (shadow zone) or absorption. That approach has been confirmed by a recent NASA study (McCurdy *et al.*, 2004) that concludes that the A-SEL metric is the second "best" metric (best correlated to annoyance) after Stevens Mark VII Perceived Level, and performs significantly better than C-SEL, even for *indoor* booms and not only (as could be expected) for *outdoor* booms.

Figure 6 shows simulated profiles of $p_{\max}(x)$, $t_r(x)$, and $L_A(x)$ for the condition of the ICAO standard atmosphere. The speed of the aircraft is set to Mach 2. Because the standard atmosphere is calm, the results are symmetric with respect to the ground track and independent of the flight direction. The variations are caused by the different altitudes and the different angles of attack due to different weights (see Table II). The lowest levels can be observed for the highest aircraft with low weight (TA1, eastbound), and the loudest boom for the lowest, heaviest aircraft (TA1, westbound). The

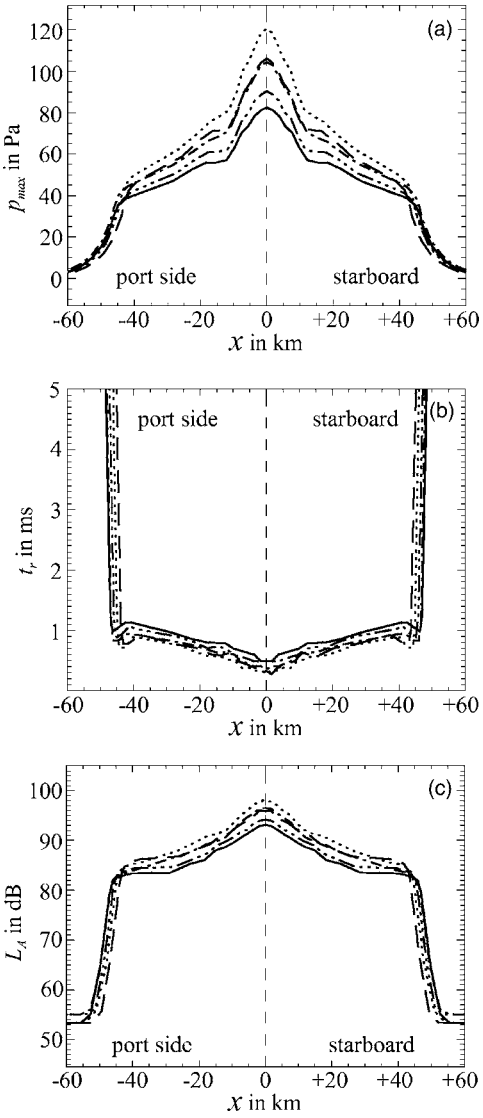


FIG. 6. Simulated horizontal profiles perpendicular to the flight ground track of (a) maximum sound pressure $p_{\max}$, (b) rise time t_r, and (c) A-weighted sound pressure level L_A for the condition of the ICAO standard atmosphere. The different curves refer to the cruising altitudes and angles of attack as given in Table II for the target areas and flight directions (solid: TA1 eastbound, dots: TA1 westbound, dashes: TA2 southbound, dash-dots: TA2 northbound, dash-dot-dot-dots: TA3 eastbound, long dashes: TA3 westbound). The solid and dashed curves overlap.

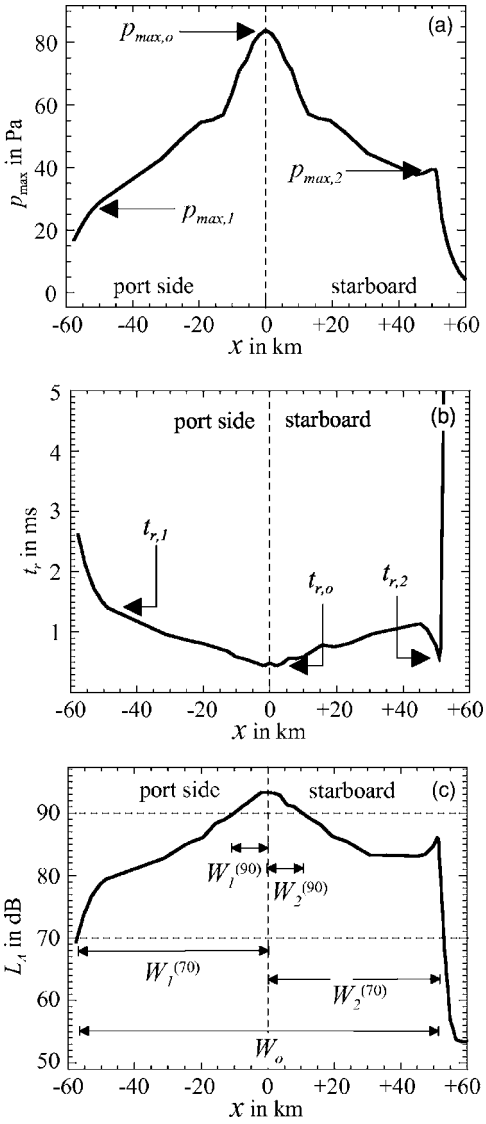


FIG. 7. Same as Fig. 6 but for westbound flights over TA2 at the local average meteorological condition. In addition, the figure visualizes the definition of acoustical shape parameters (Table V).

curves for TA1 eastbound and TA2 southbound overlap because the input parameters are identical. There is a clear pressure maximum at the ground track, then a smooth decay towards both sides (due to a combined decrease of lift effects in the source and longer propagation) until the geometrical cutoff at around $x=\pm 45$ km. The rise time due to atmospheric absorption increases slightly but remains below 1 ms. Beyond the cutoff, the amplitude decays exponentially and the rise time increases around one order of magnitude (from less than 1 ms to several tens of milliseconds).

In a real atmosphere with cross-wind the profiles of the sonic-boom parameters become asymmetric and show different behavior in the port and starboard directions. Figure 7 presents the simulated profiles of $p_{\max}(x)$, $t_r(x)$, and $L_A(x)$ for the average meteorological condition above target area TA1 and for eastbound flights (heading 106.5°). Because of prevailing winds towards the port direction a shadow zone with a sharp decay forms on the starboard side (upwind propagation) similarly to the ICAO standard case, while on the port side (downwind propagation with no geometrical cutoff) the carpet extends much farther (up to 60 km) and the maximum pressure decreases continuously. In the latter case, the shock amplitude decays significantly, and the rise time due to molecular relaxation is larger (a few milliseconds).

For statistical evaluations acoustical shape parameters a have been defined (Table V). Their definition is visualized in Fig. 7. In the following A denotes the entirety of the acoustical shape parameters with $a \in A$. As a measure of the strength of the sonic-boom event the maximum sound pressure was determined at the ground track of the aircraft ($p_{\max,0}$) as well as at the port-side and starboard edges of the sonic-boom carpet ($p_{\max,1}$ and $p_{\max,2}$). As a psychoacoustically relevant measure of the wave characteristic the rise time was calculated, again at the ground track ($t_{r,0}$) and at both edges of the carpet ($t_{r,1}, t_{r,2}$). As a measure of the

TABLE V. Acoustic shape parameters.

a	Symbol	Acoustic shape parameter
01	$p_{\max,0}$	maximum sound pressure at ground track in Pa
02	$p_{\max,1}$	port side maximum sound pressure in Pa
03	$p_{\max,2}$	starboard maximum sound pressure in Pa
04	$t_{r,0}$	rise time at ground track in ms
05	$t_{r,1}$	port side rise time at cutoff in ms
06	$t_{r,2}$	starboard rise time at cutoff in ms
07	$W_0^{(70)}$	LA70 carpet width $L_A \geqslant 70$ dB in km
08	$W_1^{(70)}$	port side carpet width $L_A \geqslant 70$ dB in km
09	$W_2^{(70)}$	starboard carpet width $L_A \geqslant 70$ dB in km
10	$W_1^{(90)}$	port side carpet width $L_A \geqslant 90$ dB in km
11	$W_2^{(90)}$	starboard carpet width $L_A \geqslant 90$dB in km

spatial dimension of the sonic-boom carpet, we define the LA70 carpet width $W_0^{(70)}$, i.e., the total width of the stripe along the aircraft ground track with $L_A \geqslant 70$ dB. In addition, we consider the widths of the stripes on either side within which the A-weighted sound exposure level reaches or exceeds 70 dB ($W_1^{(70)}, W_2^{(70)}$) or 90 dB ($W_1^{(90)}, W_2^{(90)}$), respectively.

IV. SONIC-BOOM STATISTICS

The acoustical shape parameters were separately evaluated for each season and for each target area. For TA3, the time of day (00 UTC=16 Local Time and 12 UTC=04 Local Time) was distinguished as well. For each target area two flight directions were considered (Table II). Figures 8–11 show the frequency distributions of the acoustical shape parameters. The variability of these parameters for a certain target area, time of the day, and flight direction is exclusively due to the local meteorological variability in the given season. The figures also indicate the results obtained for the ICAO standard atmosphere (see Fig. 6). When comparing the results for different target areas or different flight directions at the same target area, keep in mind that the flight altitude and angle of attack vary according to Table II.

For a given target area and flight direction the values of the maximum sound pressure at the ground track $p_{\max,0}$ vary only by about 5 to 15 Pa. However, different target areas and different flight directions show significantly different distributions. Most striking differences in the distribution of $p_{\max,0}$ are found for eastbound and westbound flights TA1 (due to different flight altitudes and weights). The maximum sound pressure at the outer edges of the sonic-boom carpet ($p_{\max,1}$ and $p_{\max,2}$) exhibits much broader frequency distributions with a variability range from 20 to 150 Pa. This higher variability is caused by the higher sensitivity of the laterally grazing boom to refraction effects due to wind gradients.

Though the mean maximum sound pressure is lower than at the ground track, larger values frequently occur due to ray "pre-focusing" (the ray tube cross section, though not vanishing, is very small). As the present model cannot handle such a situation, the numerical simulation probably overestimates the sound pressure in these cases; nevertheless, it indicates a tendency toward focusing. That tendency has been observed also at several boom recording experiments (Hubbard *et al.*, 1971, Parmentier *et al.*, 1973, Haglund and Kane, 1974), where lateral booms frequently displayed a "U" shape which is typical of a focused boom.

Most of the statistical distributions for $p_{\max,1}$ and $p_{\max,2}$ look more or less like a rather smooth and asymmetrical distribution; some seasonal variations are observable, with a strong tendency to "prefocus" laterally in autumn for TA3 12UTC eastbound. The ICAO standard atmosphere results in rather small overpressures compared to the simulations based on real meteorological data, as the absence of strong gradient near the ground smoothes lateral propagation.

The rise time of the sound pressure at the ground track ($t_{r,0}$) shows rather narrow distributions at the target areas TA1 and TA2 with almost no difference between the two flight directions. At TA1 the most frequent rise time is similar to the rise time for the standard atmosphere, while at TA2 rise times are mostly shorter than that of the standard atmosphere. A completely different behavior was simulated for the continental target area TA3. Here the variability range of the rise time is much larger with values up to 3 ms. The long rise times are caused both by low humidity and low temperature. This combination is much more frequent in the continental subpolar climate (TA3) than in the maritime midlatitude (TA1) or tropical one (TA2). At 00 UTC (local afternoon) rise times of more than 1 ms are more frequent than at 12 UTC (local morning).

The LA70 carpet width of the sonic-boom carpet ($W_0{}^{70}$) varies between 60 and 200 km. The variability range of $W_0^{(70)}$ for TA2 is smaller (60–120 km) than for TA1 and TA3. As expected, at TA1 and TA3 where westerly winds are most frequent, the width of the carpet is smaller for westbound flights than for eastbound flights, since wind astern favors downward refraction and therefore wide carpets. In the majority of the cases the port side and starboard width of the carpet (W_1^{70} and W_2^{70}) are smaller than the respective standard atmosphere values.

Seasonal differences are not well pronounced. A few exceptions are that at TA1 and for eastbound flights the maximum sound pressure $p_{\max,0}$ is significantly higher in spring than in other seasons. For the same flights (TA1, eastbound) sonic-boom carpet widths ($W_0^{(70)}$) of more than 120 km are most frequent in summer. For TA1 and westbound flights the sonic-boom carpets are generally wider in summer and narrower in winter. This can be explained by the more frequent strong westerly winds in winter which favor upward refraction and therefore narrow carpets. As expected, seasonal variations are unimportant in the tropical area TA2. In the continental area TA3 seasonal differences appear mainly at 00 UTC (local afternoon) where in spring the maximum sound pressure ($p_{\max,0}$) tends to be higher and the carpet width ($W_0^{(70)}$) tends to be larger than in other seasons.

V. DETERMINATION OF SPECIFIC INFLUENCES

In order to identify those meteorological parameters which have a significant influence on specific sonic-boom characteristics, a linear correlation analysis was performed. The statistics of the meteorological and acoustical shape parameters were evaluated using appropriate library routines

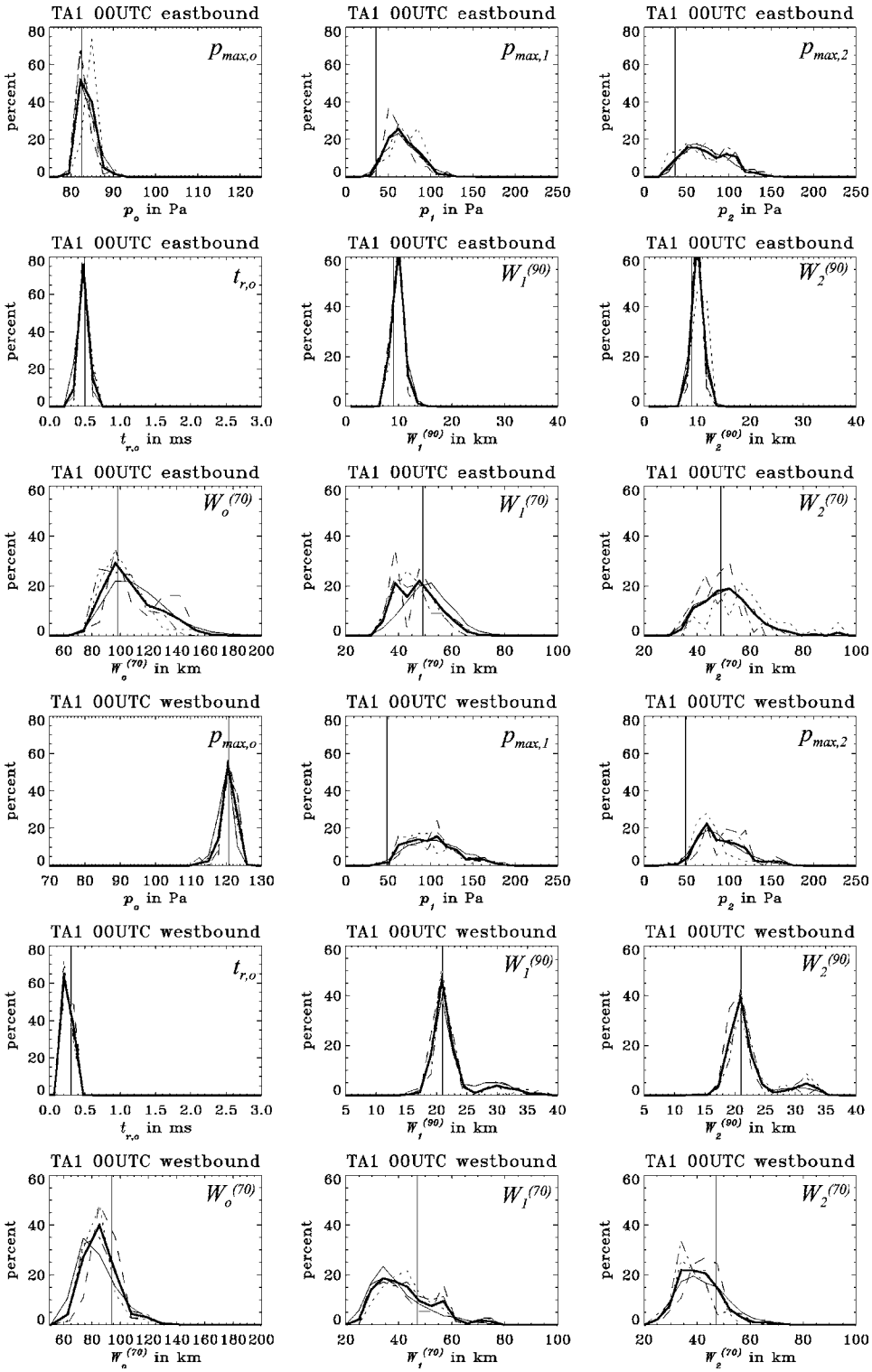


FIG. 8. Frequency distributions of a selection of simulated acoustical shape parameters (Table V) for target area TA1, 00 UTC, eastbound and westbound flights as indicated. The different lines refer to the full year and to single seasons (see caption of Fig. 2).

which are based on Neter *et al.* (1988). The correlation analyses always refer to the complete time series without considering single seasons. Different target areas and flight directions were investigated separately because the assumed flight altitudes and angles of attack differ.

At first, the linear correlation between single meteorological parameters and single acoustic parameters was determined. It was calculated by means of the linear correlation coefficient r_{ma}. It describes the linear relationship between a specific acoustical shape parameter $a \in A$ and a specific meteorological shape parameter $m \in M$, although the respective acoustical parameter might be influenced also by all other meteorological parameters. The different influences may compensate or amplify each other.

Figure 12 visualizes the correlation coefficients r_{ma} between all combinations of the meteorological and acoustical

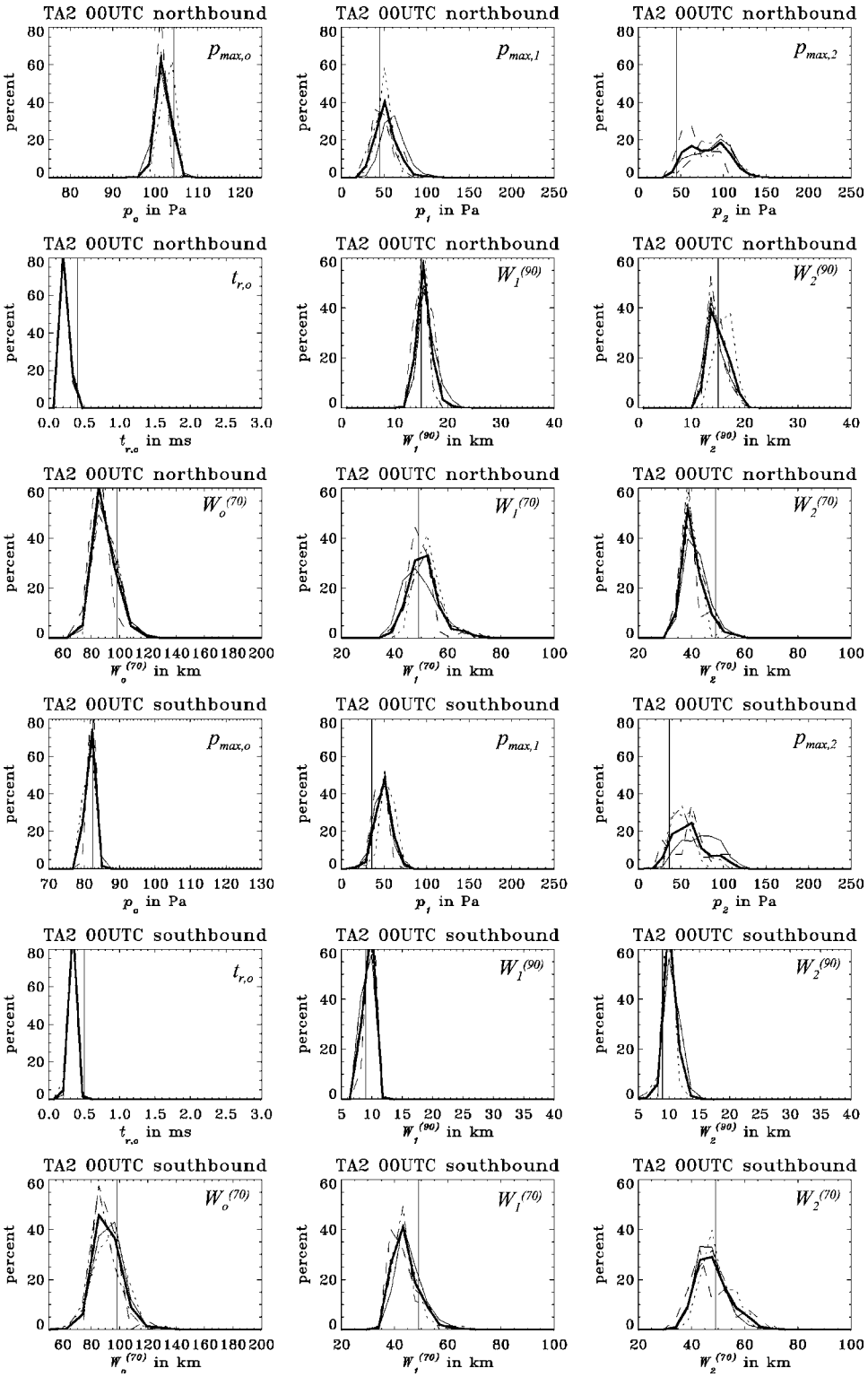


FIG. 9. Same as Fig. 8 but for target area TA2, 00 UTC.

shape parameters (see Tables III–V) for target area TA1 and westbound flights. Correlation coefficients that are not significantly different from zero at a significance level of 0.01 are indicated by symbols (×). Relatively high correlation coefficients are only found for the meteorological parameters of the second set (Table IV), i.e., the mean vertical gradients of the effective speed of sound relative to flight direction. The peak pressure amplitudes $p_{\max,1}$ and $p_{\max,2}$ at the outer edges of the sonic-boom carpet (a: 02, 03) are correlated or anticorrelated (positive or negative values of r_{ma}) with the vertical gradients of c_{eff} near the ground $\overline{\chi}_{-x,B}$, $\overline{\chi}_{+x,B}$, $\overline{\chi}_{-y,B}$, $\overline{\chi}_{-y,B}$ (m: 21A, 21B, 21C, 21D). This reflects the strong sensitivity of grazing sonic-boom waves to vertical c_{eff} gradients near the ground. The meteorological shape parameters of Table IV are also (anti-)correlated with the pressure rise times $t_{r,1}, t_{r,2}$ at the outer edges of the carpet (a: 05, 06),

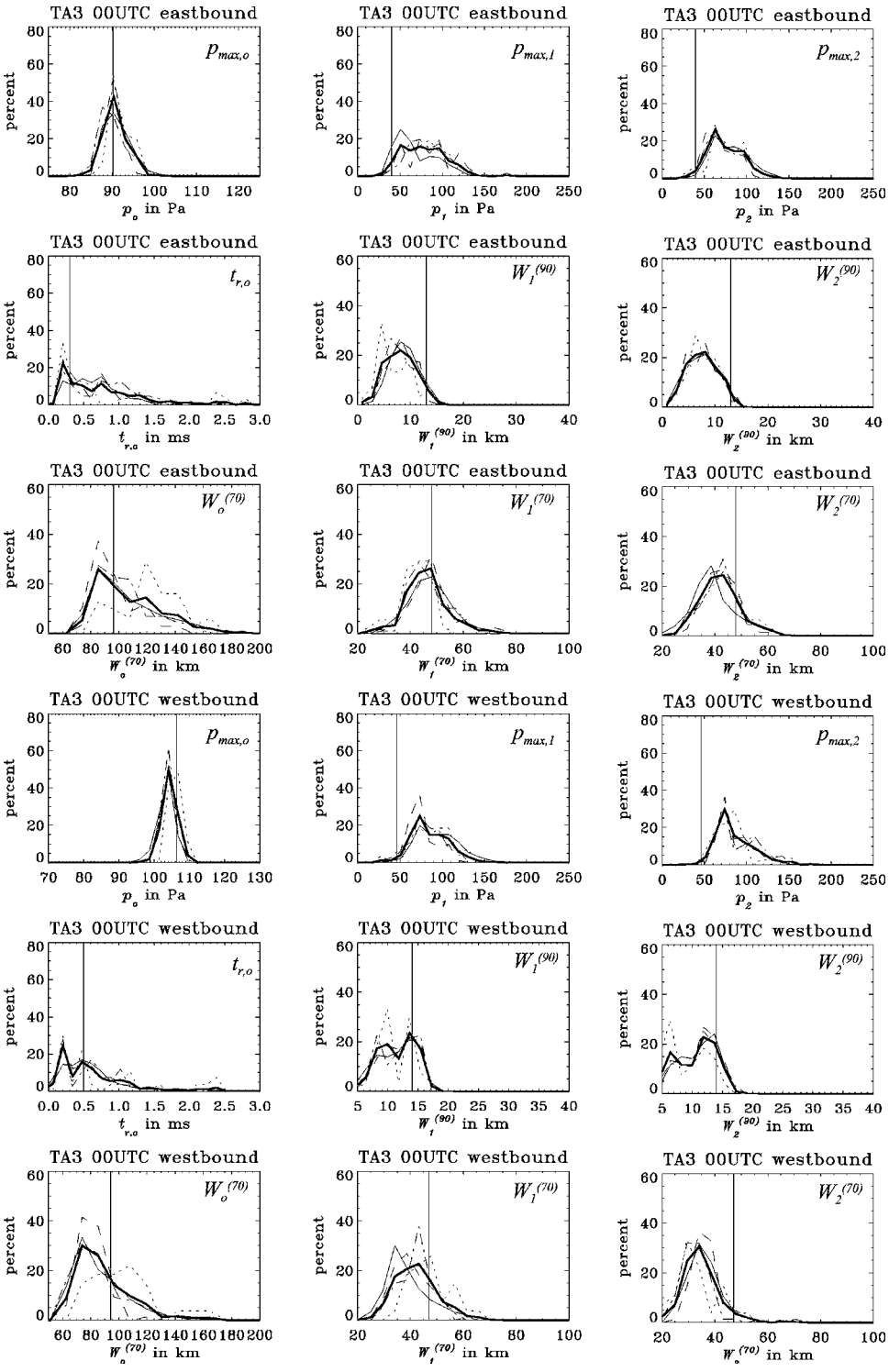


FIG. 10. Same as Fig. 8 but for target area TA3, 00 UTC.

which can be explained by the dependence of the rise time on the amplitude. The LA70 carpet width $W_0^{(70)}$ (a:07) is (anti-)correlated with the vertical gradients of the effective sound speed $\bar{\chi}_{-y,B}$, $\bar{\chi}_{+y,B}$, $\bar{\chi}_{-y,T}$, $\bar{\chi}_{+y,T}$, $\bar{\chi}_{-y,S}$, $\bar{\chi}_{+y,S}$ in tail and head directions (m: 21C, 22C, 23C, 21D, 22D, 23D). The LA70 carpet width is anticorrelated with the tail gradient near the ground and in the troposphere and the head gradient in the stratosphere.

The correlation with the stratospheric gradients is weaker than with those in the troposphere and near the ground. The lateral carpet widths $W_1^{(70)}$ and $W_2^{(70)}$ on the port and starboard sides (a: 08, 09) are correlated with the c_{eff} gradients $\bar{\chi}_{-x,B}$, $\bar{\chi}_{+x,B}$, $\bar{\chi}_{-x,T}$, $\bar{\chi}_{+x,T}$, $\bar{\chi}_{-x,S}$, $\bar{\chi}_{+x,S}$ with respect to left and right propagation (m: 21A, 22A, 23A, 21B, 22B, 23B). For the same direction (port or starboard) positive correlation is found between the gradients near the ground and in the troposphere while negative correlation is found for the gradients in the stratosphere. This means that upward refraction of sound rays in the stratosphere due to negative gradi-

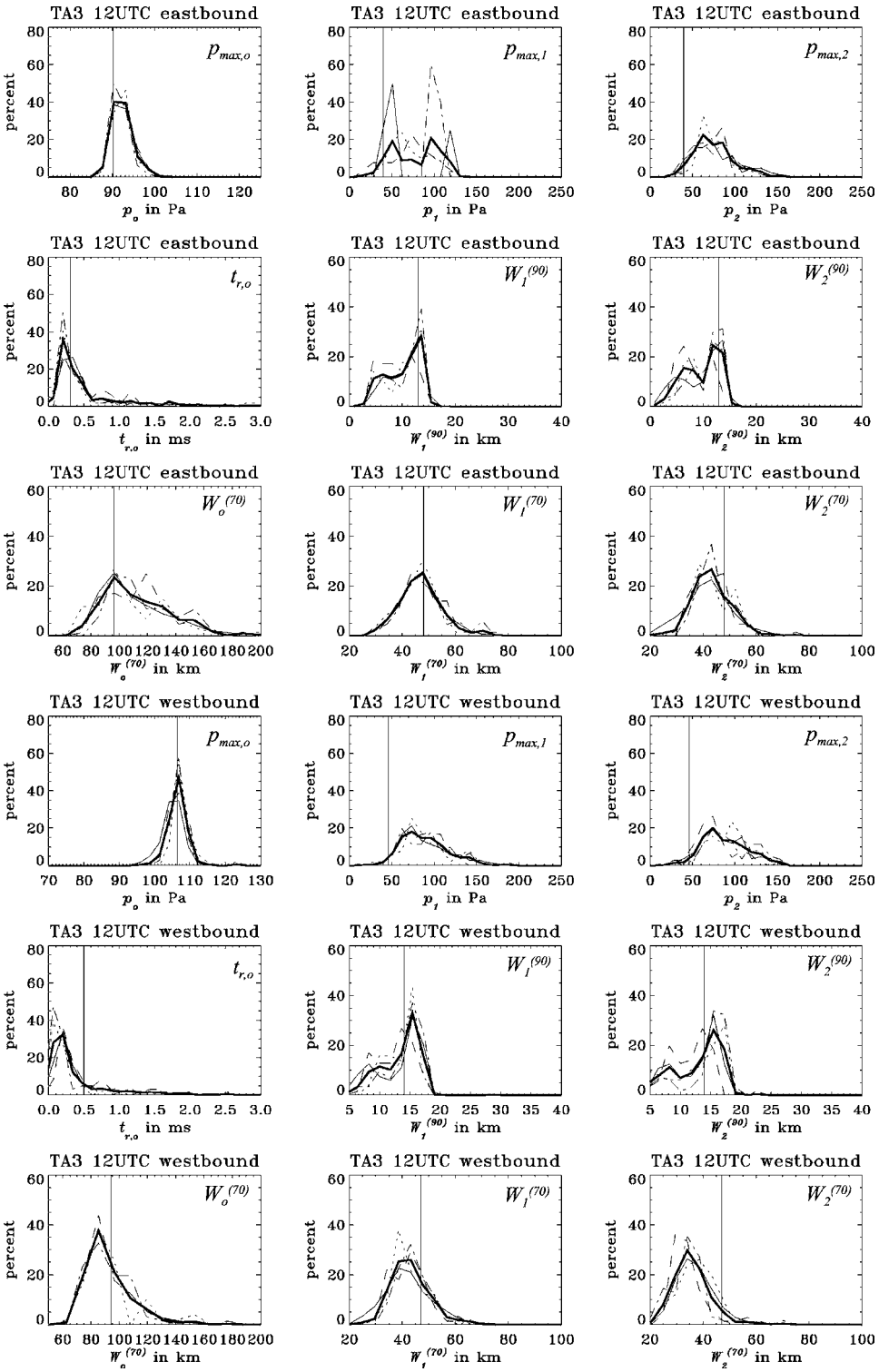


FIG. 11. Same as Fig. 8 but for target area TA3, 12 UTC.

ents ($\bar{\chi}_{-x,S}<0$) contributes to a wider carpet width $W_1^{(70)}$ like downward refraction due to positive gradients ($\bar{\chi}_{-x,T}>0$) in the troposphere. While the first condition bends the sound rays upward, the second condition prevents them from forming shadow zones which eventually would limit the carpet width.

As mentioned above the correlation coefficient r_{ma} does not describe the pure relationship between the acoustical parameter a and the meteorological parameter m, because the acoustical parameter a usually is also influenced by other meteorological parameters. Therefore partial correlation coefficients $r_{ma.M'}$ and multiple correlation coefficients r_{Ma} were determined as well. The partial correlation coefficients describe the linear correlation between the acoustical parameter a and the specific meteorological parameter m while the influence of all other meteorological parameters out of $M'=M\backslash m$ is disregarded by assuming the parameters in M' are constant. The multiple correlation coefficients are a measure

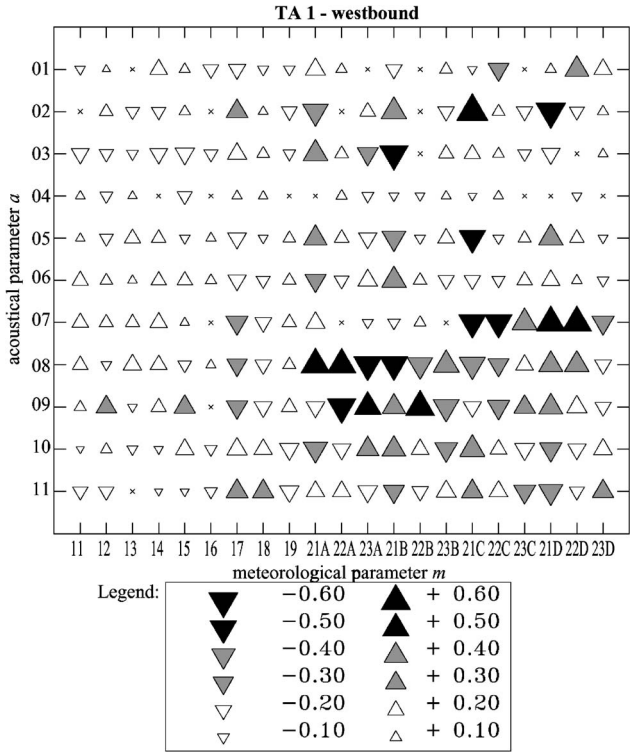


FIG. 12. Correlation coefficients r_{ma} of all combinations of meteorological (Tables III and IV) and acoustical (Table V) shape parameters. Symbols "×" indicate correlation coefficients which are not significantly different from zero at a significance level of 0.01. Read text for further explanations.

of the linear correlation between an specific acoustical parameter a and the entirety of all meteorological parameters $m \in M$. The partial and multiple correlation coefficients can attain values between zero (no correlation) and one (exact linear relationship). Correlation and anti-correlation are not distinguished.

The partial correlation coefficients $r_{ma.M'}$ are shown in Fig. 13 for all target areas and flight directions. At TA1 (St. Georges Channel) the maximum wind speed of the jet stream u_{JS} (m: 17) and the mean wind speed gradient above the jet stream $\bar{\eta}_2$ (m:19) are correlated with the maximum sound pressure $p_{\max,0}$ (a: 01) if the influence of the remaining meteorological parameters is eliminated. The partial correlation coefficients at target area TA2 (off-shore Vietnam) are similarly distributed as at TA1. In addition, a correlation between the specific humidity in the boundary layer q_B (m:15) and the rise time at ground track $t_{r,0}$ (a: 04) is evident. At TA3 (Mackenzie/Canada) the distribution of partial correlation coefficients are partially different from those at the other two target areas. In particular, the mean vertical temperature gradient $\bar{\gamma}_B$ (m:12) and the specific humidity in the boundary layer q_B turned out to be important for the carpet widths $W_1^{(90)}$ and $W_2^{(90)}$ (a:10,11). This can be explained by the relatively high variability of the boundary-layer parameters at this continental site.

VI. METEOROLOGICAL CLASSIFICATION

As mentioned earlier, performing a larger number of sonic-boom propagation simulations is rather time consuming. In order to assess representative sonic-boom statistics for a large number of areas or flight conditions, or to predict the sonic-boom behavior for given meteorological situations without extra simulations, it is desirable to define meteorological classes. Then, simulations could be restricted to situations which are representative of those classes. Once we know to which class a certain meteorological situation be-

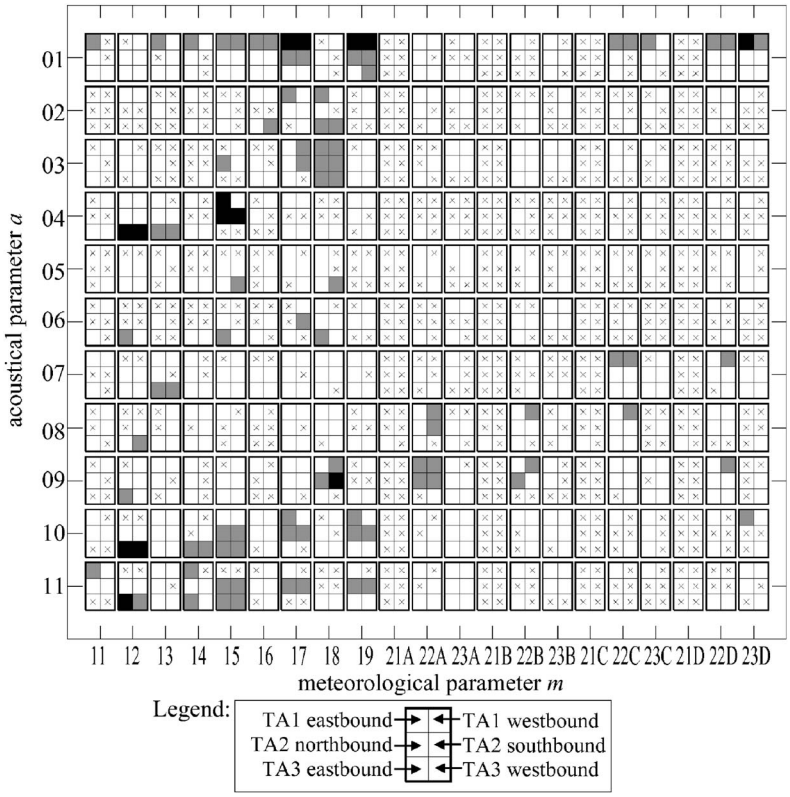


FIG. 13. Partial correlation coefficients $r_{ma.M'}$ of all combinations of meteorological (Tables III and IV) and acoustical (Table V) shape parameters for all target areas and flight directions. Symbols "×" indicate partial correlations that are not significantly different from zero at a significance level of 0.01. Blank fields: $0 \leqslant r_{ma.M'} \leqslant 0.2$, gray fields: $0.2 < r_{ma.M'} \leqslant 0.4$, black fields: $r_{ma.M'} > 0.4$.

longs, its sonic-boom characteristics could be used as an approximation without further simulations. In the following we investigate whether the meteorological shape parameters are suitable for classification. We show at what level of quality the long-term time series or statistics (mean, standard deviation) of sonic-boom parameters can be reproduced by a feasible number of presimulated class representatives.

The daily sets of 21 meteorological shape parameters (Tables III and IV) were first normalized (subtraction of the mean value) and standardized (division by the standard deviation) and then grouped into classes by using a k-means clustering algorithm (Hartigan, 1975). The number of classes n_c has to be prescribed. In the following we use $n_c=2^k$ with $k=2,\ldots,9$. For each class the most typical member was determined as a representative. It is the member whose meteorological shape parameters are least different (using the Euclidian distance measure) from the class average. The corresponding sets of 11 acoustical shape parameters (Table V), i.e., the model result for the meteorological situation of the typical members, were used to represent the sonic-boom characteristics of each meteorological class.

As a test the 10-year time series of acoustical shape parameters were reproduced by using the representative sets of shape parameters. At each date the original shape vector was replaced by the one that represents the meteorological class the date belongs to. The reproduced time series is now compared with the original time series. As quality measures, we use the time correlation coefficient and the standardized rms error for each acoustical shape parameter a. The standardized rms error is the rms error divided by the standard deviation of the original time series of the respective acoustical shape parameter. To estimate the benefit of the meteorological classification a simple time interval classification was performed as reference. Here, the complete time series was divided into n_c equal intervals and the "typical member" was randomly selected out of the interval. The number of time intervals (the classes of the reference classification) was varied continuously in the range $n_c=1,\ldots,N$, where N is the number of data points of the time series ($N=3650$ for a 10 -year set of daily data). The meteorological classification can be assessed as successful if the quality measures for a given number of classes are better than those of the simple time-interval classification.

Figure 14 shows the result for the acoustical shape parameters $p_{\max,0}$ (a: 01) and $W_0^{(70)}$ (a: 07) and westbound flights at TA1. The abscissa shows the percentage of data points involved. This percentage is given by $100n_c/N$. It expresses the fraction of the computational effort needed for the reproduction of the time series through classification. For $p_{\max,0}$ the meteorological classification fails because neither the standardized rms error is smaller nor the correlation coefficient is higher than the respective values of the simple time-interval classification. For $W_0^{(70)}$, however, the use of the meteorological classification is an improvement. The results also show that the quality grows as the number of classes increases. On the other hand, the computational effort increases with the number of classes. Nevertheless, the absolute quality of the reproduced time series is rather poor. While raising the number of classes, the rms error decreases

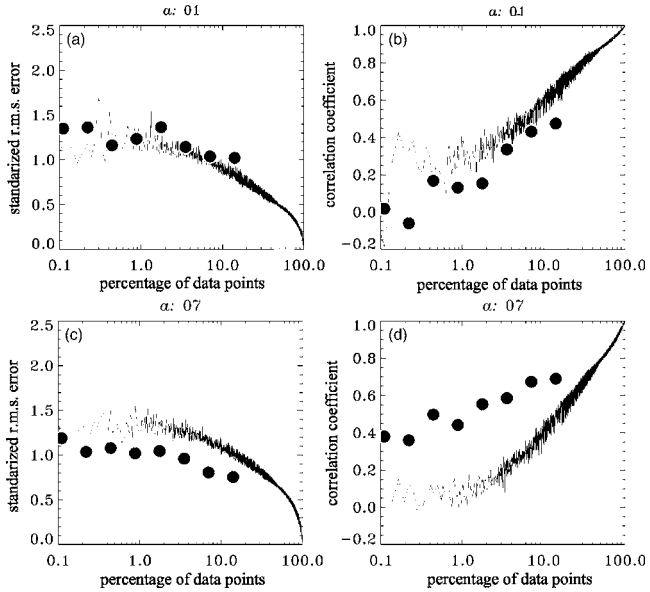


FIG. 14. Standardized rms error and time correlation coefficient of reproduced time series based on the representatives of n_c classes with respect to the respective true time series with N original data points. Top: maximum sound pressure $p_{\max,0}$ (a: 01), bottom: carpet width $W_0^{(70)}$ (a: 07). The abscissa shows $100n_c/N$. The dots indicate the results for the meteorological classification, while the solid lines present the results for the simple time-interval classification. All results refer to westbound flights over TA1.

rather slowly and it remains in the order of the standard deviation. The correlation merely reaches 0.5–0.6 even for a large number of classes.

Figures 15(a) and 15(b) show the results for all acoustical shape parameters, again for TA1 and westbound flights.

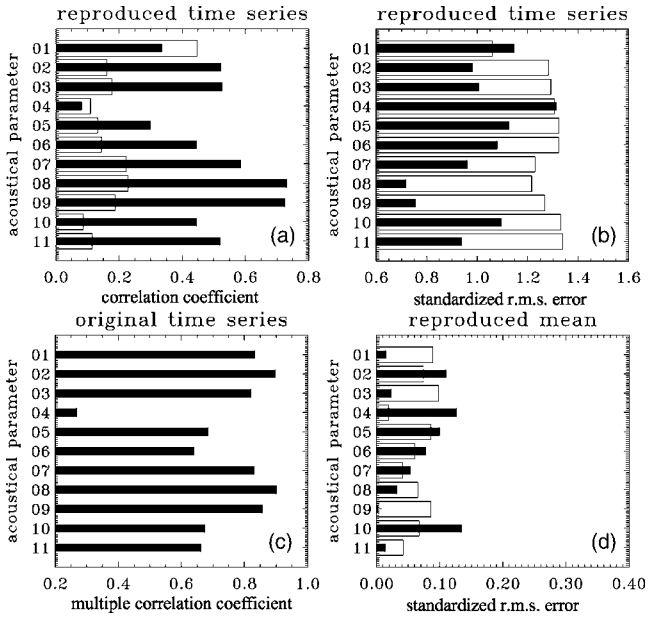


FIG. 15. (a) and (b) Correlation coefficient and standardized error of the class-based time series (black: meteorological classification, white: simple time-interval classification; $n_c=64$) with respect to the "true" time series based on all data points. (c) Distribution of the multiple correlation coefficients r_{Ma} between the acoustical sonic-boom shape parameters and the entirety of the meteorological shape parameters. (d) Standardized error of the class-based temporal mean values. All results refer to westbound flights over TA1.

The number of classes was fixed to $n_c=64$. Hence, the respective typical members correspond to 1.7% of all data points. For this number of meteorological classes the quality measures are better than for the same number of time intervals of the simple time-interval classification. Exceptions are the acoustical parameters $p_{\max,0}$ (a: 01) and the associated rise time $t_{r,0}$ (a: 04). The meteorological classification yields a particularly high quality (high correlation, low rms error) and a significant benefit with respect to the simple time-interval classification for the carpet width $W_0^{(70)}$, $W_1^{(70)}$, $W_2^{(70)}$ (a: 07,08,09). The success of the meteorological classification is determined by comparison with the multiple correlation coefficients r_{Ma} [Fig. 15(c)]. If an acoustical shape parameter is highly correlated with the entirety of the meteorological shape parameters, the reproduction of a time series with the typical members of meteorological classes works sufficiently well.

The classification was also used to determine the statistical properties of acoustical shape parameters. With 64 classes, i.e., 1.7% of the total number of data points, it is possible to estimate the temporal mean value of most of the acoustical shape parameters with a deviation from the "true" mean (based on daily samples) of less than $\frac{1}{10}$ of the standard deviation of the original time series [Fig. 15(d)]. The reduction of data leads to accurate results, yet the meteorological classification is not superior to the simple time-interval classification. Hence, the temporal mean of many acoustical shape parameters (e.g., $t_{r,0}$, $W_0^{(70)}$) could be predicted fairly well by simulating only every 50th day and averaging the results. Similar results were attained for the standard deviation of a class-based time series (not shown in the figure).

VII. CONCLUSIONS

The application of a sonic-boom propagation model to a 10-year time series of meteorological conditions at three different target areas (six flight conditions) shows that the atmospheric variability and the climatic difference between regions cannot be neglected without significantly impairing the accuracy of sonic-boom predictions. While ground-track boom amplitude does not vary much and is mostly dominated by flight altitude and weight, many sonic-boom properties, in particular the carpet width and the sound-pressure amplitude at the edges of the carpet, are rather sensitive to varying meteorological conditions. Relatively important meteorological parameters are the vertical gradients of the effective speed of sound in the troposphere and lower stratosphere perpendicular to the flight direction (for the half carpet widths) and parallel to the flight direction (for the LA70 carpet width). Over land surfaces, the gradients near the ground are also decisive. To some extent the maximum sound pressure and the rise time are correlated with the specific humidity (here only near the ground because of the steady-state approximation of the model). The tropopause height and the height of the jet stream turned out to be of minor importance.

The present study has been limited to a single Mach number (Mach 2). Similar studies would be necessary to explore the variability at lower Mach numbers. For sonic-boom focusing due to the acceleration phase around Mach 1.2, a complementary study (Blumrich *et al.*, 2005) has shown that variability is much larger than for the cruise phase at Mach 2. This might be related to the fact that rays are more grazing at lower Mach, and therefore are much more sensitive to sharp atmospheric gradients near the ground. That trend is also observed in the present study where a boom is much more variable when its propagation is lateral and therefore grazing. Though this would need a quantitative confirmation, one could already expect that decreasing the Mach number would increase the meteorologically induced variability.

The selected meteorological profile shape parameters are of only limited value if they are used to define acoustically relevant meteorological classes. The profile shape parameters perhaps oversimplify the atmospheric situation because they refer to rather thick layers. Therefore, the relationship between the meteorological parameters and the respective acoustical response is not very strong. As a consequence, the meteorological classes overlap in terms of their acoustical properties and the number of classes must be rather high ($\geqslant$64) to attain reasonable benefit from a meteorological, classification. Nevertheless, it is possible to predict at least some acoustical parameters (e.g., carpet width and overpressure near the carpet edges) by using the representative meteorological situation of the class into which a certain situation falls. Hence, these sonic-boom properties could be precalculated for meteorological classes and later used to forecast actual situations without extra simulations. This would allow flight-track optimization as a function of weather with respect to high flight efficiency and low sonic-boom impact to sensitive regions without the need for significant computing resources.

For the determination of basic sonic-boom statistics (e.g., local mean values, standard deviations) it is sufficient to reduce the number of considered situations down to approximately 2% of the original number of data points without losing too much accuracy, i.e., keeping the deviation of the estimated mean from the true mean below one tenth of the true temporal standard deviation for most of the parameters. However, the use of meteorological profile classes does not yield additional benefit over a random selection of as many situations as classes. The potential to improve the meteorological classification exists, for example by a refinement and more specific selection of the profile shape parameters. However, this could not be completed within this study and is left to future investigations.

ACKNOWLEDGMENTS

The study was carried out as part of the project "Sonic Boom European Research Programme: Numerical and Laboratory-Scale Experimental Simulation" (SOBER) which was co-funded by the European Commission under Contract No. G4RD-CT-2000-00398. The authors are obliged to Dr. Klaus-Peter Hoinka (DLR) who extracted the ERA-15 model level data from ECMWF within the Special Project "The Climatology of the Global Tropopause." We thank Dr. Günther Zängl (University of Munich) for providing his FORTRAN code to determine the thermal tropopause height.

Roland Etchevest (Airbus France S.A.S., Toulouse) and Stephane Illa (Transiciel Technologies, Toulouse) patiently helped us implement the sonic boom propagation code at DLR. Airbus France S.A.S. also contributed by providing target trajectories and input Whitham functions.

Bass, H. E., Sutherland, L. C., Piercy, J., and Evans, L. (**1984**). "Absorption of sound by the atmosphere," in *Physical Acoustics*, edited by W. P. Mason and R. N. Thurston (Academic, Orlando) Vol. **XVII**, pp. 145–232.

Blokhintzev, D. I. (**1946**). "The propagation of sound in an inhomogeneous and moving medium I," J. Acoust. Soc. Am. **18**, 322–328.

Blumrich, R., Coulouvrat, F., and Heimann, D. (**2005**). "Variability of Focused Sonic Booms from Accelerating Supersonic Aircraft in Consideration of Meteorological Effects," J. Acoust. Soc. Am. **118**, 676–686.

Candel, S. (**1977**). "Numerical solution of conservation equation arising in linear wave theory: application to aeroacoustics," J. Fluid Mech. **83**, 465–493.

Cleveland, R. O. (**1995**). "Propagation of sonic booms through a real, stratified atmosphere," Ph.D. dissertation, The University of Texas at Austin.

Cleveland, R. O., Chambers, J. P., Bass, H. E., Raspet, R., Blackstock, D. T., and Hamilton, M. F. (**1996**). "Comparison of computer codes for the propagation of sonic boom waveforms through isothermal atmospheres," J. Acoust. Soc. Am. **100**, 3017–3027.

Coulouvrat, F. (**2002**). "Sonic boom in the shadow zone: A geometrical theory of diffraction," J. Acoust. Soc. Am. **111**, 499–508.

Coulouvrat, F., and Auger, Th. (**1996**). "Influence of molecular relaxation on the rise time of sonic booms," *7th Int. Symp. Long Range Sound Propagation*, Lyon, 25–26 July 1996, Ecole Centrale de Lyon, Actes du Symposium, pp. 177–191.

Dancer, A., and Naz, P. (**2004**). "Sonic boom: ISL studies from the 60's to the 70's," in *Proceedings of 7ème Congrès Français d'Acoustique/30. Deutsche Jahrestagung für Akustik DAGA*, Strasbourg, France, 22–25 March 2004, pp. 1067–1068.

Esclangon, E. (**1925**). *L'acoustique des canons et des projectiles* (Imprimerie Nationale, Paris) (in French).

Gibson, R., Kallberg, P., Uppala, S., Hernandez, A., Nomura, A., and Serrano, E. (**1997**). ERA description, ECMWF ReAnalysis Project Report Series 1. Available from ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, U.K.

Guiraud, J.-P. (**1965**). "Acoustique géométrique, bruit balistique des avions supersoniques et focalisation," J. Mec. **4**, 215–267 (in French).

Haglund, G. T., and Kane, J. (**1974**). "Flight test measurements and analysis of sonic boom phenomena near the shock wave extremity," AIAA Pap. 74–6.

Hartigan, J. A. (**1975**). *Clustering Algorithms* (Wiley, New York).

Hayes, W. D., Haefeli, R. C., and Kulsrud, H. E. (**1969**). "Sonic boom propagation in a stratified atmosphere with computer programme," NASA CR–1299.

Heimann, D. (**2001**). "Effects of long-term atmospheric variability on the width of a sonic-boom carpet produced by high-flying supersonic aircraft," ARLO **2**, 73–78.

Hodgson, J. P. (**1973**). "Vibrational relaxation effects in weak shock waves in air and the structure of sonic bangs," J. Fluid Mech. **58**, 187–196.

Hubbard, H. H., Maglieri, D. J., and Huckel, V. (**1971**). "Variability of sonic boom signatures with emphasis on the extremities of the ground exposure patterns," Third Conference on Sonic Boom Research, NASA SP–255, pp. 351–359.

ICAO (**1993**). Manual of the ICAO Standard Atmosphere (extended to 80 kilometres), Doc 7488/3rd ed.

Le Pichon, A., Garcés, M., Blanc, E., Barthélémy, M., and Drob, D. P. (**2002**). "Acoustic propagation and atmosphere characteristics derived from infrasonic waves generated by the Concorde," J. Acoust. Soc. Am. **111**, 629–641.

Lundberg, W. R. (**1994**). "Seasonal sonic boom propagation prediction," Armstrong Laboratory, AL/OE-TR-1994-0132.

Maglieri, D. J., and Plotkin, K. J. (**1995**). "Sonic boom," in *Aeroacoustics of Flight Vehicles*, edited by H. H. Hubbard, (Acoustical Society of America, Woodbury, NY), Vol. **1**, pp. 519–561.

McCurdy, D. A., Brown, S. A., and Hilliard, R. D. (**2004**). "Subjective response of people to simulated sonic booms in their homes," J. Acoust. Soc. Am. **116**, 1573–1584.

Neter, J., Wasserman, W., and Whitmore, G. A. (**1988**). *Applied Statistics*, 3rd ed. (Allyn and Bacon, Boston).

Parmentier, G., Mathieu, G., Schaffar, M., and Johe, Ch. (**1973**). "Bang sonique de Concorde: enregistrement hors trace des variations de pression au sol. Centre d'Essais des Landes, 13–15 June 1973," Institut Franco-Allemand de Recherches de Saint-Louis, Rapport Technique RT19/73.

Pierce, A. D. (**1989**). *Acoustics, an Introduction to its Physical Principles and Applications* (Acoustical Society of America, New York) (1st ed. in **1981**).

Pierce, A. D., and Kang, J. (**1990**). "Molecular relaxation effects on sonic boom waveforms," in *Frontiers of Nonlinear Acoustics*, Proceedings of the 12th Int. Symp. Nonlinear Acoust., edited by M. F. Hamilton and D. T. Blackstock (Elsevier, London), pp. 165–170.

Plotkin, K. J. (**2002**). "State of the art of sonic boom modelling," J. Acoust. Soc. Am. **111**, 530–536.

Plotkin, K. J., and Page, J. A. (**2002**). "Extrapolation of sonic boom signatures from CFD solutions," AIAA Pap. 2002–0922.

Shepherd, K. P., and Sullivan, B. M. (**1991**). "A loudness calculation procedure applied to shaped sonic booms," NASA Technical Paper TP–3134.

Sutherland, L. C., and Bass, H. E. (**2004**). "Atmospheric absorption in the atmosphere up to 160 km," J. Acoust. Soc. Am. **115**, 1012–1032.

Thomas, C. L. (**1972**). "Extrapolation of sonic boom pressure signatures by the waveform parameter method," NASA TN D-6832.

Walkden, F. (**1958**). "The shock pattern of a wing-body combination, far from the flight path," Aeronaut. Q. **IX**, 164–194.

Whitham, G. B. (**1952**). "The flow pattern of a supersonic projectile," Commun. Pure Appl. Math. **5**, 301–348.

Whitham, G. B. (**1956**). "On the propagation of weak shock waves," J. Fluid Mech. **1**, 290–318.

Whitham, G. B. (**1974**). *Linear and Nonlinear waves* (Wiley/Interscience, New York).

WMO (**1957**). "Meteorology—A three-dimensional science: Second session of the commission for aerology," WMO Bull. **IV(4)**, 134–138.

Zängl, G., and Hoinka, K. P. (**2000**). "The tropopause in the polar regions," J. Clim. **14**, 3117–3139.

Experimental evidence of three-dimensional acoustic propagation caused by nonlinear internal waves

Scott D. Frank[a)]
Department of Mathematical Sciences, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180

Mohsen Badiey
Ocean Acoustics Laboratory, College of Marine Studies, University of Delaware, Newark, Delaware 19716

James F. Lynch
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

William L. Siegmann
Department of Mathematical Sciences, Rensselaer Polytechnic Institute, 110 8th Street, Troy, New York 12180

(Received 30 September 2004; revised 5 May 2005; accepted 5 May 2005)

The 1995 SWARM experiment collected high quality environmental and acoustic data. One goal was to investigate nonlinear internal wave effects on acoustic signals. This study continues an investigation of broadband airgun data from the two southwest propagation tracks. One notable feature of the experiment is that a packet of nonlinear internal waves crossed these tracks at two different incidence angles. Observed variations for the lower angle track were modeled using two-dimensional parabolic equation calculations in a previous study. The higher incidence angle is close to critical for total internal reflection, suggesting that acoustic horizontal refraction occurs as nonlinear internal waves traverse this track. Three-dimensional adiabatic mode parabolic equation calculations reproduce principal features of observed acoustic intensity variations. The correspondence between data and simulation results provides strong evidence of the actual occurrence of horizontal refraction due to nonlinear internal waves. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1942428]

PACS number(s): 43.30.Zk, 43.30.Pc, 43.30.Es [AIT] Pages: 723–734

I. INTRODUCTION

Acoustic interaction with nonlinear internal waves in shallow-water regions has received significant attention in recent years. Nonlinear internal wave packets are known to affect amplitude[1] and phase[2] of single frequency and broadband acoustic signals at 200 Hz and above. For example, an early paper explains anomalous frequency dependent transmission losses using classical wave-wave interactions,[1] with resonance occurring when a horizontal wavenumber of a nonlinear internal wave packet is close to a difference between wave numbers of dominant acoustic modes. Recently a theoretical study generalized this condition to relate acoustic wavenumber differences to peaks in the horizontal wavenumber spectrum of an internal wave packet.[3] Other energy-exchange mechanisms are discussed in a recent review.[4]

The SWARM 95 experiment was a multi-institutional effort which acquired high quality environmental and acoustic data.[5] Two of its goals were to observe nonlinear internal waves and to understand and describe their effects on acoustic signals. For instance, results from this experiment show that arrival time variations[6] and mode amplitude decorrelation[7] of broadband transmissions occur when internal wave packets intersect propagation paths. Numerical studies investigate mode coupling caused by range dependence in nonlinear internal wave packets[8,9] and the additional influence of a diffuse internal wave field on single frequency acoustic signals.[10]

For six days during the SWARM experiment the R/V Cape Hatteras was southwest of the two vertical linear arrays (VLAs). A 20 in.[3] Bolt airgun was suspended at various depths from the Cape Hatteras during this time. On 4 August 1995, while the source was at 12 m depth, a packet of strong nonlinear internal waves crossed the two acoustic propagation tracks with different incidence angles. If the wave fronts are assumed planar, then one angle was near 45°, and the second was near the ray theory estimate of total internal reflection. Acoustic data from both tracks show significant variations while the nonlinear internal wave packet traverses the region, and some features of the variations are correlated with those of the nonlinear internal wave packet. However, acoustic observations from the two tracks do show differences that represent azimuthal dependence of the acoustic field (when viewing from cylindrical coordinates centered on the source). Two-dimensional (2D) parabolic equation (PE) calculations have been used to illustrate[11] this azimuthal dependence on pulse shape and amplitude. Prior to SWARM,

[a)]Currently at Department of Mathematics, Marist College, 3399 North Rd., Poughkeepsie, NY 12601; electronic mail: scott.frank@marist.edu

Rubenstein and Brill[12] used a 2D PE to model observed 400 Hz continuous wave intensity variations at a horizontal array caused by an internal wave packet. These computations accurately represented the presence of acoustic features and their phase speeds, but the authors concluded the model variations did not have sufficient amplitude. Since SWARM, other 2D calculation studies of azimuthal sound-speed profile dependence from internal waves have emphasized multiple resonance effects[13] and proposed the idea that internal waves cause acoustic focusing.[14] The international experiment ASIAEX, recently performed in the South and East China Seas, includes investigation of azimuthal-dependent variability.[15]

Even though azimuth-dependent acoustic effects have usually been examined with 2D (or $N\times 2D$) propagation techniques, genuinely three-dimensional (3D) mechanisms have been conjectured for some time and have received increasing attention recently. Theoretical studies[16,17] suggest strong horizontal sound-speed gradients, as introduced by nonlinear internal waves, could cause significant refraction of horizontal acoustic rays. In the Barents Sea, acoustic phase variations were observed on a horizontal array while nonlinear internal waves were believed present.[2] Computational investigations also concluded that 3D effects of nonlinear internal waves should be observable via intensity loss[3,18] or beamforming effects,[19] although under certain conditions these effects may be minimal.[20] Only small horizontal refraction was demonstrated theoretically for deep-water situations involving, for example, strong eddies and currents.[21,22]

This paper contains three principal contributions. First, the dominant signals generated by the airgun source are broadband pulses below 100 Hz, with relatively less energy up to 180 Hz. In contrast, recent studies of acoustic propagation influenced by nonlinear internal waves focus on signals at or above 200 Hz. A comparison of data from the two VLAs suggest that nonlinear internal waves cause much larger intensity variability in these frequency bands when the acoustic propagation direction is nearly parallel to internal wave fronts. Second, a full broadband PE propagation model of the Woods Hole Oceanographic Institution (WHOI) track and data is presented. This model includes a treatment of the airgun source, specification of geoacoustic parameters and ocean sound-speed variability, as well as comparison to a corresponding model for the Naval Research Laboratory (NRL) track.[23] The calculations here use the same nonlinear internal wave parameters selected for the NRL track, and the consistency between computations and data on two independent propagation paths supports these selections. Finally, time-averaged intensity variations of similar magnitude are shown to occur in computational results that account for horizontal refraction, while results from 2D models do not demonstrate variations with sufficient amplitude. The agreement between experimental data and a 3D computational model for this portion of the SWARM site is noteworthy. The correlation between an observed nonlinear internal wave packet and horizontal refraction effects in both model results and data provide, as far as can be determined, the first strong evidence that horizontal refraction due to nonlinear internal waves can cause significant signal variability in the ocean.

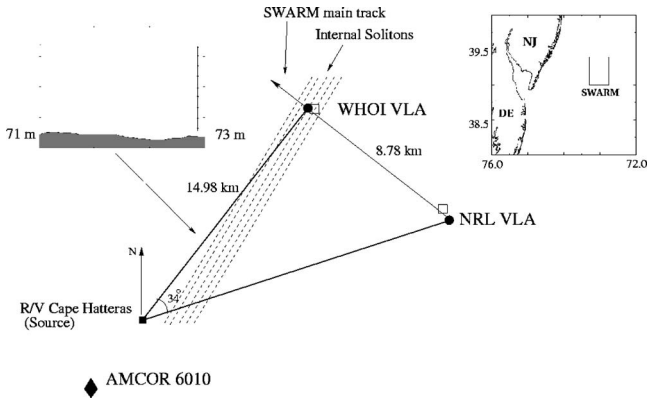


FIG. 1. Geometry of southwest portion of SWARM experiment. Signals from airgun at R/V Cape Hatteras (black square) received by WHOI and NRL VLAs (black circles). Thermistor strings (white squares) associated with each array. AMCOR 6010 site (black diamond) located several kilometers southwest of source. Bathymetry of WHOI waveguide is nearly flat. Two acoustic tracks (dark lines) intersected by internal soliton packet (dashed lines) observed on 4 August 1995.

The Barents Sea study[2] described observed horizontal acoustic phase fluctuations but did not have internal wave packet observations to correlate with these fluctuations.

The paper is organized as follows. Section II briefly reviews environmental and acoustic data available from the two southwest tracks of SWARM. Significant time-frequency and pulse-averaged intensity variability occur at both arrays. Section III provides internal wave and geoacoustic parameters used for 2D modeling of the track waveguides. Data and simulation comparisons show 2D methods appear unable to model acoustic variability at the WHOI VLA. Some aspects of 3D propagation and the adiabatic mode PE method used for 3D computations are reviewed. Section IV shows that the relatively large variations observed in data can be reproduced by incorporating the mechanism of horizontal refraction. Section V contains a summary and discussion of main results.

II. EXPERIMENT BACKGROUND

The full SWARM experiment[5] was conducted in 1995 off the New Jersey coast (see Fig. 1 inset). Figure 1 shows the southwest tracks which are the focus of this study. Suspended at 12 m depth from the R/V Cape Hatteras (indicated by a black square in the lower left) was a 20 in.[3] Bolt airgun which was fired every minute for several hours on 4 August 1995. The source signal was extremely repeatable, with usable bandwidth between 10 and 180 Hz and large energy peaks at 32 Hz and several harmonics.[11]

The airgun signal was received by two VLAs, indicated by black circles in Fig. 1. The primary subject of this study is data from the WHOI telemetered array consisting of 16 hydrophones spaced approximately 3.5 m apart with the top and bottom phones at 14.9 and 67.5 m depths. This array was suspended in 70.5 m of water and was about 15 km from the source. Details of the NRL VLA are provided for comparison. This array consisted of 32 elements spaced 2 m apart with the top and bottom phones at 21 and 85 m depths.

It was suspended in 88 m of water about 18 km northeast of the source and 9 km southeast of the WHOI array, as shown in Fig. 1.

Several packets of nonlinear internal waves passed through the SWARM region on 4 August and were monitored at several locations.[6] CTD readings performed by the R/V Cape Hatteras provide data about sound-speed profile variations at the source. Two thermistor strings (with five thermistors each) collected temperature data at locations of white squares in Fig. 1. The first was attached to the WHOI VLA and the second was attached to the NRL VLA. In addition the R/V Oceanus was near the WHOI array and used radar images of nonlinear internal wave packets to estimate their bearing.[5] During a particular 2 h period on 4 August a large packet of nonlinear internal waves passed between the airgun source and the two VLAs. This packet was recorded by the WHOI thermistor string and observed on CTD readings performed at the R/V Cape Hatteras. It was also observed near the WHOI VLA on radar by the R/V Oceanus. The propagation direction of the packet was estimated from these radar images. Dashed lines in Fig. 1 represent this packet of waves with linear wave fronts. This linear-front idealization is a modeling assumption since satellite images suggest the SWARM region contained a great deal of internal wave activity and that the solitary wave fronts have curvature.[5,24]

For this study we define the incidence angle ϕ as the angle from the propagation direction of the acoustic signal to that of the nonlinear internal wave measured positive counterclockwise. Thus, if $\phi=90°$ the wave fronts are parallel to the acoustic track and are moving from right to left across the track. Note the different incidence angles for the two acoustic propagation tracks on Fig. 1, about 45° for the NRL track and about 85° for the WHOI track (see Fig. 1 of Ref. 11).

Gabor wavelet transforms of broadband signals influenced by nonlinear internal waves reveal complex variations at the NRL VLA.[23] Corresponding wavelet transform results are shown in Fig. 2 for phone 2 (19 m depth) of the WHOI VLA. Phone 2 was chosen because its position at about one-third of the water depth records observations of higher mode arrivals. Results from two signals are shown in three-panel figures. The top panel is a normalized representation of the time domain signal, the right panel is the normalized Fourier spectrum, and the large central panel is the graph of a wavelet transform (scalogram). Figure 2(a) shows the 1945 GMT shot, the spectrum for which shows peak frequencies at 32, 64, and 95 Hz. Group velocity curves are visible in the scalogram and most acoustic energy is in the first two modes of the 32 Hz band and in the late-arriving third mode of the 64 Hz band. Figure 2(b) shows the 1952 GMT shot, 7 min later. Its spectrum indicates more high frequency harmonics, confirmed by the scalogram showing at least three modes near 96 Hz and four or more near 120 Hz. The scalogram also shows most energy in the 64 Hz band arrives as part of the second mode. These variations in the patterns of energy distribution are quasiperiodic and correlate with the passage of the nonlinear internal wave packet. Similar variations occurred in signals at the NRL array.

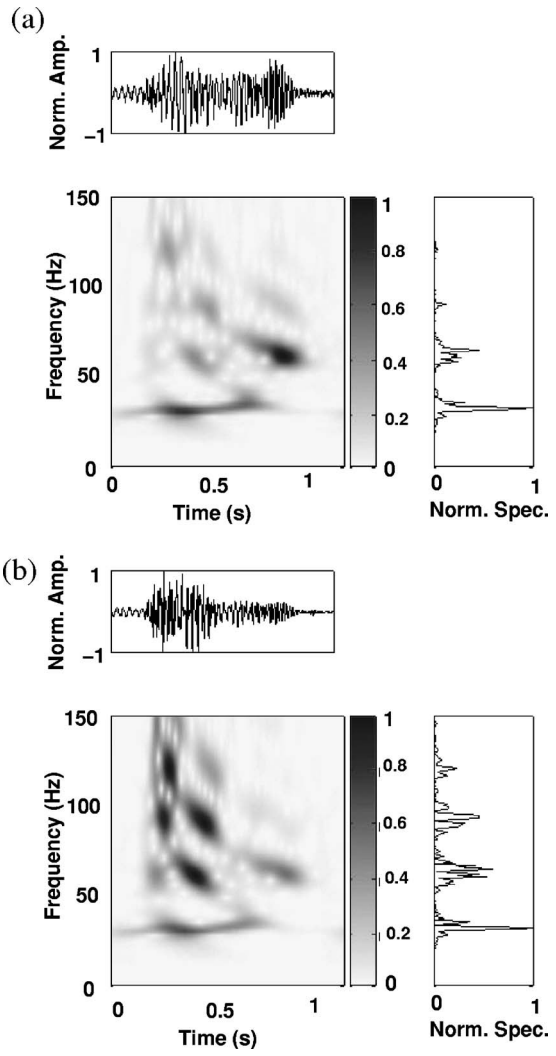


FIG. 2. Time-frequency analysis of WHOI data at phone 2 (19 m depth). (a) 1944 GMT shot, received time domain signal (top panel). Fourier transform (side panel) shows dominant energy peaks at 32, 64, and 95 Hz. Scalogram (shaded panel) represents Gabor wavelet analysis of signal in top panel. Group velocity curves are visible; two, three, and four modes appear near 32, 64, and 95 Hz. Signal is dominated by 32 Hz first mode and 64 Hz third mode energy. (b) 1951 GMT shot (7 min later). Spectrum shows more energy in high frequency bands compared with (a), which is confirmed by scalogram. The 64 Hz band is now dominated by second mode energy.

The pulse-averaged intensity in $\mathrm{W/m^2}$ of a broadband signal $p(t)$, converted to decibel measure, is

$$I_T = 10 \log_{10}\left(\frac{1}{T}\int_0^T \frac{|p(t)|^2}{\rho c} dt\right) \quad \text{dB re: } 1\,\mu\text{Pa}, \tag{1}$$

where T is the the interval of integration, t is time, ρ is water density, and c is an average sound speed. The interval of integration was chosen separately for each array to contain a complete pulse, 1.6 s for the NRL VLA and 1.2 s for the WHOI VLA. The quantity I_T is calculated for the entire hour of airgun shots beginning at 1901 GMT on 4 August for both arrays. Figure 3(a) shows calculations for the WHOI VLA and Fig. 3(b) for the NRL VLA. All hydrophones on each array are shown with depth-averaged variations indicated by thick curves at the bottom. The "geotime" label is used to emphasize that environmental

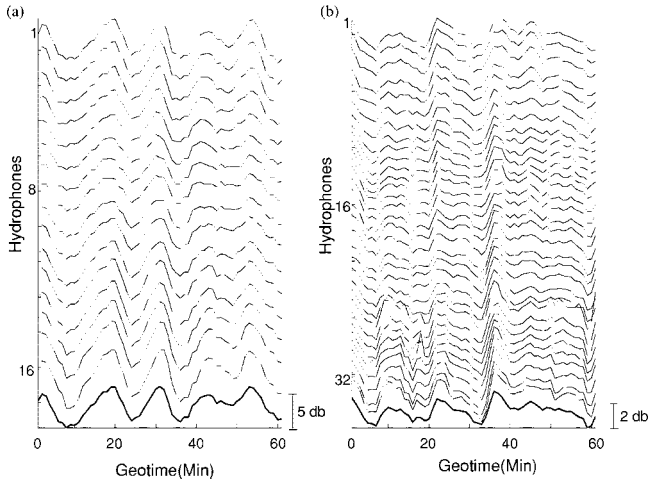


FIG. 3. Intensity I_T for each hydrophone in (a) WHOI and (b) NRL VLA vs geotime starting at 1901 GMT. Quasiperiod of variations at each phone is consistent with observed internal waves. Depth-averaged I_T (dark curves) shown at bottom. Note amplitude of depth-averaged variations in (a) is nearly twice that in (b).

variations occur on scales much larger than the acoustic pulse. Multiple-decibel variations occur at all hydrophones at each array. The occurrence and phase of these variations is independent of depth, but a weak depth dependence can be seen in the amplitude. The variations are quasiperiodic over about 12–14 min, which are well correlated with observations of the traveling nonlinear internal wave packet.[11] The maximum amplitude of depth-averaged variations is approximately 5.9 dB at the WHOI array, which is nearly twice the 3.1 dB amplitude of oscillations at the NRL VLA. This indicates that the nonlinear internal wave packet has an even greater influence on the acoustics propagating along the WHOI track. Figure 3 has been reproduced from Ref. 23 to allow easy comparison of the difference between the I_T variations at each VLA.

III. MODELING

The following describes sound-speed profile, geoacoustic, and nonlinear internal wave models used for the analysis of SWARM data. Then results of 2D PE modeling of the broadband signals from the WHOI track are compared with data. Theoretical estimates for several characteristics of 3D acoustic propagation are obtained. Finally, the method used for 3D simulations is summarized.

A. Environment

Data from CTD casts taken at the Cape Hatteras while the airgun was firing provided 25 sound-speed profiles at the source location on 4 August 1995.[11] The average of these profiles is used as a range-independent mean sound-speed profile in the water column. The profile is downward refracting with a 1534 m/s upper layer and strong thermocline transition, between 10 and 30 m depth, to a 1484 m/s lower layer. Temperature and salinity data collected by the R/V Oceanus during this time period suggests the presence of an upwardly refracting bottom layer along the main track of SWARM, especially off the shelf break (see Fig. 4 in Ref. 5). Due to the location of the WHOI array and the absence of this feature in the Cape Hatteras data, this study focuses on the effects of the nonlinear internal waves by using the range-independent mean sound-speed profile. Bathymetry data for the WHOI track was obtained from the National Geophysical Data Center.[25] The nearly flat bathymetry is shown in an inset of Fig. 1.

The nonlinear internal wave packet is assumed to perturb the mean sound speed. As a model of the packet's effect we use a piecewise linear approximation of the first internal wave gravity mode $\Phi_1(z)$ multiplied by a sum of six sech^2 waves:

$$\eta(r,z) = \Phi_1(z)\sum_{n=0}^{5} A_n \,\mathrm{sech}^2\left[\frac{2\pi(r - r_n + v_n t)}{\Lambda_n}\right], \tag{2}$$

where (r,z) are cylindrical coordinates, A_n represents the amplitude of the nth wave, Λ_n is its width, r_n is its starting position, and v_n is its speed. Time t corresponds to geotime in Fig. 3. While packet dispersion can be included in this model by assigning each wave in the packet a distinct v_n, time-evolution of the packet will not be addressed in this paper. Good estimates of the other parameters can be obtained by matching packet spectral characteristics to those of data. The spectra can be matched well[23] by using evenly spaced waves with $\Lambda_n = 140$ m and $r_n = 2.3n\Lambda_n$ m, although these certainly are not a unique set of parameters. With the assumption of linear internal wave fronts and track geometry described in Sec. II, these values are projected onto each acoustic track to obtain effective wavelength values of $\Lambda_{\mathrm{NRL}} = 195$ m and $\Lambda_{\mathrm{WHOI}} = 1600$ m for 2D simulations. In addition, an internal wave packet speed of $v_n = v = 0.42$ m/s is used to be consistent with the NRL model. Note that this value is lower than propagation speeds reported on different days of the experiment. However, this value—when projected onto the NRL track—is consistent with observations of this packet at the Cape Hatteras,[11,23] and was used to model acoustic data from the NRL VLA. It is important to note that in this analysis procedure of the WHOI VLA data, the *same* water sound speed profiles and nonlinear internal wave parameters are used as for the NRL VLA analysis. Thus, a positive comparison between data and computations on this track will also lend support to the choice of parameters describing this particular nonlinear internal wave packet.

Data from the AMCOR 6010 core is indicated by circles on the dashed-dotted curves in Figs. 4(a)–4(c) for sound speed, attenuation, and density. Distinctive characteristics of this profile are the shallow reflector within 10 m of the ocean-sediment interface and a low sound speed channel above a deep, high sound speed reflector. The presence of the deep reflector is widely recorded for this region.[26] To obtain geoacoustic parameters that modeled results at the NRL VLA, the core data were perturbed until acceptable visual matches were achieved between mode amplitudes and relative arrival times of observed data and broadband PE calculations. The light solid curves in Fig. 4 show the geoacoustic profiles that were obtained for the NRL track using a match-

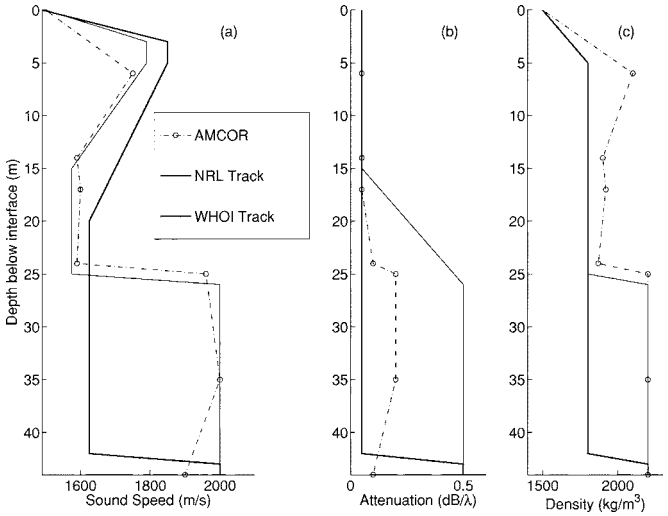


FIG. 4. AMCOR 6010 data (dashed curve with circles), model geoacoustic profiles used in Ref. 23 (solid curve), and model geoacoustic profiles used here (heavy solid curve): (a) sound speed (m/s), (b) attenuation (dB/λ), and (c) density (kg/m^3). Shallow reflector, low sound speed waveguide, and deep strong reflector are preserved in both models. To match modal characteristics in WHOI data, strong reflector must be deeper than indicated by AMCOR.

ing procedure.[23] However, these parameter values do not accurately reproduce characteristics of the broadband data at the WHOI VLA. In each panel of Fig. 5, the top curve represents broadband data from phone 2 of the WHOI VLA and the middle curve represents PE simulations using geoacoustic parameters obtained for the NRL track. Figure 5(a) shows full band signals, while (b), (c), and (d) show signals that have been filtered with a 10 Hz bandwidth order 10 Butterworth filter centered at 32, 64, and 95 Hz, respectively. The strongest observable differences between the data and the middle curves are the late arrival time of the second mode in the 32 Hz band [Fig. 5(b)] and the small amplitude of the third mode in the 64 Hz band [Fig. 5(c)].

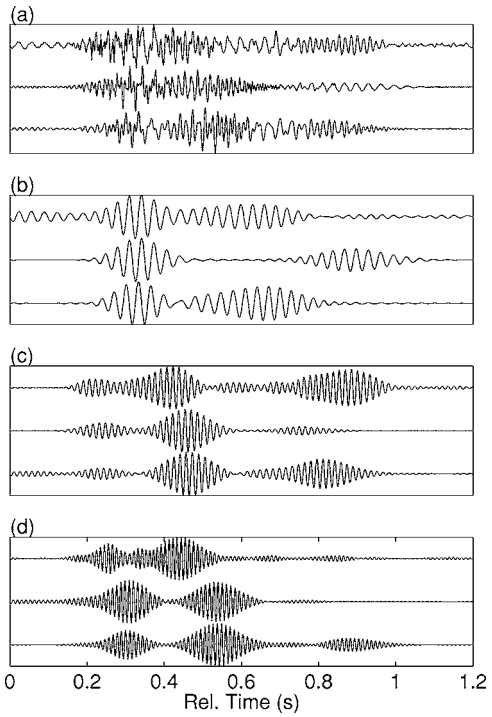


FIG. 5. Comparisons showing modal characteristics among time series of data (top curve in each panel), RAMGEO simulations using Ref. 23 geoacoustic model (middle curves), and simulations using Fig. 4 model (bottom curves). (a) Full frequency band. (b) Results from applying 10 Hz bandwidth Butterworth filter centered at 32 Hz to corresponding full band curves in (a). (c) Same, but filter centered at 64 Hz. (d) Same, but filter centered at 95 Hz. In all panels the bottom curves provide better matches of data modal occurrence, dispersion, and strength than the middle curves.

Figure 5 implies that the geoacoustic profiles along all or part of the WHOI track differ from those along the NRL track. Of course no claim can be made that the NRL profile parameters describe the actual geoacoustics along that track; these parameters provide only a range-independent representation that is useful for acoustic modeling. It is not surprising that actual geophysical variability along the WHOI track requires a different range-independent representation. Thus, the matching procedure is repeated by making physically based perturbations to the light solid profiles in Fig. 4.

Using geoacoustic profiles shown by heavy solid curves in Fig. 4, the bottom curves in each panel of Fig. 5 are obtained. These curves are more consistent with the data. In particular the arrival time of the second mode at 32 Hz is much closer to the observations, as is the amplitude of the higher modes at 64 Hz. More detailed agreement could be sought, but this improvement is sufficient for our modeling. The new geoacoustic profiles preserve principal characteristics of the AMCOR profile, but with a deeper strong reflector and a consequently broader soft-sediment channel between the two reflectors. Note that increased attenuation values in the deep reflector are expected since elastic effects are not incorporated explicitly.[27,28] These bottom parameters will be used for WHOI track models in the remainder of the study. We stress that the profiles in Fig. 4 are not intended as mathematical inversions of the bottom in this region, as that work is being pursued by others.[29]

B. 2D acoustics

For broadband computations and data analysis, a reasonably faithful representation of the source is essential. As described in Ref. 23, the signal available from the source-monitor hydrophone could not be used because it was corrupted by surface and bottom echo returns. A model source representation was developed by adding Gaussian pulses centered at 32 Hz and several harmonics in the frequency domain so that the spectrum of broadband computations was consistent with data spectra at both VLAs. By using the inverse Fourier transform, a signature was obtained that maintains the impulsive character of the airgun and contains acoustic energy in the appropriate frequency bands. Due to its success modeling data at the NRL array, and to maintain consistency between the two tracks, the identical source signature is used for propagation computations on the WHOI track.

Two-dimensional PE calculations successfully reproduced fluctuations in time-frequency behavior and pulse-averaged intensity at the NRL VLA.[23] Geoacoustic layers were assumed to be of constant thickness and follow the bathymetry, so the RAMGEO propagation model is

TABLE I. Column 1 shows Fourier spectrum peak frequencies for nonlinear internal wave packet of 4 August 1995. Internal wave speed of 0.42 m/s gives wavenumber values in column 2 and wavenumber values projected onto the WHOI track in column 4.

WHOI thermistor data			WHOI track	
f_{iw} (mHz)	κ_{iw} (rad/m)	λ_{iw} (m)	κ_{WHOI} (rad/m)	λ_{WHOI} (m)
1.32	0.0192	319	0.001 72	3660
2.43	0.0382	173	0.003 17	1985
3.95	0.0620	106	0.005 16	1216

appropriate.[30] Additionally, it was shown that acoustic wavenumber estimates for the NRL track satisfied the internal wave-acoustic resonance condition

$$\kappa_{iw} \approx k_n - k_m, \tag{3}$$

where κ_{iw} is an effective peak in the horizontal wavenumber spectrum of internal wave disturbances and k_n and k_m are wavenumbers of acoustic modes n and m. Theory and numerical simulations suggest significant mode coupling occurs when this condition is satisfied, leading to observable variations in strength and modal composition of the signal.[1,31] Table I shows frequency (column 1) and wavenumber estimates (column 2) of peak locations in the spectrum of the nonlinear internal wave packet of interest. Spectral information was obtained from WHOI thermistor string data. Wavenumbers in column 2 and wavelengths in column 3 are obtained using an internal wave packet speed estimate of $v=0.42$ m/s. Column 4 of Table I shows the effective spectral peak locations when the packet is projected onto the WHOI track using the angle estimate $\phi=85°$. Table II shows acoustic wavenumbers for the three lowest dominant frequencies in the airgun signal. These values were calculated using geoacoustic parameters given by the dark curves in Fig. 4 and the normal mode program COUPLE.[32] Table III shows differences between selected wavenumbers in Table II. From the corresponding table for the NRL track, numerous opportunities for resonant interaction exist from Eq. (3). However, when κ_{iw} in Eq. (3) is a value of κ_{WHOI} from Table I, Table III indicates how rare the possibilities for resonance interaction are. Only one wavenumber difference (underlined) is within 10% of an effective wavenumber spectrum peak. Consequently, we expect that 2D simulations will not reproduce observed I_T variation. Figure 6 shows this situation, with dashed curves representing depth-averaged I_T from the data and solid curves representing depth-averaged I_T for 60 min of 2D PE simulations using the sound-speed, geoacoustic, and internal wave parameters discussed for an 85° incidence angle as in Sec. III A. Figure 6(a) shows the calculated I_T for computed full band signals, while Fig. 6(b) displays I_T of bandpass filtered signals centered at 32 Hz. The data (dashed line) show large oscillations for each frequency band, while the simulations undergo little or no variation. Similar results occur for the 64 and 95 Hz frequency bands. These results are shown for consistency with the SWARM geometry, and similar graphs are obtained in the case of perpendicular propagation (90° incidence).

TABLE II. Acoustic wavenumbers calculated at source using COUPLE for model environment with no internal waves. Three, six, and eight propagating modes occur near 32, 64, and 95 Hz.

	32 Hz	64 Hz	95 Hz
1	0.131	0.268	0.399
2	0.117	0.258	0.392
3	0.108	0.244	0.381
4		0.240	0.366
5		0.220	0.364
6		0.204	0.345
7			0.332
8			0.316

Several possibilities could explain the inability of 2D PE simulations to model the observed variations. Among these are frequency-dependent bottom attenuation in upper sediment layers[33] or shear processes in the sediment.[34] These and other possibilities were investigated numerically, but with no success in reproducing variations in the data. In addition a large number of different values were examined for internal wave and geoacoustic profile parameters of the models in Sec. III A, also without obtaining significant acoustic variability. Another possibility is that acoustic scattering out of the vertical propagation plane is not negligible. This implies sound speed variations from the nonlinear internal waves cause significant acoustic horizontal refraction,[8] which can be investigated using 3D computational methods.

C. 3D acoustics

In this section we indicate why 3D propagation modeling is expected to resolve the mismatch between data and 2D simulations that was discussed in Sec. III B.

We show first that horizontal sound speed gradients in the internal wave environment are sufficient to produce regions of intensity focusing and defocusing over ranges and angular spreads of interest in the SWARM experiment. For simplicity ray theory is used here, since an approach based on adiabatic modes and horizontal rays[16] is a reasonable approximation to the propagation physics. Also, instead of us-

TABLE III. Differences between selected acoustic wavenumbers from Table II. Comparisons with κ_{WHOI} wavenumbers in Table I shows essentially no opportunities (except for one possibility, underlined) for acoustic mode coupling within the internal wave packet.

	k_1-k_2	k_1-k_3	k_2-k_3	k_2-k_4	k_3-k_4	k_3-k_5	k_4-k_5
32 Hz	0.013 7	0.0224	0.008 76				
64 Hz	0.009 36	0.0239	0.014 5	0.0187	0.004 22	0.0238	0.019 6
95 Hz	0.007 63	0.0184	0.010 8	0.0259	0.015 1	0.0167	<u>0.001 61</u>

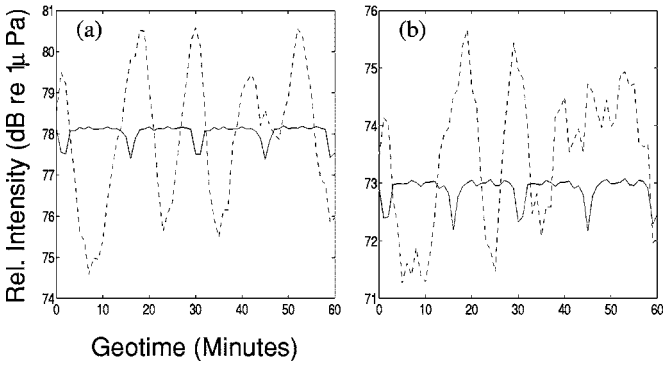


FIG. 6. Depth-averaged intensity variations for data (dashed curves) and 2D PE simulations (solid curves) of WHOI waveguide: (a) full band, and (b) Butterworth filtered band centered at 32 Hz. Relative intensity represents I_T converted to dB re: 1 μPa. Simulations are unable to reproduce large variations in WHOI data in any frequency band.

ing a horizontal sound speed channel formed from one of the nonlinear internal waves in Eq. (2), we use a symmetric approximation consisting of piecewise linear segments. The advantage is that ray paths over constant gradient segments are circular arcs. The sound speed profile for a "focusing" case is shown in Fig. 7(a), where δc is the maximum change produced by the internal wave over a horizontal distance W. The source S is located at horizontal coordinate $y=0$, which we take as the position of the sound speed minimum $c_{\min}$. Consequently the right half of the sound speed profile in the simplified model is given by

$$c(y) = c_{\min} + \delta c \frac{y}{W}, \quad 0 \leqslant y \leqslant W. \tag{4}$$

The ray paths shown represent the trapped rays with maximum horizontal excursions, and they determine the angular spread produced by the channel. It follows from Eq. (4) that the radius of these rays is[35] $R \approx (c_{\min}/\delta c)W$. From geometry half the distance to the focus point is $L = \sqrt{2WR}$ since W/R is small. Therefore, the focus distance is $2W\sqrt{2c_{\min}/\delta c} \approx 22W$, or about every 3.5 km for our parameter choices. The total angular spread of the channel is $2\phi_{\text{crit}} \approx 2W/L \approx 2\sqrt{\delta c/2c_{\min}} \approx 12°$. These estimates provide a preview for simulation results in Sec. IV.

Though the intensity of focusing created will depend on the exact "focusing profile" encountered, we can still estimate the average amount of intensity increase versus range created by soliton ducting by using a simple physical argument.[36] Specifically, acoustic rays encountering the solitons at angles less than or equal to the critical grazing angle will be trapped between the solitons, which are separated by distance $2W$ (as in Fig. 7). For these rays, no cylindrical spreading loss will occur. On the other hand, in the absence of soliton ducting, rays at the critical angle or below will suffer cylindrical spreading, subtending an area $\Delta = R\phi_{\text{crit}}$ at range R (note that we implicitly assume the depth dimension H for areas.) By taking the ratio of the areas, we obtain a mean intensity increase due to ducting at range R, $\Delta/2W = R\phi_{\text{crit}}/2W$, or about 10 for our parameters. This simple "sonar equation" type calculation gives reasonable agreement to experimentally observed numbers.[36]

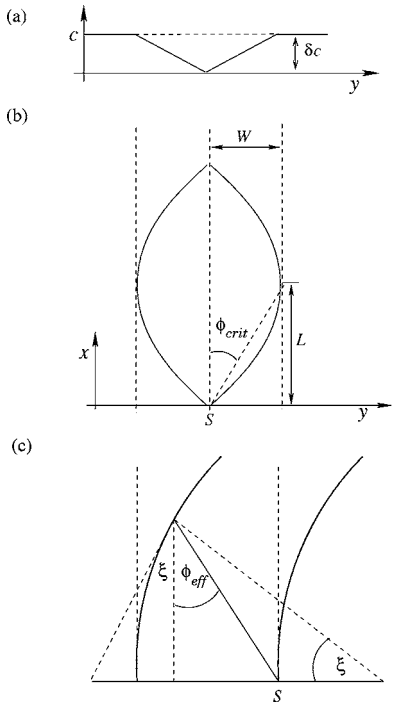


FIG. 7. (a) Idealized representation of horizontal sound-speed variation in focusing waveguide (between two nonlinear internal waves). (b) Variables used for estimate of focal range. Acoustic source is located at S, W is the horizontal distance from middle of trough to peak of nonlinear internal wave, L is half the distance to the focal point, and ϕ represents the acoustic angular spread. (c) Schematic for curved nonlinear internal waves. Acoustic source is located at S, and ξ is determined by radius of curvature.

The above ducting scenario is a reasonable first approximation to many continental shelf internal wave systems, but does not represent all the possible wave systems seen in nature, and also has an "infinite horizontal correlation length" assumption built in that is obviously an approximation.

One geometry that is commonly seen in nature, when solitons are generated by a submarine canyon or valley, is a circular wave front. For this case, acoustic energy ducted between solitary waves will reflect back and forth from the waves at angles which are alternately increased and decreased by an angle ξ. This angle is obtained from the law of sines using ϕ_{crit}, W, and the radius of curvature of the internal waves r_c. The curvature will cause acoustic rays that are pushed above the usual critical grazing angle to leak out of the waveguide, i.e., become untrapped. This results in an "effective critical angle" which is $\phi_{\text{eff}} = \phi_{\text{crit}} - \xi$ for ducted rays in a curved wavefront, as shown in Fig. 7(c).

The finite horizontal correlation length of the internal waves leads to acoustic energy leaking out of the duct between two internal waves to "neighboring ducts," a type of horizontal diffusion process. Though the way this diffusion acts is likely somewhat complicated to model exactly, we can estimate it in a simple fashion. If we assume the model that one duct splits into two ducts of the same area every horizontal correlation length of the waves, then the area subtended by the duct versus range will simply be $\Delta_{\text{diffuse}} = 2WN$, where N is the number of correlation lengths in range and again we have implicitly included the water depth H. A slight variant of this can be considered where the en-

ergy starts its ducting between the leading edge soliton of a wave train and the second wave. In this case, the energy can diffuse only to the back of the wave train, i.e., in one direction, and so the above-quoted area should be reduced by one half. This horizontal diffusion estimate is admittedly crude, but should be able to provide at least order of magnitude numbers for the effect.

D. Adiabatic mode parabolic equation

Since 2D full wave calculations cannot reproduce the large I_T variations observed at the WHOI VLA, it is necessary to determine if acoustic horizontal refraction can. This section summarizes the adiabatic mode parabolic equation (AMPE), which is an efficient method for solving range- and azimuth-dependent propagation problems. The method relies on the adiabatic mode approximation, so that unlike the PE calculations no coupling of acoustic modes is handled. Examples for 3D bathymetry environments are given in Refs. 37 and 38, and more recently, Ref. 39 suggests using a similar method for internal wave environments. The method computes mode coefficients in a full circular region about the source rather than enforcing boundary conditions on a wedge shaped region.[18,20,40]

The method uses a local mode representation of the Helmholtz equation solution in cylindrical coordinates (r,θ,z),

$$p(r,z,\theta)=\sum_n \frac{u_n(r,\theta)}{\sqrt{k_n(r,\theta)}}\psi_n(z;r,\theta), \tag{5}$$

where the modes $\psi_n(z;r,\theta)$ and wave numbers $k_n(r,\theta)$ satisfy the depth operator equation

$$\left(\frac{\rho}{\alpha}\frac{\partial}{\partial z}\frac{1}{\rho}\frac{\partial}{\partial z}\alpha+k^2(r,\theta)\right)\psi_n=k_n^2\psi_n. \tag{6}$$

The factor $\alpha=\sqrt{\rho c}$ is used to conserve energy at vertical interfaces. By substituting Eq. (5) into the Helmholtz equation, using Eq. (6), and neglecting mode coupling terms and wavenumber derivatives, a parabolic equation for the mode coefficients u_n can be obtained:

$$\frac{\partial u_n}{\partial r}=ik_0\sqrt{1+\frac{1}{k_0^2r^2}\frac{\partial^2}{\partial\theta^2}+k_0^{-2}(k_n^2-k_0^2)}u_n, \tag{7}$$

where $k_0=\omega/c_0$ and c_0 is a reference sound speed. The square root of the operator in Eq. (7) can be approximated by Padé coefficients.[37] When accurate wavenumbers and mode shapes are available, AMPE then solves for the u_n's for all values of θ at each range. The necessary wavenumber and mode parameters in Eq. (5) are obtained for the average sound speed profile and the geoacoustic parameters shown by the dark curves in Fig. 4 using COUPLE. The AMPE code was then modified to accept this output from COUPLE and to account for thermocline variations (in addition to bathymetric variations). The code was further modified to compute received complex pressure for construction of the waveguide transfer function for broadband synthesis.

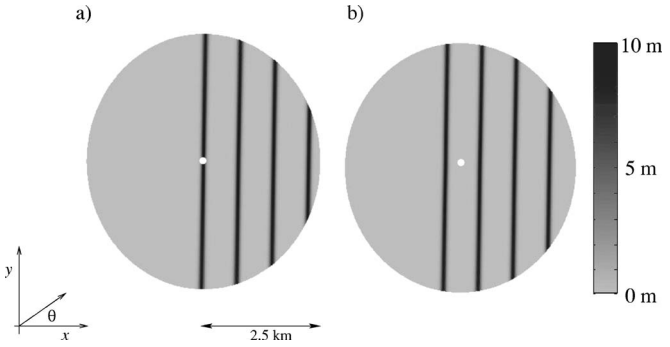


FIG. 8. Top view of portion of 3D simulation environment. Source (white dot) in center of circle. Gray scale shows thermocline variation from nonlinear internal wave model in Eq. (2) using coordinates shown. (a) Internal wave peak near source location produces defocusing of acoustic energy. (b) Source in trough between two internal waves produces acoustic focusing.

IV. THREE-DIMENSIONAL CALCULATION RESULTS

The adiabatic mode PE method is used to determine the horizontal refraction from nonlinear internal waves past a source at 12 m depth. Figure 8 shows the circular region used for AMPE calculations. The assumption of linear wave fronts makes it convenient to introduce a Cartesian (x,y) coordinate system originating at the source with the x direction pointing to the right. The nonlinear internal wave model from Eq. (2) is used with $r\rightarrow x$ and $r_n\rightarrow x_n$ (initial peak locations), $v=0.42$ m/s is the packet velocity, and $\Lambda_n=\Lambda$ is the width of each wave. Shade variations in Fig. 8 represent thermocline depressions from the nonlinear internal waves in Eq. (2). Figure 8(a) shows a situation where an internal wave peak is passing over the source, while Fig. 8(b) shows the source in a trough between two waves.

This coordinate system is also used with the standard (r,θ) coordinates when performing AMPE calculations. Thus, for our simulations, the nonlinear internal waves propagate in the negative x direction and over the source. To model the situation at the WHOI VLA, we examine acoustic results for a range of angles near $\theta=90°$.

A. Single frequency

Because the upper water depths have higher sound speeds than the lower, acoustic waves refract away from nonlinear internal wave peaks. Thus when the source is positioned as in Fig. 8(a), defocusing of acoustic energy occurs near the 90° azimuth in our coordinates. When the source is positioned as in Fig. 8(b), acoustic waves refract into the low sound speed trough between nonlinear internal waves and acoustic focusing results. These effects have been predicted using ray theory,[3] exhibited computationally,[18] and mentioned in connection with experimental results.[14]

Figure 9 shows transmission loss results from AMPE computations for three dominant peak frequencies of the airgun source in a 25 km radius, wedge-shaped region centered at the 90° azimuth. In all panels, light to dark represents high to low transmission loss, and results for the two situations in Fig. 8 are shown. Figure 9(a) displays calculations for a 32 Hz source between internal waves and low loss occurs at the peak of the nonlinear internal wave, with somewhat

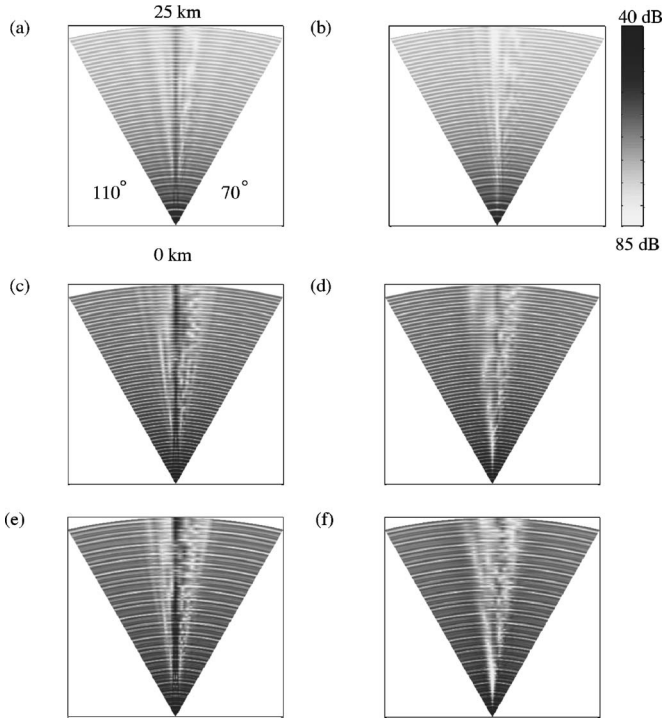


FIG. 9. Wedge-shaped regions are top portions of circles shown in Fig. 8. Contours are CW results at 19 m depth from AMPE calculations. Light shades represent high loss, and dark low. Acoustic energy at 32.5 Hz is (a) focused and (b) defocused. (c), (d) Same for 65 Hz. (e), (f) Same for 95 Hz.

larger losses on either side of the peak. Similar focusing occurs for a 64 Hz source in Fig. 9(c) and for a 95 Hz source in Fig. 9(e). At the higher frequencies it is possible to see acoustic rays converging near 15–18 km. Figure 9(b) shows large losses occur along the 90° azimuth for a 32 Hz source at a nonlinear internal wave peak. Defocusing is also present in Figs. 9(d) and 9(f) for a 64 and 95 Hz source. The higher frequencies exhibit ray-like refraction near the 90° azimuth.

B. Broadband

AMPE was used to compute the waveguide transfer function near three peak frequencies of 32, 64, and 95 Hz at all azimuths of the 3D environment. Broadband pulses were obtained using the source representation developed in Ref. 23 and standard Fourier synthesis techniques. Broadband pulses obtained from AMPE from an environment with no internal waves preserve the relevant amplitude and arrival characteristics for the dominant modes of each frequency band shown in Fig. 5.

Simulated broadband pulses from the 87° azimuth are shown in Fig. 10 for defocused (dark curves) and focused (light curves) situations. Figure 10 shows pulses normalized by the maximum of the defocused signal for (a) full band, and Butterworth filtered signals centered at (b) 32, (c) 64, and (d) 95 Hz. I_T calculations are shown next to each curve. Large dB differences exist in all frequency bands. Figure 10(b) shows a notable change in the 32 Hz band second mode arrival. Modal arrival times in Figs. 10(c) and 10(d) show small differences, but the primary effect is the decreased amplitude of focused curves.

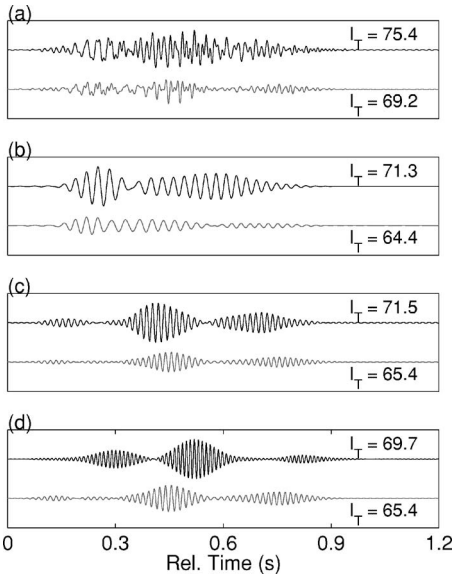


FIG. 10. Broadband AMPE simulations for defocused (dark curve) and focused (light curve) internal wave situations. (a) Full band. (b) Butterworth filter centered at 32 Hz. (c) Same, but filter centered at 64 Hz. (d) Same, but filter centered at 95 Hz. I_T calculations are shown next to each curve and indicate several dB differences occur between defocused and focused situations.

Figure 11(a) shows Gabor wavelet analysis of the defocused acoustic pulse in Fig. 10(a). The spectrum and scalogram both show expected energy peaks in the 64 and 95 Hz bands. Figure 11(b) shows the spectrum and scalogram for the focused pulse. The spectrum shows less energy arriving in the 32 Hz band and more arriving in higher frequency 95 and 120 Hz bands, consistent with patterns in the data. Time-frequency variations in higher frequency bands are not as prevalent as they appear in data, though modal arrival times are affected. In an adiabatic setting we do not anticipate significant modal interaction. However, Fig. 11 shows that horizontal refraction effects appear to have caused observable time-frequency variations of these pulses and suggests this is a possible mechanism for similar variations in the data.

A comparison between the complete evolution of data and computation acoustic pulses over geotime is shown in Fig. 12. Figure 12(a) shows geotime behavior for 31 consecutive 32 Hz band pulses recorded at the WHOI VLA phone 2 (19 m). Clear amplitude variations occur that correspond to observed fluctuations in I_T. As geotime increases, both first and second mode amplitudes increase and the second mode tends to arrive closer to the first mode. Once the maximum amplitudes are achieved, the magnitudes of both modes decrease to a minima near minute 12, then increase, and decrease again, with the second mode arrivals exhibiting a pattern similar to the first several minutes. Figure 12(b) shows 31 simulated geotime minutes at 15 km range, 19 m depth, and 87° azimuth for a packet with $\Lambda = 140$ m, $x_n = 2.3n\Lambda$ m, and $v = 0.42$ m/s. The amplitude variation patterns of both modes are fully consistent with data in Fig. 12(a), as are variations in relative second mode arrival time.

Figure 13 compares pulse-averaged intensity I_T for data (dashed curve) at hydrophone 2 (19 m) in the 32 Hz band and the corresponding computation (solid curve with circles)

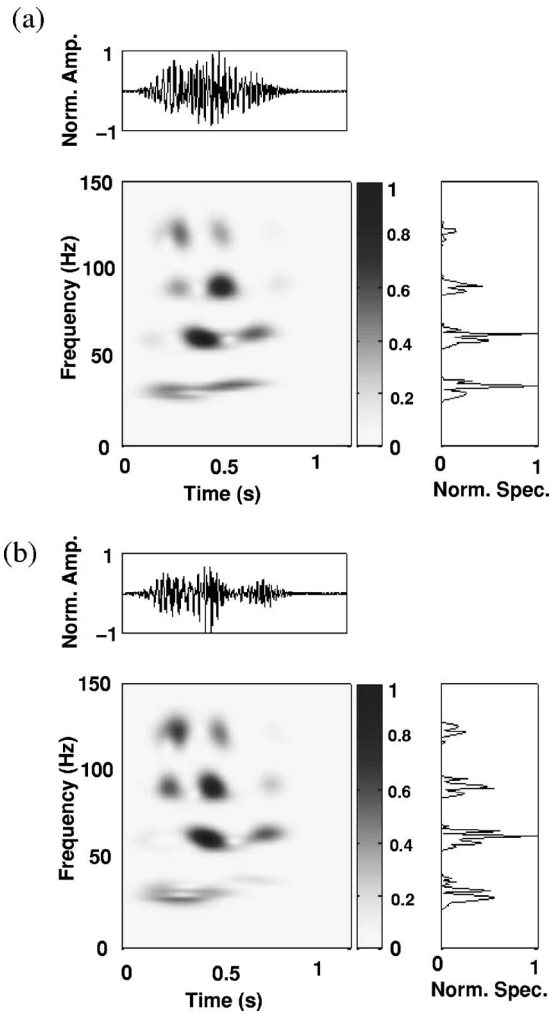


FIG. 11. Gabor wavelet analysis of AMPE simulations at 19 m depth for (a) focused and (b) defocused pulses in Fig. 10. Focused pulse shows more acoustic energy at 32 Hz, while defocused pulse has more energy in higher frequency bands. These features are consistent with time-frequency analyzed data.

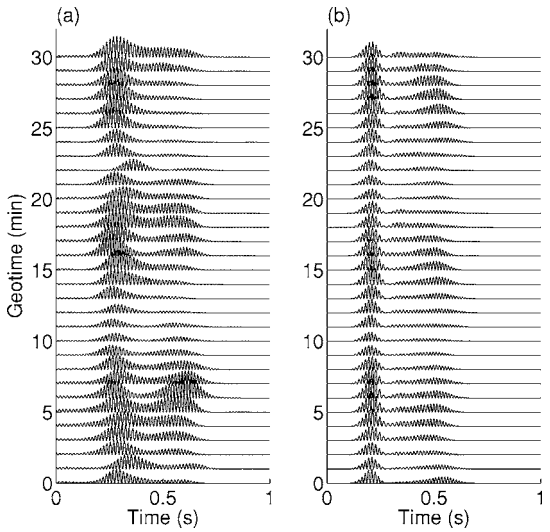


FIG. 12. (a) Pulse time-series data at WHOI phone 2 (19 m depth) in 32 Hz band for 30 min. Variations in signal amplitude correspond with I_T oscillations. Second mode arrival time also varies. (b) Corresponding results from broad-band AMPE simulations centered at 32 Hz and received at 15 km on the 87° azimuth. Amplitude variation of dominant mode arrival times are consistent with data.

from the 87° azimuth. The occurrence of maximum and minimum I_T values are consistent between data and simulations, as suggested by Fig. 12. The amplitude of the data fluctuations ranges up to nearly 6 dB. The simulation curves have a maximum variation of about 10 dB and shows three peaks, with the time between peaks essentially the same as the data. This figure clearly shows that the large variations observed in the data can result from horizontal refraction. It is emphasized that the nonlinear internal wave model and parameters are the same as those used for the other southwest track,[23] although, of course, neither these values nor the model should be considered unique. The main point is that the same ocean environmental characterization can account for the observed acoustic variations on two distinct tracks.

Calculated broadband geotime pulse-averaged intensity variations for the 32 Hz-centered band are visible for up to ±10° on either side of the 90° azimuth. In Fig. 14 shadings toward white indicate I_T values for a broadband pulse (at any particular geotime and azimuth) that are above the mean, while shadings toward black indicates values below the mean. Two prominent white stripes near 90° result from focusing and defocusing effects of the nonlinear internal wave channel. The amplitudes of the variations tend to diminish in amplitude further from 90°. The solid curve in Fig. 13 corresponds to the cross section at 87°.

V. SUMMARY AND DISCUSSION

Broadband data from the WHOI VLA of SWARM shows time-frequency variation patterns that have similar patterns to those at the NRL VLA. However, pulse-averaged intensity variations at the WHOI VLA have considerably larger amplitudes than those at the NRL array. This suggests that the nonlinear internal wave packet which traverses both tracks has a more substantial influence on the WHOI VLA

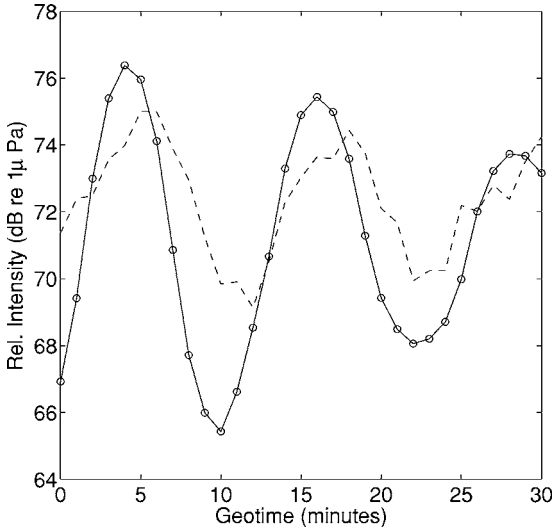


FIG. 13. I_T variations for data (dashed curve) and AMPE simulations (solid curve) at phone 2 (19 m depth) for signals centered at 32 Hz. Adiabatic-mode horizontal refraction can account for large I_T variations observed in data. Peak locations support selected values for internal wave parameters.

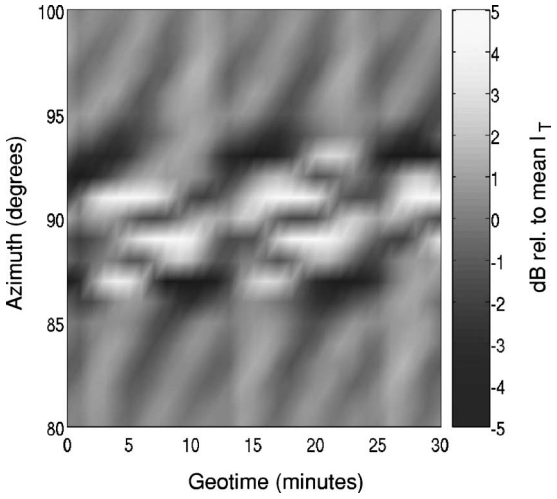


FIG. 14. Contour plot of variations from geotime mean I_T vs azimuth and geotime for 32 Hz broadband computations. Light represents increased I_T and dark represents decreased I_T. Oscillatory patterns arise for at least 10° on both sides of the 90° azimuth.

signals than those on the NRL VLA. Traditional 2D PE simulations evidently do not model fluctuations in broadband signals when nonlinear internal wave fronts are nearly parallel to the acoustic propagation direction. In this particular "high incidence angle" situation, 2D mechanisms such as mode coupling and internal wave-acoustic resonance (which are effective for lower incidence angles) apparently cannot produce sufficiently strong intensity variations. For high incident angles the environmental gradients are large enough that significant energy propagates out of vertical planes, as previous numerical computations have demonstrated. Consequently, acoustic horizontal refraction must be taken into account for the low frequency bands in this experiment.

Rough theoretical estimates calculated for the critical deflection angle and focal range of horizontally refracted acoustic rays are consistent with both experimental observations and model calculations. Similarly, an acoustic intensity amplitude estimate also provides overall agreement. The effect of small curvature of nonlinear internal wave fronts is addressed briefly.

Calculations using the adiabatic mode PE for the WHOI track show focusing and defocusing of acoustic energy as nonlinear internal waves pass a source. Large azimuthal variations occur in CW transmission loss for 32, 65, and 95 Hz. Broadband synthesis for dominant frequency bands in the airgun source show that variations of about 6 dB in pulse-averaged I_T arise using a model packet of nonlinear internal waves. Fluctuations in the time-frequency behavior of these calculations are also consistent with data from the WHOI VLA. The principal result is that the propagation model calculations provide strong evidence that observations at the WHOI VLA of SWARM show the influence of horizontal refraction due to nonlinear internal waves. To our knowledge, this is the first evidence of this type for three-dimensional refraction effects in the literature. Results in Ref. 41, which emphasize horizontal rays and vertical modes rather than a computational model for the track, conclude that horizontal refraction is the likely cause of observed intensity variations. A critical feature of the analysis here is that the "tuning" is done for the ocean environmental model of the nonlinear internal waves. Exactly the same assumptions and parameter values are used as for the previous analysis of the 4 August 1995 packet along a distinct acoustic track.[23] This does not imply that the environmental model is either optimal or unique, but rather that the mechanisms behind the acoustic variability on both tracks appear to be robust. Indeed, frequency-dependent results and related analysis reiterate the robustness of horizontal refraction effects on the WHOI track.

Three-dimensional broadband calculations are computationally intensive even with the relatively efficient AMPE algorithm, so geotime simulations remain to be performed for higher frequency bands from the airgun. A more complete sensitivity study of geoacoustic and internal wave parameters on horizontal refraction effects would be worthwhile. While the assumption of linear internal wave fronts is reasonable,[24] further investigation of how wave front curvature affects horizontal refraction and subsequent intensity and time-frequency variations would extend this work. The interaction between 2D mechanisms such as mode-coupling or resonance and horizontal refraction, which is an inherently 3D effect, is an important question. Earlier work[8] suggests that this type of interaction may occur for incidence angles above about 70°. Due to the adiabatic nature of the 3D model in this study, this question was not addressed. However, it would be interesting to obtain estimates of the azimuths where mode coupling effects become negligible and horizontal refraction effects become significant. Finally, seafloor bathymetry interactions with nonlinear internal wave packets are not addressed here, and recent work suggests that this combination can be important.[42]

ACKNOWLEDGMENTS

The authors gratefully acknowledge Dr. Yongke Mu for assistance reading the acoustic data and Dr. Michael Collins for providing the original AMPE code. We also acknowledge the work of the scientists associated with the SWARM 95 experiment which received support from NRL base funds. This work was supported by an ONR Ocean Acoustics Graduate Traineeship Award and by ONR grants to Rensselaer, the University of Delaware, and the Woods Hole Oceanographic Institution. The principal contributions of this work were taken from a Ph.D. thesis by S. D. Frank, submitted to Rensselaer Polytechnic Institute in July 2003. This is WHOI contribution No. 11257.

[1]J. Zhou, X. Zhang, and P. H. Rogers, "Resonant interaction of sound waves with internal solitons in the coastal zone," J. Acoust. Soc. Am. **90**, 2042–2054 (1991).
[2]A. Y. Shmelerv, A. A. Migulin, and V. G. Petnikov, "Horizontal refraction of low-frequency acoustic waves in the Barents Sea stationary acoustic track experiment," J. Acoust. Soc. Am. **92**, 1003–1007 (1992).
[3]B. G. Katsnel'son and S. A. Pereselkov, "Low-frequency horizontal acoustic refraction caused by internal wave solitons in a shallow sea," Acoust. Phys. **46**, 779–788 (2000).
[4]J. F. Lynch, M. H. Orr, and S. N. Wolf, "Low frequency acoustic propagation through shallow water internal waves," in *Sound Propagation Through Internal Waves* (unpublished).
[5]J. R. Apel, M. Badiey, C.-S. Chiu, S. Finette, R. H. Headrick, J. Kemp, J.

F. Lynch, A. E. Newhall, M. H. Orr, B. H. Pasewark, D. Tielbuerger, A. Turgut, K. von der Heydt, and S. N. Wolf, "An overview of the 1995 SWARM shallow-water internal wave acoustic scattering experiment," IEEE J. Ocean. Eng. **22**, 465–499 (1996).
[6]R. H. Headrick, J. F. Lynch, J. N. Kemp, A. E. Newhall, K. von der Heydt, J. R. Apel, M. Badiey, C.-S. Chiu, S. Finette, M. H. Orr, B. Pasewark, A. Turgut, S. N. Wolf, and D. Tielbuerger, "Acoustic normal mode fluctuation statistics in the 1995 SWARM internal wave scattering experiment," J. Acoust. Soc. Am. **107**, 201–220 (2000).
[7]D. Rouseff, A. Turgut, S. N. Wolf, S. Finette, M. H. Orr, B. H. Pasewark, J. R. Apel, M. Badiey, C.-S. Chiu, R. H. Headrick, J. F. Lynch, J. N. Kemp, A. E. Newhall, K. von der Heydt, and D. Tielbuerger, "Coherence of acoustic modes propagating through shallow water internal waves," J. Acoust. Soc. Am. **111**, 1655–1666 (2002).
[8]J. C. Preisig and T. F. Duda, "Coupled acoustic mode propagation through continental-shelf internal solitary waves," IEEE J. Ocean. Eng. **22**, 256–269 (1997).
[9]T. F. Duda and J. C. Preisig, "A modeling study of acoustic propagation through moving shallow water solitary wave packets," IEEE J. Ocean. Eng. **24**, 16–32 (1999).
[10]D. Tielbuerger, S. Finette, and S. N. Wolf, "Acoustic propagation through an internal wave field in a shallow water waveguide," J. Acoust. Soc. Am. **101**, 789–808 (1997).
[11]M. Badiey, Y. Mu, J. F. Lynch, J. R. Apel, and S. N. Wolf, "Temporal and azimuthal dependence of sound propagation in shallow water with internal waves," IEEE J. Ocean. Eng. **27**, 117–129 (2002).
[12]D. Rubenstein and M. H. Brill, "Acoustic variability due to internal waves and surface waves in shallow water," in *Ocean Variability and Acoustic Propagation*, edited by J. Potter and A. Warn-Varnas (Kluwer Academic, Boston, 1991), pp. 215–228.
[13]D. Rubenstein, "Observations of cnoidal internal waves and their effect on acoustic propagation in shallow water," IEEE J. Ocean. Eng. **24**, 346–357 (1999).
[14]O. C. Rodriguez, S. Jesus, Y. Steephan, X. Demoulin, M. Porter, and E. Coelho, "Nonlinear soliton interaction with acoustic signals: Focusing effects," J. Comput. Acoust. **8**, 347–363 (2000).
[15]J. F. Lynch and P. H. Dahl, "Overview of ASIAEX field experiments in the South and East China Seas," J. Acoust. Soc. Am. **112**, 2360 (2002).
[16]R. Burridge and H. Weinberg, "Horizontal rays and vertical modes," in *Wave Propagation and Underwater Acoustics*, Lecture Notes in Physics, Vol. 70, edited by J. Keller and J. S. Papadakis (Springer, New York, 1977), Chap. 2.
[17]Y. A. Kravtsov, V. M. Kuzkin, and V. G. Petnikov, "Perturbation calculation of the horizontal refraction of sound waves in a shallow sea," Sov. Phys. Acoust. **30**, 45–47 (1984).
[18]R. Oba and S. Finette, "Acoustic propagation through anisotropic internal wave fields: Transmission loss, cross-range coherence, and horizontal refraction," J. Acoust. Soc. Am. **111**, 769–784 (2002).
[19]S. Finette and R. Oba, "Horizontal array beamforming in an azimuthally anisotropic internal wave field," J. Acoust. Soc. Am. **114**, 131–144 (2003).
[20]K. B. Smith, C. W. Miller, A. F. D'Agostino, B. Sperry, J. H. Miller, and G. R. Potty, "Three-dimensional propagation effects near the Mid-Atlantic Bight shelf break (L)," J. Acoust. Soc. Am. **112**, 373–376 (2002).
[21]R. F. Henrick, M. J. Jacobson, and W. L. Siegmann, "General effects of currents and sound-speed variations on short-range acoustic transmission in cyclonic eddies," J. Acoust. Soc. Am. **67**, 121–134 (1980).
[22]K. G. Hamilton, W. L. Siegmann, and M. J. Jacobson, "Simplified calculation of ray-phase perturbations due to ocean-environment variations," J. Acoust. Soc. Am. **67**, 1193–1206 (1980).
[23]S. D. Frank, M. Badiey, J. F. Lynch, and W. L. Siegmann, "Analysis and modeling of broadband airgun data influenced by nonlinear internal waves," J. Acoust. Soc. Am. **116**, 3404–3422 (2004).
[24]A. K. Liu, "Analysis of nonlinear waves in the New York Bight," J. Geophys. Res. **93**, 12317–12329 (1988).
[25]National Geophysical Data Center, NOAA, *Hydrographic Survey Data*, Vol. 1, version 3.3.
[26]J. D. Milliman, A. Jiezao, L. Anchun, and J. I. Ewing, "Late quaternary sedimentation on the outer and middle New Jersey Continental shelf: Result of two local deglaciations?," J. Geol. **98**, 966–976 (1990).
[27]J. M. Hovem and A. Kristensen, "Reflection loss at a bottom with a fluid sediment layer over a hard solid half-space," J. Acoust. Soc. Am. **92**, 335–340 (1992).
[28]C. T. Tindle and Z. Y. Zhang, "An equivalent fluid approximation for a low shear speed ocean bottom," J. Acoust. Soc. Am. **91**, 3248–3256 (1992).
[29]A. Turgut and S. N. Wolf, "Matched-field inversion of seabed geoacoustic properties complemented by chirp sonar surveys," J. Acoust. Soc. Am. **110**, 2661(A) (2001).
[30]M. D. Collins, "RAMGEO 1.5," URL ftp://albacore.nrl.navy.mil/RAM.
[31]B. G. Katsnel'son and S. A. Pereselkov, "Resonance effects in sound scattering by internal wave packets in a shallow sea," Acoust. Phys. **44**, 684–689 (1998).
[32]R. B. Evans, "A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom," J. Acoust. Soc. Am. **74**, 188–195 (1983), URL http://oalib.saic.com/Modes/couple/.
[33]I. Rozenfeld, W. M. Carey, P. G. Cable, and W. L. Siegmann, "Modeling and analysis of sound transmission in the Strait of Korea," IEEE J. Ocean. Eng. **26**, 809–819 (2001).
[34]Z. Y. Zhang and C. T. Tindle, "Improved equivalent fluid approximations for a low shear speed ocean bottom," J. Acoust. Soc. Am. **98**, 3391–3396 (1995).
[35]F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics* (Springer, Berlin, 2000).
[36]J. F. Lynch, J. A. Colosi, G. Gawarkiewicz, T. F. Duda, A. D. Pierce, M. Badiey, B. Katsnelson, J. E. Miller, W. L. Siegmann, C. Chiu, and A. Newhall, "Inclusion of finescale coastal oceanography and 3-D acoustics effects into the ESME sound exposure model," IEEE J. Ocean. Eng. (submitted).
[37]M. D. Collins, "The adiabatic mode parabolic equation," J. Acoust. Soc. Am. **94**, 2269–2278 (1993).
[38]B. Coury, "Energy conservation and interface conditions for parabolic approximations to the Helmholtz Equation," Ph.D. thesis, Rensselaer Polytechnic Institute, 1996.
[39]B. G. Katsnel'son, S. A. Pereselkov, V. G. Petnikov, K. D. Sabinin, and A. N. Serebryanyi, "Acoustic effects caused by high-intensity internal waves in a shelf zone," Acoust. Phys. **47**, 424–429 (2001).
[40]K. B. Smith, "A three-dimensional propagation algorithm using finite azimuthal aperture," J. Acoust. Soc. Am. **106**, 3231–3239 (1999).
[41]M. Badiey, B. G. Katsnelson, J. F. Lynch, S. Pereselkov, and W. L. Siegmann, "Measurement and modeling of three-dimensional sound intensity variations due to shallow-water internal waves," J. Acoust. Soc. Am. **90**, 613–625 (2005).
[42]A. D. Pierce and J. F. Lynch, "Whispering-gallery-mode trapping of sound in shallow water between an up-slope region and internal wave solitons," J. Acoust. Soc. Am. **113**, 2279(A) (2003).

Time-reversal focusing of elastic surface waves

Pelham D. Norville[a] and Waymond R. Scott, Jr.[b]

School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 3 December 2004; revised 7 April 2005; accepted 6 May 2005)

Time-reversal focusing is experimentally applied in an elastic medium in the presence of multiple scattering objects on the order of a wavelength in size. The effectiveness of time-reversal is compared to time-delayed focusing and uniform excitation of the transducer array for focusing energy to a desired location within the medium. A filter is also designed to improve the bandwidth of the excitation signal. Time-reversal focusing is investigated in the context of an elastic-wave landmine detection system. Results are presented demonstrating the advantages and limitations of time-reversal in excitation of a resonance in a TS-50 landmine buried in the medium. A special case is presented for a landmine buried in shadow regions where uniform excitation fails to illuminate the target while time-reversal focusing yields improved target illumination. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1945468]

PACS number(s): 43.35.Pt, 43.20.El, 43.20.Fn, 43.20.Tb [PEB] Pages: 735–744

I. INTRODUCTION

A landmine detection system (Fig. 1) is under development at the Georgia Institute of Technology that functions by exciting elastic waves that propagate through the soil.[1] A feature of this system is a noncontact sensor that is used to measure ground motion, making it possible to sense motion directly above a landmine. While multiple wave types are generated by the system's excitation signal, the wave of primary importance in detecting landmines is the Rayleigh surface wave. This wave propagates near the surface along the boundary between the air and the soil and interacts with objects buried in the medium. For most objects, this interaction is observed as scattering of the Rayleigh wave front off of the object. When the buried object is a landmine, due to its structure, and the depth at which it is usually buried, the Rayleigh wave may excite a resonance in the layer of soil between the surface and the flexible top of a landmine. This resonance enhances the surface displacements and is the primary detection cue for buried landmines.[1]

Scattering off clutter objects in the medium causes the Rayleigh wave to become disorganized. If a large number of objects are present, the scattering can interfere with the Rayleigh wave to the point that it no longer effectively illuminates the buried landmine. Any resonance that is excited will be difficult to detect in the presence of the numerous scattered waves reflecting off objects in the medium. By applying time-reversal focusing methods to seismic detection techniques, energy can be focused to a specific location within the medium, irrespective of the presence of clutter. This allows one to focus energy to a certain spot in order to excite a resonance in any target that may be present there.

The primary experimental and numerical investigations of time-reversal focusing have been in the ultrasound frequency range and in the far field.[2–4] Time-reversal focusing has been investigated in both fluid and solid media, including liquid-solid interfaces. In fluid media, homogeneous backgrounds have been augmented with scattering objects to create high order scattering of incident waves.[5] In these scenarios, time-reversal focusing has been shown to be effective in inhomogeneous media,[6] even producing super-resolution effects in some cases.[5,7,8] In solid media, inhomogeneity from microstructural defects has been examined. These experimental investigations form a solid foundation in the exploration of time-reversal focusing, but the experimental cases that have been examined are still significantly different from those encountered in the landmine or buried target detection problem.

The research presented in this paper considers the effectiveness of time-reversal in the significantly different problem of seismic buried object detection.[9] In this detection problem, sources are in the near field, only a few wavelengths from the targets and scattering objects. The number of scattering objects per unit area is also significantly less than in previously considered scenarios.[6,10,11] The seismic system differs significantly from ultrasound systems in that energy is coupled directly into the soil, rather than through a fluid. The coupling of the transducer motion into the soil significantly alters the frequency response of the excited wave.

This paper first presents the basic theory of time-reversal focusing with respect to its application to elastic media. This is followed by a description of the method used to implement elastic wave time-reversal focusing in a laboratory at the Georgia Institute of Technology. This description will include a discussion of the experimental facility and the measurement techniques. The design of a filter used to improve the bandwidth of the excitation signals is also presented. Results from several experiments are shown and analyzed. An ensuing evaluation of the experimental results demonstrates the effectiveness as well as the limitations of time-reversal focusing in buried object detection.

[a]Electronic mail: norville@ece.gatech.edu
[b]Electronic mail: waymond.scott@ece.gatech.edu

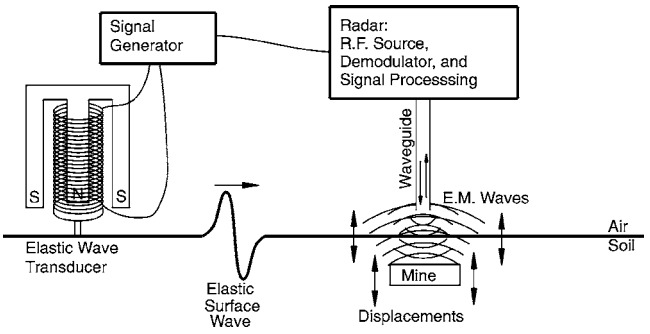


FIG. 1. Schematic of the elastic wave landmine detection system.

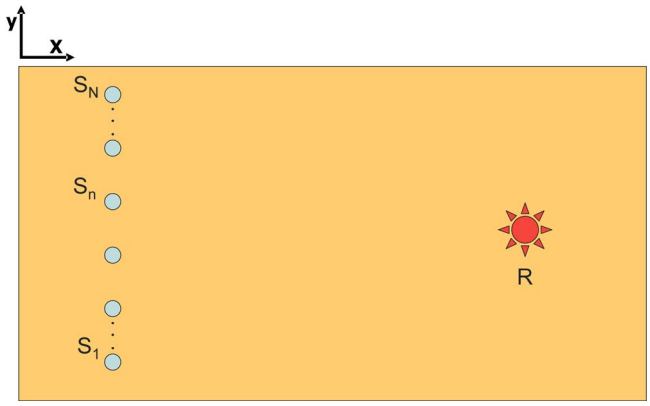


FIG. 2. (Color online) Geometry of the implementation of time-reversal focusing.

II. BASIC TIME-REVERSAL THEORY

The governing wave equation for elastic waves in solids serves as the starting point for a basic analysis of time-reversal,

$$\rho_s \frac{\partial^2 \boldsymbol{u}}{\partial t^2} = (\lambda + 2\mu)(\nabla(\nabla \cdot \boldsymbol{u})) - \mu(\nabla \times (\nabla \times \boldsymbol{u})), \tag{1}$$

where $\boldsymbol{u}$ is displacement, λ and μ are the Lamé constants of the medium and ρ_s is the density.

This equation is valid for the case for which there are no external forces (body forces) present on the medium. It also assumes that the medium is lossless with respect to wave propagation. The assumption of a lossless medium is not physical, but if the losses are very small, the additional terms in the equation have a negligible effect and can be ignored.

An examination of the wave equation shows that there are only second-order time derivatives present. Because of the lack of odd-order time derivatives, if there is a solution to this equation $\boldsymbol{u}(\boldsymbol{r},t)$, then $\boldsymbol{u}(\boldsymbol{r},-t)$ must also be a solution to this equation. Because experimentally it is necessary to work with reverse time in a causal fashion, a finite time duration must be selected over which the equation will be considered. The formulation $\boldsymbol{u}(\boldsymbol{r},T-t)$ over the interval $(T,0)$ satisfies the causality requirement. If all energy in the spatial region of interest is small outside of this time interval, then this solution should be almost exactly equal to $\boldsymbol{u}(\boldsymbol{r},-t)$.

A time-reversal cavity is a three-dimensional (3D) surface which is constructed around a location of interest, usually a source location. All waves impinging on this surface are recorded, time-reversed, and re-transmitted. Classical time-reversal focusing further simplifies this to a time-reversal mirror (TRM) where only a portion of the time-reversal cavity is realized. The TRM concept is well documented in the literature.[6]

In the case of elastic surface waves, the principle wave mode of interest is the Rayleigh wave, a surface wave that decays exponentially with depth. Though some energy is lost from mode conversion and from scattering objects in the soil, most of the Rayleigh wave's energy remains near the surface. Given that landmines are buried near the surface and the energy in the Rayleigh wave is concentrated in that region, the landmine detection problem is approached here as a quasi-two-dimensional problem.

To construct a TRM, receivers are realized as a simple array. The array subtends some angle of the 3D surface that would be necessary to surround the focus point. This array is much more practical to implement than a time-reversal cavity, but it is subject to the limitations of array techniques. The number and spacing of the array elements will have effects on grating lobes. The spot size of the focus point is also limited by the TRM aperture and diffraction effects proportional to wavelength.[2] The TRM concept has been verified in the ultrasound regime for fluid and elastic media.[10,12]

III. THE TIME-REVERSAL FOCUSING METHOD

A. Implementation

For the experimental implementation of time-reversal focusing, elastic wave sources are located in an array $S_n=(x_{S_n},y_{S_n}|n=1,2,\ldots N)$ (Fig. 2). First, consider the effect of time-reversal from a single source, S_n.

Step 1: Transmit an excitation signal, $\epsilon(t)$, from source S_n.

Step 2: Receive a signal, $f_n(t)$, at the desired focusing location, R. Propagation through the medium is described by a Greens function, $G(S_n,R,t)$ such that

$$f_n(t) = \epsilon(t) * G(S_n,R,t). \tag{2}$$

Step 3: Time-reverse the received signal: $f(t) \Rightarrow f(-t)$.

Step 4: Transmit the time-reversed signal, $f(-t)$, from S_n and record at any location on the surface, $\boldsymbol{r}$, such that the signal recorded at $\boldsymbol{r}$ is

$$U_n(\boldsymbol{r},t) = [\epsilon(-t) * G(S_n,R,-t)] * G(S_n,\boldsymbol{r},t). \tag{3}$$

Recalling the associative property of convolution, $U_n(\boldsymbol{r},t)$ then is the cross correlation of the two Greens functions convolved with the time-reversed excitation function, $\epsilon(-t)$. In the special case when $\boldsymbol{r}=R$, this becomes the autocorrelation function. This yields a mathematical explanation for the observed focusing of the signal that occurs at R.

This process can be extended to include additional transmitters in the array such that

$$U(\boldsymbol{r},t) = \sum_{n=1}^{N} [\epsilon(-t) * G(S_n,R,-t)] * G(S_n,\boldsymbol{r},t). \tag{4}$$

In the experimental implementation of this method, **steps 1–3** are performed once for each transmitter S_n in the

array. **Step 4** is performed simultaneously for all transmitters $S_{1\ldots N}$.

Traditional time-reversal focusing using a TRM requires that either a source be located at the desired focus location (R) or that an excitation be launched from the transducer array. In the latter case, after the excitation is launched from the transducer array, reflections off a target at the focal location act as a passive source. These reflections are recorded at the TRM, time-reversed, and retransmitted. In the landmine or buried target detection problem, the signal reflected off a target is not strong enough to be significantly above the noise floor. This makes it impractical to use reflected signals as a source for time-reversal focusing. Further, in the case of landmine detection, it would be unwise to place a seismic source at a location where a landmine is believed to be buried.

While the time-reversal focusing method used in the experiments presented in this paper (Fig. 2) is similar to the concept of a TRM, there is noteworthy difference. A TRM relies on reciprocity of the propagation from the source to the focus point, $G_n(R,S_n,t)=G_n(S_n,R,t)$. Applying reciprocity to $U(\boldsymbol{r},t)$ will yield the autocorrelation function for the case of $\boldsymbol{r}=R$. In the case of an anisotropic propagation medium, reciprocity may not be valid, and traditional TRM implementation could fail to yield the autocorrelation function for the special case of $\boldsymbol{r}=R$.

B. Excitation methods

To investigate the relative effectiveness of time-reversal focusing in elastic media, time-reversal excitation methods will be compared to time-delay focusing methods. Uniform excitation of the transducer array will also be used to serve as a baseline measurement to demonstrate the improvement of each focusing method over a non-focused excitation method.

The results are presented with respect to a differentiated Gaussian pulse excitation [Eqs. (5) and (6)], with center frequency, ω_c, and time delay, t_d. For collection of the data, the excitation signal is a chirp, $\epsilon(t)$, described by [Eq. (7)] where A_1, A_2, P_a, P, t_p, f_1, and f_2 are constants which define amplitude, amplitude change rate, frequency change rate, total length of the chirp, and frequency range of the chirp, respectively. For the excitation used in the experiments presented in this paper, those values are: $A_1=1$, $A_2=0.25$, $P_a=0.15$, $P=0.75$, $t_p=3.596$ s, $f_1=30$ Hz, and $f_2=2$ kHz. The signal is quiescent for 0.5 s for a total duration of 4.096 s.

A chirp signal is used since it is a more effective signal for building up a sufficient signal to noise ratio.[13] After collecting the data, $U(\boldsymbol{r},t)$, the chirp signal is removed via deconvolution and the data is convolved with a differentiated Gaussian pulse yielding $D(\boldsymbol{r},t)$. This exchange of the chirp signal for the differentiated Gaussian pulse is best described mathematically in the frequency domain [Eq. (8)], where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the standard Fourier and inverse Fourier transforms, respectively.[14] Care should be taken that the frequency range of the Gaussian pulse is chosen to be within the frequency range of the initial chirp signal such that the pulse contains useful information over the entire frequency range of interest,

$$\gamma(t) = (t - t_d)\exp\left(\frac{(t-t_d)^2}{\tau_w}\right), \tag{5}$$

$$\tau_w = \frac{\sqrt{2}}{\omega_c}, \tag{6}$$

$$\epsilon(t) = \left\{A_1 + \left[(A_2 - A_1)\frac{t}{t_p}\right]^{P_a}\right\} \times \sin\left\{t(2\pi)\left[f_1 + \left(\frac{f_2 - f_1}{p+1}\right)\left(\frac{t}{t_p}\right)^P\right]\right\}, \tag{7}$$

$$D(\boldsymbol{r},t) = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}\{U(\boldsymbol{r},t)\}}{\mathcal{F}\{\epsilon(t)\}}\mathcal{F}\{\gamma(t)\}\right\}. \tag{8}$$

Uniform excitation: All sources are excited with identical differentiated Gaussian pulses. This excitation method is simple to create, and requires no *a priori* knowledge of the physical characteristics of the medium. In the presence of clutter, the wave front may be scattered, reducing the uniformity of the excitation throughout the medium.

Time-delayed focusing: Here the pulses are time-delayed such that all pulses arrive at a focus location at the same time. Ideally, this method focuses energy to a specific point, creating a larger excitation at the focal point, but this effect is sensitive to variations in wave propagation speed. Calculation of the time-delays for each pulse requires knowledge of the propagation speed throughout the entire medium. Propagation speeds can be affected by the presence of inhomogeneity. When the Rayleigh wave speed is known, along with the distance from source S_n to the target, the time delays can be calculated such that all the pulses arrive at the same time.

Time-reversal focusing: Separate measurements are performed in which a pulse is propagated from one of the sources, recorded at the focal point and then time-reversed ($t \Rightarrow -t$). The time-reversed signals are then transmitted from their corresponding source locations (Fig. 2). Unlike time-delayed focusing, time-reversal requires no knowledge of the propagation speed in the medium.

IV. EXPERIMENTAL SETUP

The experimental results are obtained in a laboratory at the Georgia Institute of Technology (Fig. 3).[1] A large concrete wedge-shaped tank is filled with approximately 50 tons of damp compacted sand. Sand is chosen as the background medium because its seismic properties are similar to many types of soil, and because it is straightforward to recondition disturbed sand. This allows for easy burial and removal of scattering objects and targets in the tank.

The seismic waves are generated by an array of 12 electrodynamic shakers. A short metal bar foot is attached to each electrodynamic shaker. The shaker and metal foot are placed in contact with the sand and the 12.5 cm $\times$ 1.27 cm $\times$ 2.54 cm aluminum bar foot couples seismic energy into the sand.

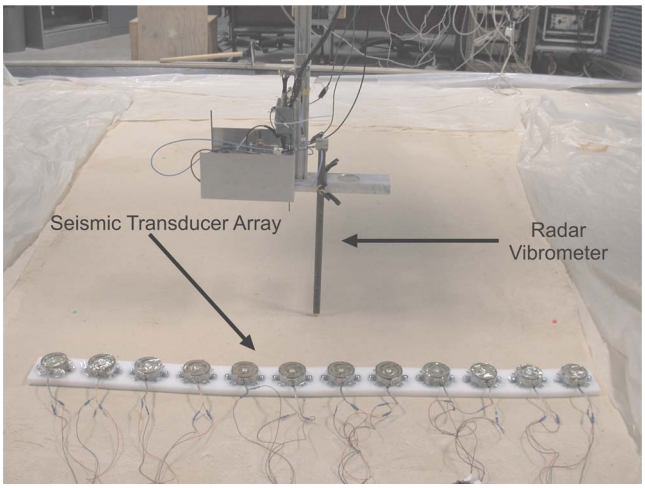


FIG. 3. (Color online) The experimental facility. The seismic transducer array and the antenna are positioned over the sand tank.

FIG. 4. (Color online) The layout of the rocks before being buried below the surface.

Once the shakers are used to excite elastic waves in the sand tank, a noncontact electromagnetic sensor (radar vibrometer) is used to record the displacement of the surface of the ground. The vibrometer is scanned across the surface of the sand using a computer controlled positioning system. The surface is sampled at 2 cm increments ($\Delta x = \Delta y = 2$ cm) over a 1.2 m $\times$ 0.8 m area. The radar has a spot size of approximately 2 cm $\times$ 2 cm and records data at each location for 4.096 s at a sampling rate of 8 kHz. By making many measurements, each at a different location on the surface, the displacement of the entire scan region can be constructed synthetically. After the entire scan has been completed, a data array of displacement information is available, $D(x_i, y_j, t_k)$, where

$$x_i = i\Delta x, \quad i = 0, 1, \ldots, \frac{X \text{ cm}}{\Delta x}$$

$$y_j = j\Delta y, \quad j = 0, 1, \ldots, \frac{Y \text{ cm}}{\Delta y}, \tag{9}$$

$$t_k = k\Delta t, \quad k = 0, 1, \ldots, \frac{T}{\Delta t}$$

and where X and Y are the dimensions of the scan region and T is the duration of time for which each measurement is recorded.

The theoretical development of time-reversal focusing using a TRM assumes that all sources and receivers are infinitesimally small points. In the physical experiment, the sources are distributed due to their use of a foot to couple energy into the ground. Similarly, the receivers are also distributed since the smallest resolvable area is limited by the spot size of the radar. While these distributed elements represent a deviation from the theoretical development of time-reversal using a TRM, the time-reversal method developed in Sec. III is insensitive to this change. Empirical observations demonstrate that the effectiveness of time-reversal focusing is not impacted by the use of distributed sources and receivers.

A total of 113 rocks are buried in the sand tank (Fig. 4) in order to introduce inhomogeneities into the sand. The rocks are randomly distributed throughout the tank both in location on the surface and burial depth. The burial region extends far beyond the scan region; rocks are buried to within 0.5 m of the edges of the sand tank. The maximum burial depth of the top of any rock is limited to approximately 20 cm. The size of the rocks varies from 10 cm in diameter to approximately 35 cm in diameter (Fig. 5).

V. WIENER FILTER

In order to effectively illuminate a buried target using time-reversal focusing, the excitation pulse that reaches the target should be both broadband and compact in time. In addition to being useful for time-reversal focusing, a compact pulse allows for better separation of incident pulses and those reflected off a target. This separation is important for affiliated detection techniques such as time-reversal imaging.[15,16]

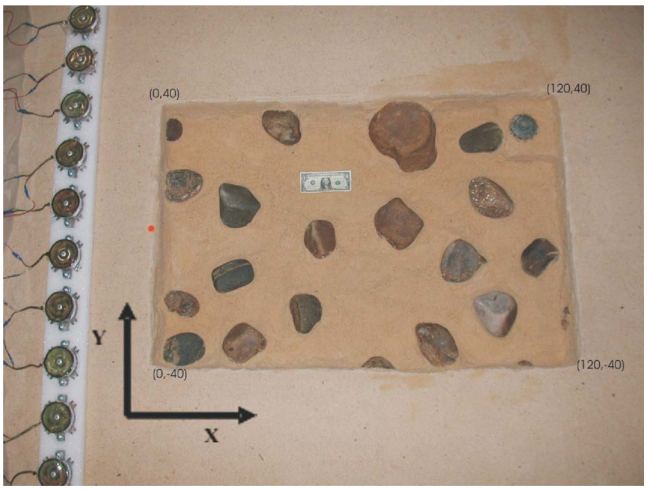


FIG. 5. (Color online) The scan region has been excavated to show the final buried rock distribution. The TS-50 landmine and the dollar bill are for scale.

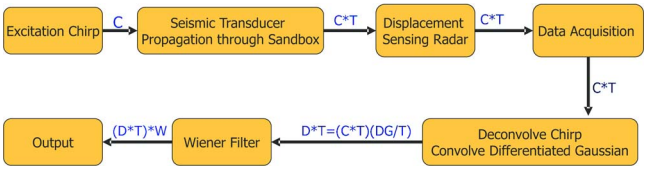


FIG. 6. (Color online) Flow graph showing the effect on the signal of its propagation through the experimental system.

In order to excite elastic waves in the ground, an excitation signal is formed digitally, passed through a D/A converter, a power amplifier, and into the electrodynamic shaker. The shaker foot then couples the transducer motion into the sand, and the excitation signal propagates through the sand and interacts with objects buried in the tank (Fig. 6). The transfer function of the electrodynamic shaker and the coupling of the shaker foot to the ground modify the excitation signal from its original temporal shape and frequency content such that the signal that arrives at the target location in the medium is significantly different from the electrical signal which is transmitted to the seismic transducer. The other elements in the signal path (A/D, amplifier, etc.) have a negligible effect on modification of the signal. The most dramatic alteration of the original excitation signal is caused by the coupling between the shaker and the ground. In the case of time-reversal focusing, this effect is more pronounced because the signal passes through the system twice, doubling the effect of the shaker-ground coupling.

To achieve the best results from time-reversal focusing, it is important to ensure that the pulse that arrives at the target is broadband and temporally compact. The practical way to do this is to design an inverse filter to restore the original response of the excitation signal. The propagating wave in the sand contains several different wave types, but the one of principal interest in the detection of buried targets is the Rayleigh surface wave. In order to most effectively design a filter that makes the Rayleigh wave temporally compact and broadband, a signal processing technique[17] is used to extract the Rayleigh wave mode from the total propagating wave.

A Wiener filter is designed that conditions the observed Rayleigh wave mode excitation signal resulting in a filtered excitation signal that is very similar to the desired temporally compact, broadband excitation pulse. A post-emphasis filter implementation is chosen because of the slightly nonlinear nature of the coupling between the shaker foot and the ground. A pre-emphasis filter would excite large amplitude displacements of the seismic transducer, which would drive the sand into an undesired nonlinear response. The recorded signal remains above the noise floor over the entire frequency range of interest, thereby making the post-emphasis filter an acceptable filter implementation scheme.

The filter coefficients are determined by recording signal outputs in an uncluttered medium, and extracting the Rayleigh wave mode. This information is used to design the Wiener filter using the Stieglitz-McBride method. The Stieglitz-McBride method iteratively minimizes the difference between the desired and designed filter impulse responses for computation of the optimal least-mean-square

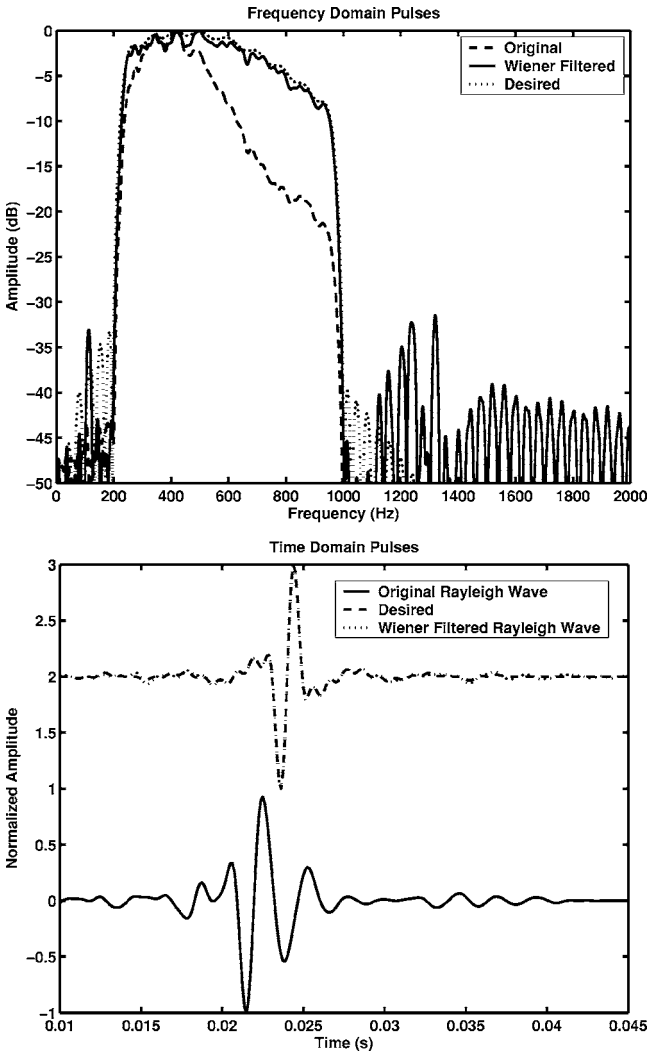


FIG. 7. Wiener filter design. (a) Frequency domain responses. (b) Time domain responses.

filter coefficients.[18] The frequency and time-domain responses of the unfiltered excitation signal, the desired signal, and the Wiener-filtered signal are displayed in Fig. 7.

VI. EXPERIMENTAL RESULTS

Focusing results for all three excitation methods are first presented for a focus location near the center of the scan region. Subsequently, three additional focusing locations are chosen. These particular locations are deliberately chosen in order to examine the relative effectiveness of time-reversal focusing when it is impeded by scattering, or very near or far from the source array. Time-reversal focusing is also used to illuminate a TS-50 landmine in two of these locations. The results of this excitation are compared to the results when the transducer array is uniformly excited.

The first results to be presented (Fig. 8) are time snapshot images comparing the effectiveness of the different focusing methods for a location near the center of the scan region. These images are formed by considering the displacement array, $D(x,y,t)$ [Eq. (9)] at a particular time. The results are presented as pseudo-color graphs of the magnitude of the vertical component of the particle displacement at

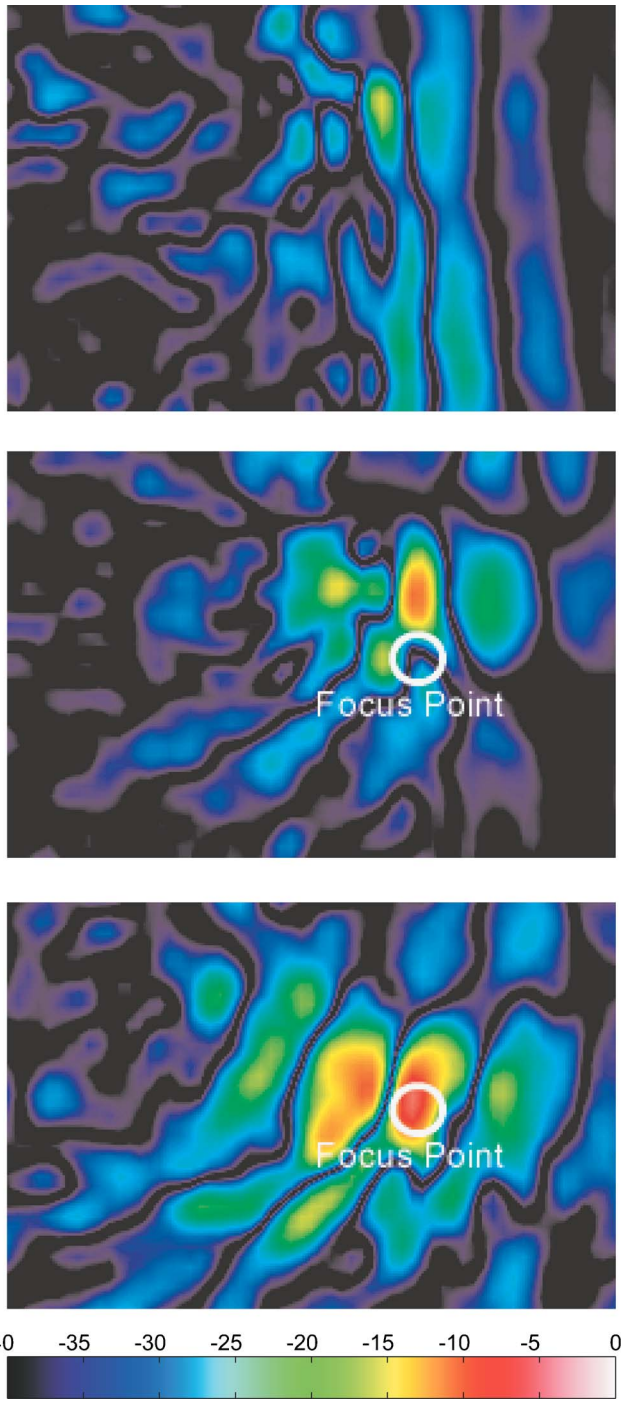


FIG. 8. (Color online) Time-snapshots for Focus Point 1. Images are on a 40 dB pseudo-color scale: 0 dB (white) to −40 dB (black). (a) Uniform excitation. (b) Time-delayed excitation. (c) Time-reversed excitation. (d) Color-amplitude scale.

the surface. The pseudo-color scale used in the viewgraphs is a 40 dB logarithmic scale from white (0 dB) to black (−40 dB). The images are normalized such that the excitation signals for each method have equal energy in the frequency band of interest, 200 Hz$<f<$1 kHz.

The first case is uniform excitation of the transducer array [Fig. 8(a)]. An excitation pulse is launched from the source array, located to the left of the scan region. As the pulse propagates through the cluttered scan region, the wave fronts are broken up by the scattering objects in the medium. An excitation pulse not modified by scattering would appear as a set of parallel, straight wave fronts propagating away from the sources. Observation of the provided time snapshot for the uniform excitation case demonstrates that the wave fronts are significantly altered by the scattering objects.

For the time-delayed excitation, an attempt is made to focus to a point near the center of the scan region, indicated by the label, *Focus Point* [Fig. 8(b)]. The speed of the Rayleigh wave is estimated from the uniform excitation experiment and used to calculate the appropriate time delays. In this case, the time-delayed focusing attempt misses the desired focus point. The most likely reason for this is the propagation velocity gradient across the surface of the tank. Due to the gradient in the direction normal to the propagation direction, the wave front moves faster on one side than the other, causing asymmetrical arrival at the focus point. A second factor is the proximity of the desired focal location to a large rock. This rock also alters the propagation speed and path of the pulses. The cumulative effect of these conditions is that the components from each of the sources add coherently, but in the wrong location.

An examination of an attempt to focus on the same location using time-reversal focusing demonstrates significant improvement over the time-delayed focusing case [Fig. 8(c)]. The time-reversal focusing method is relatively insensitive to propagation velocity gradients and the presence of inhomogeneities in the medium. This indicates that time-reversal offers a distinct advantage in focusing when the propagation medium contains unknown variations in the propagation speed, and un-catalogued scattering objects.

A second method of presenting the results from Fig. 8, is shown in Fig. 9. This presentation of the data displays the maximum displacement at each location over the entire time record. This image is formed by creating and displaying the array, $M(x,y)$ where

$$M(x_i,y_i) = \max_k |D(x_i,y_i,t_k)|. \tag{10}$$

The results are presented as pseudo-color graphs of the magnitude of the vertical component of the particle displacement at the surface. The pseudo-color scale used in the viewgraphs is a 40 dB logarithmic scale from white (0 dB) to black (−40 dB).

The scattering effects of rocks and other objects are visible in the uniform excitation case [Fig. 9(a)]. There are also areas of the scan region that are not effectively excited by the pulse, which will be referred to as shadow regions. An examination of the time-delayed excitation graph [Fig. 9(b)], shows that it focuses energy to a small area near the desired excitation point, but not on top of it. As discussed previously, this is due to propagation velocity gradients in the medium, and the presence of scattering objects. In a highly cluttered and inhomogeneous environment, time-delayed focusing fails to excite the focus point effectively. This makes time-delayed focusing excitation only marginally useful for detection of near surface targets in the presence of large scale clutter and inhomogeneity.

The time-reversal focusing result [Fig. 9(c)] is qualitatively similar to the time-delayed excitation focusing graph.

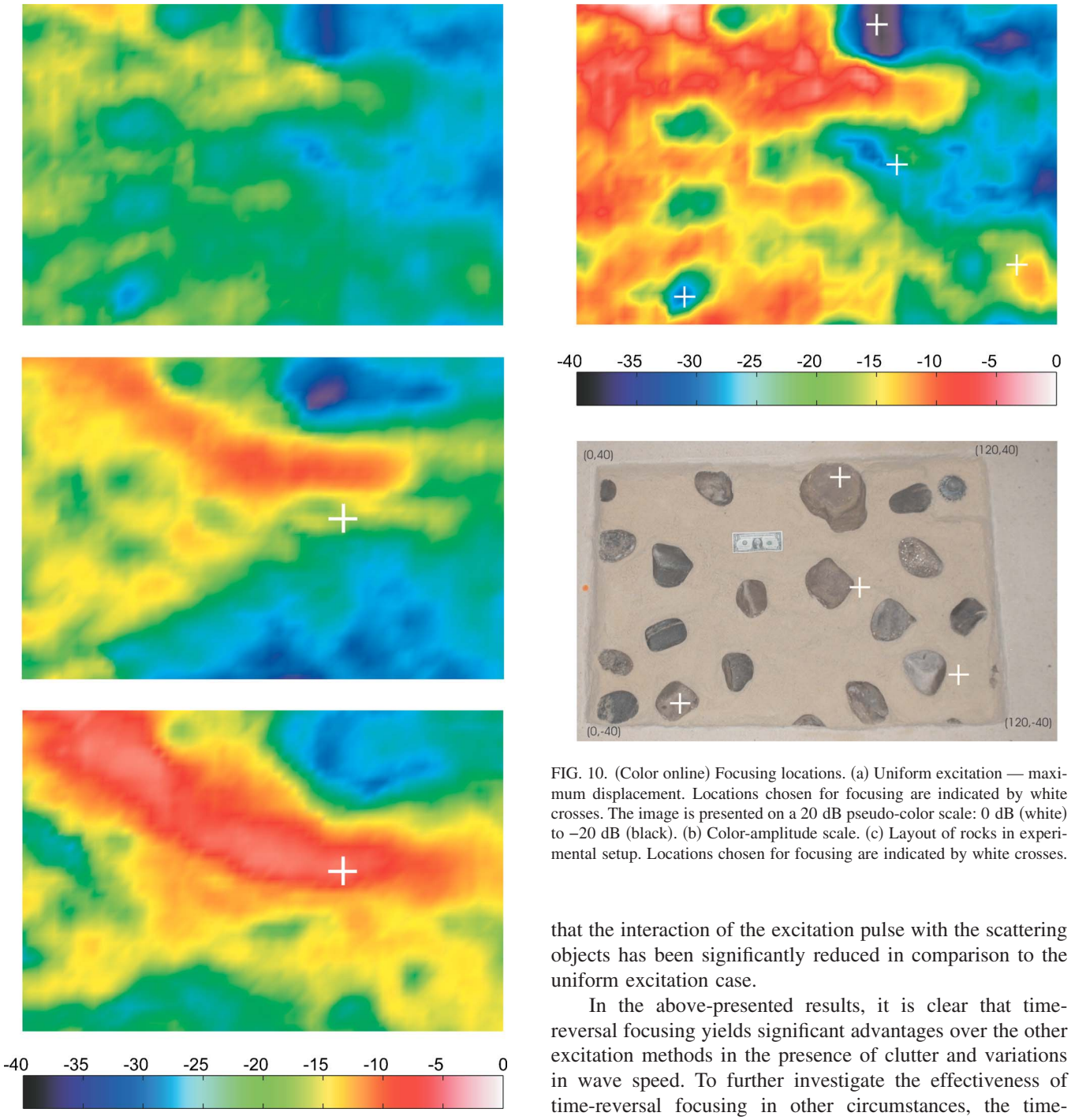


FIG. 9. (Color online) Maximum displacement for Focus Point 1. Images are on a 40 dB pseudo-color scale: 0 dB (white) to −40 dB (black). The desired focusing location is indicated by a white cross. (a) Uniform excitation. (b) Time-delayed excitation. (c) Time-reversed excitation. (d) Color-amplitude scale.

FIG. 10. (Color online) Focusing locations. (a) Uniform excitation — maximum displacement. Locations chosen for focusing are indicated by white crosses. The image is presented on a 20 dB pseudo-color scale: 0 dB (white) to −20 dB (black). (b) Color-amplitude scale. (c) Layout of rocks in experimental setup. Locations chosen for focusing are indicated by white crosses.

A notable exception is that the maximum displacement occurs at the desired focus point in the time-reversal case. The reason for this improvement is that the time-reversal method inherently incorporates the effects of scatterers and variations in propagation velocity when calculating the time-reversed excitation pulse. It should also be noted that the displacement at the focus point is much larger than the displacement throughout the rest of the medium. This means that the interaction of the excitation pulse with the scattering objects has been significantly reduced in comparison to the uniform excitation case.

In the above-presented results, it is clear that time-reversal focusing yields significant advantages over the other excitation methods in the presence of clutter and variations in wave speed. To further investigate the effectiveness of time-reversal focusing in other circumstances, the time-reversal focusing method is applied to several new locations. These locations are selected by examining the maximum displacement graph for the uniform excitation experiment (Fig. 10). Two locations are chosen that are in shadow regions, where very little energy arrives. A third point is chosen that is far from the source, with a relatively high level of excitation. By examining this location, it is possible to study the improvement afforded by time-reversal when signal levels are already high at the desired focus point.

It has already been demonstrated (Figs. 8 and 9) that time reversal can be an effective method of excitation in regions that are poorly illuminated by traditional excitation methods. The first point chosen [Fig. 11(a)] attempts to focus energy on top of a large rock. Time-reversal increases the excitation level at the desired point, but appears to focus in

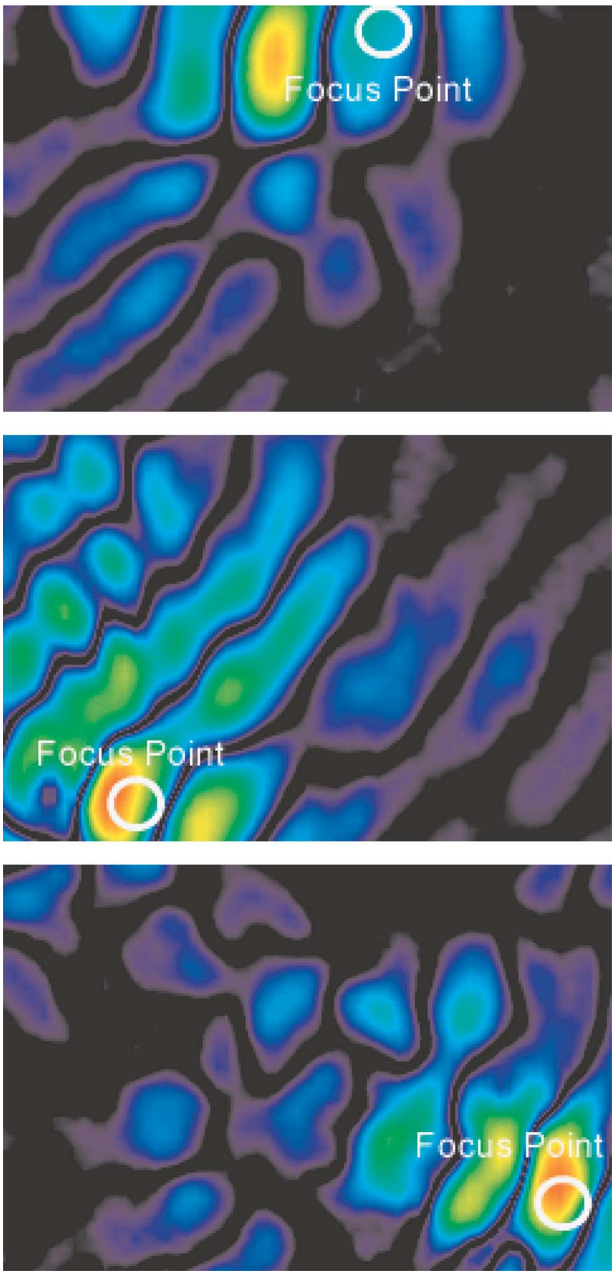


FIG. 11. (Color online) Time-reversal excitation. The desired focus point is indicated in each image. Images are on a 40 dB pseudo-color scale: 0 dB (white) to −40 dB (black). (a) Shadow region: focus point 1. (b) Shadow region: focus point 2. (c) Normal excitation region focus point.

front of the rock. In the second shadow-region focus location, similar results are observed [Fig. 11(b)]. The primary difference in this case is that the actual focus location is closer to the desired one, and the excitation level is higher.

The results presented in Fig. 8(a) measure a snapshot of displacement at a particular time. These images do not take into account the varying impedance of the materials present in the sand tank. In the case of Fig. 8(a), time-reversal appears to miss the focus point and focus in front of the desired location. If this image were a measure of energy, one would be able to account for the greatly increased stiffness of the rock present at the desired focus point and demonstrate that the time-reversal signal does in fact focus significant energy to the desired location.

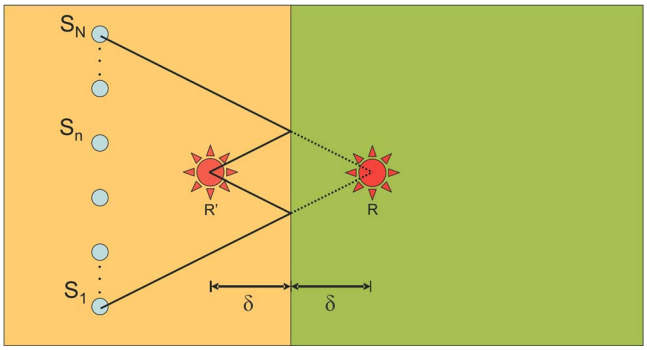


FIG. 12. (Color online) The pseudo-focus point (R') created by an infinite half-space rock.

The explanation above accounts for the absence of significant displacements at the focus point. The observation of a large apparent focus between the source array and the desired focus point can be described using a simple model. For the coordinate system assumed in these experiments, consider a rock of an arbitrary thickness in X, but of infinite extent in Y and Z (Fig. 12). As the wave propagating from each of the sources arrives at the medium interface, the majority of the energy in the Rayleigh wave will be reflected off the interface between the rock and the sand. This creates a pseudo-focal point, R', in front of the rock-soil interface. In this simple case, R' will be the same distance (δ) away from the interface as the desired focus, R.

In the actual, less simplistic case, as the extent of the rock becomes finite, a larger portion of the incident wave is unaffected by the abrupt change in material properties. Combining this with inhomogeneities in the medium causes variation in the strength, size, and location of the pseudo-focal point. These effects are apparent when comparing Figs. 11(a) and 11(b). In the former, where the scattering rock is larger in the Y dimension (Fig. 10), a more distinct pseudo-focus point is apparent. This indicates that more of the incident waves are partially reflected and refocus in front of the desired focus location. This effect is diminished for Fig. 11(b), where the scattering rock is significantly smaller.

In the final case [Fig. 11(c)], time-reversal focuses almost exactly at the desired location and some improvement in the excitation level is observed in comparison to the uniform excitation case. The small offset of the actual focus point from the desired focus point can be attributed to the phenomenon discussed above and described in Fig. 12.

The motivation for pursuing high excitation levels at a specified location is to excite greater resonances in a landmine buried at the focus location. To that end, the effect of time-reversal on resonance excitation in a landmine is presented. Two locations are chosen, one of which is in a shadow region.

For each point, the maximum displacement level over time is used as a basis for performance comparison. For the focus location in a shadow region (Fig. 13), time-reversal focusing provides an approximately 18 dB improvement over the uniform excitation case. In addition to raising the relative amplitude of the displacement at the location of the

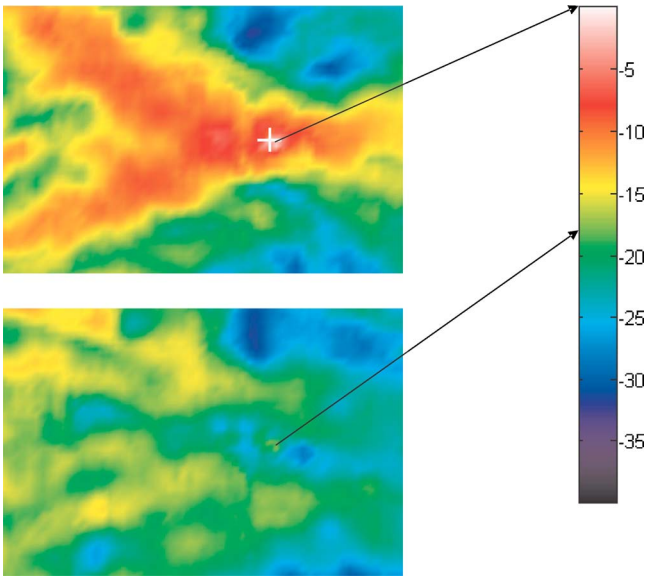


FIG. 13. (Color online) A comparison of the maximum displacement at the location of a buried TS-50 for time-reversal and uniform excitation cases.

mine, the relative signal levels over the majority of the scan region are reduced significantly, providing better contrast between the landmine and its background.

In the second position (Fig. 14), the displacement levels are high enough in the uniform excitation case to excite a substantial resonance in the landmine. While time-reversal focusing does focus energy to the location of the landmine and drop the relative displacement in the background, the increase in the displacement at the location of the resonating landmine is somewhat less, at approximately 12 dB.

VII. CONCLUSIONS

Time-reversal behavior in an elastic medium has been examined with particular interest in its application to the problem of seismic landmine detection. In the context of the detection system, an implementation of time-reversal focusing and a method for ensuring broadband excitation signals were developed and implemented.

Attempts were made to focus energy into various regions of interest including shadow regions and regions where signal levels were already high. An explanation was offered for the reason time-reversal focusing appears to miss its target focal point in some of these cases. The excitation of a resonance in a TS-50 landmine was compared for time reversal and uniform excitation methods.

The experimental results for time-reversal in an inhomogeneous medium at the frequency range of interest are consistent with the theoretically predicted behavior. Time-reversal has been demonstrated to be a superior excitation under certain conditions, but it has also been observed that there are situations in which time-reversal may not yield the expected result in a complex environment (Fig. 11). The conditions under which time-reversal shows the most dramatic improvement over other focusing methods are when a strong wave speed gradient is present in the medium, normal to the direction of propagation. The specific advantage of time-reversal over other methods is that time-reversal requires no *a priori* knowledge of the characteristics of the background medium.

The potential and limitations of time-reversal should continue to be investigated in future work. Super-resolution effects were not apparent in these experiments. Further studies might look to classify more specifically what the effect is of various types, configurations, and sizes of scattering objects on the effectiveness of time-reversal focusing, and under what conditions super-resolution effects can be expected to occur.

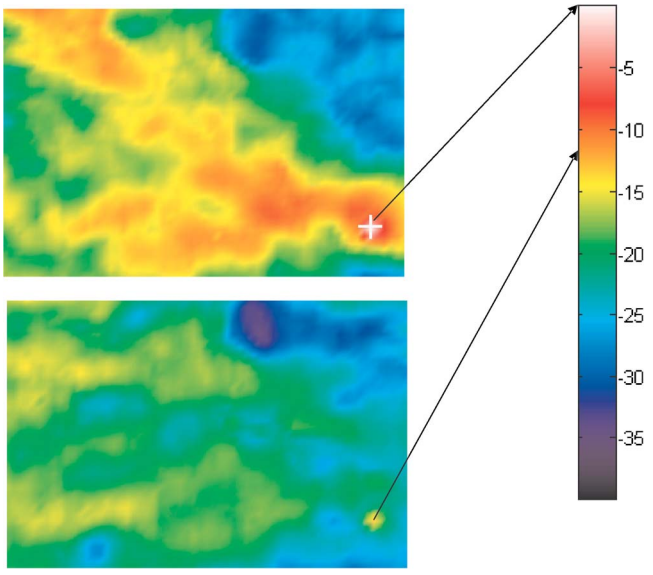


FIG. 14. (Color online) A comparison of the maximum displacement at the location of a buried TS-50 for time-reversal and uniform excitation cases.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Gregg D. Larson for his assistance in designing the LABVIEW code for the experimental measurements and Dr. James H. McClellan and Mubashir Alam for their assistance in designing the Wiener filter.

[1]W. R. Scott Jr., J. S. Martin, and G. D. Larson, "Experimental model for a seismic landmine detection system," IEEE Trans. Geosci. Remote Sens. **39**, 1155–1164 (2001).
[2]M. Fink, "Time-reversal of ultrasonic fields. i. Basic principles," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 555–556 (1992).
[3]F. Wu, J.-L. Thomas, and M. Fink, "Time-reversal of ultrasonic fields. ii. Experimental results," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**(5), 567–578 (1992).
[4]D. Cassereau and M. Fink, "Time-reversal of ultrasonic fields. iii. Theory of the closed time-reversal cavity," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**, 579–592 (1992).
[5]A. Derode, P. Roux, and M. Fink, "Robust acoustic time reversal with high-order multiple scattering," Phys. Rev. Lett. **75**, 4206–4209 (1995).
[6]M. Fink, C. Prada, F. Wu, and D. Cassereau, "Self focusing in inhomogeneous media with time reversal acoustic mirrors," IEEE Ultrason. Symp. **2**, 681–686 (1989).
[7]P. Blomgren, G. Papanicolaou, and H. Zhao, "Super-resolution in time-reversal acoustics," J. Acoust. Soc. Am. **111**, 230–248 (2002).
[8]L. Borcea, G. Papanicolaou, and C. Tsogka, "Theory and applications of time reversal and interferometric imaging," Inverse Probl. **19**(6), 139–164 (2003).
[9]P. D. Norville, W. R. Scott Jr., and G. D. Larson, "An investigation of time reversal techniques in seismic landmine detection," Proc. SPIE **5415**,

1310–1322 (2004).
[10]R. Ing, M. Fink, and O. Casula, "Self-focusing rayleigh wave using a time reversal mirror," Appl. Phys. Lett. **68**, 161–163 (1996).
[11]J.-L. Thomas, F. Wu, and M. Fink, "Self focusing on extended objects with time reversal mirror, applications to lithotripsy," IEEE Ultrason. Symp. **3**, 1809-1814 (1993).
[12]D. C. Carsten Draeger and M. Fink, "Theory of time-reversal process in solids," J. Acoust. Soc. Am. **102**, 1289–1295 (1997).
[13]J. S. Martin, W. R. Scott Jr., G. D. Larson, P. H. Rogers, and G. S. M. II, "Probing signal design for seismic landmine detection," Proc. SPIE **5415**, 133–144 (2004).
[14]R. N. Bracewell, *The Fourier Transform and Its Applications*, 3rd ed. (McGraw-Hill, New York, NY, 2000).
[15]M. Alam and J. H. McClellan, "Near field imaging of subsurface targets using active arrays and elastic waves," IEEE Digital Signal Processing Workshop, 2004.
[16]M. Alam, J. H. McClellan, P. D. Norville, and W. R. Scott, Jr., "Time-reverse imaging for the detection of landmines," Proc. SPIE **5415**, 167–174 (2004).
[17]M. Alam, J. H. McClellan, and W. R. Scott Jr., "Multi-channel spectrum analysis of surface waves," 37th Asilomar Conference on Signals, Systems and Computers, 2003.
[18]K. Steiglitz and L. McBride, "A technique for the identification of linear systems," IEEE Trans. Autom. Control **AC-10**, 461–464 (1965).

Frequency shift of a rotating mass-imbalance immersed in an acoustic fluid

Stephen R. Novascone[a] and David M. Weinberg
Idaho National Laboratory, Idaho Falls, Idaho 83415-3760

Michael J. Anderson
Department of Mechanical Engineering, University of Idaho, Moscow, Idaho 83844-0902

(Received 8 March 2004; revised 14 March 2005; accepted 10 May 2005)

In this paper, we describe a physical mechanism that relates a measurable behavior of a rigid oscillator device to the physical properties of a surrounding acoustic medium. The device under consideration is a rotating mass imbalance within an enclosed shell that is immersed in an unbounded acoustic fluid. It is assumed that the rotating mass imbalance is driven by an electromagnetic motor excited by a given dc voltage. If nonlinearities are ignored, the steady-state operational frequency of such a device is determined by a balance between the applied electromagnetic and opposing frictional torque on the rotating mass imbalance. If nonlinearities are retained, it is shown that under certain circumstances, the surrounding acoustic medium exerts an additional time-averaged opposing torque on the rotating mass imbalance that reduces the operational frequency of the device. Consequently, the operational frequency of the device becomes linked to the physical properties of the surrounding medium. Analytical calculations showed that the dissipative impedance of an acoustic fluid caused the opposing torque. The shift in frequency is proportional to the dissipative impedance and the square of the rotating mass eccentricity, but inversely proportional the total mass of the device and the damping effect of the dc motor. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1944607]

PACS number(s): 43.38.Ar, 43.58.Bh [AJZ] Pages: 745–750

I. INTRODUCTION

The interaction of a vibrating mechanical system with an acoustic medium is the basis for many acoustic sensors. The ubiquitous condenser microphone[1] couples acoustic fluid pressure to a movable diaphragm, and converts the diaphragm movement into an electrical signal. Physical properties of fluids, such as mass, density, and viscosity, can be measured by detecting the change in resonance frequency or phase speed caused by fluid loading in devices that employ bulk or surface waves in solid elastic materials.[2,3] An acoustic model of an improved Greenspan viscometer[4] has been used to fit measurements of a resonator response function to determine viscous diffusivity and speed of sound.[5]

In this paper, we describe a physical mechanism that relates a measurable behavior of a rigid oscillator to the physical properties of a surrounding acoustic medium (i.e., a driving-point acoustic impedance measurement). The device under consideration is a rotating mass imbalance that is located within an enclosed shell. The outside of the shell is in contact with an acoustic fluid, so the mass imbalance is not in direct contact with the acoustic fluid. It is assumed that the rotating mass imbalance is driven by an electromagnetic motor excited by a given dc voltage. If nonlinearities are ignored, the steady-state operational frequency of such a device is determined by a balance between the applied electromagnetic and opposing frictional torque on the rotating mass imbalance. If nonlinearities are retained, it is shown that under certain circumstances, the surrounding acoustic medium exerts an additional time-averaged opposing torque on the rotating mass-imbalance that reduces the operational frequency of the device. Consequently, the operational frequency of the device becomes linked to the physical properties of the surrounding medium.

Identification of the sensor mechanism was motivated by an observation of unexplained behavior with a device being developed for seismic measurements. This device, known as an orbital vibrator, consists of a rotating mass-imbalance driven by an electric motor within an enclosed cylinder.[6] When suspended in a liquid-filled well, vibrations of the device are coupled to the surrounding geologic media. In this mode, an orbital vibrator can be used as an efficient source "emitting" circularly polarized waves[7] in the elastic medium surrounding the borehole (i.e., a rotating dipole source). Alternately, the motion of an orbital vibrator is affected by the physical properties of the surrounding media. From this point of view, an orbital vibrator can be used as a sensor.[8] However, it has been noticed that the steady-state operational frequency of a particular orbital vibrator used in experiments changed as the device was lowered in a liquid-filled borehole. In particular, for a given applied dc voltage that caused an orbital vibrator to operate at a nominal frequency near 100 Hz, a frequency shift of approximately 20 Hz was observed when the device was lowered from a depth surrounded by limestone to a depth surrounded by shale.[9] The borehole diameter was nominally 2.5 in. larger than the or-

[a]Corresponding author. INL, MS 3760, EROB, 2525 N. Fremont Ave., Idaho Falls, Idaho 83415-3760. Electronic mail: Stephen.Novascone@inl.gov

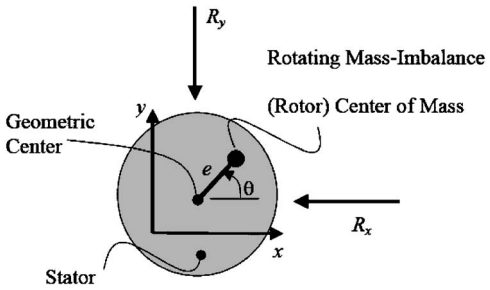


FIG. 1. Rotating-imbalance device composed of stator (mass equal to m_s) and rotor (mass equal to m) top-down view.

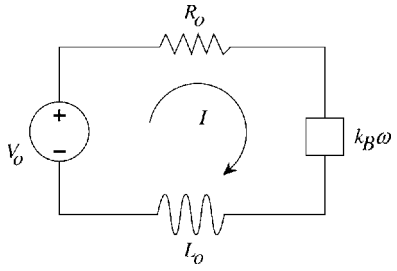


FIG. 3. Circuit for dc motor, where V_0 is the applied voltage, R_0 is the total Ohmic resistance, L_0 is the motor inductance, and $k_B\omega$ is the back-emf voltage.

bital vibrator diameter. Because the change in operational frequency did not correlate to variables that would affect the motor bearing friction, such as temperature or the surrounding hydrostatic pressure, it was suspected that the change was caused by the nature of the acoustic/elastic (borehole liquid and shale and limestone) mediums surrounding the device. To explore this hypothesis, a simpler model, that of the orbital vibrator device surrounded by an unbounded fluid medium, was considered. This model did in fact manifest a linkage between the operational frequency of an enclosed electric-powered motor rotating mass imbalance and the fluid properties of the surrounding acoustic medium. This notion could be considered in the design of an acoustic fluid property sensor.

II. THEORETICAL MODEL

A diagram of the rotating imbalance device is shown in Figs. 1 (top-down view) and 2 (side view). The device consists of two parts, a stator of mass m_s, and rotor of mass m. The rotor rotates about an axis passing through the geometric center of the stator, while the center of mass of the rotor is located a distance e from the geometric center of the stator. A mutual electromagnetic torque that acts between the stator and rotor causes the rotor to rotate. The angular displacement of the rotor is measured by the coordinate θ. Because the center of mass of the rotor is located a distance e from the axis of rotation, an imbalance force causes the geometric center of the device to undergo circular motions, as measured from a fixed coordinate system xy. The circular motion is restrained by the forces R_x and R_y that model the influence of the surrounding acoustic medium on the device. The only in-plane forces that act external to the device are R_x and R_y. Assume that the device suspension mechanism has no effect on the circular motion.

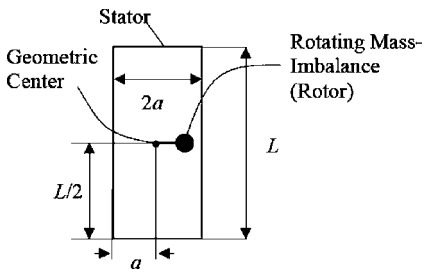


FIG. 2. Rotating-imbalance device side view.

A circuit was used to model the relationship between the dc voltage applied to the motor windings and the electromagnetic torque applied to the rotor. This circuit is presented in Fig. 3. In Fig. 3, V_0 is the voltage applied to the windings of the dc motor, I is the current flowing through the motor windings, R_0 is the total dc resistance of the motor windings and attached lead wires, L_0 is the inductance of the motor windings, k_B is the back-emf coefficient of the motor, and ω is the instantaneous angular frequency of the rotor.

Newton's second law for translation and rotation can be written for the rotor and stator components of the device. A dynamic moment equation for the rotor is

$$J\ddot{\boldsymbol{\theta}} = \mathbf{T} - k_f\dot{\boldsymbol{\theta}} - \mathbf{r} \times \mathbf{F}_i, \tag{1}$$

where J is the polar moment of inertia for the rotor, $\mathbf{T}$ is the mutual electromagnetic torque acting between the stator and rotor, k_f is the coefficient for the frictional torque that acts between the stator and rotor, $\mathbf{r}=e\cos\theta\mathbf{i}+e\sin\theta\mathbf{j}$ is the position of the rotor center of mass relative to the geometric center of the stator, $\mathbf{F}_i=F_x\mathbf{i}+F_y\mathbf{j}$ is the internal mutual force between the stator and rotor, and $\mathbf{i},\mathbf{j}$ are unit vectors in the x and y directions. The boldfaced symbols have been adopted to indicate vectors. Summing the forces on the rotor gives

$$F_x = m\{\ddot{x} - e[\ddot{\theta}\sin(\theta) + \dot{\theta}^2\cos(\theta)]\}, \tag{2}$$

$$F_y = m\{\ddot{y} + e[\ddot{\theta}\cos(\theta) - \dot{\theta}^2\sin(\theta)]\}, \tag{3}$$

where x and y locate the instantaneous position of the geometric center of the device. Summing the forces on the stator gives

$$-F_x + R_x = m_s\ddot{x} \tag{4}$$

and

$$-F_y + R_y = m_s\ddot{y}, \tag{5}$$

where R_x and R_y are the forces of the environment upon the device.

Because the external force that opposes an oscillating rigid body in an acoustic medium is typically modeled with inertial and dissipative components,[10,11] R_x and R_y are written as

$$-R_x = m_r\ddot{x} + b\dot{x} \tag{6}$$

and

$$-R_y = m_r\ddot{y} + b\dot{y}, \tag{7}$$

where m_r and b are the inertial and dissipation coefficients respectively. A voltage loop equation written for the circuit in Fig. 3 is

$$V_0 = IR_0 + L_0\dot{I} + k_B\omega. \tag{8}$$

For a permanent magnet dc motor, the relationship between current I and applied torque T is[12]

$$T = k_T I, \tag{9}$$

where k_T is the dc motor torque constant and I is the constant field current. The vector boldface on the torque has now been suppressed.

The equations of motion (1)–(7) and the circuit equations (8) and (9) can be reduced to a system of three equations and three unknowns in x, y, and θ. Neglecting the inductance L_0 (by assuming $\omega L_0 \ll R_0$),[12] and carrying out the necessary algebraic operations, one obtains the set

$$J\ddot{\theta} + me^2\ddot{\theta} - me\ddot{x}\sin(\theta) + me\ddot{y}\cos(\theta) = \frac{k_T V_0 - k_T k_B\dot{\theta}}{R_0} - k_f\dot{\theta}, \tag{10}$$

$$-me\ddot{\theta}\sin(\theta) + (m + m_s + m_r)\ddot{x} + b\dot{x} = me\dot{\theta}^2\cos(\theta), \tag{11}$$

$$me\ddot{\theta}\cos(\theta) + (m + m_s + m_r)\ddot{y} + b\dot{y} = me\dot{\theta}^2\sin(\theta). \tag{12}$$

The analytical model equations are now defined as (10)–(12), where the voltage V_0 is the input and x, y, and θ are unknowns. Note that these equations are nonlinear.

III. PERTURBATION ANALYSIS

A perturbation analysis was used to obtain an approximate solution to the model equations (10)–(12) at steady state. For this analysis, the perturbation parameter ε was defined as

$$\varepsilon = \frac{m}{M}, \tag{13}$$

where $M = m + m_s$. The following expansions were adopted for x, y, and θ to determine a steady-state solution to the analytical model equations (10)–(12):

$$x(t) = \varepsilon x_1(t) + \varepsilon^2 x_2(t) + O(\varepsilon^3), \tag{14}$$

$$y(t) = \varepsilon y_1(t) + \varepsilon^2 y_2(t) + O(\varepsilon^3), \tag{15}$$

$$\theta(t) = \omega t + \varepsilon\theta_1(t) + \varepsilon^2\theta_2(t) + O(\varepsilon^3), \tag{16}$$

where ω is the operational frequency of the device at steady state. It was further assumed that $x(t)$, $y(t)$, $\theta_1(t)$, and $\theta_2(t)\ldots$ in the expansions (14)–(16) are zero mean. Note that $\theta(t)$ is not zero mean.

The leading-order components of the mathematical model for the rotating imbalance device are found by substituting the expansions (14)–(16) into (10)–(12). The result of this operation is

$$\left[\frac{J}{M} + \varepsilon e^2\right][\varepsilon\ddot{\theta}_1] - \varepsilon e[\varepsilon\ddot{x}_1]\sin[\omega t + \varepsilon\theta_1] - \varepsilon e[\varepsilon\ddot{y}_1]\cos[\omega t + \varepsilon\theta_1] = \frac{k_T V_0}{R_0 M} - \frac{k_B k_T}{R_0 M}[\omega + \varepsilon\dot{\theta} + \varepsilon^2\dot{\theta}] - \frac{k_f}{M}[\omega + \varepsilon\dot{\theta} + \varepsilon^2\dot{\theta}] + O(\varepsilon^3). \tag{17}$$

$$\varepsilon\frac{(M + m_r)}{M}\ddot{x}_1 + \varepsilon\frac{b}{M}\dot{x}_1 = \varepsilon e\omega^2\cos(\omega t) + O(\varepsilon^2). \tag{18}$$

$$\varepsilon\frac{(M + m_r)}{M}\ddot{y}_1 + \varepsilon\frac{b}{M}\dot{y}_1 = \varepsilon e\omega^2\sin(\omega t) + O(\varepsilon^2). \tag{19}$$

A time average operation was applied to the moment equation (17) to obtain a solution at steady state. Applying this operation gives

$$-\varepsilon^2\overline{e\ddot{x}_1\sin(\omega t)} + \varepsilon^2\overline{e\ddot{y}_1\cos(\omega t)} = \frac{k_T V_0}{R_0 M} - \frac{k_B k_T}{R_0 M}\omega - \frac{k_f}{M}\omega + O(\varepsilon^3), \tag{20}$$

where the overbar indicates the time-average operation. It is apparent that there is an $O(\varepsilon^2)$ correction to the steady-state frequency ω if the time-averaged terms are nonzero. This correction is smaller than the first-order terms εx_1, εy_1 for small ε.

The operational steady-state frequency ω is found by solving for the motions x_1 and y_1 and substituting them into the steady-state torque balance (20). The motions x_1 and y_1 determined from (18) and (19) are

$$x_1 = -\frac{(M + m_r)^2 e\omega^2}{(M + m_r)^2\omega^2 + b^2}\cos\omega t + \frac{bMe\omega}{(M + m_r)^2\omega^2 + b^2}\sin\omega t, \tag{21}$$

$$y_1 = -\frac{bMe\omega}{(M + m_r)^2\omega^2 + b^2}\cos\omega t - \frac{(M + m_r)^2 e\omega^2}{(M + m_r)^2\omega^2 + b^2}\sin\omega t. \tag{22}$$

Placing the expressions (21) and (22) into the time-averaged torques in (20), and assuming the damping b is small, one obtains

$$\frac{\varepsilon^2 bMe^2\omega}{(M + m_r)^2} = \frac{k_T V_0}{MR_0} - \frac{k_B k_T}{MR_0}\omega - \frac{k_f}{M}\omega. \tag{23}$$

Then the steady-state frequency ω becomes

$$\omega = \frac{\dfrac{k_T V_0}{R_0}}{\dfrac{k_B k_T}{R_0} + k_f} - \frac{\varepsilon^2 e^2\omega bM^2}{\left(\dfrac{k_B k_T}{R_0} + k_f\right)(M + m_r)^2}. \tag{24}$$

Denoting ω_0 as the operational frequency when nonlinearities are ignored, i.e., when the device is located in a vacuum, one calculates

$$\omega_0 = \frac{\dfrac{k_T V_0}{R_0}}{\dfrac{k_B k_T}{R_0} + k_f}. \tag{25}$$

Using the definition of ω_0, the expression (24) can be simplified to

$$\omega = \omega_0 \left[1 - \frac{\varepsilon^2 e^2 \omega b R_0 M^2}{k_T V_0 (M + m_r)^2} \right]. \tag{26}$$

The constants in (26) can be further grouped by defining K as

$$K = \frac{e^2 b R_0 M^2}{k_T V_0 (M + m_r)^2}. \tag{27}$$

With this definition, the expression (26) for the steady-state frequency of a rotating mass-imbalance sensor located in an acoustic medium with inertial and dissipative acoustic properties becomes

$$\omega = \frac{\omega_0}{1 + \varepsilon^2 K \omega_0}. \tag{28}$$

When $\varepsilon^2 K$ is small relative to unity, $\omega \approx \omega_0$. When $\varepsilon^2 K$ increases, a frequency shift occurs resulting in a lower steady-state frequency of operation. The increase or decrease of $\varepsilon^2 K$ depends on the geometric and electrical properties of the device and the properties of the medium surrounding the device. Specifically, the factors that affect the frequency shift are the product of eccentricity mass and size me, the voltage V_0 and resistance R_0, the motor torque constant k_T, the total mass of the orbital vibrator M, and the inertial and dissipative coefficients m_r and b of the surrounding medium.

For the three-dimensional case at steady state, inertia m_r and dissipative b coefficients can be modeled using the inertial and dissipative terms of the fluid force acting on a rigid sphere oscillating in an acoustic medium, where viscous and nonviscous effects are included. The impedance for such a situation, as derived by Temkin,[11] is

$$Z = -\frac{A_1}{\alpha^2} \exp(i\alpha)[4\pi\rho\omega a^2(\alpha + i) + 8\pi i\mu(3\alpha - \alpha^3 + 3i - 2i\alpha^2)] - 8\pi i\mu \frac{B_1}{\beta^2} \exp(i\beta)(3\beta + \beta^3 + 3i). \tag{29}$$

In Eq. (29) a is the sphere radius, ρ is the fluid mass density, and μ is the fluid shear viscosity. The remaining terms in (29) are

$$\alpha = ka, \tag{30a}$$

$$y = \sqrt{\omega a^2 / 2\nu}, \tag{30b}$$

$$\beta = (1 + i)y, \tag{30c}$$

where $k = \omega / c$, and ν is fluid kinematic viscosity. In Eq. (29), A_1 and B_1 are coefficients obtained from boundary conditions, and are functions of α, β, and wave number k. The expressions for A_1 and B_1 are

$$A_1 = \frac{\alpha^3 \exp(-i\alpha)}{3ik} \times \frac{3y + 2y^2 + 3i(1 + y)}{4\alpha y^2 + \alpha^2(1 + y) - i[y\alpha^2 - 2y^2(2 - \alpha^2)]},$$

$$B_1 = \frac{-\beta^2 \alpha \exp(-i\beta)}{3ik} \times \frac{3\alpha + i(3 - \alpha^2)}{4\alpha y^2 + \alpha^2(1 + y) - i[y\alpha^2 - 2y^2(2 - \alpha^2)]}. \tag{31}$$

Equations (29)–(31) apply when the radius of the sphere is very small compared to the acoustic wavelength λ. The coefficients m_r and b for inertial and dissipation are found from

$$m_r = -\frac{\mathrm{Im}[-Z]}{\omega}, \quad b = \mathrm{Re}[-Z]. \tag{32}$$

To estimate the coefficients m_r and b, it was assumed that the frequency could be approximated by $\omega \approx \omega_0$.

IV. DISCUSSION

The motivation of this work was to understand the frequency shift phenomenon that was observed in a borehole experiment, which is briefly described in the Introduction. In the present work, however, we focus on describing a mechanism that could cause a frequency shift for the case of a rigid oscillator immersed in an acoustic fluid when the oscillatory motion is due to a rotating mass-imbalance internal to the rigid oscillator. A model of such a device coupled to an unbounded acoustic medium predicts that a steady-state torque will be developed that alters the operational frequency (frequency shift). This steady-state torque will always act to decrease the operational frequency, and the magnitude of the shift is dependent upon the properties of the device and the acoustic properties of the surrounding medium.

Some insight can be gained from a rearrangement of the expression (28) for the operational frequency ω. Assuming that the frequency shift is small, one may approximate (28) as

$$\omega = \frac{\omega_0}{1 + \varepsilon^2 K \omega_0} \approx \omega_0 - \varepsilon^2 K \omega_0^2 = \omega_0 - \frac{(me)^2}{(M + m_r)^2} \frac{b}{k_T (V_0 / R_0)} \omega_0^2,$$

followed by a further substitution for ω_0 from (25) to obtain

$$\omega \approx \omega_0 \left[1 - \frac{(me)^2}{(M + m_r)^2} \frac{b}{\dfrac{k_B k_T}{R_0} + k_f} \right]. \tag{33}$$

The above expression (33) shows that the frequency shift is composed of two factors. The first factor shows that the frequency shift is proportional to the square of the eccentricity me normed against the addition of the total mass M of the device and the inertial impedance of the acoustic medium. The second factor shows that the interaction of the device with the medium must contain a dissipative component for a

TABLE I. Orbital vibrator device parameters.

Property	Value
Eccentricity *me* (kg m)	$3(10)^{-4}$
Cylinder Radius a (m)	0.0445
Sphere Radius a_{sp} (m)	0.111
Length L (m)	0.432
Total Mass M (kg)	4.1
Torque Constant k_T (N m/A)	0.007
Back-emf Coefficient (k_B) (V s)	0.026
Winding Resistance R_0 (Ω)	10
Winding Inductance L_0 (mH)	0.5
Friction Coefficient k_f (N s)	$1.8(10)^{-5}$

frequency shift to exist, and that the dissipative coefficient b is compared to the total damping effect of the dc motor, including that caused by back-emf. Consequently, if one wishes to design an acoustic property sensor based upon this principle, this analysis predicts that a strong eccentricity me/M and low-loss motor would be desired.

Equation (33) was used to explore the theoretical sensitivity of steady-state frequency to changes in the eccentricity (*me*) and the total damping effect of the dc motor ($k_B k_T/R_0 + k_f$) for the impedance defined in (32). The mass density and sound speed of water ($\rho = 998$ kg/m^3, $c = 1481$ m/s) were used. Estimated device parameters of the orbital vibrator are in Table I. These device parameters were used to baseline Eq. (33). Because a spherical geometry was used for the inertial and dissipation coefficients in (33) and the geometric device parameters are cylindrical, the radius used for calculation was determined by setting the lateral projected area of the cylinder equal to the projected area of a sphere and solving for the radius of the sphere,

$$a_{sp} = \sqrt{2aL/\pi},$$

where a_{sp} is the radius used in the calculation (0.111 m; see Table I), a is the radius of the source, and L is the length of the source. Three plots of frequency shift, shown in Fig. 4, were generated by plotting the frequency shift

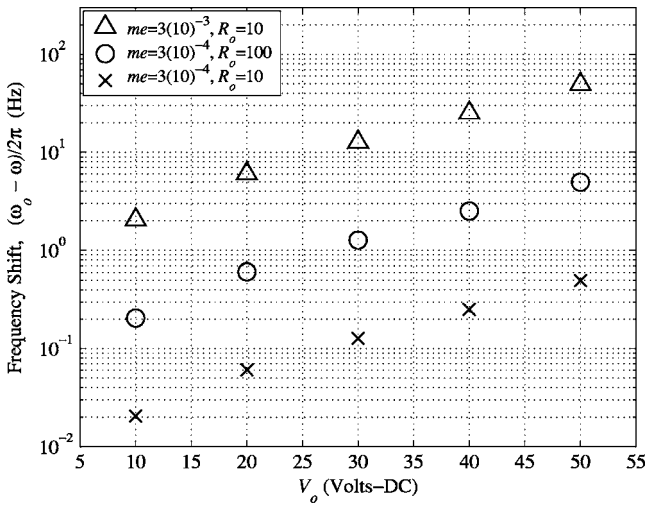


FIG. 4. Rotating imbalance steady-state frequency sensitivity to eccentricity and damping effect of dc motor.

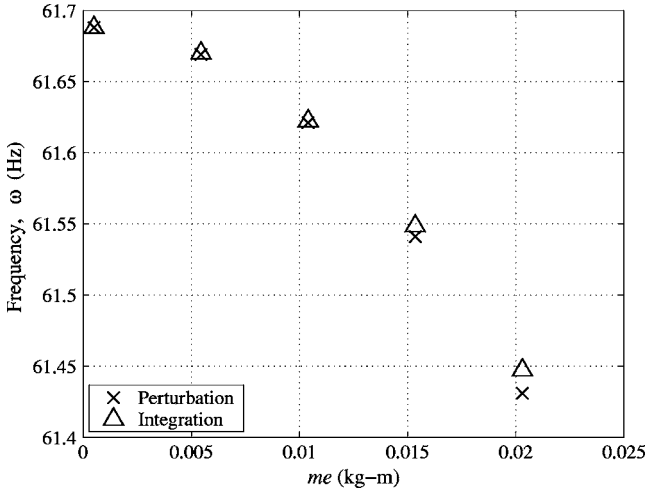


FIG. 5. Steady-state frequency calculations from perturbation and numerical integration at $V_0 = 10$ V dc for a range of eccentricity (*me*) values.

on the vertical axis against applied dc voltage on the horizontal axis, where frequency shift is defined here as the steady-state frequency in a vacuum ω_0 (25) minus the steady-state frequency in an unbounded fluid ω (33). One curve was generated using the parameters in Table I. The remaining two curves were generated with the values in Table I, but with the eccentricity increased by a factor of 10 and the total damping effect of the dc motor decreased by a factor of 10. For calculation purposes the total damping effect of the dc motor was decreased by increasing the total Ohmic resistance. For the device parameters in Table I, the frequency shift was small, however, for an applied voltage of 50 V dc, $me = 3(10)^{-3}$ kg m, and $R_0 = 10\ \Omega$ the frequency shift was about 50 Hz, which is about a 16% change from the nominal steady-state frequency from Eq. (25) that was 308 Hz.

To support the validity of the perturbation approach, an independent method was used to calculate the steady-state frequency of an oscillating sphere in water. The system equations [(8) and (10)–(12)] were integrated numerically with respect to time using the MATLAB ordinary differential equations algorithm ode45, where ode45 is based on an explicit Runge-Kutta (4) and (5) formula, the Dormand-Prince pair.[13] The device parameters of the orbital vibrator seismic source, shown in Table I, were used for calculations (with sphere radius, a_{sp}). The mass density and sound speed of water ($\rho = 998$ kg/m^3, $c = 1481$ m/s) were used. Figure 5 shows the steady-state frequency at 10 V dc for the perturbation and integration calculations plotted on the vertical axis with eccentricity (*me*) plotted on the horizontal axis. Figure 5 shows that the perturbation approximation (33) through a range of large eccentricity values is consistent with numerical integration. According to the data in Fig. 5, the largest difference between the perturbation and numerical integration calculations is within 0.1% at an eccentricity (*me*) value of 0.203 kg m.

V. CONCLUSIONS

A physical mechanism that relates a measurable behavior of an enclosed rotating mass-imbalance device at steady

state to the physical properties of a surrounding acoustic medium has been described. For the case of a rotating mass-imbalance enclosed within a rigid shell that is oscillating in an unbounded acoustic fluid, the shift in frequency is proportional to acoustic dissipation and the square of the rotating eccentricity, but inversely proportional the total mass of the device and the damping effect of the dc motor. Also, the interaction of the device with the medium must contain a dissipative component for a frequency shift to exist. The acoustic dissipation acts to oppose the motion of the rotating mass imbalance through a countertorque and decreases the steady-state frequency. This information may be useful if one wished to design an acoustic-fluid property sensor based upon this principle.

ACKNOWLEDGMENTS

This work was supported by the INL LDRD program under DOE Contract No. DE-AC07-99ID13727.

[1]G. S. K. Wong and T. F. W. Embleton, *AIP Handbook of Condensor Microphones: Theory, Calibration and Measurements* (American Institute of Physics, Woodbury, NY, 1995).
[2]D. S. Ballantine, Jr., S. J. Martin, A. J. Ricco, G. C. Frye, H. Wohltjen, R. M. White, and E. T. Zellers, *Acoustic Wave Sensors, Theory, Design and Physico-Chemical Applications* (Academic, New York, 1997).
[3]J. D. N. Cheeke, *Fundamentals and Applications of Ultrasonic Waves* (CRC Press, City, 2000).
[4]J. Wilhelm, K. A. Gillis, J. B. Mehl, and M. R. Moldover, "An improved Greenspan acoustic viscometer," Int. J. Thermophys. **21**, 983–997 (2000).
[5]K. A. Gillis, J. B. Mehl, and M. R. Moldover, "Theory of the Greenspan viscometer," J. Acoust. Soc. Am. **114**, 166–173 (2003).
[6]J. H. Cole, "The orbital vibrator, a new tool for characterizing interwell reservoir space," The Leading Edge **16**, 281–283 (1997).
[7]T. M. Daley and D. Cox, "Orbital vibrator seismic source for simulations *P*- and *S*-wave crosswell acquisition," Geophys. J. **66**, 1471–1480 (2001).
[8]D. M. Weinberg, J. H. Cole, R. R. Reynolds, C. D. Christensen, and S. R. Novascone, "Driving Point Impedance—A new paradigm for fracture detection and *in situ* stress detection in boreholes," *Proceedings of the Society of Mining Engineers Annual Meeting*, Phoenix, February, 2002.
[9]S. R. Novascone, "Analysis of seismic transducer motion sensitivity to constitutive properties of the surrounding medium," Ph.D. dissertation, University of Idaho, Moscow, ID, 2003.
[10]P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1968).
[11]S. Temkin, *Elements of Acoustics*, Acoustical Society of America through the Institute of Physics, 2001, ISBN 1-56396-997-1.
[12]R. C. Dorf and R. H. Bishop, *Modern Control Systems* (Addison-Wesley, New York, 1995).
[13]J. R. Dormand and P. J. Prince, "A family of Runge-Kutta formulae," J. Comput. Appl. Math. **6**, 19–26 (1980).

Theoretical and experimental vibration analysis for a piezoceramic disk partially covered with electrodes

Chi-Hung Huang[a)]
Department of Mechanical Engineering, Ching Yun University, 229, Chien-Hsin Road, Chung-Li, Taiwan 320, Republic of China

(Received 8 September 2004; revised 14 March 2005; accepted 4 May 2005)

Using the linear two-dimensional electroelastic theory, the vibration characteristics of partially electrode-covered thin piezoceramic disks with traction-free boundary conditions are investigated by theoretical analysis, numerical calculation, and experimental measurement. Four types of piezoceramic disks are discussed owing to the different electrode-covered distribution and electrical boundary conditions. It is found that when an alternating electrical potential is applied to the piezoceramic disk with axisymmetric partial electrodes, not only extensional but also transverse vibrations occur in resonance. The electrode-covered distribution influences the extensional vibration characteristics of piezoceramic disks; however, this phenomenon is not present for the transverse vibration. In this paper the optical speckle interferometry and electrical impedance analysis are employed to validate the theoretical analysis. Numerical calculations based on the finite element method are performed to make comparison with the theoretical and experimental results. According to the theoretical calculation, the variations of extensional resonant frequencies and dynamic electromechanical coupling coefficients depending on the various electrode-covered ratios are also investigated in this work. For the type with circular partial electrodes being connected, it is shown that the coupling coefficients will reach a maximum at the fundamental mode and drop to zero at the overtone mode for certain electrode-covered ratios. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1940468]

PACS number(s): 43.40.At, 43.40.Dx, 43.38.Fx [JGM] Pages: 751–761

I. INTRODUCTION

Since Pierre and Jacques Curie first noted the piezoelectric phenomena in 1880, there have been many industrial applications, such as ultrasonic transducers, accelerometers, nondestructive testing devices, and electro-optic modulators. The piezoelectric effect can be applied to many modern-engineering applications because it connects the electrical and mechanical fields. Piezoelectricity describes the phenomenon in which the material generates electric charge when subjected to stress and, conversely, generates strain when an electric field is applied. The vibration characteristics of piezoelectric materials can be determined by the linear piezoelasticity, the Maxwell equation, and piezoelectric constitutive equations;[1,2] nevertheless, analytical theoretical solutions have been obtained only for simple types. Because the transducers made of piezoceramics are usually circular, the vibration characteristics of piezoceramic disks are important in transducer design and application. Kunkel *et al.*[3] studied the vibration modes of PZT-5H ceramics disks concerning the diameter-to-thickness ratio ranging from 0.2 to 10. Both the resonant frequencies and effective electromechanical coupling coefficients were calculated for the optimal transducer design. Guo *et al.*[4] presented the results for PZT-5A piezoelectric disks with diameter-to-thickness ratios of 20 and 10. There were five types of modes being classified according to the mode shape characteristics, and the physical interpretation was well clarified. Ivina[5] studied the symmetric modes of vibration for circular piezoelectric plates to determine the resonant and antiresonant frequencies, radial mode configurations, and the optimum geometrical dimensions to maximize the dynamic electromechanical coupling coefficient. For the discussion of partial-electroded piezoceramic disks, Schmidt[6] employed the linear piezoelectric equations to investigate the extensional vibrations of a thin, partly electroded piezoelectric plate. The theoretical calculations were applied to the circular piezoceramic plate with partial concentric electrodes for the first fundamental frequency. Rogacheva[7] used the cases of piezoceramic disk and cylindrical shells to discuss the dependence of the electromechanical coupling coefficient (EMCC) on the size and position of the electrodes. By theoretical verifications an appropriate choice of the position and size of electrodes can increase the EMCC, which is easier than changing the geometrical parameters. Ivina[8] analyzed the thickness-symmetric vibrations of piezoelectric disks with partial axisymmetric electrodes by using the finite element method. According to the spectrum and value of the dynamic electromechanical coupling coefficient (DCC) of quasi-thickness vibrations, the piezoceramics can be divided into two groups. Only for the first group can the DCC be increased by means of the partial electrodes, which depends on the vibration modes.

In general, there are two experimental methods that are usually used to study the vibration problem of piezoelectric materials, one is the equivalent circuit measurement (called admittance analysis) and the other is optical interferometric technique. Shaw[9] used an optical interference technique in

[a)]Electronic mail: chhuang@cyu.edu.tw

which a stroboscopically illuminated multiple beam was applied to measure the surface motion of thick barium titanate disks. However, only normal modes having symmetry with respect to the axis and to the central plane were observed. Minoni and Docchio[10] proposed an optical self-calibrating technique, which was based on the signal-processing chain, to measure the vibration amplitude of PZTs for different operating frequencies. Chang[11] employed dual-beam speckle interferometry to measure the in-plane vibration amplitude on the PZT surface. As the measured displacement spectrum reaches the local maximum under some driven voltages, the resonant frequencies of in-plane modes were determined. To obtain the vibration mode shapes simultaneously, the technique with full-field measurement is employed to perform the work. Koyuncu[12] used ESPI with reference beam modulation to observe the vibration amplitudes and vibration modes of PZT-4 transducers in air and water. Oswin *et al.*[13] utilized electronic speckle pattern interferometry (ESPI) to validate the finite element model of flextensional transducer with an elliptical shape. Both in-plane and out-of plane vibrations were studied and discussed. Ma and Huang[14,15] used amplitude-fluctuation electronic speckle pattern interferometry (AF-ESPI) to investigate the three-dimensional vibration of piezoelectric rectangular parallelepipeds and cylinders, and presented both the resonant frequencies and mode shapes.

In this study, the vibration characteristics of a thin piezoceramic disk with axisymmetric partial electrodes are provided in detail, including the resonant frequencies and electrical current intensity. For the previous studies regarding partial-electroded piezoceramic disks, the materials usually focused the attention on extensional vibration and less on transverse vibration. Moreover, the case of the region with electrodes being disconnected is not discussed in the literature. Due to the electrode distribution and alternating potential connection, there are four different types of partial-electroded piezoceramic disks presented herein. If the piezoceramic disk is thin, the out-of-plane (transverse) vibration and the in-plane (extensional) vibration are uncoupled. The nonaxisymmetric modes are investigated for the transverse vibration, while extensional vibrations are restricted to axisymmetric modes. To validate the theoretical results, the optical techniques AF-ESPI and electrical impedance analysis are employed to measure the vibration characteristics of piezoceramic disks in resonance. The advantage of using the AF-ESPI method is that not only the resonant frequencies but also the corresponding mode shapes for transverse and extensional modes can be obtained. According to experimental results obtained in this study, it is shown that only the radial extensional vibration for the piezoceramic disk can be measured by the impedance analysis. For the transverse vibration, the effect of partial-electroded distribution can be neglected and served as the completely electrode-covered case. The finite element method (FEM) is also employed and good agreement is obtained for experimental, theoretical and numerical results. Finally, to understand the influences on the electrode-covered ratio, the variations of extensional resonant frequencies and dynamic electromechanical coupling coefficient are calculated in this work. It is found that the dynamic electromechanical coupling coefficients can be increased or even equal to zero, which depends on the vibration modes and proper electrode distribution.

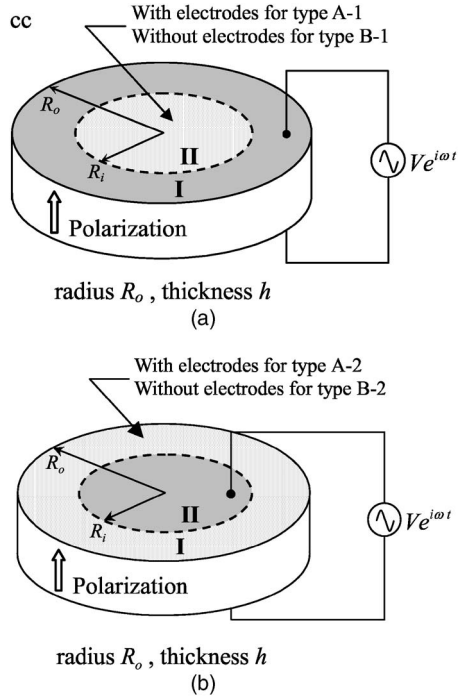


FIG. 1. Configuration of the piezoceramic disk with partial electrodes.

II. THEORETICAL ANALYSIS OF THE PARTIAL-ELECTRODED PIEZOCERAMIC DISK

Figure 1 shows the geometrical configuration of the piezoceramic disk with radius R_o and thickness h, where the cylindrical coordinates (r,θ,z) with origin is located at the center of the disk. The piezoceramic disk is assumed to be thin $(h \ll R_o)$ and polarized in the thickness direction. Both of the main surfaces are divided into regions I and II by a circumferential circle at $r=R_i$. For the theoretical analysis, the system of governing equations needed to determine the vibration characteristics of an axisymmetric partial-electroded piezoceramic disk is presented following Rogacheva.[16] The differential equations of equilibrium are

$$\frac{\partial \sigma_{rr}}{\partial r}+\frac{1}{r}\frac{\partial \sigma_{r\theta}}{\partial \theta}+\frac{\partial \sigma_{rz}}{\partial z}+\frac{1}{r}(\sigma_{rr}-\sigma_{\theta\theta})=\rho\frac{\partial^2 u}{\partial t^2}, \tag{1a}$$

$$\frac{\partial \sigma_{r\theta}}{\partial r}+\frac{1}{r}\frac{\partial \sigma_{\theta\theta}}{\partial \theta}+\frac{\partial \sigma_{\theta z}}{\partial z}+\frac{2}{r}\sigma_{r\theta}=\rho\frac{\partial^2 v}{\partial t^2}, \tag{1b}$$

$$\frac{\partial \sigma_{rz}}{\partial r}+\frac{1}{r}\frac{\partial \sigma_{\theta z}}{\partial \theta}+\frac{\partial \sigma_{zz}}{\partial z}+\frac{1}{r}\sigma_{rz}=\rho\frac{\partial^2 w}{\partial t^2}, \tag{1c}$$

where σ_{rr}, $\sigma_{r\theta}$, and $\sigma_{zz} \rightarrow \sigma_{rr}, \sigma_{r\theta}, \ldots, \sigma_{zz}$ are the components of stress; u, v, and w are the displacement field in the r, θ, and z directions, respectively; and ρ is the density. The strain-mechanical displacement relations are

$$e_{rr}=\frac{\partial u}{\partial r},\quad e_{\theta\theta}=\frac{u}{r}+\frac{1}{r}\frac{\partial v}{\partial\theta},$$

$$e_{zz}=\frac{\partial w}{\partial z},\quad e_{r\theta}=\frac{1}{r}\frac{\partial u}{\partial\theta}+\frac{\partial v}{\partial r}-\frac{v}{r}, \tag{2}$$

$$e_{rz}=\frac{\partial u}{\partial z}+\frac{\partial w}{\partial r},\quad e_{\theta z}=\frac{\partial v}{\partial z}+\frac{1}{r}\frac{\partial w}{\partial\theta},$$

where e_{rr}, $e_{r\theta}$, and $e_{zz}\rightarrow e_{rr},e_{r\theta},\ldots,e_{zz}$ are the components of strain. The linear piezoceramic constitutive equations for a piezoceramic material with crystal symmetry class $C_{6\ \mathrm{mm}}$ are

$$\begin{bmatrix} e_{rr}\\ e_{\theta\theta}\\ e_{zz}\\ e_{\theta z}\\ e_{rz}\\ e_{r\theta}\\ D_r\\ D_\theta\\ D_z \end{bmatrix} = \begin{bmatrix} s_{11}^E & s_{12}^E & s_{13}^E & 0 & 0 & 0 & 0 & 0 & d_{31}\\ s_{12}^E & s_{11}^E & s_{13}^E & 0 & 0 & 0 & 0 & 0 & d_{31}\\ s_{13}^E & s_{13}^E & s_{33}^E & 0 & 0 & 0 & 0 & 0 & d_{33}\\ 0 & 0 & 0 & s_{44}^E & 0 & 0 & 0 & d_{15} & 0\\ 0 & 0 & 0 & 0 & s_{44}^E & 0 & d_{15} & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & s_{66}^E & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & d_{15} & 0 & \varepsilon_{11}^T & 0 & 0\\ 0 & 0 & 0 & d_{15} & 0 & 0 & 0 & \varepsilon_{11}^T & 0\\ d_{31} & d_{31} & d_{33} & 0 & 0 & 0 & 0 & 0 & \varepsilon_{33}^T \end{bmatrix} \times \begin{bmatrix} \sigma_{rr}\\ \sigma_{\theta\theta}\\ \sigma_{zz}\\ \sigma_{\theta z}\\ \sigma_{rz}\\ \sigma_{r\theta}\\ E_r\\ E_\theta\\ E_z \end{bmatrix}, \tag{3}$$

where s_{11}^E,s_{12}^E, and $s_{66}^E\rightarrow s_{11}^E,s_{12}^E,\ldots,s_{66}^E$ are the compliance constants at constant electrical field; d_{15},d_{31},d_{33} are the piezoelectric constants; $\varepsilon_{11}^T,\varepsilon_{33}^T$ are the dielectric constants at constant stress; D_r,D_θ,D_z are the electrical displacement components, and E_r,E_θ,E_z are the electrical fields.

The charge equations of electrostatics is given by

$$\frac{\partial D_r}{\partial r}+\frac{1}{r}\frac{\partial D_\theta}{\partial\theta}+\frac{1}{r}D_r+\frac{\partial D_z}{\partial z}=0. \tag{4}$$

The electric field–electric potential relations are

$$E_r=-\frac{\partial\varphi}{\partial r},\quad E_\theta=-\frac{1}{r}\frac{\partial\varphi}{\partial\theta},\quad E_z=-\frac{\partial\varphi}{\partial z}, \tag{5}$$

where φ is the electrical potential.

To simplify the analysis for thin piezoceramic disks, some basic hypotheses are also employed in the analysis.[16] (a) For the thin disk, normal stress σ_{zz} is very small and can be neglected relative to the principal stresses σ_{rr} and $\sigma_{\theta\theta}$, i.e., $\sigma_{zz}=0$. (b) The rectilinear element normal to the middle surface ($z=0$) before deformation remains perpendicular to the strained surface after deformation and the elongation of which can be neglected, i.e., the shear strains $e_{rz}=e_{\theta z}=0$. (c) When the time dependence is suppressed, the electrical potential φ varies quadratically along the z direction; that is, $\varphi(r,\theta,z)=\varphi_0(r,\theta)+z\varphi_1(r,\theta)+z^2\varphi_2(r,\theta)$, where φ_0, φ_1, and φ_2 are unknown parameters. (d) Electrical displacement D_z is a constant with respect to plate thickness; in fact, this is a consequence of Eq. (4) for thin piezoceramic disks. It is noted that the hypotheses (a) and (b) correspond to the first and second Kirchhoff–Love hypotheses, respectively.

According to hypothesis (a), the electroelasticity relation of Eq. (3) can be simplified as

$$\sigma_{rr}=\frac{1}{s_{11}^E(1-\nu_p^2)}(e_{rr}+\nu_p e_{\theta\theta})-\frac{d_{31}}{s_{11}^E(1-\nu_p)}E_z, \tag{6a}$$

$$\sigma_{\theta\theta}=\frac{1}{s_{11}^E(1-\nu_p^2)}(e_{\theta\theta}+\nu_p e_{rr})-\frac{d_{31}}{s_{11}^E(1-\nu_p)}E_z, \tag{6b}$$

$$\sigma_{r\theta}=\frac{1}{s_{66}^E}e_{r\theta}=\frac{e_{r\theta}}{2s_{11}^E(1+\nu_p)}, \tag{6c}$$

$$D_z=d_{31}(\sigma_{rr}+\sigma_{\theta\theta})+\varepsilon_{33}^T E_z, \tag{6d}$$

where $\nu_p=-s_{12}^E/s_{11}^E$ is the planar Poisson's ratio. From the hypothesis (b), the displacement fields can be expressed as

$$u=u(r,\theta,z,t)=u_r(r,\theta,t)+z\frac{\partial w_z(r,\theta,t)}{\partial r},$$

$$v=v(r,\theta,z,t)=v_\theta(r,\theta,t)+\frac{z}{r}\frac{\partial w_z(r,\theta,t)}{\partial\theta},$$

$$w=w(r,\theta,z,t)=w_z(r,\theta,t).$$

It is noted that $u_r=u|_{z=0}$, $v_\theta=v|_{z=0}$, and $w_z=w|_{z=0}$ represent the radial, the tangential, and the transverse displacements of the middle surface of the disk, respectively. The strain-mechanical displacement relations presented in Eq. (2) can be rewritten as

$$e_{rr}=\frac{\partial u_r}{\partial r}+z\frac{\partial^2 w_z}{\partial r^2}, \tag{7a}$$

$$e_{\theta\theta}=\frac{u_r}{r}+\frac{1}{r}\frac{\partial v_\theta}{\partial\theta}+\frac{z}{r}\left[\frac{\partial w_z}{\partial r}+\frac{\partial}{\partial\theta}\left(\frac{1}{r}\frac{\partial w_z}{\partial\theta}\right)\right], \tag{7b}$$

$$e_{r\theta}=\frac{1}{r}\frac{\partial u_r}{\partial\theta}+\frac{z}{r}\frac{\partial^2 w_z}{\partial r\partial\theta}+\frac{\partial v_\theta}{\partial r}+z\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial w_z}{\partial\theta}\right)-\frac{v_\theta}{r}-\frac{z}{r^2}\frac{\partial w_z}{\partial\theta}. \tag{7c}$$

Due to the different electrical connections and electrode distributions, as indicated in Fig. 1, there are four types of axisymmetric partial-electroded piezoceramic disks that will be presented in this paper. When the region of electrode-covered surfaces is connected to an alternating electrical potential $Ve^{i\omega t}$, the other region with electrodes being disconnected is termed type A and that without electrodes is termed type B. We analyze in detail the dynamic radial extensional vibration characteristics, including the resonant frequencies and electrical currents. The nonaxisymmetric modes are also

investigated for transverse vibration, while the extensional vibrations are restricted to axisymmetric modes.

A. Type A—With electrodes disconnected

For the type A-1 configuration, as shown in Fig. 1(a), the piezoceramic disk with annular electrode-covered surfaces on region I ($R_i < r < R_o$) is connected to the voltage $Ve^{i\omega t}$ and the other circular region II ($0 < r < R_i$) with electrodes is disconnected.

Suppose that the extensional vibration is axisymmetric, the radial extensional displacement field of the middle plane can be assumed to be

$$u_r(r,t) = U(r)e^{i\omega t}, \tag{8}$$

where ω is the angular frequency. If the time-dependent term $e^{i\omega t}$ is uniformly suppressed in the analysis, the stress-displacement relations for the extensional vibration are given by

$$\sigma_{rr} = \frac{1}{s_{11}^E(1-\nu_p^2)}\left(\frac{dU}{dr} + \nu_p\frac{U}{r}\right) + \frac{d_{31}}{s_{11}^E(1-\nu_p)}\cdot\frac{2V}{h}, \tag{9a}$$

$$\sigma_{\theta\theta} = \frac{1}{s_{11}^E(1-\nu_p^2)}\left(\frac{U}{r} + \nu_p\frac{dU}{dr}\right) + \frac{d_{31}}{s_{11}^E(1-\nu_p)}\cdot\frac{2V}{h}, \tag{9b}$$

in which the electrical potential boundary condition on region I

$$\varphi|_{z=\pm h/2} = \pm V \tag{10}$$

has been employed. Substituting Eq. (9) into the equilibrium equation (1a) and integrating over the disk thickness, we have the well-known governing equation of extensional vibration

$$\frac{d^2U}{dr^2} + \frac{1}{r}\frac{dU}{dr} - \frac{1}{r^2}U - \rho\omega^2 s_{11}^E(1-\nu_p^2)U = 0. \tag{11}$$

According to Eq. (11), the displacement field of the piezoceramic disk for type A-1 is found to be

$$U(r) = C_1^{(A)}J_1(\beta_1 r), \quad 0 < r < R_i, \tag{12a}$$

$$U^{(e)}(r) = C_2^{(A)}J_1(\beta_1 r) + C_3^{(A)}Y_1(\beta_1 r), \quad R_i < r < R_o, \tag{12b}$$

where the superscript (e) represents the electrode-covered region being connected and

$$\beta_1^2 = \rho s_{11}^E(1-\nu_p^2)\omega^2. \tag{12c}$$

In Eq. (12), J_1 and Y_1 are first-order Bessel functions of the first and second kinds, respectively.

From the continuum and boundary conditions,

$$U(r) = U^{(e)}(r) \quad \text{at } r = R_i, \tag{13a}$$

$$\int_{-h/2}^{h/2}\sigma_{rr}dz = \int_{-h/2}^{h/2}\sigma_{rr}^{(e)}dz \quad \text{at } r = R_i, \tag{13b}$$

$$\int_{-h/2}^{h/2}\sigma_{rr}^{(e)} = 0 \quad \text{at } r = R_o, \tag{13c}$$

the constants $C_1^{(A)}$, $C_2^{(A)}$, and $C_3^{(A)}$ are found to be

$$C_1^{(A)} = \frac{2Vd_{31}(1+\nu_p)R_o}{h\Delta_1^{(A)}}\cdot\{aJ_1(a\eta)[\eta Y_0(\eta) - \eta Y_0(a\eta) - (1-\nu_p)Y_1(\eta)] - aY_1(a\eta)[\eta J_0(\eta) - \eta J_0(a\eta) - (1-\nu_p)J_1(\eta)]\}, \tag{14a}$$

$$C_2^{(A)} = \frac{2Vd_{31}(1+\nu_p)R_o}{h\Delta_1^{(A)}}\cdot\left\langle J_1(a\eta)\left\{a\eta[Y_0(\eta) - Y_0(a\eta)] - a(1-\nu_p)Y_1(\eta) + \frac{k_p^2(1+\nu_p)}{1-k_p^2}Y_1(a\eta)\right\} + a\eta J_0(a\eta)Y_1(a\eta)\right\rangle, \tag{14b}$$

$$C_3^{(A)} = \frac{2Vd_{31}(1+\nu_p)R_o}{h\Delta_1^{(A)}}\cdot J_1(a\eta)\left\{a(1-\nu_p)J_1(\eta) - a\eta J_0(\eta) - \frac{k_p^2(1+\nu_p)}{1-k_p^2}J_1(a\eta)\right\}, \tag{14c}$$

where $\eta = \beta_1 R_o$ and

$$\Delta_1^{(A)} = \frac{k_p^2(1+\nu_p)}{1-k_p^2}J_1(a\eta)[\eta J_1(a\eta)Y_0(\eta) - \eta J_0(\eta)Y_1(a\eta) + (1-\nu_p)J_1(\eta)Y_1(a\eta) - (1-\nu_p)J_1(a\eta)Y_1(\eta)] + a\eta[J_1(a\eta)Y_0(a\eta) - J_0(a\eta)Y_1(a\eta)]\cdot[\eta J_0(\eta) - (1-\nu_p)J_1(\eta)]. \tag{15}$$

Equation (14) denotes that $a = R_i/R_o$ is the electrode-covered ratio and

$$k_p = \sqrt{\frac{2d_{31}^2}{\varepsilon_{33}^T s_{11}^E(1-\nu_p)}}$$

is the planar electromechanical coupling coefficient.

By employing the constants $C_1^{(A)}$, $C_2^{(A)}$, and $C_3^{(A)}$, the electrical current $I_1^{(A)}$ for type A-1 can be expressed as

$$I_1^{(A)} = \frac{\partial}{\partial t}\int_\Omega D_3 d\Omega = \frac{i\omega\varepsilon_{33}^T k_p^2}{d_{31}}\cdot\pi R_o\{C_2^{(A)}[J_1(\eta) - aJ_1(a\eta)] + C_3^{(A)}[Y_1(\eta) - aY_1(a\eta)]\} + \frac{i\omega V\varepsilon_{33}^T(k_p^2-1)}{h}\cdot 2\pi R_o^2(1-a^2). \tag{16}$$

From Eq. (16), the resonant frequencies of extensional vibration can be found whenever the current $I_1^{(A)}$ approaches infinity. Referring to Eqs. (14b), (14c), and (15), the characteristic equation of resonant frequencies for extensional vibration is given for type A-1 as

$$\frac{k_p^2(1+\nu_p)}{1-k_p^2}J_1(a\eta)[\eta J_1(a\eta)Y_0(\eta)-\eta J_0(\eta)Y_1(a\eta)+(1-\nu_p)\times J_1(\eta)Y_1(a\eta)-(1-\nu_p)J_1(a\eta)Y_1(\eta)]$$
$$=a\eta[J_0(a\eta)Y_1(a\eta)-J_1(a\eta)Y_0(a\eta)]\cdot[\eta J_0(\eta)-(1-\nu_p)J_1(\eta)]. \quad (17)$$

Following the similar procedure as type A-1, the radial extensional displacements of type A-2 can be expressed as

$$U^{(e)}(r)=C_4^{(A)}J_1(\beta_1 r),\quad 0<r<R_i, \quad (18a)$$

$$U(r)=C_5^{(A)}J_1(\beta_1 r)+C_6^{(A)}Y_1(\beta_1 r),\quad R_i<r<R_o. \quad (18b)$$

By the continuum and boundary conditions for type A-2, the constants $C_4^{(A)}$, $C_5^{(A)}$, and $C_6^{(A)}$ can be obtained as

$$C_4^{(A)}=\frac{2Vd_{31}(1+\nu_p)R_i}{h\Delta_2^{(A)}}\cdot\left\{Y_1(a\eta)[\eta J_0(\eta)-(1-\nu_p)J_1(\eta)]-J_1(a\eta)[\eta Y_0(\eta)-(1-\nu_p)Y_1(\eta)]+\frac{k_p^2(1+\nu_p)}{(1-k_p^2)(1-a^2)}[J_1(\eta)Y_1(a\eta)-J_1(a\eta)Y_1(\eta)]\right\}, \quad (19a)$$

$$C_5^{(A)}=\frac{2Vd_{31}(1+\nu_p)R_i}{h\Delta_2^{(A)}}\cdot J_1(a\eta)\left[(1-\nu_p)Y_1(\eta)-\eta Y_0(\eta)-\frac{k_p^2(1+\nu_p)}{(1-k_p^2)(1-a^2)}[Y_1(\eta)-aY_1(a\eta)]\right], \quad (19b)$$

$$C_6^{(A)}=\frac{2Vd_{31}(1+\nu_p)R_i}{h\Delta_2^{(A)}}\cdot J_1(a\eta)\left[\eta J_0(\eta)-(1-\nu_p)J_1(\eta)+\frac{k_p^2(1+\nu_p)}{(1-k_p^2)(1-a^2)}[J_1(\eta)-aJ_1(a\eta)]\right], \quad (19c)$$

in which

$$\Delta_2^{(A)}=\frac{a^2k_p^2(1+\nu_p)}{(1-k_p^2)(1-a^2)}J_1(a\eta)\cdot\{(1-\nu_p)[J_1(\eta)Y_1(a\eta)-J_1(a\eta)Y_1(\eta)]+\eta J_1(a\eta)[Y_0(\eta)-Y_0(a\eta)]-\eta Y_1(a\eta)[J_0(\eta)-J_0(a\eta)]\}+\frac{ak_p^2(1+\nu_p)}{(1-k_p^2)(1-a^2)}\times\eta J_1(a\eta)[J_0(\eta)Y_1(\eta)-J_1(\eta)Y_0(\eta)]+a\eta[J_0(a\eta)Y_1(a\eta)-J_1(a\eta)Y_0(a\eta)]\cdot\left\{\left[1-\nu_p-\frac{k_p^2(1+\nu_p)}{(1-k_p^2)(1-a^2)}\right]J_1(\eta)-\eta J_0(\eta)\right\}. \quad (20)$$

The electrical current $I_2^{(A)}$ can be expressed as

$$I_2^{(A)}=\frac{\partial}{\partial t}\int_\Omega D_3 d\Omega=\frac{i\omega\varepsilon_{33}^T k_p^2}{d_{31}}\cdot\pi R_o[C_4^{(A)}aJ_1(a\eta)]+\frac{i\omega V\varepsilon_{33}^T(k_p^2-1)}{h}\cdot 2\pi R_o^2a^2, \quad (21)$$

and for type A-2 the characteristic equation of resonant frequencies for extensional vibrations is

$$\Delta_2^{(A)}=0. \quad (22)$$

B. Type B—Without electrodes

For a piezoceramic disk whose surfaces are not covered with electrodes, the electrical conditions $D_z|_{z=\pm h/2}=0$ are employed and the radial displacement has the form of a first-kind Bessel function.[16] Using the preliminaries, we can obtain the radial displacement field of the piezoceramic disk with annular electrode-covered region, called type B-1 as shown in Fig. 1(a), are

$$U(r)=C_1^{(B)}J_1(\beta_2 r),\quad 0<r<R_i, \quad (23a)$$

$$U^{(e)}(r)=C_2^{(B)}J_1(\beta_1 r)+C_3^{(B)}Y_1(\beta_1 r),\quad R_i<r<R_o, \quad (23b)$$

where

$$\beta_2^2=\rho\frac{2s_{11}^E(1-\nu_p^2)(1-2k_p^2)}{2-(1-\nu_p)k_p^2}\omega^2. \quad (23c)$$

From the continuum and boundary conditions, as shown in Eqs. (13a)–(13c), the expressions of the constants $C_1^{(B)}$, $C_2^{(B)}$, and $C_3^{(B)}$ are

$$C_1^{(B)}=\frac{2Vd_{31}(1+\nu_p)R_o}{h\Delta_1^{(B)}}\cdot\{J_1(a\eta)[a\eta Y_0(\eta)-a(1-\nu_p)Y_1(\eta)-a\eta Y_0(a\eta)+(1-\nu_p)Y_1(a\eta)]-Y_1(a\eta)[a\eta J_0(\eta)-a(1-\nu_p)J_1(\eta)-a\eta J_0(a\eta)+(1-\nu_p)J_1(a\eta)]\}, \quad (24a)$$

$$C_2^{(B)}=\frac{2Vd_{31}(1+\nu_p)R_o}{h\Delta_1^{(B)}}\cdot\left\{J_1(a\zeta)[a\eta Y_0(\eta)-a(1-\nu_p)Y_1(\eta)-a\eta Y_0(a\eta)+(1-\nu_p)Y_1(a\eta)]+Y_1(a\eta)\left[\frac{2-(1-\nu_p)k_p^2}{2(1-k_p^2)}a\zeta J_0(a\zeta)-(1-\nu_p)J_1(a\zeta)\right]\right\}, \quad (24b)$$

TABLE I. Material properties of piezoceramics PIC-151.

PIC-151 ceramics	
s_{11}^E (10^{-12} m^2/N)	16.83
s_{33}^E	19.0
s_{12}^E	−5.656
s_{13}^E	−7.107
s_{44}^E	50.96
s_{66}^E	44.97
d_{31} (10^{-10} m/V)	−2.14
d_{33}	4.23
d_{15}	6.1
ε_{11}^T (10^{-9} F/m)	17.134
ε_{33}^T	18.665
ρ (kg/m^3)	7800

$$C_3^{(B)} = \frac{2Vd_{31}(1+\nu_p)R_o}{h\Delta_1^{(B)}} \cdot \left\{ J_1(a\zeta)[a\eta J_0(a\eta) - (1-\nu_p) \times J_1(a\eta) - a\eta J_0(\eta) + a(1-\nu_p)J_1(\eta)] - J_1(a\eta) \times \left[\frac{2-(1-\nu_p)k_p^2}{2(1-k_p^2)} a\zeta J_0(a\zeta) - (1-\nu_p)J_1(a\zeta) \right] \right\}, \tag{24c}$$

in which $\zeta=\beta_2 R_o$ and

$$\Delta_1^{(B)} = \eta J_1(a\zeta)Y_0(a\eta)\cdot[\eta J_0(\eta) - (1-\nu_p)J_1(\eta)] - \eta J_0(a\eta)J_1(a\zeta)\cdot[\eta Y_0(\eta) - (1-\nu_p)Y_1(\eta)] - \frac{2-(1-\nu_p)k_p^2}{2(1-k_p^2)}\zeta J_0(a\zeta)\cdot\{(1-\nu_p)[J_1(a\eta)Y_1(\eta) - J_1(\eta)Y_1(a\eta)] + \eta[J_0(\eta)Y_1(a\eta) - J_1(a\eta)Y_0(\eta)]\}. \tag{25}$$

The electrical current $I_1^{(B)}$ for type B-1 can be expressed as

$$I = \frac{\partial}{\partial t}\int_\Omega D_3 d\Omega = \frac{i\omega\varepsilon_{33}^T k_p^2}{d_{31}} \cdot \pi R_o\{C_2^{(B)}[J_1(\eta) - aJ_1(a\eta)] + C_3^{(B)}[Y_1(\eta) - aY_1(a\eta)]\} + \frac{i\omega V\varepsilon_{33}^T(k_p^2-1)}{h} \cdot 2\pi R_o^2(1-a^2), \tag{26}$$

and the characteristic equation of resonant frequencies for extensional vibration is

$$\Delta_1^{(B)} = 0. \tag{27}$$

For the case of type B-2, the radial displacement field can be expressed as

$$U^{(e)}(r) = C_4^{(B)}J_1(\beta_1 r), \quad 0 < r < R_i, \tag{28a}$$

$$U(r) = C_5^{(B)}J_1(\beta_2 r) + C_6^{(B)}Y_1(\beta_2 r), \quad R_i < r < R_o, \tag{28b}$$

and the constants $C_4^{(B)}$, $C_5^{(B)}$, and $C_6^{(B)}$ are

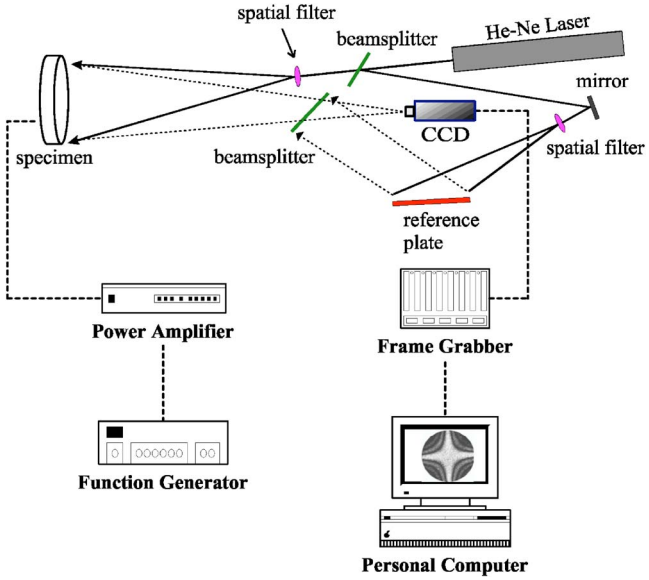


FIG. 2. Schematic diagram of AF-ESPI setup for out-of-plane measurement.

$$C_4^{(B)} = \frac{2Vd_{31}(1+\nu_p)R_i}{h\Delta_2^{(B)}} \cdot \left\{ \frac{2-(1-\nu_p)k_p^2}{2(1-k_p^2)}\zeta[J_0(\zeta)Y_1(a\zeta) - J_1(a\zeta)Y_0(\zeta)] - (1-\nu_p)[J_1(\zeta)Y_1(a\zeta) - J_1(a\zeta)Y_1(\zeta)] \right\}, \tag{29a}$$

$$C_5^{(B)} = \frac{2Vd_{31}(1+\nu_p)R_i}{h\Delta_2^{(B)}} \cdot J_1(a\eta)\left[(1-\nu_p)Y_1(\zeta) - \frac{2-(1-\nu_p)k_p^2}{2(1-k_p^2)}\zeta Y_0(\zeta) \right], \tag{29b}$$

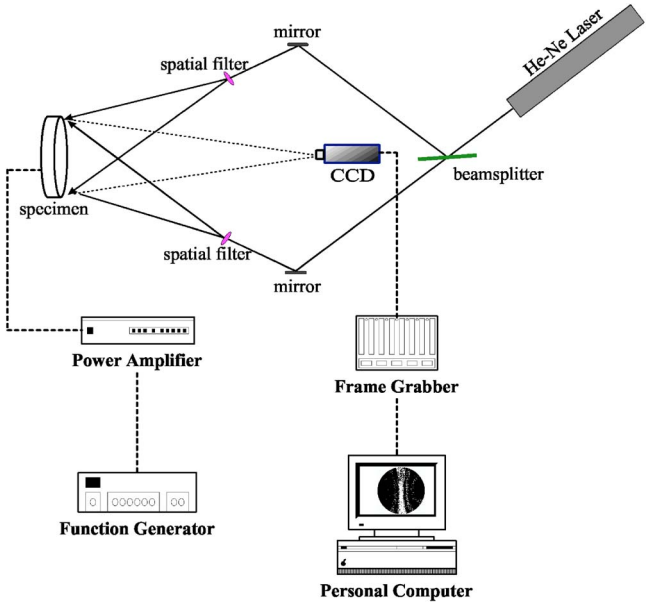


FIG. 3. Schematic diagram of AF-ESPI setup for in-plane measurement.

TABLE II. Results for the first eight modes obtained by theory, FEM, and AF-ESPI of transverse vibration. Square brackets represent the values calculated for the completely electrode-covered piezoceramic disk.

Transverse mode	Theory (Hz)	FEM (Hz)	Error	AF-ESPI (Hz)	Mode shape
1	3 224	3193 [3193]	0.97%	3 200	
2	7 040	6991 [6990]	0.70%	6 485	
3	7 638	7499 [7499]	1.85%	7 460	
4	13 602	13 238 [13 238]	2.75%	13 200	
5	15 358	15 099 [15 095]	1.72%	14 300	
6	21 094	20 371 [20 371]	3.55%	20 180	
7	25 924	25 172 [25 165]	2.99%	24 000	
8	28 786	27 983 [27 925]	2.87%	26 350	

$$C_6^{(B)} = \frac{2Vd_{31}(1+\nu_p)R_i}{h\Delta_2^{(B)}} \cdot J_1(a\eta)\left[\frac{2-(1-\nu_p)k_p^2}{2(1-k_p^2)}\zeta J_0(\zeta) - (1-\nu_p)J_1(\zeta)\right], \tag{29c}$$

where

$$\Delta_2^{(B)} = \eta J_0(a\eta)\left\{J_1(a\zeta)\left[\frac{2-(1-\nu_p)k_p^2}{2(1-k_p^2)}\zeta Y_0(\zeta) - (1-\nu_p)Y_1(\zeta)\right] - Y_1(a\zeta)\left[\frac{2-(1-\nu_p)k_p^2}{2(1-k_p^2)}\zeta J_0(\zeta) - (1-\nu_p)J_1(\zeta)\right]\right\} - \frac{2-(1-\nu_p)k_p^2}{2(1-k_p^2)}\zeta J_1(a\eta)\left\{Y_0(a\zeta)\times\left[-\frac{2-(1-\nu_p)k_p^2}{2(1-k_p^2)}\zeta J_0(\zeta) + (1-\nu_p)J_1(\zeta)\right] + J_0(a\zeta)\left[\frac{2-(1-\nu_p)k_p^2}{2(1-k_p^2)}\zeta Y_0(\zeta) - (1-\nu_p)Y_1(\zeta)\right]\right\}. \tag{30}$$

The electrical current $I_2^{(B)}$ for type B-2 can be expressed as

$$I = \frac{\partial}{\partial t}\int_\Omega D_3 d\Omega = \frac{i\omega\varepsilon_{33}^T k_p^2}{d_{31}} \cdot \pi R_o[C_4^{(B)} a J_1(a\eta)] + \frac{i\omega V\varepsilon_{33}^T(k_p^2-1)}{h} \cdot 2\pi R_o^2 a^2 \tag{31}$$

and the characteristic equation of resonant frequencies for extensional vibration is shown as

$$\Delta_2^{(B)} = 0. \tag{32}$$

Not only the radial extensional vibration, but also the transverse vibration are discussed in the analysis. Suppose that the transverse vibration is independent of the electrode-covered area and the displacement in the z direction is non-axisymmetric, the analytical solutions can be obtained by those for the completely electrode-covered piezoceramic disk.[17] Herein, the derivation is suppressed and only the theoretical and experimental results are presented in Sec. III.

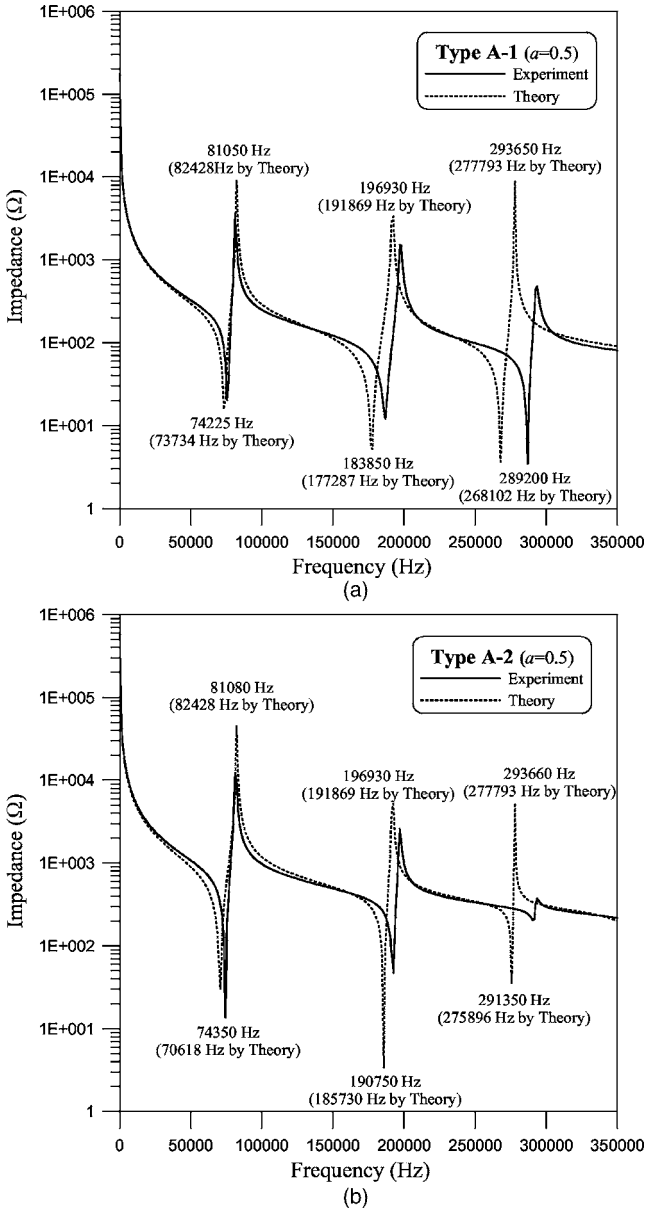


FIG. 4. Experimental and simulated impedance variation curves of the (a) type A-1 and (b) type A-2 piezoceramic disks.

III. EXPERIMENTAL MEASUREMENT AND THEORETICAL INVESTIGATION

The piezoceramic disk with R_o=15 mm and h=1 mm, and with the modal number of PIC-151 (Physik Instrument Company, Germany) is selected for experimental measurement. The polarization is in the z direction, as shown in Fig. 1, and the faces at $z=\pm h/2$ of the disk are coated with electrodes, with an electrode cutting circle at R_i=7.5 mm. Consequently, the specimen configurations for the experiments

TABLE III. Results of the first three modes obtained by theory, FEM, impedance analysis, and AF-ESPI of the extensional vibration for (a) type A-1 and (b) type A-2 piezoceramic disks.

Extensional mode	(a) Theory (Hz)	(b) FEM (Hz)	Error (a)/(b)	(c) Impedance (Hz)	(d) AF-ESPI (Hz)	Difference (c)/(d)	Mode shape
(a) Square brackets represent the values calculated for the type B-1.							
1	73 734 [73 789]	⋯ [73 097]	⋯ [0.95%]	74 225	74 220	0.01%	
2	177 287 [185 876]	⋯ [185 329]	⋯ [0.30%]	183 850	183 450	0.22%	
3	268 102 [294 230]	⋯ [289 906]	⋯ [1.49%]	289 200	288 350	0.29%	
(b) Square brackets represent the values calculated for the type B-2.							
1	70 618 [71 732]	⋯ [71 152]	⋯ [0.82%]	74 350	74 300	0.07%	
2	185 730 [189 642]	⋯ [188 510]	⋯ [0.60%]	190 750	190 800	–0.03%	
3	275 896 [291 592]	⋯ [288 659]	⋯ [1.02%]	291 350	291 480	–0.04%	

are available for both types A-1 and A-2 herein. The electroelastic properties of the specimen PIC-151 are listed in Table I.

The schematic layout of time-averaged AF-ESPI optical systems, as shown in Figs. 2 and 3, are used to perform the out-of-plane and in-plane experimental measurements for resonant frequencies and corresponding mode shapes. A continuous-wave He–Ne laser (35 mW, Melles Griot) with $\lambda = 632.8$ nm is used as the coherent light source. A CCD camera (Pulnix Company, TM-7CN) and frame grabber (Dipix Technologies Inc., P360F) with an onboard digital signal processor are applied to record and process the images. To achieve the sinusoidal output, a digitally controlled function generator (Hewlett Packard, HP33120A) connected to a power amplifier (NF Electronic Corporation, Type 4005) is used. Ma and Huang[14] provided a detailed discussion of the AF-ESPI method to investigate the out-of-plane and in-plane vibrations, including the theoretical analysis and experimental procedure.

Based on the experimental results obtained by the AF-ESPI optical technique, the resonant frequencies and corresponding mode shapes can be measured for both the transverse and extensional vibrations. In addition to the theoretical analysis and experimental measurement, numerical calculations are performed by the commercially available software ABAQUS finite element package,[18] in which 20-node three-dimensional solid piezoelectric elements (C3D20E) are selected to analyze the problem. The electrical potential on the connecting partial-electroded surfaces of piezoceramic disks is specified as "zero" to simulate the closed-circuit condition for the resonant frequency extraction. Table II shows the experimental, theoretical, and FEM results for the first eight modes of transverse vibration. The zero-order fringe of the image pattern, which is the brightest fringes for the mode shapes obtained by AF-ESPI, represents the nodal lines of the vibrating piezoceramic disk at resonance. It is noted that the piezoceramic disk, which with crystal symmetry class $C_{6\ mm}$, can be indeed served as the transverse isotropic disk for transverse vibration. Consequently, both the nonaxisymmetric and axisymmetric modes are present in the analytical and experimental investigations. The resonant frequencies of transverse vibration are calculated on the assumption that they are independent of the electrode-covered distribution. According to the FEM calculations for the completely electrode-covered piezoceramic disk, as shown in Table II, this assumption seems to be acceptable in respect of transverse vibration.

Because the electrical impedance of piezoceramic materials drops to a local minimum when it vibrates in resonance, the resonant frequency can also be determined by impedance analysis. Experimental impedance measurement of the piezoceramic disk is obtained by using an impedance/gain-phase analyzer (Hewlett Packard, HP4194A). The experimental and simulated impedance variation curves for types A-1 and A-2 piezoceramic disks are shown in Figs. 4(a) and 4(b) respectively. The local minima and maxima appearing in the impedance variation curves correspond to resonance and antiresonance, respectively. The simulated results shown in Figs. 4(a) and 4(b) are calculated by Eqs. (16) and (21), respectively. It is found that only the resonant frequencies of the radial extensional modes are indicated in Fig. 4, i.e., those of the transverse modes cannot be obtained by the impedance analysis. This phenomenon can be explained qualitatively as follows by means of the characteristics of piezoelectricity. When the piezoceramic disk vibrates at resonance, the charge will be strongly induced on the electrode surfaces due to the vibration deformation, called the direct piezoelectric effect, and the impedance will drop to a local minimum value. This demonstrates that the resonant frequencies of piezoceramic disks can be determined by using the impedance analyzer. However, if the summation of the induced charge distributed over the electrode surfaces is

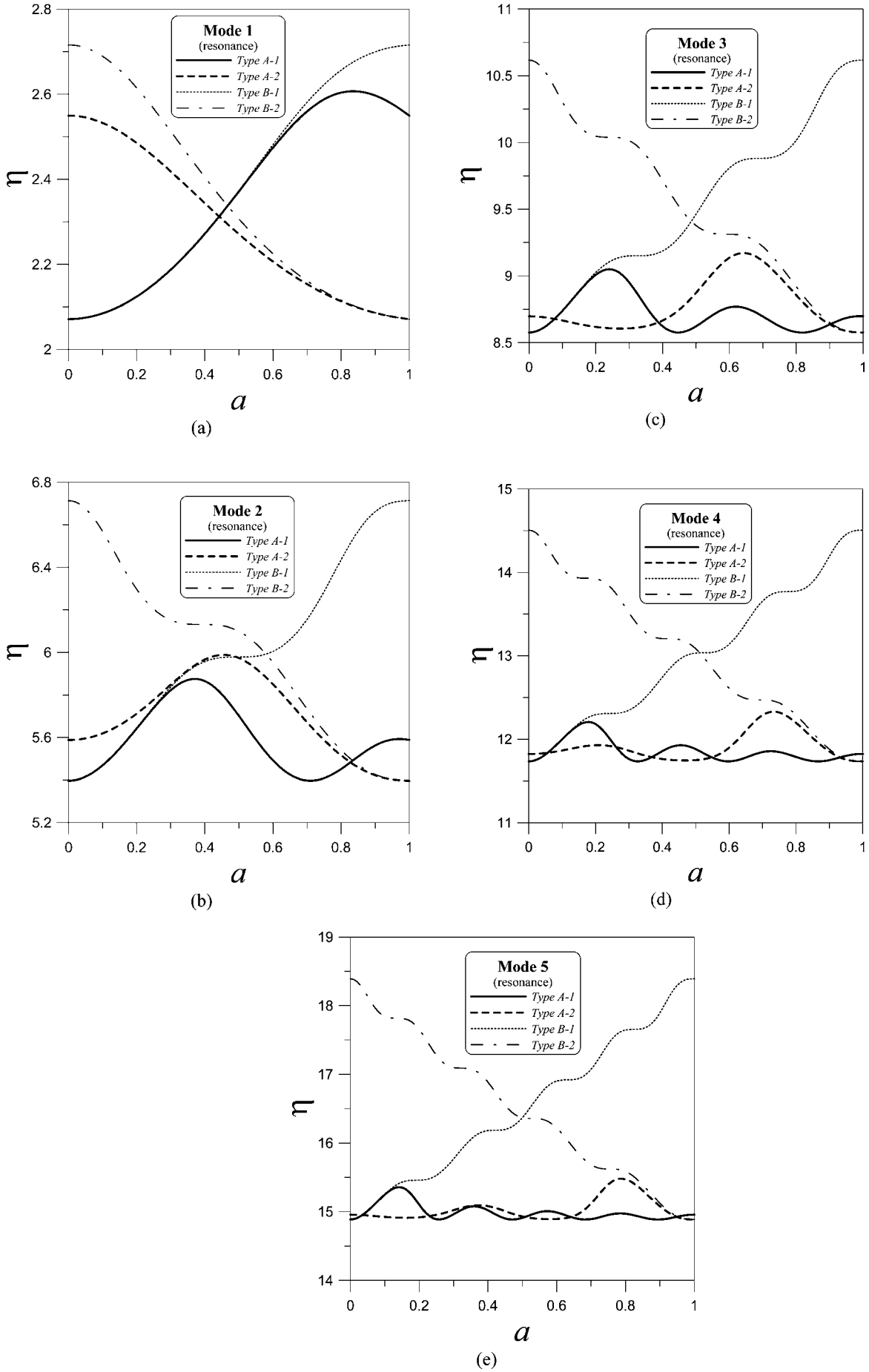


FIG. 5. The variation of frequency parameter of the extensional vibration modes for different values of α.

proportional to the input alternating potential $Ve^{i\omega t}$ only, we are not able to find the large variation of impedance at the resonant frequency. The other important feature, as indicated in Figs. 4(a) and 4(b), is that the antiresonant frequencies obtained by the experiment for type A-1 are almost the same as that for type A-2; in fact, the antiresonant frequencies of these two types are identical according to the theoretical calculations. Tables III(a) and III(b) show the first three extensional resonant frequencies of the piezoceramic disk obtained by using the AF-ESPI, impedance analysis, theoretical calculation, and FEM. It is found that there is nearly no difference between types A-1 and A-2 for the mode shapes of extensional vibration. The discrepancy of resonant frequencies between AF-ESPI and impedance analysis is smaller than 0.3%. The theoretical predictions for type B are in excellent agreement with the finite element results, and the discrepancies are within 1.5%. The errors between the theoretical and numerical results of transverse resonance are greater than that of extensional resonance. This consequence implies that thickness of the disk exhibits significant influence on the theoretical investigation for transverse vibration. The difference between the experimental measurements and analytical results may also result from the determination of the material properties or the defects of the piezoceramic disk.

From the solutions for characteristic equations of resonant frequencies, the dependence of frequency parameter (η) on electrode-covered ratio (a) is elucidated. The results shown in Figs. 5(a)–5(e) are the first five extensional modes for the four types of partial-electroded piezoceramic disks. Note that $a=0$ represents the completely electrode-covered piezoceramic disk for types A-1 and B-1, whereas $a=1$ is for types A-2 and B-2. It is found that the extreme values of frequency parameters for type B will occur at $a=0$ and $a=1$, but this situation does not happen for type A. For the higher modes of types A-1 and A-2, the locations of maximum resonant frequencies gradually move toward $a=0$ and $a=1$, respectively. Except for the first mode, the maximum resonant frequency of type A-2 is greater than that of type A-1, as shown in Figs. 5(a)–5(e). It is recognized that, for type A, the face with electrodes disconnected will induce an extra voltage in resonance, which will effect the dynamic behavior of the piezoceramic disk. Observations of the frequency variation curve of types A-1 and B-1 in Figs. 5(a)–5(e) show that there is no difference between them when the electrode-covered ratio a is below the critical value. For instance, $a<0.53$ of mode 1 and $a<0.26$ of mode 2 make the resonant frequency difference within 0.1% between the types A-1 and B-1. It is also shown that the critical value will decrease as the mode number increases. When the ratio a exceeds the critical value, the resonant frequency of type B-1 is clearly larger than that of type A-1. On the other hand, the resonant frequency of type B-2 is always larger than that of type A-2.

The electromechanical coupling coefficient is an important characteristic of piezoceramic elements for converting mechanical energy into electrical energy, or vice versa. The dynamic electromechanical coupling coefficient for near resonant frequency is introduced as[19]

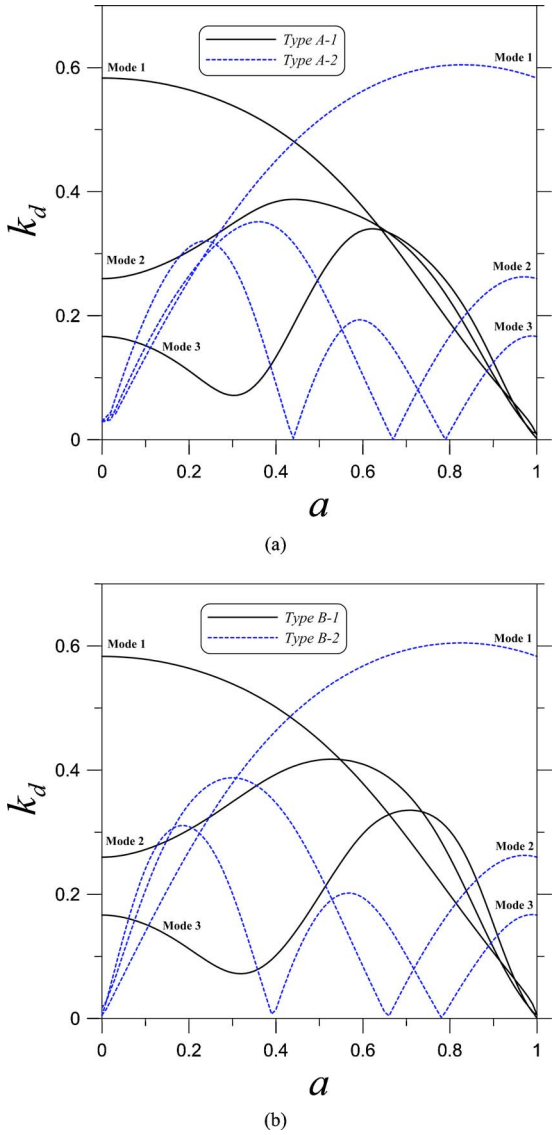


FIG. 6. The value k_d vs a for the (a) type A and (b) type B piezoceramic disks.

$$k_d = \frac{\sqrt{f_a^2 - f_r^2}}{f_a}, \tag{33}$$

where f_r is the resonant frequency and f_a is the antiresonant frequency. Equation (33) is usually employed to determine the electromechanical coupling coefficient by measuring the resonance and antiresonance. The antiresonant frequency f_a is defined as the frequency of maximum resistance, which can be evaluated by equating the electrical current to zero. According to the theoretical analysis derived in Sec. II, the value of k_d is plotted as a function of a for the first three extensional modes of types A and B in Figs. 6(a) and 6(b), respectively. As shown in Fig. 6 for types A-2 and B-2 with circular partial electrodes being connected, the values k_d of the second and third modes will oscillate and drop to zero for some values of a. Under the circumstances of k_d vanishing, the resonant frequency f_r is equal to the antiresonant frequency f_a and the energy conversion characteristics of piezoceramics will be ineffective. It can also be seen that for $a=0.83$ of types A-2 and B-2, the value k_d of mode 1

reaches 0.605 that is the maximum for all the four types discussed in this analysis. In comparison with the completely electrode-covered piezoceramic disk, by which $k_d = 0.583$, the value can be increased 3.43% by taking advantage of partial electrode manipulation.

IV. CONCLUSIONS

Many of the previous works regarding vibration analysis of piezoelectric disks are available for disks completely covered with electrodes, but there is less research on partially covered electrodes. With the aid of the electroelastic theory, the vibration characteristics of axisymmetric partial-electroded thin piezoceramic disks with traction-free boundary conditions are investigated, which comprise the radial extensional and transverse vibration modes. Based on the theoretical and experimental results by the impedance analysis, it is shown that only the resonant frequencies of radial extensional vibration can be obtained. However, both the resonant frequencies and corresponding mode shapes for transverse and extensional vibrations are measured by the AF-ESPI method. The theoretical, numerical, and experimental results are all in good agreement. The resonant frequencies of transverse vibration can be assumed to be independent of electrode-covered area, which has been verified by experimental and numerical results. The axisymmetric partial-electroded distribution significantly influences the resonant frequencies and dynamic electromechanical coupling coefficients for extensional vibration. The resonant frequencies of type A-1 are the same as those of type B-1 when the electrode-covered ratio is below the critical value; however, the resonant frequencies of type B-2 are always greater than those of type A-2. Moreover, the antiresonant frequencies of types A-1 and A-2 are identical for the theoretical calculation and demonstrated by the experimental results. The electrode-covered distribution is actually influential on the dynamic electromechanical coupling coefficient k_d. For the piezoceramic disk with circular partial electrodes being connected, the value k_d of the second and third modes will oscillate and drop to zero for certain electrode-covered ratios. It is also indicated that the maximum value of k_d can be increased by 3.43% in comparison with the completely electrode-covered piezoceramic disk.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support of this research by the National Science Council (Republic of China) under Grant No. NSC 91-2212-E-231-002.

[1] H. F. Tiersten, *Linear Piezoelectric Plate Vibrations* (Plenum, New York, 1969).

[2] "IEEE Standard on Piezoelectricity," ANSI-IEEE Std. 176, IEEE New York, 1987.

[3] H. A. Kunkel, S. Locke, and B. Pikeroen, "Finite-element analysis of vibrational modes in piezoelectric ceramics disks," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **37**, 316–328 (1990).

[4] N. Guo, P. Cawley, and D. Hitchings, "The finite element analysis of the vibration characteristics of piezoelectric discs," J. Sound Vib. **159**, 115–138 (1992).

[5] N. F. Ivina, "Numerical analysis of the normal modes of circular piezoelectric plates of finite dimensions," Sov. Phys. Acoust. **35**, 385–388 (1990).

[6] G. H. Schmidt, "Extensional vibrations of piezoelectric plates," J. Eng. Math. **6**, 133–142 (1972).

[7] N. N. Rogacheva, "The dependence of the electromechanical coupling coefficient of piezoelectric elements on the position and size of the electrodes," J. Appl. Math. Mech.0021-8928 **65**, 317–326 (2001).

[8] N. F. Ivina, "Analysis of the natural vibrations of circular piezoceramic plates with partial electrodes," Acoust. Phys. **47**, 714–720 (2001).

[9] E. A. G. Shaw, "On the resonant vibrations of thick barium titanate disks," J. Acoust. Soc. Am. **28**, 38–50 (1956).

[10] U. Minoni and F. Docchio, "An optical self-calibrating technique for the dynamic characterization of PZT's," IEEE Trans. Instrum. Meas. **40**, 851–854 (1991).

[11] M. Chang, "In-plane vibration displacement measurement using fiber-optical speckle interferometry," Precis. Eng. **16**, 36–41 (1994).

[12] B. Koyuncu, "The investigation of high frequency vibration modes of PZT-4 transducers using ESPI techniques with reference beam modulation," Opt. Lasers Eng. **1**, 37–49 (1980).

[13] J. R. Oswin, P. L. Salter, F. M. Santoyo, and J. R. Tyrer, "Electronic speckle pattern interferometric measurement of flextensional transducer vibration patterns: in air and water," J. Sound Vib. **172**, 433–448 (1994).

[14] C. C. Ma and C. H. Huang, "The investigation of three-dimensional vibration for piezoelectric rectangular parallelepipeds by using the AF-ESPI method," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 142–153 (2001).

[15] C. H. Huang and C. C. Ma, "Vibration characteristics for piezoelectric cylinders using amplitude-fluctuation electronic speckle pattern interferometry," AIAA J. **36**, 2262–2268 (1998).

[16] N. N. Rogacheva, *The Theory of Piezoelectric Shells and Plates* (CRC, Boca Raton, 1994).

[17] C. H. Huang, Y. C. Lin, and C. C. Ma, "Theoretical analysis and experimental measurement for resonant vibration of piezoceramic circular plates," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 12–24 (2004).

[18] Hibbit, Karlsson, and Sorensen, Inc., *ABAQUS User's Manual, ver. 6.4*, Pawtucket, RI, 2003.

[19] W. P. Mason, *Piezoelectric Crystals and Their Application to Ultrasonics* (Plenum, New York, 1950).

Realization of mechanical systems from second-order models

Wenyuan Chen[a] and Pierre E. Dupont[b]
Department of Aerospace and Mechanical Engineering, College of Engineering, Boston University, Boston, Massachusetts 02215

(Received 3 October 2004; revised 22 May 2005; accepted 23 May 2005)

Congruent coordinate transformations are used to convert second-order models to a form in which the mass, damping, and stiffness matrices can be interpreted as a passive mechanical system. For those systems which can be constructed from interconnected mass, stiffness, and damping elements, it is shown that the input–output preserving transformations can be parametrized by an orthogonal matrix whose dimension corresponds to the number of internal masses—those masses at which an input is not applied nor an output measured. Only a subset of these transformations results in mechanically realizable models. For models with a small number of internal masses, complete discrete mapping of the transformation space is possible, permitting enumeration of all mechanically realizable models sharing the original model's input–output behavior. When the number of internal masses is large, a nonlinear search of transformation space can be employed to identify mechanically realizable models. Applications include scale model vibration testing of complicated structures and the design of electromechanical filters. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1953227]

PACS number(s): 43.40.At, 43.40.Sk [JGM] Pages: 762–773

I. INTRODUCTION

The mechanical realization problem is the conversion of a passive input–output dynamic model to a form that is recognizable as an interconnected system of mechanical components. Applications of mechanical realization arise in those situations for which it is desirable to fabricate a mechanical system possessing specified input–output behavior.

An important example is the scale model testing of complicated structures in naval and aircraft design. While the major structural elements can be easily scaled and fabricated, scale models of other components, such as electronic equipment and machinery, are not easily manufactured. In these cases, the most efficient solution can be to model the input–output behavior of the equipment where it attaches to the major structure and to build a simple structure which is dynamically equivalent. Similarly, in the design of electromechanical filters,[1] the desired input–output behavior is specified and its mechanical realization is sought.

The mechanical realization problem starts with the specification of a dynamic model describing input–output behavior. In the case of scale modeling, this model may be obtained through finite-element analysis or estimation from experimental data. In filter design, the model will depend on the purpose of the filter. While both time-domain and frequency-domain model descriptions are possible, this paper examines the realization problem for time-domain models specified in the second-order form

$$\begin{aligned} M\ddot{q} + C\dot{q} + Kq &= Fu, \\ y &= H_d q + H_v \dot{q} + H_a \ddot{q}. \end{aligned} \tag{1}$$

The $n \times 1$ vector q is the set of displacement coordinates, the $m \times 1$ vector u is the input vector, which is often an external force vector, and the $p \times 1$ vector y is the output vector. The mass matrix is $M = M^T > 0$, the damping matrix is $C = C^T \geqslant 0$, and the stiffness matrix is $K = K^T \geqslant 0$. F is the $n \times m$ input influence matrix, which is determined by the location of the input forces or torques. H_d, H_v, and H_a are the output influence matrices of displacement, velocity, and acceleration, respectively. In many circumstances, only accelerations need be considered as outputs and so $H_d = H_v = 0$, while $H_a \neq 0$. This is the case considered in this paper.

The mechanical realization problem for undamped or proportionally damped systems in the form of (1) has been widely studied. For these systems, the mass matrix can be reduced to diagonal form, while the damping and stiffness matrices can be converted to either tridiagonal or border diagonal form. The former consists of a realization in which the masses are connected in series while, in the latter, they are connected in parallel. For example, a serial model can be obtained by Falk's algorithm using a congruent transformation computed from the given mass and stiffness matrices.[2,3] Parallel realizations can be obtained using the normal mode theory of O'Hara and Cunniff.[4] Their results were generalized to a mechanical system undergoing three-dimensional vibration by Pierce.[5]

The existence of structure-preserving transformations which result in diagonal mass, damping, and stiffness matrices has been demonstrated for most real second-order systems.[6,7] While this form is amenable to numerical computation of input–output response by superposition, it is not

[a]Currently at Servo Dynamics Corporation, 21541 Nordhoff Street, Chatsworth, CA 91311. Electronic mail: wychen@alum.bu.edu
[b]Electronic mail: pierre@bu.edu

appropriate for mechanical realization which requires any superposition of responses to be performed mechanically.

A related body of work addresses inverse eigenvalue and inverse vibration problems.[8–11] The former is concerned with constructing a matrix with specified eigenvalues, and so applies to the realization of mass normalized systems. The inverse vibration problem involves the reconstruction of mass and stiffness matrices from prescribed frequency response data, such as resonance frequencies. This approach can be extended to include proportional damping.

The question of whether or not an arbitrary $\{M,C,K,F,H\}$ corresponding to a passive system can be transformed to mechanically realizable form has not been addressed in the literature. It remains an open question, although one might anticipate that a result similar to the positive realness requirement of electrical network synthesis[12] also holds for mechanical systems. Furthermore, a recipe for transforming a system to mechanically realizable form is unknown. As a result, the realizability problem must be solved numerically using optimization algorithms.

The contribution of this paper is to characterize the set of transformations by which a class of models with viscous, but nonproportional damping can be converted to mechanically realizable form. The approach taken is to parametrize the set of transformations relating all input–output equivalent models which could result in a mechanically realizable form. Using this parametrization, mechanical realizations can be found by mapping or selectively searching the set of transformations. They can also guide future efforts seeking closed-form solutions. These topics and examples are presented in the following sections.

II. STRUCTURE OF MECHANICALLY REALIZABLE SECOND-ORDER MODELS

Motivated by the application of scale modeling equipment and machinery, this paper considers a specific subset of mechanically realizable systems consisting only of interconnected mass, stiffness, and damping elements. Other types of elements, such as transmissions, are precluded. It is also assumed that there are no isolated masses in the system and that the system is statically stable, i.e., each mass is connected to the rest of the realization by at least one spring. Furthermore, the models are constrained to include a rigid-body mode, i.e., they cannot employ skyhook connections comprised of springs and dashpots attached to a fixed ground.

For the intended applications, model simplicity drives the choice of mechanical elements, while ease of implementation precludes the use of skyhook attachments. The results presented here can be adapted to permit additional model elements or to eliminate the rigid-body mode. Both of these cases are less restrictive than the one considered since, for the former, the solution space is enlarged and, for the latter, the number of constraints is reduced.

Finally, only realizations corresponding to diagonal mass matrices are considered here. Block-diagonal mass matrices involving, e.g., coupling between linear and rotational coordinates, may arise in practical applications, but are beyond the scope of this paper.

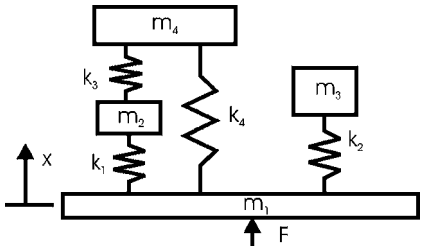


FIG. 1. Simple mechanical model.

A. Realizable stiffness and damping matrices

In addition to enforcing diagonality of the mass matrix, the conditions above also constrain the form of the stiffness and damping matrices. The simple mechanical system of Fig. 1 is used to illustrate these properties, which are well known. Since these requirements are the same for both types of matrices, a realizable stiffness matrix is used to demonstrate them. The mass and stiffness matrices are expressed as follows:

$$M=\begin{bmatrix} m_1 & & & \\ & m_2 & & \\ & & m_3 & \\ & & & m_4 \end{bmatrix}, \tag{2}$$

$$K=\begin{bmatrix} k_1+k_2+k_4 & -k_1 & -k_2 & -k_4 \\ -k_1 & k_1+k_3 & 0 & -k_3 \\ -k_2 & 0 & k_2 & 0 \\ -k_4 & -k_3 & 0 & k_3+k_4 \end{bmatrix}.$$

The stiffness matrix can be decomposed into the following form:[13]

$$K=C_K K_D C_K^T, \tag{3}$$

where the connectivity matrix C_K and non-negative diagonal matrix K_D are given by

$$C_K=\begin{bmatrix} 1 & 1 & 1 & 0 & 1 \\ 1 & -1 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & -1 \end{bmatrix}, \tag{4}$$

$$K_D=\begin{bmatrix} 0 & & & & \\ & k_1 & & & \\ & & k_2 & & \\ & & & k_3 & \\ & & & & k_4 \end{bmatrix}.$$

The connectivity matrix C_K encodes the interconnection of masses by springs. The first column of C_K is chosen arbitrarily to represent the rigid-body mode of the system in Fig. 1, corresponding to the zero element of K_D. The other col-

umns of C_K indicate connections between pairs of masses. For example, the third column represents the connection between m_1 and m_3 by stiffness k_2 in K_D. The opposite signs on the nonzero elements of these columns $(+1,-1)$ together with $K_{D_{11}}=0$ ensures that K will have a nullspace vector $[1\ 1\ \cdots\ 1]^T$ corresponding to a rigid-body mode. For a mechanical system with n masses and n_k springs, C_K is an $n \times (n_k+1)$ matrix and K_D is an $(n_k+1)\times(n_k+1)$ diagonal matrix.

Similarly, a mechanically realizable damping matrix C can be decomposed as $C=C_C C_D C_C^T$, where C_C is a connectivity matrix and C_D is a diagonal matrix with non-negative diagonal elements. It should be noted that, while C_C is not necessarily equal to C_K, both will share the nullspace vector $[1\ 1\ \cdots\ 1]^T$, ensuring the prohibition against skyhook springs and dashpots.

It can be summarized that the mechanically realizable mass, damping, and stiffness matrices must satisfy the following realization conditions:

$$\begin{aligned} &M=\mathrm{diag}([m_1\ m_2\ \cdots\ m_n]), \quad m_i>0,\\ &C=C^T, \quad C_{ii}\geqslant 0, \quad C_{ij}\leqslant 0, \quad C[1\ 1\ \cdots\ 1]^T=0,\\ &K=K^T, \quad K_{ii}>0, \quad K_{ij}\leqslant 0, \quad K[1\ 1\ \cdots\ 1]^T=0, \end{aligned} \tag{5}$$

where $i,j=1,2,\ldots,n$ and $i\neq j$.

B. Realizable input and output influence matrices

The input and output influence matrices can be categorized in terms of both the number of inputs and outputs as well as their relative locations. In the case of single-input, single-output (SISO) systems, the influence matrices are vectors, while for multi-input, multi-output (MIMO) systems, they are matrices. If the inputs and outputs are collocated then the system can be further classified as a driving-point realization, while those systems with noncollocated inputs and outputs are termed transfer realizations.

Without loss of generality, it is assumed that the desired input and output influence vectors or matrices are given by

(1) SISO driving-point accelerance

$$F=H^T=e_1. \tag{6}$$

(2) SISO transfer accelerance.

$$\begin{aligned} &F_f=e_1,\\ &H_f^T=e_2. \end{aligned} \tag{7}$$

(3) MIMO driving-point accelerance

$$F_f=H_f^T=[e_1\ e_2\ \cdots\ e_m]. \tag{8}$$

(4) MIMO transfer accelerance

$$\begin{aligned} &F_f=[e_1\ e_2\ \cdots\ e_m],\\ &H_f^T=[e_{m+1}\ e_{m+2}\ \cdots\ e_{m+p}]. \end{aligned} \tag{9}$$

(5) MIMO driving-point and transfer accelerance

$$\begin{aligned} &F_f=[e_1\ e_2\ \cdots\ e_m],\\ &H_f^T=[e_1\ e_2\ \cdots\ e_r\ e_{m+1}\ e_{m+2}\ \cdots\ e_{m+(p-r)}]. \end{aligned} \tag{10}$$

Here, e_i is an element of the standard basis for $\Re^n$, which has a 1 at the ith component and zeros elsewhere. The excitation forces are applied at a set of m coordinates and the accelerations are measured at a set of p coordinates. Both sets share r common coordinates.

III. TRANSFORMATIONS RELATING REALIZATIONS

Following the form of (1), an initial second-order model describing the accelerance of a mechanical system is given by

$$\begin{aligned} &M_0\ddot{x}+C_0\dot{x}+K_0x=F_0u,\\ &y=H_0\ddot{x}. \end{aligned} \tag{11}$$

The goal of this paper is to convert this initial model to one possessing the same input–output dynamic behavior, but which also satisfies the mechanical realizability conditions defined in the previous section. Congruent coordinate transformations can be seen to maintain the input–output behavior of the model while also preserving the symmetry of the mass, damping, and stiffness matrices. Consider the coordinate transformation

$$x=Tq, \tag{12}$$

where T is a nonsingular matrix. A congruence transformation converts the initial model (11) to the following form:

$$\begin{aligned} &M_f\ddot{q}+C_f\dot{q}+K_fq=F_fu,\\ &y=H_f\ddot{q}. \end{aligned} \tag{13}$$

Here, M_f, C_f, K_f, F_f, and H_f are, respectively, the final mass, damping, and stiffness matrices, and the input and output influence matrices. They are defined as

$$\begin{aligned} &M_f=T^TM_0T,\\ &C_f=T^TC_0T,\\ &K_f=T^TK_0T,\\ &F_f=T^TF_0,\\ &H_f=H_0T. \end{aligned} \tag{14}$$

Thus, the set of invertible matrices $T\in\Re^{n\times n}$ describes the family of all second-order models satisfying input–output equivalence with (11) while preserving mass, damping, and stiffness matrix symmetry. Only a subset of matrices T may result in a mechanically realizable model in which the final mass, damping, and stiffness matrices satisfy (5) and the final input and output influence matrices satisfy one of (6)–(10).

While necessary and sufficient conditions for an initial model to be transformable to mechanically realizable form are not available, the following is a necessary condition for an initial model to possess a rigid-body mode:

$$C_0 v_0 = K_0 v_0 = 0. \tag{15}$$

This equation states that the initial damping and stiffness matrices must share the same nullspace vector, v_0. This follows from $C[1\ 1\ \cdots\ 1]^T=0$ and $K[1\ 1\ \cdots\ 1]^T=0$ in (5), and the fact that congruence transformations preserve the signature of a matrix.

A. Decomposition of the transformation

The coordinate transformation in (12) can be decomposed into a product of three components as follows:

$$T = M_0^{-1/2} R M_f^{1/2}. \tag{16}$$

The first component, the inverse square root of the initial mass matrix, is used to mass normalize the initial second-order model (11). The second component R is an orthogonal matrix, which preserves mass normalization. To obtain mechanically realizable form, it must perform two tasks. First, it should convert the input and output influence matrices to one of the desired forms (6)–(10). Second, from (5), it must ensure that all off-diagonal components of the damping and stiffness matrices are nonpositive. Since the congruent transformation preserves definiteness of a symmetric real matrix, the diagonal elements of the damping and stiffness matrices are always non-negative.[14]

The last component of the transformation is the square root of the final mass matrix M_f. As will be shown, if an orthogonal matrix can be found such that the realizability conditions mentioned above are satisfied, the final mass matrix can be computed explicitly.

Given the decomposition of the transformation in (16), obtaining realizable form reduces to solving for an appropriate orthogonal matrix R. Orthogonal matrices are comprised of rotations, with determinant +1, and reflections, with determinant −1. In addition, permutation matrices constitute a subset of both rotation and reflection matrices. Used in a congruence transformation, permutation matrices simply reorder the coordinates.

A basis for orthogonal matrices can be constructed from the rotation matrices plus a single arbitrary reflection. Choosing this reflection as a permutation matrix reduces the basis, without loss of generality, to the rotation matrices. The $n \times n$ rotation matrices constitute the special orthogonal group, $SO(n)$. In the remainder of the paper, rotation matrices will be used as a basis for R.

B. Parametrization of the orthogonal transformation

The component R of the transformation (16) must perform two tasks, aligning the input and output influence matrices as well as ensuring that the off-diagonal elements of the mass-normalized stiffness and damping matrices are nonpositive. These tasks can be performed sequentially by writing R as the product of two rotation matrices

$$R = R_i R_o, \tag{17}$$

where the component R_i aligns the influence matrices. R_o ensures nonpositive off-diagonal elements of the stiffness and damping matrices while preserving the form of the influence matrices obtained with R_i.

1. Aligning input and output influence matrices

For the first task, denote the coordinate transformation as

$$x = M_0^{-1/2} R_i \tilde{z}. \tag{18}$$

Substituting (18) into the initial model (11) and premultiplying by $R_i^T M_0^{-1/2}$ yields the following model:

$$\begin{aligned} &\ddot{\tilde{z}} + C_{\tilde{z}} \dot{\tilde{z}} + K_{\tilde{z}} \tilde{z} = F_{\tilde{z}} u, \\ &y = H_{\tilde{z}} \ddot{\tilde{z}}, \end{aligned} \tag{19}$$

in which the matrices are defined by

$$\begin{aligned} C_{\tilde{z}} &= R_i^T M_0^{-1/2} C_0 M_0^{-1/2} R_i, \\ K_{\tilde{z}} &= R_i^T M_0^{-1/2} K_0 M_0^{-1/2} R_i, \\ F_{\tilde{z}} &= R_i^T F_z, \\ H_{\tilde{z}} &= H_z R_i, \end{aligned} \tag{20}$$

where $F_z = M_0^{-1/2} F_0$ and $H_z = H_0 M_0^{-1/2}$.

R_i can be obtained by QR factorization of the mass-normalized input and output influence matrices, F_z and H_z. In this QR factorization, a matrix is decomposed into a product of an orthogonal matrix and an upper triangular matrix. A property of this method is that a matrix whose column vectors are perpendicular to each other can be factored as a product of an orthogonal matrix and a diagonal matrix.

In the most general case, the input and output influence matrices in the final realizable model (13) must have the form given by (10). With consideration of (16), it can be proved that the columns of F_z are mutually orthogonal. Thus, after the rotation R_i, the input and output influence matrices $F_{\tilde{z}}$ and $H_{\tilde{z}}$ in (19) should satisfy the following relationships:

$$\begin{aligned} F_{\tilde{z}} &= R_i^T F_z = [\|f_{z_1}\| e_1\ \|f_{z_2}\| e_2\ \cdots\ \|f_{z_m}\| e_m], \\ H_{\tilde{z}} &= H_0 M_0^{-1/2} R_i = \begin{bmatrix} \|f_{z_1}\| e_1^T \\ \|f_{z_2}\| e_2^T \\ \vdots \\ \|f_{z_r}\| e_r^T \\ \|h_{z_{(r+1)}}\| e_{m+1}^T \\ \|h_{z_{(r+2)}}\| e_{m+2}^T \\ \cdots \\ \|h_{z_p}\| e_{m+(p-r)}^T \end{bmatrix}. \end{aligned} \tag{21}$$

To fulfill these requirements, the component R_i can be decomposed as a product of two rotations

$$R_i = R_{F_z} R_{H_z}. \tag{22}$$

In (22), the first component R_{F_z} satisfies

$$R_{F_z}^T F_z = [\|f_{z_1}\| e_1 \ \|f_{z_2}\| e_2 \ \cdots \ \|f_{z_m}\| e_m]. \tag{23}$$

Suppose the QR factorization of F_z is given by

$$Q_{F_z}[\|f_{z_1}\| e_1 \ \|f_{z_2}\| e_2 \ \cdots \ \|f_{z_m}\| e_m] = F_z, \tag{24}$$

where Q_{F_z} is an orthogonal matrix. The first component R_{F_z} then can be chosen as

$$R_{F_z} = Q_{F_z}. \tag{25}$$

From (23), the first m column vectors of R_{F_z} (or Q_{F_z}) should be equal to $f_{z_i}/\|f_{z_i}\|$ $(i=1,2,\ldots,m)$, respectively. According to (10), $H_z R_{F_z}$ should have the following form:

$$H_z R_{F_z} = \begin{bmatrix} \|f_{z_1}\| e_1^T \\ \|f_{z_2}\| e_2^T \\ \vdots \\ \|f_{z_r}\| e_r^T \\ \bar{H}_z \end{bmatrix}, \tag{26}$$

where $\bar{H}_z = [0_{(p-r)\times m} \ \tilde{H}_z]$ and $\tilde{H}_z$ is a $(p-r)\times(n-m)$ matrix.

The second component R_{H_z} of the transformation R_i needs to preserve e_j's $(j=1,2,\ldots,m)$ and should convert (26) to the following form:

$$(H_z R_{F_z}) R_{H_z} = \begin{bmatrix} \|f_{z_1}\| e_1^T \\ \|f_{z_2}\| e_2^T \\ \vdots \\ \|f_{z_r}\| e_r^T \\ \|h_{z_{(r+1)}}\| e_{m+1}^T \\ \|h_{z_{(r+2)}}\| e_{m+2}^T \\ \cdots \\ \|h_{z_p}\| e_{m+(p-r)}^T \end{bmatrix}. \tag{27}$$

Suppose the QR factorization of $\tilde{H}_z^T$ is given by

$$Q_{H_z}[\|h_{z_{(r+1)}}\| \tilde{e}_1 \ \|h_{z_{(r+2)}}\| \tilde{e}_2 \ \cdots \ \|h_{z_p}\| \tilde{e}_{(p-r)}] = \tilde{H}_z^T, \tag{28}$$

where Q_{H_z} is an orthogonal matrix and $\tilde{e}_i$ is an element of the standard basis for $\Re^{n-m}$, which has a 1 at its i'th component and zeros elsewhere. Equation (28) can be rewritten as

$$\begin{bmatrix} \|h_{z_{(r+1)}}\| \tilde{e}_1^T \\ \|h_{z_{(r+2)}}\| \tilde{e}_2^T \\ \vdots \\ \|h_{z_p}\| \tilde{e}_{(p-r)}^T \end{bmatrix} Q_{H_z}^T = \tilde{H}_z, \tag{29}$$

or equivalently

$$\begin{bmatrix} \|h_{z_{(r+1)}}\| \tilde{e}_1^T \\ \|h_{z_{(r+2)}}\| \tilde{e}_2^T \\ \vdots \\ \|h_{z_p}\| \tilde{e}_{(p-r)}^T \end{bmatrix} = \tilde{H}_z Q_{H_z}. \tag{30}$$

Thus, the second component R_{H_z} in (22) is given by

$$R_{H_z} = \begin{bmatrix} I_{m\times m} & 0 \\ 0 & Q_{H_z} \end{bmatrix}. \tag{31}$$

In summary, from (22), (25), and (31), the component R_i of the transformation R in (17) is given by

$$R_i = R_{F_z} R_{H_z} = Q_{F_z} \begin{bmatrix} I_{m\times m} & 0 \\ 0 & Q_{H_z} \end{bmatrix}. \tag{32}$$

2. Achieving nonpositive off-diagonal damping and stiffness elements

After aligning the input and output influence matrices, a second rotation R_o is needed to convert the damping and stiffness matrices in (19) to mechanically realizable form in which all off-diagonal elements are nonpositive. An explicit solution for R_o is not available; however, its form and the number of its free parameters can be derived as follows. Denote the coordinate transformation

$$\tilde{z} = R_o w. \tag{33}$$

Substituting this transformation into (19) and premultiplying by R_o^T yields the following second-order model:

$$\begin{aligned} &\ddot{w} + C_w \dot{w} + K_w w = F_w u, \\ &y = H_w \ddot{w}, \end{aligned} \tag{34}$$

in which

$$\begin{aligned} C_w &= R_o^T C_{\tilde{z}} R_o, \\ K_w &= R_o^T K_{\tilde{z}} R_o, \\ F_w &= R_o^T F_{\tilde{z}}, \\ H_w &= H_{\tilde{z}} R_o. \end{aligned} \tag{35}$$

The input and output influence matrices in (19) are already in the desired form, given by (21), only with a lack of scaling, and the transformation R_o should preserve this form. To do so, it can be expressed as

$$R_o = \begin{bmatrix} I_{n_i \times n_i} & 0 \\ 0 & \tilde{R}_o \end{bmatrix}, \tag{36}$$

with $I_{n_i \times n_i}$ as the $n_i \times n_i$ identity matrix and the rotation matrix $\tilde{R}_o \in SO(n-n_i)$. The value n_i is the number of masses at which input forces are applied and/or accelerations are measured

$$n_i = m + p - r. \tag{37}$$

It follows that $n-n_i$ is the number of internal masses of the system, i.e., those masses to which an input is not applied nor at which an output is measured.

The free parameters of R_o are those of $\tilde{R}_o \in SO(n-n_i)$, which number $(n-n_i)(n-n_i-1)/2$. Given R_o, an explicit solution exists for the final mass matrix; this is also the number of free parameters of the transformation space defined by (16). This number, quadratic in the number of internal masses in the model, represents the dimension of the space which must be mapped or searched for mechanically realizable models.

C. Solving for the final mass matrix

An explicit solution for the final mass matrix, M_f, can be derived from the realization conditions of (5) requiring the damping and stiffness matrices in (13) to satisfy

$$\begin{aligned} K_f[1\ 1\ \cdots\ 1]^T &= 0, \\ C_f[1\ 1\ \cdots\ 1]^T &= 0. \end{aligned} \tag{38}$$

Since the model (13) is related to the model (34) by the congruent transformation $M_f^{1/2}$, $C_f = M_f^{1/2} C_w M_f^{1/2}$ and $K_f = M_f^{1/2} K_w M_f^{1/2}$. Substituting these expressions into (38) reduces to

$$\begin{aligned} C_w \sqrt{m_f} &= 0, \\ K_w \sqrt{m_f} &= 0, \end{aligned} \tag{39}$$

where $\sqrt{m_f}$ is a vector of the square roots of the final masses, i.e., $\sqrt{m_f} = [\sqrt{m_{f_1}}\ \sqrt{m_{f_2}}\ \cdots\ \sqrt{m_{f_n}}]^T$.

The vector $\sqrt{m_f}$ is a scaled version of the shared nullspace vector of C_w and K_w. The final masses are obtained by scaling the nullspace vector according to the following theorem, presented for the most general case of input and output influence matrices (MIMO drivepoint and transfer accelerance) given by (10).

Theorem 1. *The input masses, to which excitation forces are applied, and the output masses, at which accelerations are measured, are given by*

$$\begin{aligned} m_{f_i} &= \frac{1}{\|f_{z_i}\|^2}, \quad i = 1,2,\ldots,m, \\ m_{f_{(m+j)}} &= \frac{1}{\|h_{z_{(r+j)}}\|^2}, \quad j = 1,2,\ldots,p-r \end{aligned} \tag{40}$$

where f_{z_i} $(i=1,2,\ldots,m)$ is the ith column vector of $M_0^{-1/2} F_0$ and h_{z_j} $(j=1,2,\ldots,p)$ is the jth row vector of $H_0 M_0^{-1/2}$.

Proof. According to (10), (14), and (16)

$$F_f = T^T F_0 = M_f^{1/2} R^T M_0^{-1/2} F_0 = M_f^{1/2} R^T F_z. \tag{41}$$

This is equivalent to

$$M_f^{-1/2}[e_1\ e_2\ \cdots\ e_m] = R^T[f_{z_1}\ f_{z_2}\ \cdots\ f_{z_m}], \tag{42}$$

which simplifies to

$$\left[\frac{1}{\sqrt{m_{f_1}}} e_1\ \frac{1}{\sqrt{m_{f_2}}} e_2\ \cdots\ \frac{1}{\sqrt{m_{f_m}}} e_m\right] = [R^T f_{z_1}\ R^T f_{z_2}\ \cdots\ R^T f_{z_m}] \tag{43}$$

Since rotation matrices preserve vector length, equating the magnitude of columns yields

$$m_{f_i} = \frac{1}{\|f_{z_i}\|^2}, \quad i = 1,2,\ldots,m \tag{44}$$

The second equation of (40) follows similarly. □

This theorem states that each member of the set of mechanically realizable models which are input–output equivalent to the initial model (11) has the same input and output masses. The following theorem proves the invariance of total system mass for all mechanical realizations. Physically, this result follows from input–output equivalence at zero frequency to preserve the rigid-body mode.

Theorem 2. *All mechanical realizations which are input–output equivalent to the original second-order model (11) possess the same total mass*.

Proof. Recall from (15) the necessary condition for realizability that C_0 and K_0 share the same nullspace vector, and let this vector v_0 be of unit length

$$C_0 v_0 = K_0 v_0 = 0. \tag{45}$$

By (5) and (14), $T^T K_0 T[1\ 1\ \cdots\ 1]^T = 0$ and, since T^T is invertible, $T[1\ 1\ \cdots\ 1]^T = \alpha v_0$, where α is a scalar constant.

Substituting $T = M_0^{-1/2} R M_f^{1/2}$ yields

$$\sqrt{m_f} = \alpha R^T M_0^{1/2} v_0, \tag{46}$$

and an expression for total mass is given by

$$\sqrt{m_f}^T \sqrt{m_f} = \sum_{i=1}^{i=n} m_{f_i} = \alpha^2 v_0^T M_0 v_0. \tag{47}$$

To compute α, it is known from Theorem 1 that $\sqrt{m_{f_1}} = 1/\|f_{z_1}\|$. Since a system must have at least one input and output, the first mass can always be used in this expression, and combining it with (46) yields

$$\alpha = \frac{1/\|f_{z_1}\|}{(R^T M_0^{1/2} v_0)_1}, \tag{48}$$

where the subscript 1 in the denominator indicates the first element of the column vector.

Recall (17), in which R_i aligns the inputs and outputs and R_o has the structure of (36). Since all mechanical realizations share the same R_i and, furthermore, since R_o cannot change the first element of $M_0^{1/2} v_0$, the constant α is independent of R_o and thus the same for all mechanical realizations. □

Taken together, the preceding theorems indicate that only the internal masses of the system can differ between realizations and that the total internal mass is constant.

IV. OBTAINING REALIZABLE MODELS

In the preceding section, it has been demonstrated that congruent coordinate transformations T for converting a model to mechanically realizable form can be expressed as

$$T = M_0^{-1/2} R M_f^{1/2} = M_0^{-1/2} (R_i R_o) M_f^{1/2}, \tag{49}$$

The first component $M_0^{-1/2}$ is known from the initial second-order model (11) and the second component R_i can be obtained via QR factorization of the input and output influence matrices in (19). The last component $M_f^{1/2}$ can be obtained from (39) and Theorem 1.

An explicit solution for the remaining component R_o is only available for SISO systems with no damping or proportional damping. In all other cases, a solution for R_o must be sought through mapping or selectively searching the special orthogonal group $SO(n-n_i)$, in which $n-n_i$ is the number of internal masses. $SO(n-n_i)$ can be described by n_p parameters, where

$$n_p = (n-n_i)(n-n_i-1)/2, \tag{50}$$

and each parameter corresponds to a two-dimensional rotation angle.

These parameters must be selected to satisfy the $n(n-1)$ inequality constraints that the off-diagonal components of the stiffness and damping matrices be nonpositive. Since the number of constraints exceeds the number of parameters in (50), it is not clear that a solution will exist in the general case. If the initial model is derived from either experiment or FEM, however, it is likely that these constraints will be dependent and mechanically realizable solutions will exist.

To search for a solution, R_o in (36) can be written as the product of n_p two-dimensional rotation matrices involving the last $n-n_i$ coordinates

$$R_o = \prod_{\substack{i=n_i+1 \\ j=n_i+2}}^{\substack{i=n-1 \\ j=n}} R_{ij}, \tag{51}$$

in which R_{ij} is the two-dimensional rotation matrix in the ith and jth coordinates. Two-dimensional rotations, also known as Givens rotations, have been widely used to convert symmetric matrices to tridiagonal matrices in solving symmetric matrix eigenvalue problems.[15]

The elements of these rotation matrices correspond to those of an identity matrix except for the following four:

$$\begin{aligned} &R_{ij}(i,i) = \cos(\theta_{ij}), \quad R_{ij}(i,j) = -\sin(\theta_{ij}), \\ &R_{ij}(j,i) = \sin(\theta_{ij}), \quad R_{ij}(j,j) = \cos(\theta_{ij}). \end{aligned} \tag{52}$$

To obtain a bijection (one-to-one and onto map) between θ_{ij} and $SO(n-n_i)$ where $n-n_i \geqslant 3$, it is not necessary for all θ_{ij} to vary as $0 \leqslant \theta_{ij} < 2\pi$. For example, in $SO(3)$, all rotation matrices can be generated from the product $R_{12}(\theta_{12}) R_{13}(\theta_{13}) R_{23}(\theta_{23})$, in which $0 \leqslant \theta_{12} < 2\pi$, $0 \leqslant \theta_{13} < \pi$, and $0 \leqslant \theta_{23} < 2\pi$. Allowing $0 \leqslant \theta_{13} < 2\pi$ would result in a two-to-one map.

Even when the angle ranges are appropriately restricted so that the map from θ_{ij} to $SO(n-n_i)$ is a bijection, the map from $SO(n-n_i)$ to the system model (34) is many-to-one. This is due to the equivalence class of models corresponding to permutations of the internal masses.

Recall that a permutation matrix congruence transformation swaps pairs of rows and columns of the matrix to which it is applied. This operation results purely in a renumbering of the internal mass coordinates of the model. There are $(n-n_i)!$ possible permutations of the internal masses. Half of these correspond to swapping an even number of pairs of rows and columns and so result from rotation permutation matrices. As a result, the mapping from θ_{ij} to the system model (34) will be $(n-n_i)!/2$-to-one.

The following sections describe how realizable models can be found by mapping or selectively searching the space of transformations. For each example, the initial second-order model was generated by applying a random congruent transformation to a realizable second-order model. The first two examples involve mapping the entire transformation space, and so the initial models are recovered as members of the sets of realizable models.

A. Mapping transformation space

When the number of internal masses is small, the number of free parameters of the transformation space, given by n_p in (50), is also small. In this case, a complete mapping of transformation space is feasible and the results can be easily visualized. Two examples with three internal masses are presented here.

1. Example 1: Four-mass driving-point accelerance

An initial second-order model satisfying (15) is given by

$$\begin{bmatrix} 1.5182 & 3.8080 & -1.0679 & 1.8792 \\ 3.8080 & 10.4989 & -2.9426 & 5.2055 \\ -1.0679 & -2.9426 & 0.9188 & -1.4534 \\ 1.8792 & 5.2055 & -1.4534 & 2.7745 \end{bmatrix} \ddot{x} + \begin{bmatrix} 0.1836 & 0.3089 & 0.0234 & 0.0841 \\ 0.3089 & 1.1514 & -0.0858 & 0.4348 \\ 0.0234 & -0.0858 & 0.0728 & 0.0187 \\ 0.0841 & 0.4348 & 0.0187 & 0.2717 \end{bmatrix} \dot{x} + \begin{bmatrix} 6818 & 16814 & -2200 & 2656 \\ 16814 & 63328 & -11552 & 14063 \\ -2200 & -11552 & 3827 & -910 \\ 2656 & 14063 & -910 & 6627 \end{bmatrix} x = \begin{bmatrix} -0.4326 \\ -1.1465 \\ 0.3273 \\ -0.5883 \end{bmatrix} u \tag{53}$$

$$y = [-0.4326 \ -1.1465 \ 0.3273 \ -0.5883]\ddot{x}.$$

After mass normalization, the initial model becomes

$$\ddot{w} + \begin{bmatrix} 0.7316 & -0.1283 & 0.3312 & -0.1787 \\ -0.1283 & 0.4246 & 0.2858 & -0.1413 \\ 0.3312 & 0.2858 & 1.0947 & 0.2789 \\ -0.1787 & -0.1413 & 0.2789 & 0.5134 \end{bmatrix} \dot{w} + \begin{bmatrix} 17792 & -6182 & 5980 & -2606 \\ -6182 & 27522 & 6075 & -20547 \\ 5980 & 6075 & 21560 & 05432 \\ -2606 & -20547 & 5432 & 28812 \end{bmatrix} w = \begin{bmatrix} -0.1401 \\ -0.2679 \\ 0.0935 \\ -0.1728 \end{bmatrix} u \tag{54}$$

$$y = [-0.1401 \ -0.2679 \ 0.0935 \ -0.1728]\ddot{w}.$$

By QR factorization of the vector $[-0.1401 \ -0.2679 \ 0.0935 \ -0.1728]^T$, the orthogonal matrix R_i is obtained as

$$R_i = \begin{bmatrix} -0.3886 & -0.7430 & 0.2592 & -0.4793 \\ -0.7430 & 0.6025 & 0.1387 & -0.2565 \\ 0.2592 & 0.1387 & 0.9516 & 0.0895 \\ -0.4793 & -0.2565 & 0.0895 & 0.8345 \end{bmatrix}. \tag{55}$$

After aligning the input and output influence vectors, the mass-normalized second-order model (54) is

$$\ddot{\tilde{z}} + \begin{bmatrix} 0.0490 & -0.0882 & -0.2034 & 0.0236 \\ -0.0882 & 0.6629 & -0.0759 & 0.1068 \\ -0.2034 & -0.0759 & 1.3181 & 0.0127 \\ 0.0236 & 0.1068 & 0.0127 & 0.7344 \end{bmatrix} \dot{\tilde{z}} + \begin{bmatrix} 0.1877 & -0.5873 & -0.4114 & 0.4359 \\ -0.5873 & 3.2409 & -0.0973 & -1.3774 \\ -0.4114 & -0.0973 & 2.5883 & -0.0377 \\ 0.4359 & -1.3774 & -0.0377 & 3.5517 \end{bmatrix} \tilde{z} = \begin{bmatrix} 0.3605 \\ 0.0000 \\ 0.0000 \\ 0.0000 \end{bmatrix} u \tag{56}$$

$$y = [0.3605 \ 0.0000 \ 0.0000 \ 0.0000]\ddot{\tilde{z}}.$$

The number of internal masses is $n-n_i=3$, and so the transformation space can be parametrized by $n_p=3$ two-dimensional rotations. The terms R_{ij} in (51) which preserve the driving-point input and output influence vectors are given by

$$R_{23} = \begin{bmatrix} 1 & & & \\ & \cos\theta_{23} & -\sin\theta_{23} & \\ & \sin\theta_{23} & \cos\theta_{23} & \\ & & & 1 \end{bmatrix}, \quad R_{24} = \begin{bmatrix} 1 & & & \\ & \cos\theta_{24} & & -\sin\theta_{24} \\ & & 1 & \\ & \sin\theta_{24} & & \cos\theta_{24} \end{bmatrix}, \quad R_{34} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \cos\theta_{34} & -\sin\theta_{34} \\ & & \sin\theta_{34} & \cos\theta_{34} \end{bmatrix}. \tag{57}$$

The matrix R_o is the product

$$R_o(\theta_{23}, \theta_{24}, \theta_{34}) = R_{23}R_{24}R_{34}. \tag{58}$$

A complete map relating rotation angles to realizable models is obtained by discretizing the rotation angles as shown in Fig. 2. The shaded regions correspond to mechanically realizable models. Note that the plotted angle ranges are $0 \leqslant \theta_{23} < 2\pi$, $\pi/2 \leqslant \theta_{24} < 3\pi/2$, $0 \leqslant \theta_{34} < 2\pi$ in order to obtain one-to-one coverage of $SO(3)$. Since there are three internal masses, there are six possible permutations of these masses, three of which are obtained through rotations. Consequently, the map from θ_{ij} to system models is three-to-one, resulting in three equivalent regions of mechanically realizable models. Removing equivalent realizations reduces the set to that shown in Fig. 3.

In the figures, the shading indicates the number of connecting elements (springs and dampers) in the realizations.

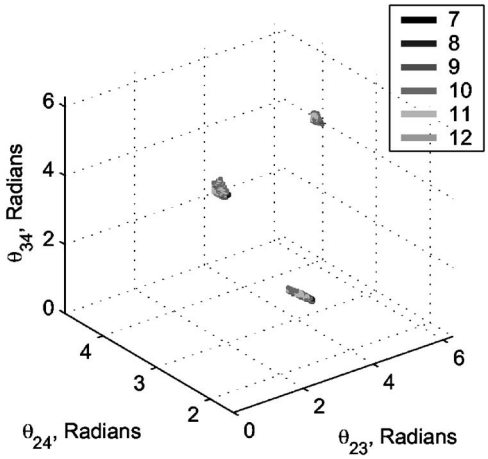


FIG. 2. Realizable regions for Example 1. Three regions correspond to cyclic permutations of the three internal masses. Legend indicates total number of connecting elements (springs and dampers) in realizations.

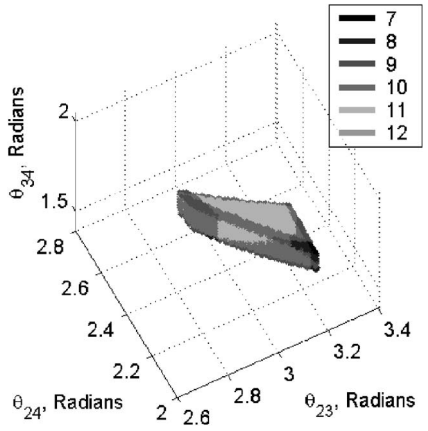


FIG. 3. Single region containing all distinct mechanical realizations for Example 1. Legend indicates total number of connecting elements (springs and dampers) in realizations.

The realizations with the fewest connecting elements are located on the boundary between realizable and unrealizable regions where off-diagonal elements of the damping and stiffness matrices change their signs.

Two realizable models from this set are presented here which differ in the number of springs and dampers.

Realization 1. Selection of rotation angles $\theta_{23}=2.9496$, $\theta_{24}=2.3387$, and $\theta_{34}=1.7104$ radians yields the mechanically realizable model

$$\begin{bmatrix} 7.6941 & & & \\ & 0.0164 & & \\ & & 0.1906 & \\ & & & 0.2268 \end{bmatrix}\ddot{q} + \begin{bmatrix} 0.3770 & -0.0058 & -0.0974 & -0.2738 \\ -0.0058 & 0.0136 & -0.0012 & -0.0066 \\ -0.0974 & -0.0012 & 0.1124 & -0.0139 \\ -0.2738 & -0.0066 & -0.0139 & 0.2943 \end{bmatrix}\dot{q} + \begin{bmatrix} 14444 & -115 & -8796 & -5533 \\ -115 & 339 & -124 & -100 \\ -8796 & -124 & 9078 & -158 \\ -5533 & -100 & -158 & 5792 \end{bmatrix} q = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u \tag{59}$$

$$y=[1\ 0\ 0\ 0]\ddot{q}.$$

Since there are no zero elements in the damping and stiffness matrices, this realization includes a dashpot and spring between each pair of masses.

Realization 2. Rotation angles $\theta_{23}=3.2484$, $\theta_{24}=2.1776$, and $\theta_{34}=1.5769$ radians produce a mechanically realizable model with the fewest springs (four) and dashpots (three), as shown in Fig. 4.

$$\begin{bmatrix} 7.6941 & & & \\ & 0.0220 & & \\ & & 0.2502 & \\ & & & 0.1616 \end{bmatrix}\ddot{x} + \begin{bmatrix} 0.3770 & -0.0177 & -0.1450 & -0.2143 \\ -0.0177 & 0.0178 & 0 & -0.0001 \\ -0.1450 & 0 & 0.1450 & 0 \\ -0.2143 & -0.0001 & 0 & 0.2144 \end{bmatrix}\dot{x} + \begin{bmatrix} 14\,444 & 0 & -10\,633 & -3\,810 \\ 0 & 474 & -474 & 0 \\ -10\,633 & -474 & 11\,528 & -421 \\ -3\,810 & 0 & -421 & 4\,232 \end{bmatrix} x = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u \tag{60}$$

$$y=[1\ 0\ 0\ 0]\ddot{x}$$

Although the initial model (53) and two realizations (59) and (60) have different mass, damping, and stiffness matrices, they possess the same driving-point accelerance.

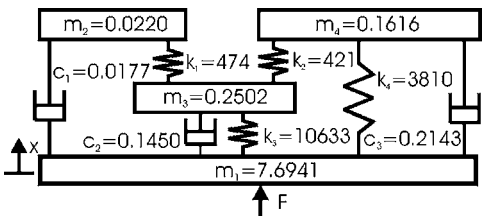


FIG. 4. Realization with the fewest springs and dashpots.

2. Example 2: Five-mass transfer accelerance

Consider the SISO second-order model satisfying (15) and given by

$$\begin{bmatrix} 36.0632 & 0.2743 & 7.0077 & 37.1531 & -14.0227 \\ 0.2743 & 16.5796 & -6.7431 & 5.5551 & 2.1107 \\ 7.0077 & -6.7431 & 17.1533 & 16.6964 & -13.3775 \\ 37.1531 & 5.5551 & 16.6964 & 61.9214 & -25.8255 \\ -14.0227 & 2.1107 & -13.3775 & -25.8255 & 17.8591 \end{bmatrix} \ddot{x} + \begin{bmatrix} 97.2937 & -57.5426 & 91.8234 & 72.4757 & -38.5575 \\ -57.5426 & 62.9875 & -56.3684 & -51.0698 & 29.9845 \\ 91.8234 & -56.3684 & 92.8620 & 64.1156 & -37.4205 \\ 72.4757 & -51.0698 & 64.1156 & 60.7210 & -27.3858 \\ -38.5575 & 29.9845 & -37.4205 & -27.3858 & 34.8577 \end{bmatrix} \dot{x}$$
$$+ \begin{bmatrix} 931.9559 & -207.5005 & 567.4285 & 768.4827 & -618.2834 \\ -207.5005 & 383.5154 & -262.6956 & -186.1123 & 172.2111 \\ 567.4285 & -262.6956 & 514.7016 & 383.3350 & -393.4070 \\ 768.4827 & -186.1123 & 383.3350 & 709.9495 & -487.2931 \\ -618.2834 & 172.2111 & -393.4070 & -487.2931 & 544.2288 \end{bmatrix} x = [1.9574 \ -0.2111 \ 0.5512 \ 0.4620 \ -1.2316]^T u$$
$$y = [0.5045 \ 1.1902 \ -1.0998 \ -0.3210 \ 1.0556]\ddot{x}. \tag{61}$$

The input and output influence vectors differ, indicating that the system represents a transfer accelerance. Following (7), a realizable model is sought in which the force excitation is applied at the first coordinate and the acceleration is measured at the second. The solution for R_i is not included here for the sake of brevity.

With $n=5$ masses and $n_i=2$ input and output masses, the transformation space is parametrized by $n_p=3$ two-dimensional rotations. The matrix R_o is given by

$$R_o(\theta_{34}, \theta_{35}, \theta_{45}) = R_{34} R_{35} R_{45}. \tag{62}$$

Figure 5 depicts the map between rotation angles and system models. As in Example 1, there are three equivalent regions of mechanically realizable models corresponding to rotational permutations of the three internal masses. The plotted angle ranges in this figure are $0 \leqslant \theta_{34} < 2\pi$, $0 \leqslant \theta_{35} < \pi$, $0 \leqslant \theta_{45} < 2\pi$.

As an example realization, rotation angles $\theta_{34}=3.0386$, $\theta_{35}=3.0048$, and $\theta_{45}=1.1158$ radians produce the following mechanically realizable model with a fully populated stiffness matrix and a damping matrix possessing a single zero dashpot:

$$\begin{bmatrix} 2.0000 & 0 & 0 & 0 & 0 \\ 0 & 5.0000 & 0 & 0 & 0 \\ 0 & 0 & 3.4340 & 0 & 0 \\ 0 & 0 & 0 & 2.3326 & 0 \\ 0 & 0 & 0 & 0 & 10.2334 \end{bmatrix} \ddot{q} + \begin{bmatrix} 8.5000 & -1.0000 & -6.8047 & -0.0596 & -0.6357 \\ -1.0000 & 13.3000 & -11.0845 & -0.6876 & -0.5280 \\ -6.8047 & -11.0845 & 31.9340 & -14.0385 & -0.0063 \\ -0.0596 & -0.6876 & -14.0385 & 14.7857 & -0.0000 \\ -0.6357 & -0.5280 & -0.0063 & -0.0000 & 1.1700 \end{bmatrix} \dot{q}$$
$$+ \begin{bmatrix} 130.0000 & -20.0000 & -49.9987 & -8.2885 & -51.7128 \\ -20.0000 & 92.0000 & -47.6153 & -20.9996 & -3.3852 \\ -49.9987 & -47.6153 & 139.9326 & -29.8486 & -12.4700 \\ -8.2885 & -20.9996 & -29.8486 & 59.6030 & -0.4664 \\ -51.7128 & -3.3852 & -12.4700 & -0.4664 & 68.0344 \end{bmatrix} q = [1 \ 0 \ 0 \ 0 \ 0]^T u$$
$$y = [0 \ 1 \ 0 \ 0 \ 0]\ddot{q}. \tag{63}$$

B. Searching transformation space

As the number of internal masses $n-n_i$ in the model grows, it becomes impractical to map the entire transformation space, as was done in the preceding examples. Instead, the transformation space can be selectively searched using a nonlinear optimization method.

Recalling that the role of R_o in obtaining a realizable model is to ensure that the off-diagonal elements of the stiffness and damping matrices are nonpositive, a cost function for optimization can be chosen as

$$J(\theta) = w_1 S_K + w_2 S_C, \tag{64}$$

where θ is the vector of rotation angles, S_K is the summation of all positive off-diagonal elements in the stiffness matrix K, and S_C is the summation of all positive off-diagonal ele-

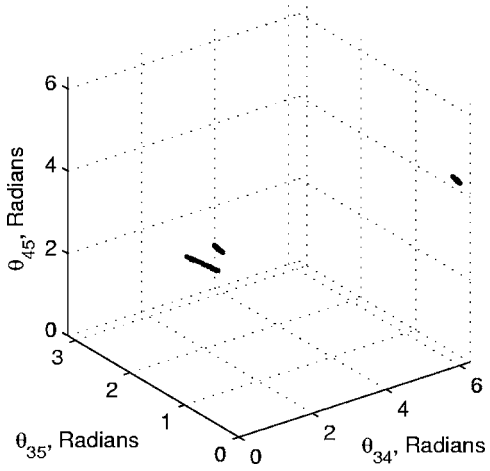


FIG. 5. Realizable regions for Example 2. Three regions correspond to cyclic permutations of the three internal masses.

ments in the damping matrix C. In order to balance the contributions from the stiffness and damping matrices, the weighting factors w_1 and w_2 are defined as

$$w_1 = 1,$$

$$w_2 = \frac{\text{trace}(K)}{\text{trace}(C)}.$$

Since a congruent orthogonal transformation does not change the trace of a matrix, the weighting factor w_2 is constant.

A wide variety of optimization techniques can be employed to search for an angle vector θ resulting in a realizable model. Since the problem is nonlinear, local minima of the cost function can exist. If such a minima is detected during optimization, a perturbation of random direction and magnitude can be applied to escape its domain of attraction.

1. Example 3: Ten-mass SISO driving-point accelerance

To illustrate the use of an optimization method in solving the mechanical realization problem, the Nelder–Mead method[16] was applied to the following ten-mass driving-point system using the cost function defined in (64). For brevity, the model is presented after mass normalization and alignment of input and output influence vectors. The damping matrix does not correspond to proportional damping.

$$C_{\tilde{z}} = 10^{-2} \times \begin{bmatrix}
2.71 & 0.59 & 0.91 & 0.31 & 0.28 & 0.04 & -0.53 & 0.38 & -0.86 & 0.10 \\
0.59 & 2.21 & -0.45 & 0.18 & -0.56 & -0.00 & 0.14 & 0.33 & 0.74 & -0.21 \\
0.91 & -0.45 & 3.24 & 0.10 & -0.72 & 0.12 & 0.23 & 0.54 & 0.83 & 0.14 \\
0.31 & 0.18 & 0.10 & 2.48 & -0.02 & -0.54 & -0.12 & -0.87 & 0.86 & -0.13 \\
0.28 & -0.56 & -0.72 & -0.02 & 2.51 & -0.53 & -0.00 & 0.32 & -0.16 & -0.01 \\
0.04 & -0.00 & 0.12 & -0.54 & -0.53 & 3.21 & -0.33 & 0.07 & 0.01 & -0.07 \\
-0.53 & 0.14 & 0.23 & -0.12 & -0.00 & -0.33 & 2.91 & -0.31 & -0.23 & 0.18 \\
0.38 & 0.33 & 0.54 & -0.87 & 0.32 & 0.07 & -0.31 & 2.74 & 0.51 & -0.68 \\
-0.86 & 0.74 & 0.83 & 0.86 & -0.16 & 0.01 & -0.23 & 0.51 & 2.60 & 0.59 \\
0.10 & -0.21 & 0.14 & -0.13 & -0.01 & -0.07 & 0.18 & -0.68 & 0.59 & 3.54
\end{bmatrix}$$

$$K_{\tilde{z}} = \begin{bmatrix}
16\,685 & 5\,279 & 2\,255 & 964 & 2\,052 & 1\,663 & -3627 & 432 & -6\,797 & 93 \\
5\,279 & 11\,712 & -615 & -742 & 952 & 664 & -1\,747 & -2\,261 & 3\,368 & -2\,184 \\
2\,255 & -615 & 13\,728 & -1\,508 & -3\,500 & 709 & 2\,709 & -669 & 3\,985 & 1\,844 \\
964 & -742 & -1\,508 & 15\,194 & -821 & -457 & 360 & -362 & 4\,447 & 249 \\
2\,052 & 952 & -3\,500 & -821 & 17\,704 & -591 & -855 & -525 & 2\,505 & -2\,485 \\
1\,663 & 664 & 709 & -457 & -591 & 19\,298 & -199 & 1\,753 & 1\,535 & 34 \\
-3\,627 & -1\,747 & 2\,709 & 360 & -855 & -199 & 13\,734 & -1\,139 & -153 & 2\,045 \\
432 & -2\,261 & -669 & -362 & -525 & 1\,753 & -1\,139 & 16\,008 & 808 & -1\,654 \\
-6\,797 & 3\,368 & 3\,985 & 4\,447 & 2\,505 & 1\,535 & -153 & 808 & 15\,770 & 3\,236 \\
93 & -2\,184 & 1\,844 & 249 & -2\,485 & 34 & 2\,045 & -1\,654 & 3\,236 & 19\,318
\end{bmatrix} \tag{65}$$

$$F_{\tilde{z}} = H_{\tilde{z}}^T = [1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]^T.$$

With nine internal masses, there are 36 two-dimensional rotation parameters describing the space of transformations. The optimization method was initiated with the rotation angles set to random numbers in the range 0 to 2π. The search terminates when a realizable model is found. The result of one trial appears below. This trial involved six iterations in which at most 2500 evaluations of the cost function were permitted for each iteration.

$$M_f = \mathrm{diag}([1.000\ 0.590\ 0.221\ 0.302\ 0.461\ 0.310\ 0.378\ 0.123\ 0.378\ 0.521])$$

$$C_f = 10^{-2} \times \begin{bmatrix} 2.71 & -0.49 & -0.05 & -0.18 & -0.60 & -0.24 & -0.29 & -0.07 & -0.45 & -0.36 \\ -0.49 & 1.84 & -0.00 & -0.07 & -0.56 & -0.02 & -0.29 & -0.02 & -0.07 & -0.34 \\ -0.05 & -0.00 & 0.77 & -0.13 & -0.07 & -0.12 & -0.05 & -0.00 & -0.18 & -0.16 \\ -0.18 & -0.07 & -0.13 & 0.88 & -0.06 & -0.16 & -0.06 & -0.00 & -0.12 & -0.09 \\ -0.60 & -0.56 & -0.07 & -0.06 & 1.45 & -0.00 & -0.00 & -0.01 & -0.04 & -0.10 \\ -0.24 & -0.02 & -0.12 & -0.16 & -0.00 & 0.88 & -0.21 & -0.00 & -0.10 & -0.03 \\ -0.29 & -0.29 & -0.05 & -0.06 & -0.00 & -0.21 & 1.27 & -0.00 & -0.22 & -0.15 \\ -0.07 & -0.02 & -0.00 & -0.00 & -0.01 & -0.00 & -0.00 & 0.10 & -0.00 & -0.00 \\ -0.45 & -0.07 & -0.18 & -0.12 & -0.04 & -0.10 & -0.22 & -0.00 & 1.24 & -0.06 \\ -0.36 & -0.34 & -0.16 & -0.09 & -0.10 & -0.03 & -0.15 & -0.00 & -0.06 & 1.29 \end{bmatrix}$$

$$K_f = \begin{bmatrix} 16\,685 & -4\,698 & -265 & -1705 & -1809 & -2239 & -1058 & -74 & -1632 & -3205 \\ -4\,698 & 10\,694 & -7 & -809 & -1432 & -702 & -2497 & -315 & -41 & -193 \\ -265 & -7 & 3609 & -754 & -129 & -278 & -135 & -422 & -718 & -900 \\ -1\,705 & -809 & -754 & 4805 & -295 & -759 & -4 & -68 & -407 & -4 \\ -1\,809 & -1\,432 & -129 & -295 & 6300 & -1016 & -522 & -1 & -1088 & -8 \\ -2\,239 & -702 & -278 & -759 & -1016 & 6269 & -876 & -359 & -8 & -33 \\ -1\,058 & -2\,497 & -135 & -4 & -522 & -876 & 6655 & -96 & -1182 & -285 \\ -74 & -315 & -422 & -68 & -1 & -359 & -96 & 1967 & -178 & -453 \\ -1\,632 & -41 & -718 & -407 & -1088 & -8 & -1182 & -178 & 5453 & -199 \\ -3\,205 & -193 & -900 & -4 & -8 & -33 & -285 & -453 & -199 & 5280 \end{bmatrix} \tag{66}$$

$$F_f = H_f^T = [1\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0\ 0]^T.$$

V. CONCLUSIONS

A congruent coordinate transformation has been developed to convert second-order models to a form interpretable as a passive mechanical system. The space of transformations has been parametrized by a set of two-dimensional rotation matrices. Complete mapping of transformation space is possible for systems with small numbers of internal masses; however, an optimization method is needed to search for realizable models in higher dimensional cases. While not demonstrated here, optimization methods allow the flexibility to search for mechanically realizable models that satisfy additional criteria. For example, the cost function could be adapted to find models with the fewest dashpots or springs. These results can be applied to the vibration testing of complicated structures and to the design of electromechanical filters.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research under Grants N00014-01-1-0155 and N00014-03-1-0881.

[1]R. A. Johnson, *Mechanical Filters in Electronics* (Wiley, New York, 1983).
[2]S. Falk, "Die Abbildung eines allgemeine schwingungssystems auf eine einfache Schwingerkette," Ingenieur-Archiv, Vol. **23**, pp. 314–328 (1955).
[3]S. L. Chen and M. Géradin, "An exact model reduction procedure for mechanical systems," Comput. Methods Appl. Mech. Eng. **143**, 69–78 (1997).
[4]G. J. O'Hara and P. F. Cunniff, "Elements of Normal Mode Theory," Naval Research Laboratory Report, 1963.
[5]A. D. Pierce, "Resonant-frequency-distribution of internal mass inferred from mechanical impedance matrices, with application to fuzzy structure theory," Trans. ASME, J. Vib. Acoust. **119**, 325–333 (1997).
[6]S. D. Garvey, M. I. Friswell, and U. Prells, "Coordinate transformations for second order systems. I. General transformations," J. Sound Vib. **258**(5), 885–909 (2002).
[7]S. D. Garvey, M. I. Friswell, and U. Prells, "Coordinate transformations for second order systems. II. Elementary structure-preserving transformations," J. Sound Vib. **258**(5), 911–930 (2002).
[8]G. M. L. Gladwell, "Inverse problems in vibration," Appl. Mech. Rev. **49**, S25–S34 (1996).
[9]M. T. Chu, "Inverse eigenvalue problems," SIAM Rev. **40**, 1–39 (1998).
[10]Y. M. Ram and S. Elhay, "An Inverse Eigenvalue Problem for the Symmetric Positive Linear Pencil with Applications to Vibrating Systems," in *Proceedings of the Computational Techniques and Applications: CTAC 97*, edited by B. J. Noye, M. D. Teubner, and A. W. Gill (World Scientific, Singapore, 1998), pp. 561–568.
[11]O. Rojo, R. Soto, and J. Egana, "A note on the construction of a positive oscillatory matrix with a prescribed spectrum," Comput. Math. Appl. **41**, 353–361 (2001).
[12]H. Baher, *Synthesis of Electrical Networks* (Wiley, New York, 1984).
[13]N. M. M. Maia and J. M. M. Silva, *Theoretical and Experimental Modal Analysis* (Research Studies, Taunton, England, 1997).
[14]R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, UK, 1990).
[15]A. Jennings, *Matrix Computation* (Wiley, New York, 1992).
[16]W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C: the Art of Scientific Computing*, 2nd ed. (Cambridge University Press, New York, 1992).

The transmission loss of curved laminates and sandwich composite panels

Sebastian Ghinet[a] and Noureddine Atalla
Department of Mechanical Engineering, Université de Sherbrooke, 2500 Boulevard Université, Sherbrooke, QC, J1K 2R1, Canada

Haisam Osman
The Boeing Company, 5301 Bolsa Avenue, Huntington Beach, California 92647

(Received 30 August 2004; revised 7 April 2005; accepted 26 April 2005)

The paper describes a model to calculate the transmission loss of both curved laminate and sandwich composite panels within statistical energy analysis (SEA) context. The vibro-acoustic problem is developed following a wave approach based on a discrete lamina description. Each lamina is considered to consist of membrane, bending, transverse shearing and rotational inertia behaviors. Moreover, the orthotropic ply angle of each lamina is considered. Using such a discrete lamina description, the dispersion behaviors of the panel are correctly represented. Using the dispersion curves, the radiation efficiency, the modal density, as well as, the nonresonant and the resonant transmission are computed. Moreover, expression for the evaluation of the ring frequency and the critical frequencies of such panels is given. The described model is shown to handle accurately, both laminate and sandwich composite shells. Additionally, a transmission loss test is presented to confirm the validity of the presented model. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1932212]

PACS number(s): 43.40.−r, 43.40.At [JGM] Pages: 774–790

I. INTRODUCTION

Laminate and sandwich composite panels and cylinders have increasingly found application in modern aerospace and aeronautical structures. The composite materials used for these constructions are generally lighter and stronger than the most advanced aluminum alloys, which are prevalent in aerospace constructions. These qualities lead however to increased radiation efficiency and lower nonresonant transmission loss which unfortunately leads, in some instances, to higher interior noise levels. In consequence, there is a need for robust and fast numerical tools to efficiently estimate and optimize the vibroacoustic behaviors of large laminate and sandwich panels.

A large amount of work has been devoted to the modeling of laminate and sandwich panels. The published models handle two classical types of constructions: laminates and sandwich. The first class refers to lay-ups of composite plies of similar (or close) physical properties. In general the laminate is symmetric and each layer is modeled using a thin plate theory[1–3] or thin shell theory.[4–9] The laminate is represented by an equivalent set of variables. It will be referred to in the paper by a smeared thin laminate. Other authors consider generalizations in which each layer is modeled using thick plate theory and the laminate is planar[10,11] or curved.[12] Still, an equivalent set of variables is used for the whole laminate. In this latter case, the proper estimation of the equivalent shear correction coefficient is of paramount importance. The majority of these formulations[1–11] assume the laminate symmetric. More advanced approaches accounting for through-thickness deformations are presented in Refs. 13, 14, and 40. Reference 14 uses spectral finite element to handle flat laminates; that is 1D finite element are used to model through-thickness deformation and exponential functions (propagating wave shape) are used to handle the in-plane displacements. Reference 40 uses a similar method and considers laminated cylindrical shells and pipes. It uses axisymmetry to handle the circumferential direction, exponential functions to model the axial displacement and solid finite element for the radial displacement. The model is applied to the calculation of the dispersion curves of the first modes of laminated cylindrical shells and pipes.

For sandwich constructions a trilayer arrangement is classically used; the core is generally soft and thick compared to the skins. Each layer can be of a composite construction. The classical models are based on two main assumptions: (i) the skins are thin and work in bending; (ii) the core is relatively thick and handles shearing effects only. Following these assumptions and assuming the sandwich symmetric, two classes of models are used. The first smears the elastic constants of the panel through the thickness.[3,15–21,37,39] The second uses a discrete layer representation.[22] This model uses a complete and mathematically coherent discrete layer theory for sandwich-type panels. The theory is developed for a symmetric singly curved sandwich made up of a bottom skin laminate, a shearing core, and a top skin laminate. Finally, note that in the context of laminated composite cylinders modeling, two models were presented and compared by Ghinet *et al.*;[12] symmetrical laminate composite and discrete thick laminate composite. The latter was shown to encompass the first and to handle accu-

[a]Author to whom correspondence should be addressed. Tel: 1 819 821 8000 # 3773; Fax: 1 819 821 7163; e-mail: Sebastian.Ghinet@USherbrooke.ca

rately, as a particular case, sandwich composite shells. In these two models, membrane, bending, transverse shearing as well as rotational inertia effects and orthotropic ply angle of the layers were considered.

The thickness and structural complexity of the sandwich panel could influence in varying degrees, the transverse shear, the rotational inertia as well as the separate bending motion of the skins at the mid-to-high frequencies. At these frequencies it was observed[12] that the two classes of modeling approaches, smearing and discrete layer, may lead to different behaviors depending on the nature of the construction (sandwich, thick laminate, etc.). It is important that a model be devised that can handle both configurations. This paper presents such a model. It is based on a discrete layer approach. Each layer is allowed to exhibit bending, shearing, and membrane behaviors. The construction can be none symmetric, laminates or sandwich. It will be referred to in this text by a general laminate.

The numerical estimation of the transmission loss of elastic structures can be accurately performed using finite elements (FE) and boundary elements (BE) methods.[23,24] These approaches require extensive computing resources and are inappropriate for large geometrical-scale structures and high frequencies calculation, when the vibration wavelength becomes much smaller than the structural dimensions. In contrast, statistical energy analysis is commonly characterized as much simpler to apply than FE/BE methods but it is known to fail at low frequencies where the number of modal resonance frequencies in the analysis band is low.[25] There are two methods for applying the SEA methodology. The first is referred to by the "modal approach" and is based on modeling each subsystem as a superposition of the resonant responses of the set of uncoupled modes of the system. The second method is based on a wave approach and models each subsystem as a superposition of waves travelling around the subsystem. It consists of deriving and solving a dispersion system of equations between wave numbers and frequency for the subsystem of interest assuming simple geometrical configurations (plates, shells, etc.). At interfaces, the coupling loss factor is usually expressed in terms of the semi-infinite system wave impedances. At low frequencies where size effects are important, it is essential to include corrections. One approach, for the specific problem of airborne transmission loss, is based on the application of spatial windowing.[26,27] An asymptotical approach is presented in Ref. 28.

One particular problem of interest in the paper is the estimation, using SEA, of the transmission loss of laminate composite and sandwich curved structures. The principal phenomena concerning the resonant and nonresonant transmission of shells as well as a comprehensive review were presented by Szechenyi.[29,30] The problem of the nonresonant transmission of a cylinder was solved by a simple geometrical argument.[30] The contribution of the stiffness-controlled region was neglected and the subcoincident region was assumed to be a circular sector for frequencies below the ring frequency. The problem of the transmission loss into finite cylinders was also discussed by Pope *et al.*[31–33] and Lesueur[35] using a modal approach. A similar approach[33,35] is successfully applied here to calculate the low frequency transmission loss of laminate and sandwich composite curved panels.

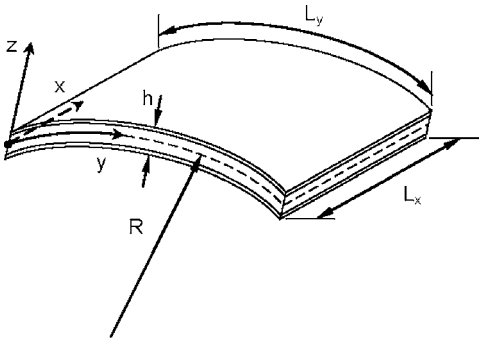


FIG. 1. The composite shell coordinates.

This paper describes the SEA modeling of the transmission loss through finite laminate and sandwich composite singly curved panels. Both laminate composite and sandwich composite are modeled using a discrete thick laminate composite theory. The studied transmission problem has three primary resonant systems: two reverberant rooms separated by the composite curved panel. The dispersion curves of the structure are derived and solved for the modal density and the radiation efficiency. Several models to compute the radiation efficiency were tested.[26,27,34] Identical results were obtained and the model of Leppington[34] was selected due to its accuracy and fast convergence. These parameters allow for the calculation of the radiation loss factor and also the resonant contribution of the transmission loss. The standard flat panel theory[35] is used to compute the nonresonant transmission but it is adapted here to the particular vibration behaviors of the curved panels (see Sec. VI). In particular, a subcoincident modes selection method is used to compute the nonresonant transmission contribution. Moreover, the classical wave approach nonresonant contribution is corrected using the spatial windowing method presented in Ref. 27. Finally, a transmission loss experimental result of a curved sandwich composite panel is successfully compared with numerical estimations.

II. GEOMETRY AND COORDINATE SYSTEM

Figure 1 represents the global geometrical configuration of the composite shell, where R is the curvature radius and h is the total thickness. The layered construction is considered, in general, asymmetrical as represented in Fig. 2(a). The origin for the z axis is defined on a reference surface passing through the middle thickness of the shell.

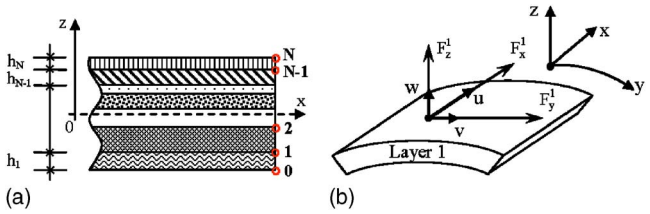


FIG. 2. The discrete laminated composite shell thickness constitution (a) and interlayer forces (b).

III. DISPERSION RELATION

The dynamic behavior of the curved panel is modeled using a discrete layer theory which allows for both thick laminate composites and sandwich shells. The displacement field of any discrete layer "*i*" of the panel is of Mindlin's type:

$$u^i(x,y,z) = u_0^i(x,y) + z\varphi_x^i(x,y),$$

$$v^i(x,y,z) = v_0^i(x,y) + z\varphi_y^i(x,y), \quad w^i(x,y,z) = w_0^i(x,y). \quad (1)$$

For any layer of the shell, Flüge's theory[36] is used to describe the strain-displacement relations. Rotational inertia, in-plane, bending as well as transverse shearing effects are accounted for in each layer. Also, orthotropic ply angle is used for any layer. The resultant stress forces and moments of any layer are defined in Appendix A [Eqs. (A1) and (A2)]. There are three interlayer forces between any two layers, as represented in Fig. 2(b). The total number of interlayer forces is $3(N-1)$, where N is the number of layers. For any layer "*i*" there are five equilibrium equations:

$$N_{x,x}^i + N_{yx,y}^i + F_x^i - F_x^{i-1} = \left[\left(m_s + \frac{I_{z2}}{R}\right)u_{,tt} + \left(\frac{I_z}{R} + I_{z2}\right)\varphi_{x,tt}\right]^i,$$

$$N_{y,y}^i + N_{xy,x}^i + \frac{Q_y^i}{R} + F_y^i - F_y^{i-1} = \left[\left(m_s + \frac{I_{z2}}{R}\right)v_{,tt} + \left(\frac{I_z}{R} + I_{z2}\right)\varphi_{y,tt}\right]^i,$$

$$Q_{x,x}^i + Q_{y,y}^i - \frac{N_y^i}{R} + F_z^i - F_z^{i-1} = \left[\left(m_s + \frac{I_{z2}}{R}\right)w_{,tt}\right]^i,$$

$$M_{x,x}^i + M_{yx,y}^i - Q_x^i + z^i F_x^i - z^{i-1}F_x^{i-1} = \left[I_z\left(\varphi_{x,tt} + \frac{u_{,tt}}{R}\right) + I_{z2}u_{,tt}\right]^i,$$

$$M_{xy,x}^i + M_{y,y}^i - Q_y^i + z^i F_y^i - z^{i-1}F_y^{i-1} = \left[I_z\left(\varphi_{y,tt} + \frac{v_{,tt}}{R}\right) + I_{z2}v_{,tt}\right]^i. \quad (2)$$

The subscript notation $\Omega_{\alpha\beta,\gamma\delta}$ in relations (2) and in the following indicates partial derivatives of $\Omega_{\alpha\beta}$ with respect to γ and δ. The external and internal surfaces of the shell are considered stress-free so that $F_x^0 = F_y^0 = F_z^0 = 0$ and $F_x^N = F_y^N = F_z^N = 0$.

The expressions of the transverse shear stress forces Q_i, the in-plane stress forces N_i, the inertial terms I_i, and the stress moments M_{ij} of each layer are presented in Eqs. (A4)–(A7). For any layer, the dynamic equilibrium equations can be rewritten, using Eqs. (2) and relations (A4)–(A6) with appropriate algebraic manipulations, as presented in Eq. (A12). The resulting dynamic equilibrium system, presented in that form, has $5N+3(N-1)$ variables regrouped in two vectors; a displacement-rotation vector $\{U\}$, and an interlayer forces vector $\{F\}$:

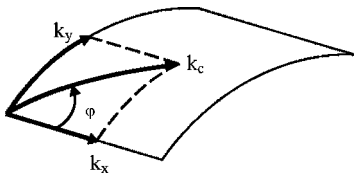


FIG. 3. The propagative wave number and the heading direction.

$$\{U\} = \{u^1; v^1; w^1; \varphi_x^1; \varphi_y^1; u^2; v^2; w^2; \varphi_x^2; \times \varphi_y^2; \cdots u^N; v^N; w^N; \varphi_x^N; \varphi_y^N\}^T,$$

$$\{F\} = \{F_x^1; F_y^1; F_z^1; F_x^2; F_y^2; F_z^2; \cdots F_x^{N-1}; F_y^{N-1}; F_z^{N-1}\}^T. \quad (3)$$

The associated $5N+3(N-1)$ equations are composed of 5 equations of dynamic equilibrium for each of the N layers plus 3 equations of interlayer continuity of displacements for each of the $N-1$ interlayer surfaces.

To solve for the dispersion relations, the system of dynamic equilibrium equations is expressed in terms of a hybrid displacement-force vector $\langle e\rangle$ defined as:

$$\langle e\rangle = \begin{Bmatrix} U \\ F \end{Bmatrix}. \quad (4)$$

Assuming a harmonic solution $\langle e\rangle = \{e\}\exp(jk_x x + jk_y y - j\omega t)$, the system is expressed in the form of a generalized polynomial complex eigenvalue problem:

$$k_c^2[A_2]\{e\} - ik_c[A_1]\{e\} - [A_0]\{e\} = 0, \quad (5)$$

where, $k_c = \sqrt{k_x^2 + k_y^2}$ is represented in Fig. 3, with $k_x = k_c \cos\varphi$ and $k_y = k_c \sin\varphi$, $i = \sqrt{-1}$ and $[A_0]$, $[A_1]$, $[A_2]$ are real square matrices (in the absence of damping) of dimension $5N+3(N-1)$ defined in (A13). Relation (5) has $2(5N+3(N-1))$ complex conjugate eigenvalues and represents the dispersion relations of the laminated composite shell. In the context of the present approach, at any heading direction the curved panel has two propagating solutions below the ring frequency. At the ring frequency a third solution becomes propagating thus, in the dispersion field context the ring frequency is mathematically perceived as a cut-off or transition frequency. Two other cut-off frequencies appear at high frequencies where two additional solutions become propagating.

IV. RING AND CRITICAL FREQUENCIES

The ring and critical frequencies of symmetrical laminate composite cylinders were presented in Ref. 12. In this section the ring frequency and the critical frequencies relations are presented in the context of the presented curved discrete laminate model. As mentioned above, the ring frequency can be considered as the first cut-off or transition frequency of the dispersion equation. Here a simple approach is used to estimate it. Considering that at the ring frequency ω_{ring}, the shell displacement is characterized by a breathing mode shape, the derivatives of the hybrid vector $\langle e\rangle$ along x and y directions are equal to zero. Consequently, Eq. (5) simplifies to

$$[A_{01}]\{e\} = \omega_{\text{ring}}^2 [A_{02}]\{e\}, \tag{6}$$

which represents a generalized eigenvalues problem with $[A_{01}]$ and $[A_{02}]$, real square matrices of dimension $5N+3(N-1)$ defined by the expression $[A_0]=[A_{01}]-\omega^2[A_{02}]$ knowing that $[A_0]$ is defined by expression (A13).

By analogy with plates, the critical frequency limits of the laminate composite curved shell are given by the particular solution of the dispersion equation (5) at coincidence; that is when the structural wave number "k_c" matches the acoustic wave number "k_0"

$$k_c^2[A_2]\{e\} - ik_c[A_1]\{e\} - [A_0]\{e\} = 0, \quad k_0 = k_c = \omega/c_0. \tag{7}$$

In the case of a discrete laminated composite flat panel ($R \to \infty$), the critical frequencies are computed numerically from (7) using $[A_0]=[A_{01}]-\omega^2[A_{02}]$,

$$\omega_c^2 \left[\frac{[A_2]}{c_0^2} + [A_{02}] \right] \{e\} - i\omega_c \frac{[A_1]}{c_0}\{e\} - [A_{01}]\{e\} = 0, \tag{8}$$

which is a second order polynomial eigenvalues problem with $[A_1]$ and $[A_2]$ defined by the relations (A13). Assuming $\lambda_c = i\omega_c$, Eq. (8) can be expressed in the form,

$$\lambda_c \begin{bmatrix} \frac{[A_1]}{c_0} & \frac{[A_2]}{c_0^2} + [A_{02}] \\ [1] & [0] \end{bmatrix} \begin{Bmatrix} e \\ \lambda_c e \end{Bmatrix} = \begin{bmatrix} -[A_{01}] & [0] \\ [0] & [I] \end{bmatrix} \begin{Bmatrix} e \\ \lambda_c e \end{Bmatrix}, \tag{9}$$

to obtain a generalized eigenvalue problem, where $[I]$ is the identity matrix and $[0]$ a zero matrix of dimension $2(5N+3(N-1))$. This problem has $2(5N+3(N-1))$ complex conjugate eigenvalues. The critical frequencies correspond to solutions which satisfies the condition $\lambda_c(\varphi) = \pm i\omega_c$, purely imaginary. This heading dependency will lead to a critical frequency region given by

$$f_c(\varphi) = \mp \frac{i\lambda_c(\varphi)}{2\pi}. \tag{10}$$

It is found that the limits of the critical frequency region are defined by $f_{c1}=f_c(\varphi=0)$ and $f_{c2}=f_c(\varphi=\pi/2)$.

V. THE MODAL DENSITY AND THE RADIATION EFFICIENCY

The angular distribution of the modal density is classically expressed in terms of the ratio of the structural wave number and the group velocity[1]

$$n(\varphi,\omega) = \frac{A}{2\pi^2} \frac{k(\varphi,\omega)}{|c_g(\varphi,\omega)|} \tag{11}$$

with A the area of the panel. The modal density is obtained numerically by integrating over all headings directions,

$$n(\omega) = \int_0^{\pi} n(\varphi,\omega) d\varphi. \tag{12}$$

The structural wave number of the shell $k(\varphi,\omega)$ and the group velocity are computed numerically from the solution of the dispersion relation (5).

The radiation efficiency of the panel $\sigma(k(\varphi,\omega))$ for a given frequency and heading is computed from Leppington's analytical formulas.[34] Assuming energy equipartition amongst the resonant modes (equal modal energy), the radiation efficiency of the composite panel is given by

$$\sigma_{\text{rad}} = \frac{1}{n(\omega)} \int_0^{\pi} \sigma(k(\varphi,\omega)) n(\varphi,\omega) d\varphi. \tag{13}$$

The corresponding band averaged values of the modal density and radiation efficiency are given, respectively, by

$$n(\omega) = \frac{\int_{k_{\min}}^{k_{\max}} \int_0^{\pi} n(\omega,\varphi) k dk d\varphi}{\int_{k_{\min}}^{k_{\max}} k dk},$$

$$\sigma_{\text{rad}}(\omega) = \frac{\int_{k_{\min}}^{k_{\max}} \int_0^{\pi} \sigma(\omega,\varphi) n(\omega,\varphi) k dk d\varphi}{\int_{k_{\min}}^{k_{\max}} \int_0^{\pi} n(\omega,\varphi) k dk d\varphi}, \tag{14}$$

where, $k_{\min}$ and $k_{\max}$ are the wave number bounds of the studied frequency band. The radiation efficiency and the modal density of each of the first three solutions are computed using relation (14). Each solution is then considered a resonant SEA subsystem.

VI. NONRESONANT TRANSMISSION

In general, for a complex construction, the nonresonant transmission is heading dependent. Thus for or a given excitation frequency band (with ω_{cen} the center band frequency and ω_1, ω_2 the frequency limits of the band), and an incidence direction (θ,φ) the structural and the forced wave numbers are calculated from the dispersion relation (5) and the following conditions is checked to ensure that the forced modes are nonresonant,

$$k_0(\omega_{\text{cen}})\sin\theta < k_s(\omega_1) \quad \text{or} \quad k_0(\omega_{\text{cen}})\sin\theta > k_s(\omega_2). \tag{15}$$

This accounts for both mass and stiffened controlled nonresonant modes. Usually, stiffness controlled modes contribution is neglected and the mass-controlled non resonant transmission coefficient is given by

$$\tau_{\text{nr}}(\omega) = \frac{1}{\pi(\cos^2\theta_{\min} - \cos^2\theta_{\max})} \times \int_0^{2\pi} \int_{\theta_{\min}}^{\theta_{\max}} \tau_{\text{nr}}(\omega,\theta,\varphi) \sin\theta\cos\theta d\theta d\varphi, \tag{16}$$

where

$$\tau_{\text{nr}}(\omega,\theta,\varphi) = \frac{4Z_0^2}{|i\omega m_s + 2Z_0|^2}, \tag{17}$$

and $(\theta_{\min},\theta_{\max})$ are the limit incidence angles describing the diffuse field, $Z_0=\rho_0 c_0/\cos\theta$ is the specific acoustic impedance of the medium and m_s is the surface mass of

the panel, and φ is the heading direction limited to nonresonant modes. The allowable heading directions are obtained using the dispersion equation (5) and the first condition in Eq. (15).

In order to improve the low frequency predictions of the non-resonant transmission coefficient, a geometrical windowing correction method is also used. The correction method used here, is detailed in Ref. 27 and examples of its validation are given in Refs. 27,37 and 39. According to this correction, the relation (17) changes as follows:

$$\tau_{\mathrm{nr}}(\omega,\theta,\varphi)=\frac{4Z_0^2}{|i\omega m_s+2Z_0|^2}\sigma(\omega,\theta,\varphi)\cos\,\theta, \tag{18}$$

where, $\sigma(\omega,\theta,\varphi)$ is the "geometric" radiation efficiency of the finite baffled window defined by[27]

$$\sigma(\omega,\theta,\varphi)=\Re\left[\frac{jk_0}{A}\int_S\int_S e^{-ik_p(x\cos\,\varphi+y\sin\,\varphi)}G(x,y,x',y')\times e^{jk_p(x'\cos\,\varphi+y\sin\,\varphi)}dS(x,y)dS(x',y')\right]. \tag{19}$$

In the above relation, $G(x,y,x',y')$ is the Green's function, A is the area of the panel, and S denotes the radiating surface of the panel. Relation (19) is evaluated using a semianalytical algorithm as presented in Ref. 27. Since this correction only affect the low frequencies, it is postulated to remain applicable for singly curved panels even if its derivation is strictly limited to flat systems. The comparison with the experimental measurement presented in Sec. VIII C corroborates the validity of this correction.

Alternatively, a modal method can also be used to calculate the diffuse field incidence transmission coefficient for mass controlled modes[33,35]

$$\tau_{\mathrm{nr}}=\frac{16\pi}{A}c_0^2\sum_{\omega_{mn}<\Delta\omega}\frac{R_{mn}^2}{m_s^2\omega^4}=2\frac{32}{\pi}\frac{A}{m_s^2}\rho^2\sum_{\omega_{mn}<\Delta\omega}[j_{mn}^2]^2. \tag{20}$$

In the above equation, $R_{mn}=\sigma_{mn}\rho_0 c_0$ denotes the modal radiation resistance and $[j_{mn}^2]$ the joint acceptance. Symbol $\omega_{mn}<\Delta\omega$ indicates that the summation is limited to nonresonant modes in the band of interest. In order to apply the modal approach, the shell is assumed simply supported. The modes are related to the wave number components by, $k_x=m\pi/L_x$; $k_y=n/R$. The corresponding natural frequencies ω_{mn} are obtained from the solution of the dispersion equation (5) recast in the symbolic form

$$(k_{mn}^2[A_2]-ik_{mn}[A_1]-[A_{01}])\{e\}=\omega_{mn}^2[A_{02}]\{e\}, \tag{21}$$

with, $k_{mn}=(k_x^2+k_y^2)^{1/2}$ and the matrices $[A_1]$, $[A_2]$, $[A_{01}]$, and $[A_{02}]$, real square matrices of dimension $5N+3(N-1)$ defined Eq. (A13) with $[A_0]=[A_{01}]-\omega^2[A_{02}]$. The modal radiation resistance or equivalently the modal joint acceptance is calculated analytically assuming the classical spatial separate form of the Green's function.[31]

VII. DIFFUSE FIELD TRANSMISSION LOSS

The modal density and the radiation efficiency presented in the above sections are used within a SEA framework to compute the transmission loss of the laminated composite shell. A simple SEA acoustic transmission scheme consists of two reverberation rooms separated by the studied curved panel. One of the rooms is excited by a diffuse field and the acoustic transmission problem is assumed to encompass two transmission contributions: resonant and nonresonant transmission. As a first approximation, and for the sole calculation of the transmission loss, the first solution wave of the dispersion relation is used to represent the dynamics of the curved panel.

The acoustic rooms (cavities) are described by the systems 1 and 3 while the curved panel is identified as system 2. Using the classical SEA equations, the noise reduction of the panel is given by

$$\mathrm{NR}=10\log_{10}\left(\frac{\bar{\alpha}A_3}{A\tau}+\left[\tau_{\mathrm{nr}}+\tau_r\frac{\bar{\eta}_2}{\eta_{\mathrm{rad}}}\right]\frac{1}{\tau}\right), \tag{22}$$

where, $\bar{\alpha}A_3$ is the surface absorption of the receiving room, $\bar{\eta}_2=\eta_2+2\eta_{\mathrm{rad}}$ is the sum of the space and band averaged radiation loss factor[35] η_{rad} of the panel and its averaged structural loss factor η_2; τ_{nr} is the field incidence nonresonant transmission coefficient (mass controlled), τ_r is the diffuse field resonant transmission coefficient, $\tau=\tau_{\mathrm{nr}}+\tau_r$, and finally $\bar{\alpha}$ is the random incidence absorption coefficient of the panel seen from the receiving room. Explicitly, in terms of modal densities (n_1,n_2,n_3), damping loss factors (η_1,η_2,η_3) and coupling loss factors,[35] ($\eta_{21}=\eta_{23}=\eta_{\mathrm{rad}}$, η_{13}), Eq. (22) is equivalent to

$$\mathrm{NR}=10\log_{10}\left(\frac{\eta_{13}+\frac{n_2}{n_1}\frac{\eta_{\mathrm{rad}}^2}{\bar{\eta}_2}}{\eta_3+\frac{n_1}{n_3}\eta_{13}+\frac{n_2}{n_3}\eta_{\mathrm{rad}}}\right). \tag{23}$$

The resonant transmission coefficient[31] is calculated from the radiation efficiency of the panel and its modal density using

$$\tau_r=\frac{8\pi c_0^2 n_2}{A\omega}\frac{\eta_{\mathrm{rad}}^2}{\bar{\eta}_2}. \tag{24}$$

The SEA total transmission loss of the panel including resonant and non resonant contributions is expressed as

$$\mathrm{TL}=\mathrm{NR}+10\log_{10}\left(\frac{A}{A_3}\right), \tag{25}$$

where, NR is the noise reduction, A the area of the panel and A_3 the absorption area of the receiving room.

A richer model was also studied. The panel was assumed to be composed of three subsystems corresponding to each of the first three propagating solutions of the dispersion relation. It was assumed that each of the three subsystems couple with the acoustic cavities. As expected, the contribution of

TABLE I. Materials' properties for diffuse field transmission loss validations.

	Material #1	Material #2	Material #3	Material #4	Material #5
E_L (Pa)	7.1×10^{10}	1.25×10^{11}	0.48×10^{11}	0.1448×10^{9}	3.0×10^{7}
E_T (Pa)	7.1×10^{10}	10^{10}	0.48×10^{11}	0.1448×10^{9}	3.0×10^{7}
G_{LT} (Pa)	2.67×10^{10}	5.9×10^{9}	0.181×10^{11}	0.5×10^{8}	1.25×10^{7}
G_{LZ} (Pa)	2.67×10^{10}	3×10^{9}	0.2757×10^{10}	0.5×10^{8}	1.25×10^{7}
G_{TZ} (Pa)	2.67×10^{10}	5.9×10^{9}	0.2757×10^{10}	0.5×10^{8}	1.25×10^{7}
ν_{LT}	0.33	0.4	0.3	0.2	0.2
ρ (kg/m^3)	2700	1600	1550	110.44	48

the first wave solution was found to be dominant. The computed resonant contributions of the other solutions were found to be unimportant for this air-borne transmission case. However, it is worth recalling that these propagating solutions (resonant subsystems) as well as the evanescent components are important in structural transmission problems (e.g., plate to plate or plate to beam junction).

VIII. NUMERICAL RESULTS AND VALIDATION

In this section, results of the acoustic transmission problem applied to curved composite laminate and sandwich panels are presented. The problem is solved in a SEA context, using a wave approach. An alternate modal approach is also used to validate the wave approach. The properties of the materials used in this study are presented in Table I. The associated notations concerning the orthotropic directions (L,T) are presented in Fig. 4.

A. Dispersion curves

In this section, the presented general discrete laminate model is compared to a discrete sandwich model[22] and to a symmetrical laminate model.[12] Recall that symmetrical laminate model assumes each layer thick and uses equivalent properties including a shear correction factor for the whole structure (= smeared physical properties over the total thickness of the panel). On the other hand, the sandwich model is based on a discrete approach using the classical assumptions for a laminate sandwich (e.g., thin laminate skins, a shear bearing core). It leads to a 47 order dispersion system. On the other hand, the presented discrete general laminate model is also based on a discrete layer approach but allow all layers (skins and core) to be thick laminate (e.g., with smeared properties through each layer's thickness), orthotropic and thick. It leads to a 42 order dispersion system for the particular case of a curved sandwich panel. It allows for both symmetrical and asymmetrical laminated composite and/or sandwich-type composite panels with thin or thick laminate skins. For the particular case of sandwich composite panels, the present discrete laminate model and the sandwich model[22] lead to identical results.

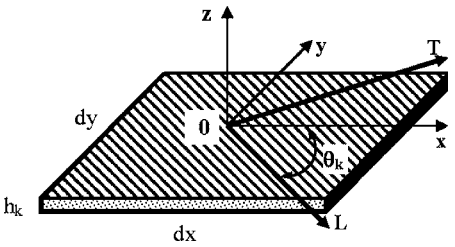


FIG. 4. Orthotropic directions of a ply.

The comparison is shown here for a singly curved sandwich composite panel. It has a 2 m radius of curvature and a projected area of 2.43 m $\times$ 2.03 m. The thickness of the skins is 1.2 mm and that of the core is 12.7 mm. It is made up of Graphite/Epoxy (Material #3) face sheets and of a rigid foam core (Material #4); the panel's orthotropic layout is (0/0/0). Figure 5 illustrates the propagating solutions of the dispersion system at three selected heading directions: $\Phi=0°$, $\Phi=45°$, $\Phi=90°$. It is observed that the presented model and the sandwich model lead to identical results. The dispersion relation, for this case, is of the 42nd order. In Fig. 5 the five propagative solutions are represented. As it can be observed in Fig. 5, below the ring frequency (f_{ring}=401.8 Hz) the dispersion relation has two propagating solutions. Note that the ring frequency is the first transition frequency of the panel. At this frequency a third solution becomes propagating. At very high frequencies two supplementary core transition frequencies are defined and at these frequencies two supplementary solutions become propagating.

The same behaviors are present for laminate composite panels. A seven layers graphite/epoxy (Material #2) laminate panel is considered as a second example. The layers have equal thickness h_i=2 mm and the panel has the orthotropic layout: 0/45/-45/90/-45/45/0. Comparisons are made here between the present discrete lamina model and a simple symmetrical laminate model.[12] The solutions of the dispersion relations in these cases are represented in Fig. 6. It is observed that the two models lead to identical results. The core transition frequencies appear in that case at frequencies above the audible range and are not represented in Fig. 6.

In order to identify the asymptotic tendencies and the solution type of the dispersion problem, a three layer isotropic thin curved panel is considered in Fig. 7. The thickness of each layer (Material #1) is 1 mm. The corresponding flat panel has the classical wave number solutions: bending, shear, and membrane. These asymptotes are also represented in Fig. 7 and are calculated using the following relations:

$$k_{\text{bending}} = \sqrt{\omega\sqrt{\frac{m_s}{D}}}; \quad k_{\text{shear}} = \omega\sqrt{\frac{m_s}{Gh}};$$

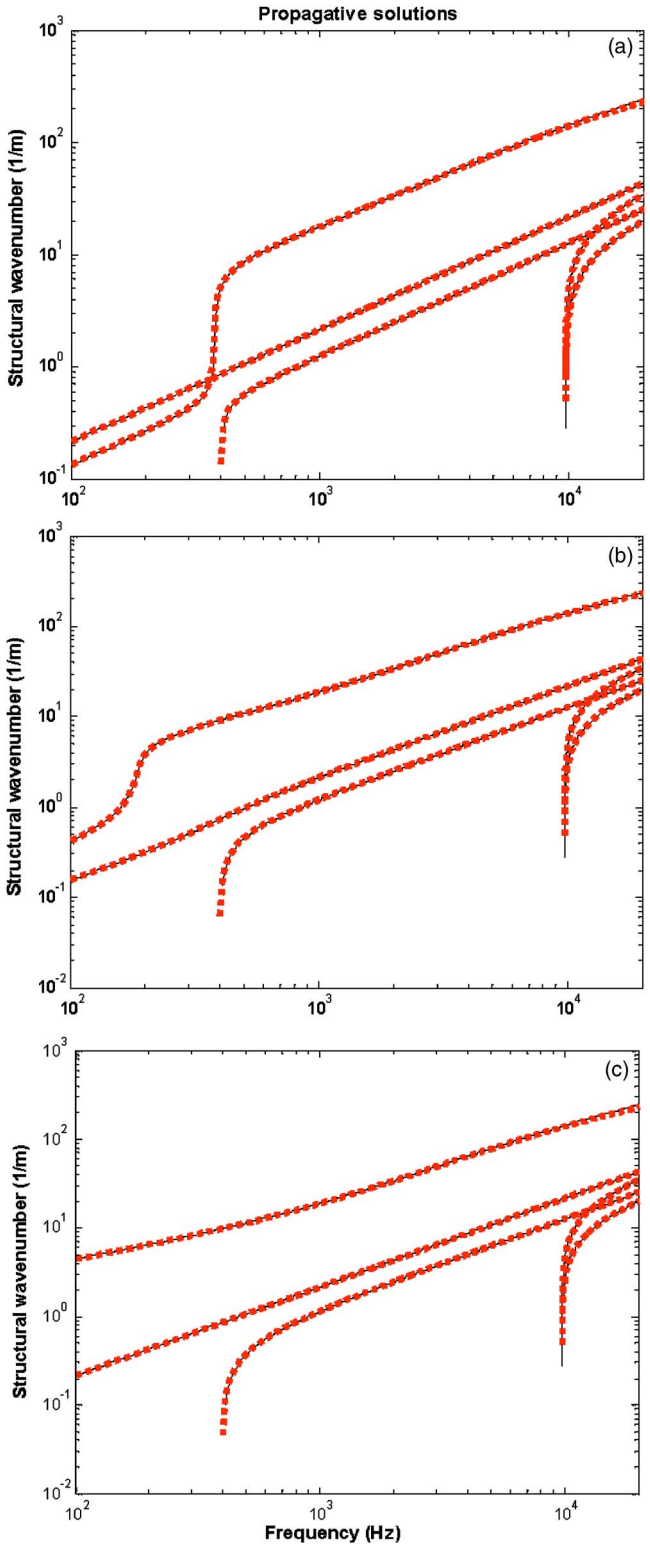


FIG. 5. Dispersion curves of a sandwich composite shell for different heading directions: (a) $\Phi=0°$; (b) $\Phi=45°$; (c) $\Phi=90°$. Modeling type: Discrete laminate composite (—); sandwich composite (···).

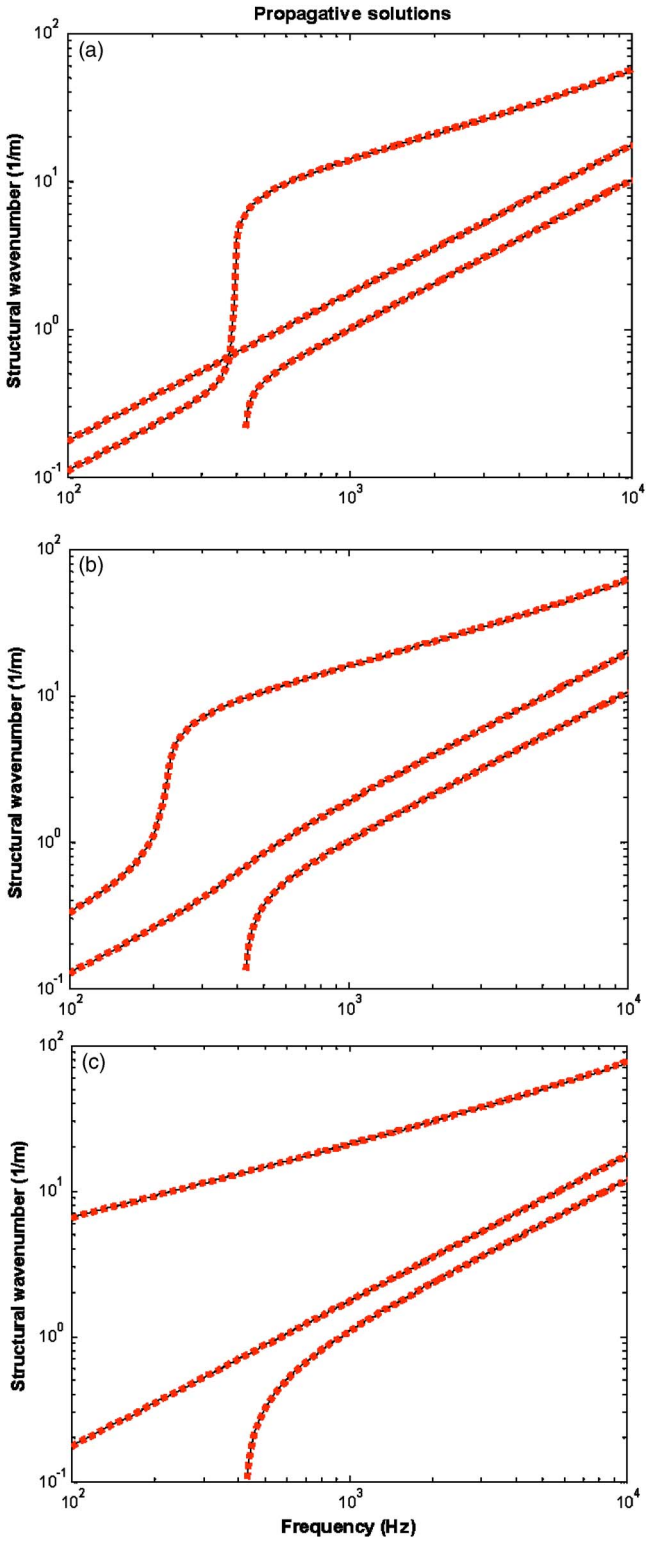


FIG. 6. Dispersion curves of a laminate composite shell for different heading directions: (a) $\Phi=0°$; (b) $\Phi=45°$; (c) $\Phi=90°$. Modeling type: Discrete laminate composite (—); symmetrical laminate composite (···).

$$k_{\text{membrane}} = \omega \sqrt{m_s \frac{(1-\nu^2)}{Eh}}, \tag{26}$$

with, m_s the surface mass, D the bending stiffness, G the core shear modulus, and E the Young's modulus.

The first propagating wave number solution of a sandwich panel has three different behaviors: pure bending of the whole panel at low frequencies; shearing of the core at mid-to-high frequencies, and pure bending of the skins at very high frequencies. This is illustrated in Fig. 8. It represents the first solution of the sandwich panel employed for the dispersion studies represented in Fig. 5 and its three dispersion

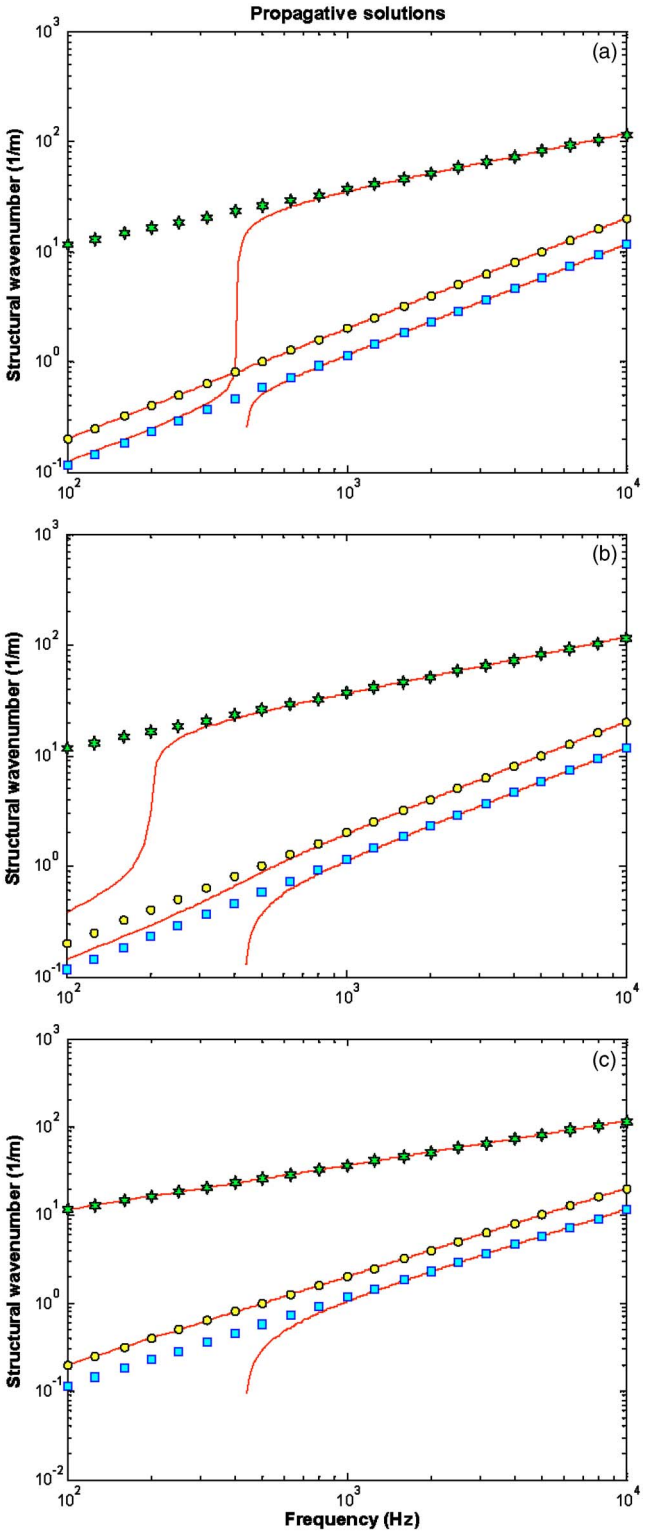


FIG. 7. Dispersion curves of a thin isotropic shell for different heading directions: (a) $\Phi=0°$; (b) $\Phi=45°$; (c) $\Phi=90°$. Analytical thin plate solutions: Bending (∗); shear (○); membrane (□).

asymptotes. For a heading direction set to 90°, the dispersion relation has equivalent flat panel behaviors; that is, the same first propagating solution is obtained for a heading of 90°, using $R \to \infty$.

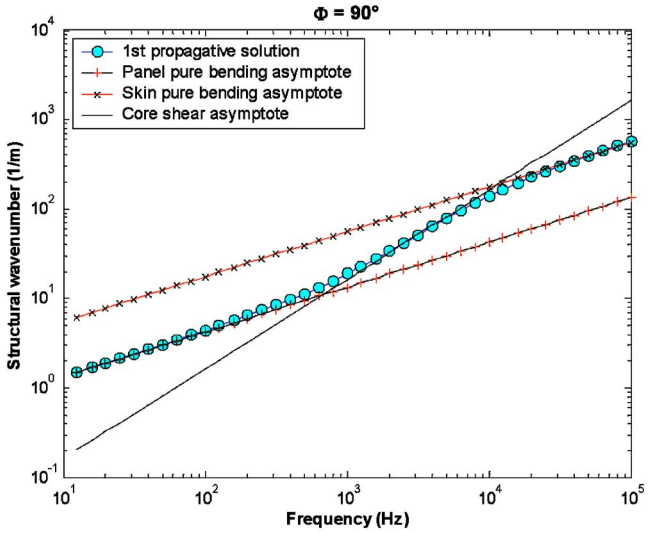


FIG. 8. Dispersion asymptotes of the first propagative solution for a composite sandwich curved panel at a heading angle $\Phi=90°$.

B. Vibroacoustic indicators and experimental validation

The dispersion results represented in Fig. 5 are used in this section to compute the associated vibro-acoustic indicators (modal density, radiation efficiency, and transmission loss) following the methodology of Secs. V–VII. Figure 9 presents the results obtained for the modal density computation. Three models are compared: the presented general discrete laminate model, the sandwich composite theory,[22] and the laminate panel theory.[12] The asymptotic behaviors observed in Fig. 8 are also seen in the modal density curve. It is observed that at low frequencies (below the ring frequencies) all three models lead to the same result. Moreover, at high frequencies the modal density of a sandwich panel is accurately calculated using a sandwich assumption modeling or a general discrete laminate modeling. However, the laminate model,[12] which uses equivalent physical properties with shear corrections, fails at high frequencies; its dispersion equation has just two correct asymptotes e.g., pure bending

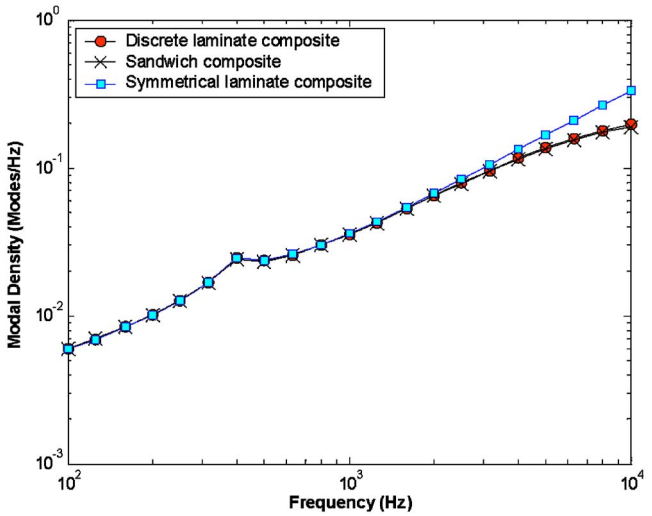


FIG. 9. Modal density of a sandwich composite curved panel.

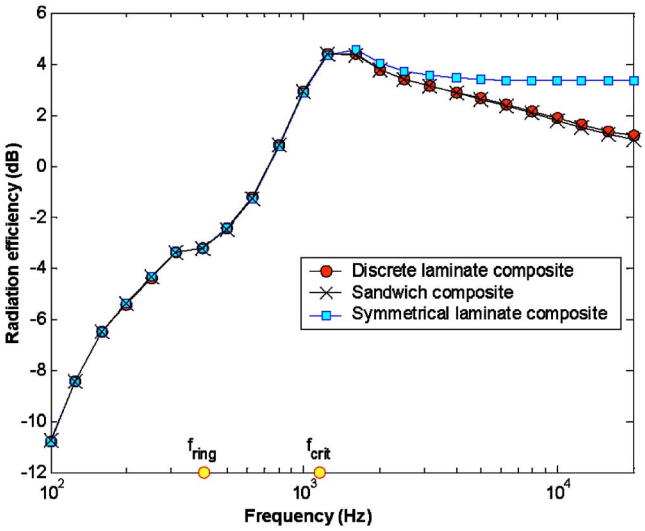


FIG. 10. Radiation efficiency of a sandwich composite curved panel.

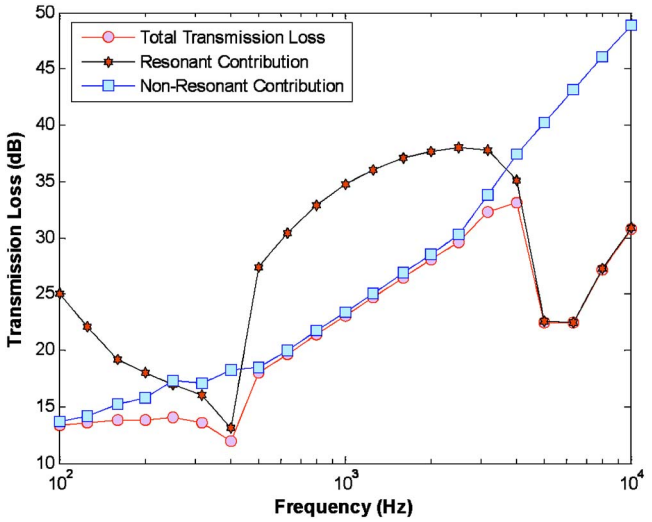


FIG. 12. Contributions of nonresonant and resonant transmissions to the total transmission loss of the structure.

of the panel and shear. It is not able to correctly capture the separate skins asymptotes at high frequencies. The separate bending of the skins behavior becomes especially important for the radiation efficiency computation. As it can be observed in Fig. 10, the composite laminate model could result in large errors in the radiation efficiency estimation for frequencies above the critical frequency zone of the panel.

The proposed general discrete laminate approach uses individual first-order shear displacement fields for each layer. While the use of high order displacement fields may seem more appropriate for very high frequencies, it is worth showing that the proposed approach is sufficient to capture the physics. An example comparing the modal density of a typical flat thick sandwich panel modeled by the proposed approach and a laminate model based on spectral finite elements[14] in which the through-thickness deformation is captured using finite elements is represented in Fig. 11. Also the analytical modal density tendencies for low frequencies (pure bending of the panel) and high frequencies (pure bending of skins) are computed and represented. It can be observed in Fig. 11 that the present approach is accurate. The studied panel has skins made up of Material #1 (0.001 m of thickness) and a core made up of Material #5 (0.003 m thickness). The materials physical properties are presented in Table I. The dimensions of the panel are 2×2.4 m^2.

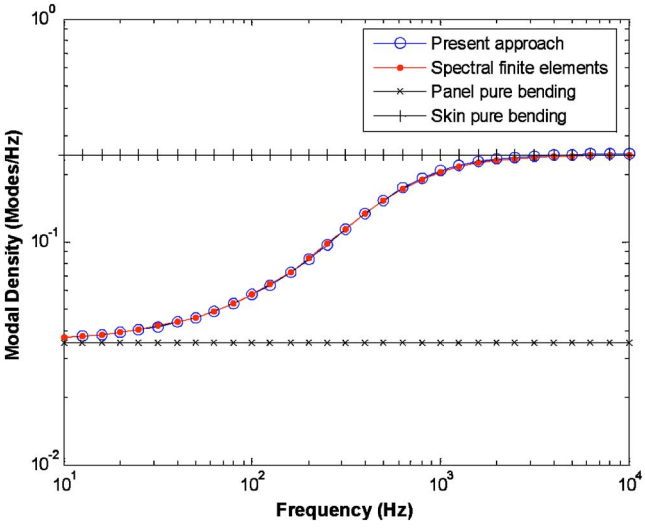


FIG. 11. Discrete laminate composite modeling validation. Comparison of the present approach with spectral finite elements prediction.

In passing, it is worth noting that the presented general discrete laminate model can also be refined using a finer subdivision of the layers to capture complicated behavior through the thickness.

Finally, the physical properties of the lay-up studied in Fig. 5 are used to illustrate the contribution of the nonresonant and resonant transmission to the total transmission loss. This time, the thickness of a skin is 1 mm while the thickness of the core is 3 mm. The lateral dimensions of the panel are 2×2.4 m^2 and the radius of curvature is 2 m. In Fig. 12 are represented the total transmission loss, the resonant, and the nonresonant contributions. The classical tendencies are observed: a combined contribution of resonant and nonresonant transmissions below the ring frequency, a nonresonant contribution between the ring and the critical frequencies, and a resonant contribution above the critical frequency.

C. Experimental validation

Transmission loss tests were performed on the singly curved sandwich composite panel described in the previous section. The tests were performed at the Canadian National Research Center transmission loss facility located in Ottawa. The test panel was installed in the opening window of a transmission loss suite comprising two reverberant chambers. It should be noted that the two chambers are mechanically isolated, and that the test specimen is mounted into a movable heavy frame. The tests were performed according to the specifications of ASTM-E90-97 and ISO 140-1:97. Both chambers are equipped with automated moving microphone position systems, which allow sampling a large volume of the rooms. The volumes of the source and receiving rooms are 140 m^3 and 250 m^3, respectively. The receiving room is

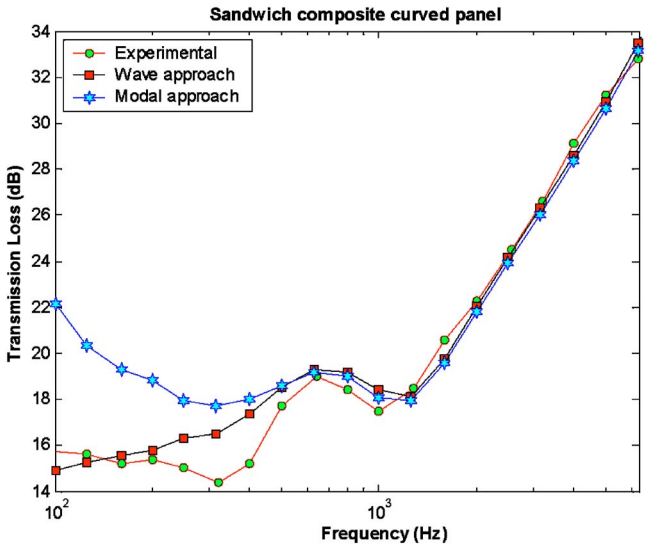


FIG. 13. Total transmission loss of a sandwich composite curved panel.

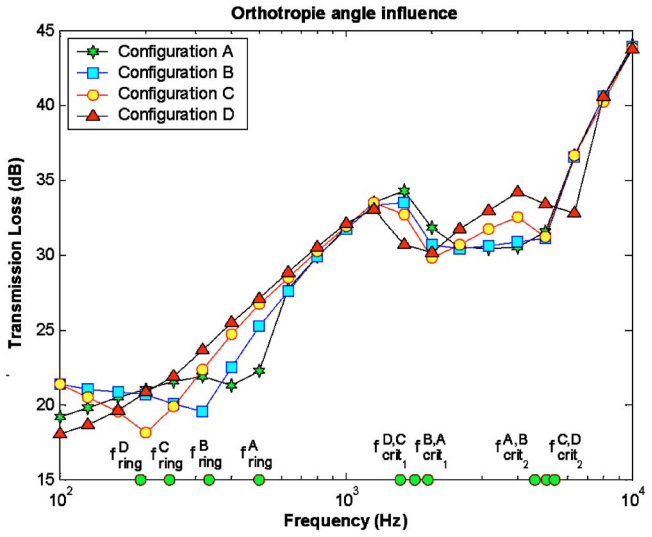


FIG. 14. Orthotropic ply angle influence on the transmission loss of a laminate composite curved panel.

equipped with both stationary and rotating diffusers. TL measurements in this facility are considered valid down to the 100 Hz 1/3 octave band. Considerable care was taken in the design of the supporting frame for the panel and the method of mounting the panel. The curved edges were inserted in a groove supported by closed cell foam and sealed by caulk. The straight edges were also supported by closed cell foam gasket attached to wood chocks and sealed by caulk and tape.

Measured transmission loss and predictions with both the wave approach and the modal approach are given in Fig. 13. In the latter method, the modes of the panel computed from the dispersion equation [Eq. (5)] with the assumption of a simply supported panel are used in conjunction with Eqs. (20) and (21) to compute the nonresonant transmission loss while the resonant transmission loss is calculated with Eq. (24). It should be noted at this stage that the mounted panel damping was not measured, and that a nominal modal damping ratio of 2.5% was assumed in the analysis. This is acceptable since the damping of the panel in free-free conditions is around 1% and increases to 3% around the critical frequency. The wave approach with the geometrical correction leads to a very good agreement throughout the frequency range of the test apart from the ring frequency region. On the other hand, the modal approach shows a higher transmission loss than the test below the panel ring frequency. Above the panel critical frequency, both the wave and modal approaches yield almost identical results. Note that the number of resonant modes is not sufficient in the first 1/3 octave bands for the modal method to be reliable at low frequencies (less than 1 mode at 100 Hz). In summary, the wave approach leads to excellent results, especially at low frequencies, with the benefit of applying the geometrical correction [Eq. (18)], even if the latter is based on the flat baffled window theory.

D. Parametric study

Laminated composite and sandwich composite curved panels of surface 2.438×2.032 m^2 are used here to study the influence of physical parameters on the transmission loss. Orthotropic ply angles, layers' thickness and constant mass per unit area effects are studied. Moreover, this study will also allow for a numerical validation of the ring and critical frequencies' expressions of the panel as calculated from relations (6) and (10).

The transmission loss of a symmetrical laminate sandwich composite panel is presented in Fig. 14. Each skin is a composite made up from three lamina of equal thickness (h_{skin}=0.003 m) made of Material #2. The core has a single layer of Material #4 and thickness h_{core}=0.01 m. The ply angle of the core is 0° but each skin has the following orthotropic layout: 90/45/−45 for configuration A; 0/45/−45 for configuration B, 0/30/−30 for configuration C and 0/0/0 for configuration D. The computed values of the panel's ring and critical frequencies are as follows: f_{ring}=458.55 Hz; f_{c1} =1938.55 Hz; f_{c2}=4586.99 Hz for Configuration A; f_{ring} =334.39 Hz; f_{c1}=1920.76 Hz; f_{c2}=4587.01 Hz for Configuration B; f_{ring}=242.61 Hz; f_{c1}=1751.74 Hz; f_{c2} =5056.42 Hz for Configuration C and f_{ring}=191.66 Hz; f_{c1} =1546.72 Hz; f_{c2}=5398.25 Hz for Configuration D. Their transmission loss is represented in Fig. 14. It is observed that the panel in Configuration A has the shortest region controlled by the mass (between the ring frequency and the first critical frequency) and consequently will better react to a damping treatment. A symbolical simplification of the ring frequency relation (6) shows that it is a function of the y-direction (circumferential) elastic behaviors, the radius and the surface mass of the panel. For this reason, the orthotropic arrangement with the highest y-direction elongation stiffness has the highest ring frequency. Moreover, the width of the critical frequency zone is strongly dependent on the orthotropic arrangement of the laminas. As an example, for a perfectly equilibrate orthotropic arrangement (= same number of plies oriented along θ and $-\theta$) the width of the critical zone tends to zero. The most equilibrate arrangement in Fig. 14 is described by configuration A while the reference configuration D shows the largest critical zone that the panel could have.

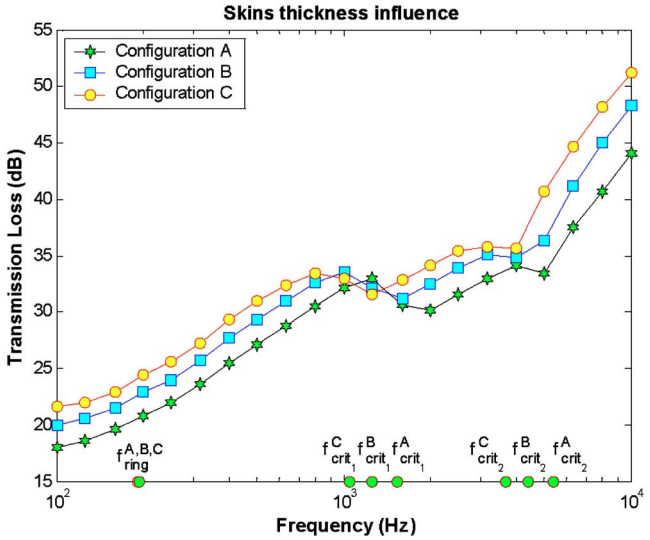


FIG. 15. Skins' thickness influence on the transmission loss of a sandwich composite curved panel.

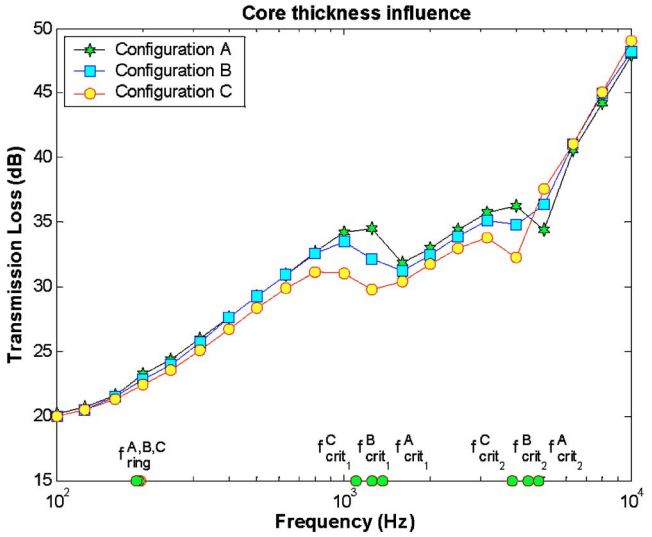


FIG. 16. Core's thickness influence on the transmission loss of a sandwich composite curved panel.

In Fig. 15 the influence of the skin's thickness is illustrated in the context of a sandwich panel. The skins are made of one layer of Material #2 while the core has a single layer of Material #4. The 0/0/0 ply angle case is considered with three different skin thicknesses: Configuration A with the skins' thickness of $h_1=h_3=0.003$ m; Configuration B with $h_1=h_3=0.004$ m and Configuration C with $h_1=h_3=0.005$ m. The core thickness remains unchanged at 0.01 m. Figure 15 shows that the TL increases with the skin's thickness. The increase is in the order of about 3.5 dB in the low frequency range and about 7 dB in the high frequencies between Configuration A and C. The ring frequencies and the critical frequencies calculated for these three configurations are: $f_r=191.66$ Hz; $f_{c1}=1543.9$ Hz; $f_{c2}=5393.32$ Hz for Configuration A, $f_r=193.38$ Hz; $f_{c1}=1260.14$ Hz; $f_{c2}=4403.52$ Hz for Configuration B and $f_r=194.41$ Hz; $f_{c1}=1050.87$ Hz; $f_{c2}=3673.34$ Hz for Configuration C.

Next the influence of the core thickness is illustrated in Fig. 16. Three configurations are considered: $h_2=0.005$ m for Configuration A, $h_2=0.01$ m for Configuration B and $h_2=0.015$ m for Configuration C. In all the three cases, the thickness of the skins is $h_1=h_3=0.004$ m. It is observed that the transmission loss decreases between the ring frequency and the first critical frequency for configurations B and C compared to Configuration A.

A detailed study of this case reveals that the mass law of the three configurations is almost identical while the resonant transmission below the highest critical frequency controls the tendencies observed in Fig. 16. The radiation efficiency in this region is found to increase with the core thickness for this particular construction. Around the highest critical frequency the influence of the separate bending and shearing of the skins starts for these configurations. In this region, the modal density is almost identical for the three configurations and so is the resonant transmission.

To clarify the results of Fig. 16 the same configurations are reinvestigated but this time the sandwich panel has skins made up of Material #3 and a core made up of Material #4 (see Case A in Fig. 17). Next, the skins' density is multiplied by two (Case B). The results are plotted in Fig. 17 for the two cases to highlight the phenomena appearing in the transmission loss nonresonant mass controlled zone (between the ring and the critical frequencies). In Case A the panel does not have a zone mainly controlled by the nonresonant transmission. The region between the ring frequency and the critical frequency is just a resonant transition zone from membrane to bending and shearing behaviors. The transmission loss is mainly controlled by the resonant contributions and for this reason increasing the core thickness results in a degradation of the transmission loss in low and mid-frequency ranges (below and inside the critical frequency region).

Finally, a comparison between three typical configurations based on previous comparisons and having the same overall mass per unit area is presented in Fig. 18. The mass per unit area of the sandwich composite shell is kept constant to $m_s=17.1044$ kg/m^2. As in the previous studies the same

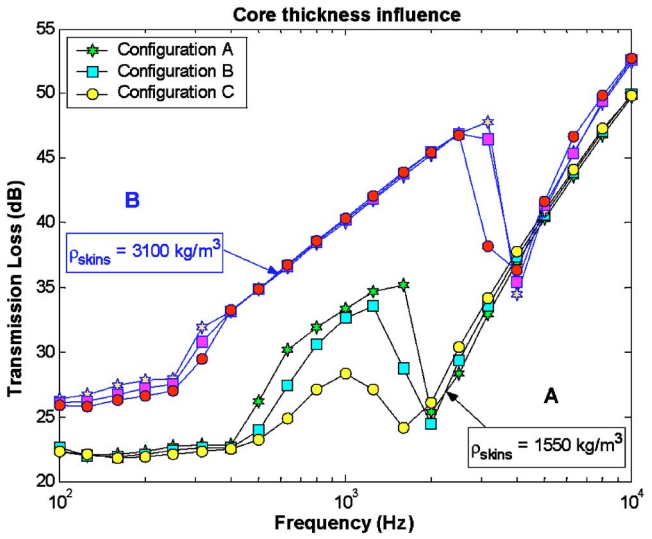


FIG. 17. Core's thickness influence on the transmission loss of a sandwich composite curved panel combined with the influence of the skin's density.

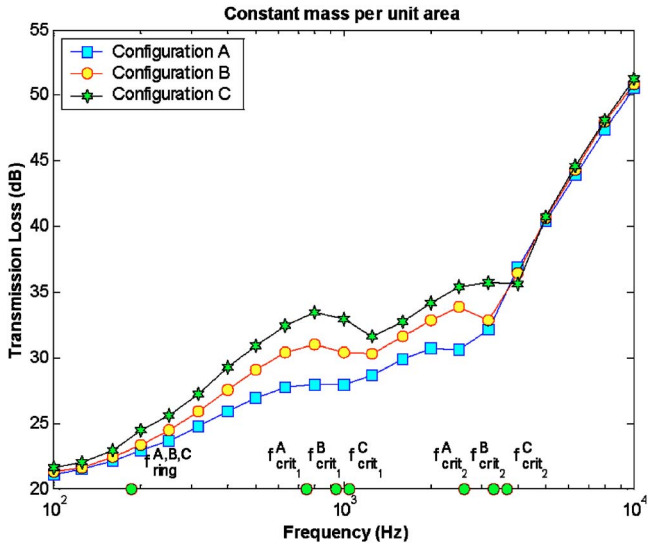


FIG. 18. Constant mass per unit area influence on the transmission loss of a sandwich composite curved panel.

materials are used here. For Configuration A the skins' thickness is $h_1=h_3=0.0045$ m and the core thickness is $h_2=0.0245$ m; for Configuration B the skins' thickness is $h_1=h_3=0.00475$ m and the core thickness is $h_2=0.017244$ m and for Configuration C the thickness of the skins is $h_1=h_3=0.005$ m and the core thickness is $h_2=0.01$ m. The computed ring and critical frequencies of the three configurations are: $f_r=186.577$ Hz; $f_{c1}=748.677$ Hz; $f_{c2}=2621.572$ Hz for Configuration A; $f_r=190.5255$ Hz; $f_{c1}=941.73$ Hz; $f_{c2}=3289.65$ Hz for Configuration B and $f_r=194.41$ Hz; $f_{c1}=1050.87$ Hz; $f_{c2}=3673.35$ Hz for Configuration C. The most interesting configuration (Configuration C) has the thicker skins and the thinnest core. It could be concluded here that the transmission loss of a sandwich composite shell can be improved by a judicious choice of the thicknesses of the layers while keeping constant the mass per unit area.

IX. DISCUSSIONS AND CONCLUSIONS

An efficient model to compute the transmission loss of sandwich and laminate composite curved panels has been presented. The physical behavior of the panel is represented using a discrete lamina description. Each lamina is represented by membrane, bending, transversal shearing and rotational inertia behaviors. The model is developed in the context of a wave approach. It is shown that the dispersion curves and the panel's ring and critical frequencies are accurately estimated. Using the dispersion relation's solutions, the modal density, the radiation efficiency as well as the resonant and nonresonant transmission loss are calculated. The acoustic transmission problem is represented within statistical energy analysis context using two different schemes for the nonresonant path. It is observed that for the presented problem, the modal energy exchange is dominated by the first wave solution. It is concluded that for classical acoustic transmission problems, the SEA scheme presented here is accurate. The results were compared successfully to the transmission loss test of a singly curved sandwich panel and to two other models. In particular, the presented model is applicable to both sandwich panels and composite laminate panels with thin and/or thick layers. Finally, a parameters study showed that the transmission loss of such panels can be improved by a judicious choice of orthotropic arrangements of plies and layer's thicknesses.

APPENDIX: MAIN EQUATIONS OF THE GENERAL DISCRETE LAYER MODEL

1. Equilibrium equations

The resultant stress forces and moments of any laminate layer are defined by, Leissa:[36]

$$Q_x^i=\int_z \tau_{xz}\left(1+\frac{z}{R}\right)dz=\sum_{k=1}^{N^i}\int_{h_{lk}^i}^{h_{uk}^i}\tau_{xz}^k\left(1+\frac{z}{R}\right)dz,$$

$$Q_y^i=\int_z \tau_{yz}dz=\sum_{k=1}^{N^i}\int_{h_{lk}^i}^{h_{uk}^i}\tau_{yz}^k dz, \tag{A1}$$

$$N_x^i=\int_z \sigma_x\left(1+\frac{z}{R}\right)dz=\sum_{k=1}^{N^i}\int_{h_{lk}^i}^{h_{uk}^i}\sigma_x^k\left(1+\frac{z}{R}\right)dz,$$

$$M_x^i=\int_z \sigma_x z\left(1+\frac{z}{R}\right)dz=\sum_{k=1}^{N^i}\int_{h_{lk}^i}^{h_{uk}^i}\sigma_x^k z\left(1+\frac{z}{R}\right)dz,$$

$$N_y^i=\int_z \sigma_y dz=\sum_{k=1}^{N^i}\int_{h_{lk}^i}^{h_{uk}^i}\sigma_y^k dz,$$

$$M_y^i=\int_z \sigma_y z dz=\sum_{k=1}^{N^i}\int_{h_{lk}^i}^{h_{uk}^i}\sigma_y^k z dz,$$

$$N_{xy}^i=\int_z \tau_{xy}\left(1+\frac{z}{R}\right)dz=\sum_{k=1}^{N^i}\int_{h_{lk}^i}^{h_{uk}^i}\tau_{xy}^k\left(1+\frac{z}{R}\right)dz,$$

$$M_{xy}^i=\int_z \tau_{xy} z\left(1+\frac{z}{R}\right)dz=\sum_{k=1}^{N^i}\int_{h_{lk}^i}^{h_{uk}^i}\tau_{xy}^k z\left(1+\frac{z}{R}\right)dz,$$

$$N_{yx}^i=\int_z \tau_{yx}dz=\sum_{k=1}^{N^i}\int_{h_{lk}^i}^{h_{uk}^i}\tau_{yz}^k dz,$$

$$M_{yx}^i=\int_z \tau_{yx} z dz=\sum_{k=1}^{N^i}\int_{h_{lk}^i}^{h_{uk}^i}\tau_{yx}^k z dz, \tag{A2}$$

The integral limits h_{uk}^i and h_{lk}^i in relations (A1) and (A2) are computed using the following relations:

$$h_{uk}^i=z^{i-1}+\sum_{j=1}^{k}h_j^i,\quad h_{lk}^i=z^{i-1}+\sum_{j=0}^{k-1}h_j^i, \tag{A3}$$

where, h_j^i is the thickness of the lamina **j** of the layer **i** ($h_0^i=0$) and z^{i-1} is the position of the $(i-1)$ surface delimiting a layer.

The transverse shear stress forces are defined by the following relations:

$$Q_x^i = \left[F_{45}\left(w_{,y} + \varphi_y - \frac{v}{R} \right) + F_{55}(w_{,x} + \varphi_x) + H_{55}\left(\frac{w_{,x}}{R} + \frac{\varphi_x}{R} \right) \right]^i,$$

$$Q_y^i = \left[F_{44}\left(w_{,y} + \varphi_y - \frac{v}{R} \right) + F_{45}(w_{,x} + \varphi_x) + H_{44}\left(\frac{v}{R^2} - \frac{\varphi_y}{R} - \frac{w_{,y}}{R} \right) \right]^i, \tag{A4}$$

and the in-plane stress forces,

$$N_x^i = \left[A_{11}u_{,x} + A_{12}\left(v_{,y} + \frac{w}{R} \right) + A_{16}(u_{,y} + v_{,x}) + B_{11} \times \left(\frac{u_{,x}}{R} + \varphi_{x,x} \right) + B_{12}\varphi_{y,y} + B_{16}\left(\frac{v_{,x}}{R} + \varphi_{x,y} + \varphi_{y,x} \right) + D_{11}\frac{\varphi_{x,x}}{R} + D_{16}\frac{\varphi_{y,x}}{R} \right]^i,$$

$$N_y^i = \left[A_{12}u_{,x} + A_{22}\left(v_{,y} + \frac{w}{R} \right) + A_{26}(u_{,y} + v_{,x}) + B_{12}\varphi_{x,x} + B_{22}\left(\varphi_{y,y} - \frac{v_{,y}}{R} - \frac{w}{R^2} \right) + B_{26}\left(\varphi_{x,y} + \varphi_{y,x} - \frac{u_{,y}}{R} \right) - D_{22}\frac{\varphi_{y,y}}{R} - D_{26}\frac{\varphi_{x,y}}{R} + D_{26}\frac{u_{,y}}{R^2} + D_{22}\frac{v_{,y}}{R^2} + D_{22}\frac{w}{R^3} \right]^i,$$

$$N_{xy}^i = \left[A_{16}u_{,x} + A_{26}\left(v_{,y} + \frac{w}{R} \right) + A_{66}(u_{,y} + v_{,x}) + B_{16}\left(\frac{u_{,x}}{R} + \varphi_{x,x} \right) + B_{26}\varphi_{y,y} + B_{66} \times \left(\frac{v_{,x}}{R} + \varphi_{x,y} + \varphi_{y,x} \right) + D_{16}\frac{\varphi_{x,x}}{R} + D_{66}\frac{\varphi_{y,x}}{R} \right]^i,$$

$$N_{yx}^i = \left[A_{16}u_{,x} + A_{26}\left(v_{,y} + \frac{w}{R} \right) + A_{66}(u_{,y} + v_{,x}) + B_{16}\varphi_{x,x} + B_{26}\left(\varphi_{y,y} - \frac{v_{,y}}{R} - \frac{w}{R^2} \right) + B_{66}\left(\varphi_{x,y} + \varphi_{y,x} - \frac{u_{,y}}{R} \right) - D_{26}\frac{\varphi_{y,y}}{R} - D_{66}\frac{\varphi_{x,y}}{R} + D_{66}\frac{u_{,y}}{R^2} + D_{26}\frac{v_{,y}}{R^2} + D_{26}\frac{w}{R^3} \right]^i, \tag{A5}$$

as well as the stress moments,

$$M_x^i = \left[B_{11}u_{,x} + B_{12}\left(v_{,y} + \frac{w}{R} \right) + B_{16}(u_{,y} + v_{,x}) + D_{11}\left(\frac{u_{,x}}{R} + \varphi_{x,x} \right) + D_{12}\varphi_{y,y} + D_{16}\left(\frac{v_{,x}}{R} + \varphi_{x,y} + \varphi_{y,x} \right) \right]^i,$$

$$M_y^i = \left[B_{12}u_{,x} + B_{22}\left(v_{,y} + \frac{w}{R} \right) + B_{26}(u_{,y} + v_{,x}) + D_{12}\varphi_{x,x} + D_{22}\left(\varphi_{y,y} - \frac{v_{,y}}{R} - \frac{w}{R^2} \right) + D_{26}\left(\varphi_{x,y} + \varphi_{y,x} - \frac{u_{,y}}{R} \right) \right]^i,$$

$$M_{xy}^i = \left[B_{16}u_{,x} + B_{26}\left(v_{,y} + \frac{w}{R} \right) + B_{66}(u_{,y} + v_{,x}) + D_{16}\left(\frac{u_{,x}}{R} + \varphi_{x,x} \right) + D_{26}\varphi_{y,y} + D_{66}\left(\frac{v_{,x}}{R} + \varphi_{x,y} + \varphi_{y,x} \right) \right]^i,$$

$$M_{yx}^i = \left[B_{16}u_{,x} + B_{26}\left(v_{,y} + \frac{w}{R} \right) + B_{66}(u_{,y} + v_{,x}) + D_{16}\varphi_{x,x} + D_{26}\left(\varphi_{y,y} - \frac{v_{,y}}{R} - \frac{w}{R^2} \right) + D_{66}\left(\varphi_{x,y} + \varphi_{y,x} - \frac{u_{,y}}{R} \right) \right]^i. \tag{A6}$$

The inertial terms derived in the equilibrium equations (2) are expressed by the following relations:

$$m_s^i = \sum_{k=1}^{N^i} [\rho_k(h_{uk} - h_{lk})]^i, \quad I_z^i = \sum_{k=1}^{N^i} \left[\rho_k \frac{(h_{uk}^3 - h_{lk}^3)}{3} \right]^i,$$

$$I_{z2}^i = \sum_{k=1}^{N^i} \left[\rho_k \frac{(h_{uk}^2 - h_{lk}^2)}{2} \right]^i, \tag{A7}$$

where, m_s^i is the mass per unit area, I_z^i and I_{z2}^i are the rational inertia, and ρ_k^i is the mass density of the lamina k of the layer i. The rotational inertia I_{z2}^i is zero for symmetrically laminated composite sandwich panels. The elastic constants derived in Eqs. (A4)–(A6) are defined by the following relations:

$$\begin{cases} A^i_{\alpha\beta} = \sum_{k=1}^{N^i} [Q^k_{\alpha\beta}(h_{uk} - h_{lk})]^i \\ B^i_{\alpha\beta} = \sum_{k=1}^{N^i} \left[Q^k_{\alpha\beta} \frac{h^2_{uk} - h^2_{lk}}{2} \right]^i \qquad \alpha,\beta = 1,2,6 \\ D^i_{\alpha\beta} = \sum_{k=1}^{N^i} \left[Q^k_{\alpha\beta} \frac{h^3_{uk} - h^3_{lk}}{3} \right]^i \end{cases}$$

$$\begin{cases} F^i_{\alpha\beta} = \sum_{k=1}^{N^i} [C^k_{\alpha\beta}(h_{uk} - h_{lk})]^i \\ H^i_{\alpha\beta} = \sum_{k=1}^{N^i} \left[C^k_{\alpha\beta} \frac{h^3_{uk} - h^3_{lk}}{2} \right]^i \end{cases} \qquad \alpha,\beta = 4,5. \tag{A8}$$

In Eq. (A8), $[Q^k_{\alpha\beta}]^i$ are the elastic constants of the kth lamina of layer i and are defined by the following relations:[38]

$$[Q^k_{11}]^i = [C^k_L \cos^4\theta_k + C^k_T \sin^4\theta_k + 2(C^k_{LT} + 2G^k_{LT})\sin^2\theta_k \cos^2\theta_k]^i,$$

$$[Q^k_{12}]^i = [(C^k_L + C^k_T - 4G^k_{LT})\sin^2\theta_k \cos^2\theta_k + C^k_{LT}(\cos^4\theta_k + \sin^4\theta_k)]^i,$$

$$[Q^k_{16}]^i = [(C^k_L - C^k_{LT} - 2G^k_{LT})\sin\theta_k \cos^3\theta_k + (C^k_{LT} - C^k_T + 2G^k_{LT})\sin^3\theta_k \cos\theta_k]^i,$$

$$[Q^k_{22}]^i = [C^k_L \sin^4\theta_k + C^k_T \cos^4\theta_k + 2(C^k_{LT} + 2G^k_{LT}) \times \sin^2\theta_k \cos^2\theta_k]^i,$$

$$[Q^k_{26}]^i = [(C^k_L - C^k_{LT} - 2G^k_{LT})\sin^3\theta_k \cos\theta_k + (C^k_{LT} - C^k_T + 2G^k_{LT})\sin\theta_k \cos^3\theta_k]^i,$$

$$[Q^k_{66}]^i = [(C^k_L + C^k_T - 2(C^k_{LT} + G^k_{LT}))\sin^2\theta_k \cos^2\theta_k + G^k_{LT}(\cos^4\theta_k + \sin^4\theta_k)]^i, \tag{A9}$$

with

$$[C^k_L]^i = \left(\frac{E^k_L}{1 - \nu^k_{LT}\nu^k_{TL}} \right)^i,$$

$$[C^k_T]^i = \left(\frac{E^k_T}{1 - \nu^k_{LT}\nu^k_{TL}} \right)^i, \quad [C^k_{LT}]^i = \left(\frac{\nu^k_{LT}E^k_T}{1 - \nu^k_{LT}\nu^k_{TL}} \right)^i. \tag{A10}$$

θ_k is the orthotropic orientation (represented in Fig. 4) and $[C^k_{\alpha\beta}]^i$ are the transverse shear elastic constants of the kth lamina of the layer i and are defined by[38]

$$[C^k_{44}]^i = [G^k_{TZ}\cos^2\theta_k + G^k_{LZ}\sin^2\theta_k]^i,$$

$$[C^k_{45}]^i = [(G^k_{LZ} - G^k_{TZ})\sin\theta_k \cos\theta_k]^i, \tag{A11}$$

$$[C^k_{55}]^i = [G^k_{LZ}\cos^2\theta_k + G^k_{TZ}\sin^2\theta_k]^i.$$

The dynamic equilibrium equations of the shell are rewritten, after appropriate algebraic manipulations as

$$\left[\left(A_{11} + \frac{B_{11}}{R}\right)u_{,xx} + 2A_{16}u_{,xy} + \left(A_{66} - \frac{B_{66}}{R}\right)u_{,yy} + \left(A_{16} + \frac{B_{16}}{R}\right)v_{,xx} + (A_{12} + A_{66})v_{,xy} + \left(A_{26} - \frac{B_{26}}{R}\right)v_{,yy} + \left(B_{11} + \frac{D_{11}}{R}\right)\varphi_{x,xx} + 2B_{16}\varphi_{x,xy} + \left(B_{66} - \frac{D_{66}}{R}\right)\varphi_{x,yy} + \left(B_{16} + \frac{D_{16}}{R}\right)\varphi_{y,xx} + (B_{12} + B_{66})\varphi_{y,xy} + \left(B_{26} - \frac{D_{26}}{R}\right)\varphi_{y,yy} + \frac{A_{12}}{R}w_{,x} + \left(\frac{A_{26}}{R} - \frac{B_{26}}{R^2}\right)w_{,y} \right]^i + F^i_x - F^{i-1}_x + \left(\left(m_s + \frac{I_{z2}}{R}\right)u + \left(\frac{I_z}{R} + I_{z2}\right)\varphi_x \right)^i \omega^2 = 0,$$

$$\left[\left(A_{16} + \frac{B_{16}}{R}\right)u_{,xx} + (A_{12} + A_{66})u_{,xy} + \left(A_{26} - \frac{B_{26}}{R}\right)u_{,yy} + \left(A_{66} + \frac{B_{66}}{R}\right)v_{,xx} + 2A_{26}v_{,xy} + \left(A_{22} - \frac{B_{22}}{R}\right)v_{,yy} + \left(B_{16} + \frac{D_{16}}{R}\right)\varphi_{x,xx} + (B_{12} + B_{66})\varphi_{x,xy} + \left(B_{26} - \frac{D_{26}}{R}\right)\varphi_{x,yy} + \left(B_{66} + \frac{D_{66}}{R}\right)\varphi_{y,xx} + 2B_{26}\varphi_{y,xy} + \left(B_{22} - \frac{D_{22}}{R}\right)\varphi_{y,yy} + \left(\frac{A_{26}}{R} + \frac{F_{45}}{R}\right)w_{,x} + \left(\frac{A_{22}}{R} - \frac{B_{22}}{R^2} + \frac{F_{44}}{R} - \frac{H_{44}}{R^2}\right)w_{,y} - \left(\frac{F_{44}}{R^2} - \frac{H_{44}}{R^3}\right)v + \frac{F_{45}}{R}\varphi_x + \left(\frac{F_{44}}{R} - \frac{H_{44}}{R^2}\right)\varphi_y \right]^i + F^i_y - F^{i-1}_y + \left(\left(m_s + \frac{I_{z2}}{R}\right)v + \left(\frac{I_z}{R} + I_{z2}\right)\varphi_y \right)^i \omega^2 = 0,$$

$$\left[\left(F_{55} + \frac{H_{55}}{R}\right)w_{,xx} + 2F_{45}w_{,xy} + \left(F_{44} - \frac{H_{44}}{R}\right)w_{,yy} - \frac{A_{12}}{R}u_{,x} - \left(\frac{A_{26}}{R} - \frac{B_{26}}{R^2}\right)u_{,y} - \left(\frac{A_{26}}{R} + \frac{F_{45}}{R}\right)v_{,x} - \left(\frac{A_{22}}{R} - \frac{B_{22}}{R^2} + \frac{F_{44}}{R} - \frac{H_{44}}{R^2}\right)v_{,y} - \left(\frac{B_{12}}{R} - F_{55} - \frac{H_{55}}{R}\right)\varphi_{x,x} - \left(\frac{B_{26}}{R} - \frac{D_{26}}{R^2} - F_{45}\right)\varphi_{x,y} - \left(\frac{B_{26}}{R} - F_{45}\right)\varphi_{y,x} - \left(\frac{B_{22}}{R} - \frac{D_{22}}{R^2} - F_{44} + \frac{H_{44}}{R}\right)\varphi_{y,y} - \left(\frac{A_{22}}{R} - \frac{B_{22}}{R^3}\right)w \right]^i + F^i_z - F^{i-1}_z + (m_s w)^i \omega^2 = 0,$$

$$\left[\left(B_{11}+\frac{D_{11}}{R}\right)u_{,xx}+2B_{16}u_{,xy}+\left(B_{66}-\frac{D_{66}}{R}\right)u_{,yy}+\left(B_{16}+\frac{D_{16}}{R}\right)v_{,xx}+(B_{12}+B_{66})v_{,xy}+\left(B_{26}-\frac{D_{26}}{R}\right)v_{,yy}+D_{11}\varphi_{x,xx}+D_{16}\varphi_{x,xy}\right.$$
$$+D_{66}\varphi_{x,yy}+D_{16}\varphi_{y,xx}+(D_{12}+D_{66})\varphi_{y,xy}+D_{26}\varphi_{y,yy}+\left(\frac{B_{12}}{R}-F_{55}-\frac{H_{55}}{R}\right)w_{,x}+\left(\frac{B_{26}}{R}-\frac{D_{26}}{R^2}-F_{45}\right)w_{,y}+\frac{F_{45}}{R}v$$
$$\left.-\left(F_{55}+\frac{H_{55}}{R}\right)\varphi_x-F_{45}\varphi_y\right]^i+z^iF_x^i-z^{i-1}F_x^{i-1}+\left(I_z\left(\varphi_x+\frac{u}{R}\right)+I_{z2}u\right)^i\omega^2=0,$$

$$\left[\left(B_{16}+\frac{D_{16}}{R}\right)u_{,xx}+(B_{12}+B_{66})u_{,xy}+\left(B_{26}-\frac{D_{26}}{R}\right)u_{,yy}+\left(B_{66}+\frac{D_{66}}{R}\right)v_{,xx}+2B_{26}v_{,xy}+\left(B_{22}-\frac{D_{22}}{R}\right)v_{,yy}+D_{16}\varphi_{x,xx}\right.$$
$$+(D_{12}+D_{66})\varphi_{x,xy}+D_{26}\varphi_{x,yy}+D_{66}\varphi_{y,xx}+2D_{26}\varphi_{y,xy}+D_{22}\varphi_{y,yy}+\left(\frac{B_{26}}{R}-F_{45}\right)w_{,x}-\left(\frac{B_{22}}{R}-\frac{D_{22}}{R^2}\right.$$
$$\left.\left.-F_{44}+\frac{H_{44}}{R}\right)w_{,y}+\left(\frac{F_{44}}{R}-\frac{H_{44}}{R^2}\right)v-F_{45}\varphi_x-\left(F_{44}-\frac{H_{44}}{R}\right)\varphi_y\right]^i+z^iF_y^i-z^{i-1}F_y^{i-1}+\left(I_z\left(\varphi_y+\frac{v}{R}\right)+I_{z2}v\right)^i\omega^2=0. \quad \text{(A12)}$$

2. Dispersion equation matrices

The matrices $[A_0]$, $[A_1]$, $[A_2]$ used in Eq. (5) are real square matrices of dimension $5N+3(N-1)$ defined as follows:

$$[A_0]=\begin{bmatrix}
[A_0]^1 & 0 & 0 & 0 & 0 & 0 & [F_0]^1 & 0 & 0 & 0 & 0 & 0\\
 & [A_0]^2 & 0 & 0 & 0 & 0 & -[F_0]^1 & [F_0]^2 & 0 & 0 & 0 & 0\\
 & & [A_0]^3 & 0 & 0 & 0 & 0 & -[F_0]^2 & [F_0]^3 & 0 & 0 & 0\\
 & & & \ddots & 0 & 0 & 0 & 0 & 0 & \ddots & 0 & 0\\
 & & & & [A_0]^{N-1} & 0 & 0 & 0 & 0 & 0 & -[F_0]^{N-2} & [F_0]^{N-1}\\
 & & & & & [A_0]^N & 0 & 0 & 0 & 0 & 0 & -[F_0]^{N-1}\\
 & & & & & & 0 & 0 & 0 & 0 & 0 & 0\\
 & & & & & & & 0 & 0 & 0 & 0 & 0\\
 & & \text{sym} & & & & & & 0 & 0 & 0 & 0\\
 & & & & & & & & & 0 & 0 & 0\\
 & & & & & & & & & & 0 & 0\\
 & & & & & & & & & & & 0
\end{bmatrix}$$

$$[A_1]=\begin{bmatrix}
[A_1]^1 & 0 & 0 & 0 & 0\\
0 & [A_1]^1 & 0 & 0 & 0\\
0 & 0 & \ddots & 0 & 0\\
0 & 0 & 0 & [A_1]^N & 0\\
0 & 0 & 0 & 0 & [0]
\end{bmatrix};\quad [A_2]=\begin{bmatrix}
[A_2]^1 & 0 & 0 & 0 & 0\\
0 & [A_2]^2 & 0 & 0 & 0\\
0 & 0 & \ddots & 0 & 0\\
0 & 0 & 0 & [A_2]^N & 0\\
0 & 0 & 0 & 0 & [0]
\end{bmatrix}; \quad \text{(A13)}$$

where

$$[A_0]^i=\begin{bmatrix}
a_{11} & 0 & 0 & a_{14} & 0\\
0 & a_{22} & 0 & a_{24} & a_{25}\\
0 & 0 & a_{33} & 0 & 0\\
a_{14} & a_{24} & 0 & a_{44} & a_{45}\\
0 & a_{25} & 0 & a_{45} & a_{55}
\end{bmatrix}^i;\quad [F_0]^i=\begin{bmatrix}
1 & 0 & 0\\
0 & 1 & 0\\
0 & 0 & 1\\
z^i & 0 & 0\\
0 & z^i & 0
\end{bmatrix};$$

$$[A_1]^i=\begin{bmatrix}
0 & 0 & \alpha_{13} & 0 & 0\\
0 & 0 & \alpha_{23} & 0 & 0\\
-\alpha_{13} & -\alpha_{23} & 0 & \alpha_{34} & \alpha_{35}\\
0 & 0 & -\alpha_{34} & 0 & 0\\
0 & 0 & -\alpha_{35} & 0 & 0
\end{bmatrix}^i;\quad [A_2]^i=\begin{bmatrix}
\beta_{11} & \beta_{12} & 0 & \beta_{14} & \beta_{15}\\
\beta_{12} & \beta_{22} & 0 & \beta_{24} & \beta_{25}\\
0 & 0 & \beta_{33} & 0 & 0\\
\beta_{14} & \beta_{24} & 0 & \beta_{44} & \beta_{45}\\
\beta_{15} & \beta_{25} & 0 & \beta_{45} & \beta_{55}
\end{bmatrix}^i; \quad \text{(A14)}$$

with coefficients $a^i_{\gamma\delta}$, $\alpha^i_{\gamma\delta}$ and $\beta^i_{\gamma\delta}$ defined as follows:

$$a^i_{11} = \left(m_s + \frac{I_{z2}}{R}\right)\omega^2; \quad a^i_{14} = \left(\frac{I_z}{R} + I_{z2}\right)\omega^2;$$

$$a^i_{22} = \left(m_s + \frac{I_{z2}}{R}\right)\omega^2 - \frac{F_{44}}{R^2} + \frac{H_{44}}{R^3};$$

$$a^i_{24} = \frac{F_{45}}{R}; \quad a^i_{25} = \frac{F_{44}}{R} - \frac{H_{44}}{R^2} + \left(\frac{I_z}{R} + I_{z2}\right)\omega^2;$$

$$a^i_{33} = \left(m_s + \frac{I_{z2}}{R}\right)\omega^2 - \frac{A_{22}}{R^2} + \frac{B_{22}}{R^3};$$

$$a^i_{44} = I_z\omega^2 - F_{55} - \frac{H_{55}}{R}; \tag{A15}$$

$$a^i_{45} = -F_{45}; \quad a^i_{55} = I_z\omega^2 - F_{44} - \frac{H_{44}}{R};$$

$$a^i_{13} = -\left(\frac{A_{12}}{R}\right)^i \cos\varphi - \left(\frac{A_{26}}{R} - \frac{B_{26}}{R^2}\right)^i \sin\varphi,$$

$$\alpha^i_{23} = -\left(\frac{A_{22}}{R} - \frac{B_{22}}{R^2} + \frac{F_{44}}{R} - \frac{H_{44}}{R^2}\right)^i \times \sin\varphi - \left(\frac{A_{26}}{R} + \frac{F_{45}}{R}\right)^i \cos\varphi,$$

$$\alpha^i_{34} = \left(\frac{B_{12}}{R} - F_{55} - \frac{H_{55}}{R}\right)^i \cos\varphi + \left(\frac{B_{26}}{R} - \frac{D_{26}}{R^2} - F_{45}\right)^i \sin\varphi,$$

$$\alpha^i_{35} = \left(\frac{B_{22}}{R} - \frac{D_{22}}{R^2} - F_{44} + \frac{H_{44}}{R}\right)^i \sin\varphi + \left(\frac{B_{26}}{R} - F_{45}\right)^i \cos\varphi, \tag{A16}$$

and

$$\beta^i_{11} = \left(A_{11} + \frac{B_{11}}{R}\right)^i \cos^2\varphi + 2A^i_{16}\cos\varphi\sin\varphi + \left(A_{66} - \frac{B_{66}}{R}\right)^i \sin^2\varphi,$$

$$\beta^i_{12} = \left(A_{16} + \frac{B_{16}}{R}\right)^i \cos^2\varphi + (A_{12} + A_{66})^i \cos\varphi\sin\varphi + \left(A_{26} - \frac{B_{26}}{R}\right)^i \sin^2\varphi,$$

$$\beta^i_{14} = \left(B_{11} + \frac{D_{11}}{R}\right)^i \cos^2\varphi + 2B^i_{16}\cos\varphi\sin\varphi + \left(B_{66} - \frac{D_{66}}{R}\right)^i \sin^2\varphi,$$

$$\beta^i_{15} = \left(B_{16} + \frac{D_{16}}{R}\right)^i \cos^2\varphi + (B_{12} + B_{66})^i \cos\varphi\sin\varphi + \left(B_{26} - \frac{D_{26}}{R}\right)^i \sin^2\varphi,$$

$$\beta^i_{22} = \left(A_{66} + \frac{B_{66}}{R}\right)^i \cos^2\varphi + 2A^i_{26}\cos\varphi\sin\varphi + \left(A_{22} - \frac{B_{22}}{R}\right)^i \sin^2\varphi,$$

$$\beta^i_{24} = \left(B_{16} + \frac{D_{16}}{R}\right)^i \cos^2\varphi + (B_{12} + B_{66})^i \cos\varphi\sin\varphi + \left(B_{26} - \frac{D_{26}}{R}\right)^i \sin^2\varphi,$$

$$\beta^i_{25} = \left(B_{66} + \frac{D_{66}}{R}\right)^i \cos^2\varphi + 2B^i_{26}\cos\varphi\sin\varphi + \left(B_{22} - \frac{D_{22}}{R}\right)^i \sin^2\varphi,$$

$$\beta^i_{33} = \left(F_{55} + \frac{H_{55}}{R}\right)^i \cos^2\varphi + 2F^i_{45}\cos\varphi\sin\varphi + \left(F_{44} - \frac{H_{44}}{R}\right)^i \sin^2\varphi,$$

$$\beta^i_{44} = D^i_{11}\cos^2\varphi + 2D^i_{16}\cos\varphi\sin\varphi + D^i_{66}\sin^2\varphi,$$

$$\beta^i_{45} = D^i_{16}\cos^2\varphi + (D_{12} + D_{66})^i \cos\varphi\sin\varphi + D^i_{26}\sin^2\varphi,$$

$$\beta^i_{55} = D^i_{66}\cos^2\varphi + 2D^i_{26}\cos\varphi\sin\varphi + D^i_{22}\sin^2\varphi. \tag{A17}$$

[1] R. S. Langley, "The modal density of anisotropic structural components," J. Acoust. Soc. Am. **99**(6), 3481–3487 (1996).
[2] L. A. Roussos, C. A. Powell, F. W. Grosveld, and L. R. Koval, "Noise Transmission Characteristics of advanced composite structural materials," J. Aircr. **21**, 528–535 (1984).
[3] J. P. D. Wilkinson, "Modal densities of certain shallow structural elements," J. Acoust. Soc. Am. **43**, 245–251 (1968).
[4] L. R. Koval, "On sound transmission into an orthotropic shell," J. Sound Vib. **63**(1), 51–59 (1979).
[5] H. C. Nelson, B. Zapatowski, and M. Bernstein, "Vibration analysis of orthogonally stiffened circular fuselage and comparison with experiment," in Proceedings of the Institute of Aeronautical Sciences National Specialist's Meeting on Dynamics and Aeroelasticity, 1958, pp. 77–87.
[6] L. R. Koval, "Sound transmission into a laminated composite cylindrical shell," J. Sound Vib. **71**(4), 523–530 (1980).
[7] C. W. Bert, J. L. Baker, and D. M. Egle, "Free vibrations of multilayer anisotropic cylindrical shells," J. Compos. Mater. **3**, 480–499 (1969).
[8] A. Blaise and C. Lesueur, "Acoustic transmission through a 2D orthotropic multilayered infinite cylindrical shell," J. Sound Vib. **155**(1), 95–109 (1992).
[9] S. B. Dong, "Free vibration of laminated orthotropic cylindrical shells," J. Acoust. Soc. Am. **44**, 1628–1635 (1968).
[10] S. Ghinet and N. Atalla, "Vibro-acoustic behavior of multilayer orthotropic panels," Can. Acoust. **30**, 72–73 (2002).
[11] Y. X. Zhang and K. S. Kim, "Two simple and efficient displacement-based quadrilateral elements for the analysis of composite laminate plates," Int. J. Numer. Methods Eng. **61**, 1771–1796 (2004).
[12] S. Ghinet, N. Atalla, and H. Osman, "Diffuse field transmission into infi-

nite sandwich composite and laminate composite cylinders," J. Sound Vib. (to be published).
[13]A. Blaise and C. Lesueur, "Acoustic transmission through a "3D" orthotropic multilayered infinite cylindrical shell, Part I: Formulation of the problem," J. Sound Vib. **171**(5), 651–664 (1994).
[14]P. J. Shorter, "Wave propagation and damping in linear viscoelastic laminates," J. Acoust. Soc. Am. **115**, 1917–1925 (2004).
[15]L. L. Erickson, "Modal densities of sandwich panels: theory and experiment," The Shock and Vibration Bulletin **39**(3), 1–16 (1969).
[16]B. L. Clarkson and M. F. Ranky, "Modal density of honeycomb plates," J. Sound Vib. **91**, 103–118 (1983).
[17]K. Renji, P. S. Nair, and S. Narayanan, "Modal density of composite honeycomb sandwich panels," J. Sound Vib. **195**(5), 687–699 (1996).
[18]G. Kurtze and B. G. Watters, "New wall design for high transmission loss or high damping," J. Acoust. Soc. Am. **31**(6), 739–748 (1959).
[19]M. A. Lang and C. L. Dym, "Optimal acoustic design of sandwich panels," J. Acoust. Soc. Am. **57**(6), 1481–1487 (1975).
[20]E. H. Baker and G. Herrmann, "Vibrations of orthotropic cylindrical sandwich shells under initial stress," AIAA J. **4**, 1063–1070 (1966).
[21]E. Nilsson and A. C. Nilsson, "Prediction and measurement of some dynamic properties of sandwich structures with honeycomb and foam cores," J. Sound Vib. **251**(3), 409–430 (2002).
[22]K. H. Heron, "Curved laminates and sandwich panels within predictive SEA," in Proceedings of the Second International AutoSEA Users Conference, 2002, Detroit, USA.
[23]R. Panneton and N. Atalla, "Numerical prediction of sound transmission through finite multilayer systems with poroelastic materials," J. Acoust. Soc. Am. **100**, 346–353 (1996).
[24]F. Sgard, N. Atalla, and J. Nicolas, "A numerical model for the low frequency diffuse field sound transmission loss of double-wall sound barriers with elastic porous linings," J. Acoust. Soc. Am. **108**(6), 2865–2872 (2000).
[25]J. N. Pinder and F. J. Fahy, "A method for assessing noise reduction provided by cylinders," in Proceedings of the Institute of Acoustics, (1993), 195–205, Vol. **15**, Part 3.
[26]M. Villot, C. Guigou, and L. Gagliardini, "Predicting the acoustical radiation of finite size multilayered structures by applying spatial windowing on infinite structures," J. Sound Vib. **245**(3), 433–455 (2001).
[27]S. Ghinet and N. Atalla, "Sound transmission loss of insulating complex structures," Can. Acoust. **29**, 26–27 (2001).
[28]F. G. Leppington, K. H. Heron, and E. G. Broadbent, "Resonant and nonresonant noise through complex plates," Proc. R. Soc. London, Ser. A **458**, 683–704 (2002).
[29]E. Szechenyi, "Modal densities and radiation efficiencies of unstiffened cylinders using statistical methods," J. Sound Vib. **19**, 65–81 (1971).
[30]E. Szechenyi, "Sound transmission through cylinder walls using statistical considerations," J. Sound Vib. **19**, 83–94 (1971).
[31]L. D. Pope and J. F. Wilby, "Band limited power flow into enclosures," J. Acoust. Soc. Am. **62**, 906–911 (1977).
[32]L. D. Pope and J. F. Wilby, "Band limited power flow into enclosures. II," J. Acoust. Soc. Am. **67**, 823–826 (1980).
[33]L. D. Pope, D. C. Rennison, C. M. Willis, and W. H. Mayes, "Development and validation of preliminary analytical models for aircraft interior noise prediction," J. Sound Vib. **82**(4), 541–575 (1982).
[34]F. G. Leppington, E. G. Broadbent, and K. H. Heron, "The acoustic radiation efficiency from rectangular panels," Proc. R. Soc. London, Ser. A **382**, 245–271 (1982).
[35]C. Lesueur, "Rayonnement acoustique des structures—Vibroacoustique, Interactions fluide-structure" (in French) *Acoustical radiation of Structures—Vibroacoustics, Interactions Fluid-Structure* (Editions Eyrolles, Paris, 1988).
[36]A. W. Leissa, *Vibration of Shells*, NASA SP 288 (U.S. Government Printing Office, Washington, D.C., 1973).
[37]H. Osman, N. Atalla, Y. Atalla, and R. Panneton, "Effects of acoustic blankets on the insertion loss of a composite sandwich cylinder," Tenth International Congress on Sound and Vibration, ICSV 10, Stockholm, Sweden, July 2003.
[38]J.-M. Berthelot, *Composite Materials, Mechanical Behavior and Structural Analysis* (Springer-Verlag, New York, 1999).
[39]N. Atalla, S. Ghinet, and H. Osman, "Transmission loss of curved composite panels with acoustic materials," in Proceedings of the 18th International Congress on Acoustics (ICA), Kyoto, 2004.
[40]Z. C. Xi, G. R. Liu, K. Y. Lam, and H. M. Shang, "Dispersion and characteristic surfaces of waves in laminated composite circular cylindrical shells," J. Acoust. Soc. Am. **108**(5), 2179–2186 (2000).

Experimental investigation of targeted energy transfers in strongly and nonlinearly coupled oscillators

D. Michael McFarland[a)]
Department of Aerospace Engineering, University of Illinois at Urbana—Champaign, 306 Talbot Laboratory, 104 S. Wright Street, Urbana, Illinois 61801

Gaetan Kerschen[b)]
Department of Materials Mechanical and Aerospace Engineering, University of Liège, 1 Chemin des Chevreuils (B52/3), B-4000 Liege, Belgium

Jeffrey J. Kowtko, Young S. Lee, and Lawrence A. Bergman[c)]
Department of Aerospace Engineering, University of Illinois at Urbana—Champaign, 306 Talbot Laboratory, 104 S. Wright Street, Urbana, Illinois 61801

Alexander F. Vakakis[d)]
Division of Mechanics, National Technical University of Athens, P.O. Box 64042, GR-157 10 Zografos, Athens, Greece, and Department of Mechanical and Industrial Engineering (adjunct), Department of Aerospace Engineering (adjunct),
University of Illinois at Urbana—Champaign, 306 Talbot Laboratory, 104 S. Wright Street, Urbana, Illinois 61801

(Received 6 October 2004; revised 11 May 2005; accepted 11 May 2005)

Our focus in this study is on experimental investigation of the transient dynamics of an impulsively loaded linear oscillator coupled to a lightweight nonlinear energy sink. It is shown that this seemingly simple system exhibits complicated dynamics, including nonlinear beating phenomena and resonance captures. It is also demonstrated that, by facilitating targeted energy transfers to the nonlinear energy sink, a significant portion of the total input energy can be absorbed and dissipated in this oscillator. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1944649]

PACS number(s): 43.40.−r, 43.40.Ga, 43.40.At [JGM] Pages: 791–799

I. INTRODUCTION

Targeted nonlinear energy transfers between coupled oscillators have been recently investigated. In Refs. 1 and 2 irreversible and almost complete energy transfers between a discrete breather in a donor nonlinear system weakly coupled to an acceptor nonlinear system sustaining another discrete breather were reported. The transfer of energy is very selective because the two oscillators must be well tuned, and the donor must have a specific amount of energy, the acceptor being initially at rest. Application examples include donor and acceptor oscillators described by Hamiltonians (e.g., discrete nonlinear Schrödinger models and a weakly coupled rotor-Morse oscillator system).

Targeted nonlinear energy transfers between an impulsively loaded linear oscillator—termed the primary system—and an essentially nonlinear attachment—termed the nonlinear energy sink (NES)—weakly coupled to it were observed numerically[3,4] and experimentally.[5] It was shown that nonlinear energy pumping (i.e., an irreversible channeling of vibrational energy to the NES) may occur in the presence of viscous dissipation. The concept of essential nonlinearity (i.e., the absence of a linear term in the stiffness-displacement relation) is central because it means that the NES has no preferential resonant frequency; it may resonate *a priori* with (and extract energy from) any mode of the primary structure,[6] which is an attractive feature for vibration absorption and shock mitigation in the presence of broadband disturbances.

However, grounded and relatively heavy nonlinear attachments were considered in these studies, which represents a limitation when the structural weight is an important design criterion. To overcome this drawback, an ungrounded and light attachment strongly nonlinearly coupled to a linear oscillator depicted in Fig. 1 was studied in detail in Refs. 7 and 8. The main conclusions from these studies were as follows.

(i) This seemingly simple two degrees of freedom system can exhibit very complicated dynamics including, for instance, the existence of a countable infinity of periodic orbits and the ability of the NES to engage in an $i:j$ internal resonance with the linear oscillator, i and j being relatively prime integers.
(ii) A nonlinear beating phenomenon can be excited with the NES initially at rest that triggers transient resonance capture on a resonant manifold, which, in turn, is responsible for an irreversible and almost complete energy transfer to the NES.
(iii) A significant percentage of the total input energy can be dissipated in the NES in spite of its modest mass.

[a)]Electronic mail: dmmcf@uiuc.edu
[b)]Currently, Postdoctoral Fellow at the University of Illinois at Urbana—Champaign. Electronic mail: g.kerschen@ulg.ac.be
[c)]Electronic mail: lbergman@uiuc.edu
[d)]Electronic mail: vakakis@central.ntua.gr; avakakis@uiuc.edu

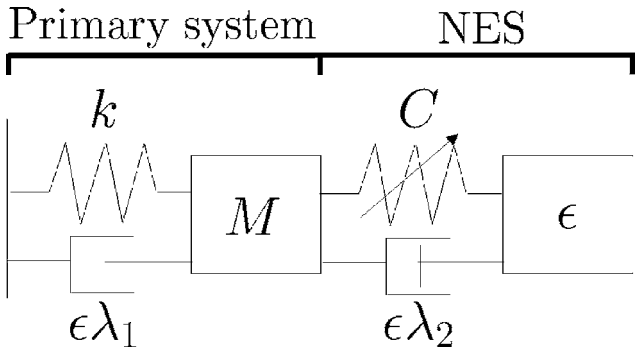


FIG. 1. Linear oscillator coupled to a lightweight NES.

It should be noted that the study of the resonance capture phenomenon has received increasing attention in recent years. Interested readers can refer to Refs. 9–12 for further details.

Our purpose in this paper is to report an experimental study of the targeted energy transfers that may occur between a linear oscillator and a lightweight NES. The paper is organized as follows: In Sec. II, the dynamics of the system of Fig. 1, together with the basic mechanisms for the energy exchanges, are briefly reviewed. The experimental setup and the parameter identification procedure are described in Sec. III. In Sec. IV we present the experimental results, and the conclusions of the present study are summarized in Sec. V.

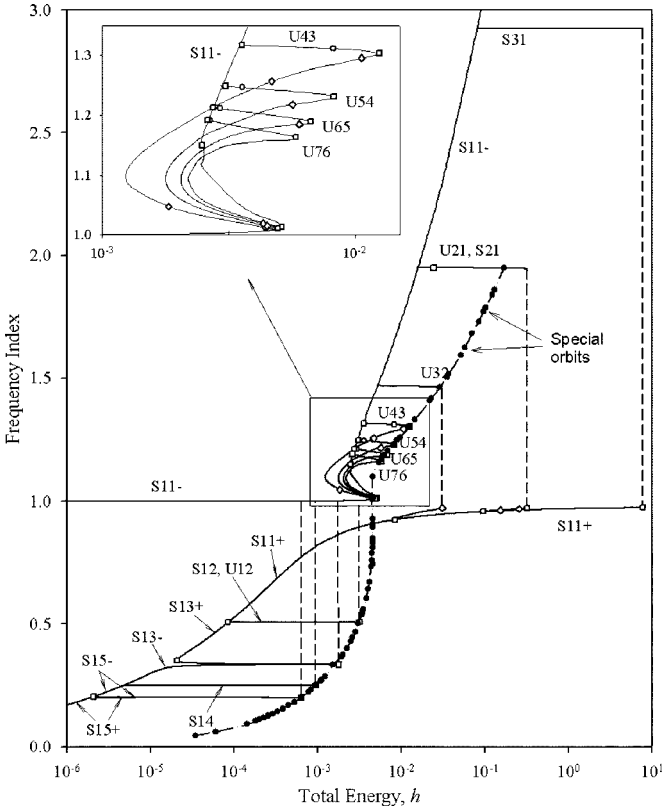


FIG. 2. Frequency-energy plot ($M=k=C=1, \epsilon=0.05$); unfilled dots represent points of stability exchange.

II. REVIEW OF THE DYNAMICS OF THE UNDAMPED SYSTEM

The system considered herein, depicted in Fig. 1, is composed of a linear oscillator strongly coupled to a lightweight NES. The equations of motion are

$$M\ddot{y} + \epsilon\lambda_1\dot{y} + \epsilon\lambda_2(\dot{y}-\dot{v}) + C(y-v)^3 + ky = 0,$$
$$\epsilon\ddot{v} + \epsilon\lambda_2(\dot{v}-\dot{y}) + C(v-y)^3 = 0. \quad (1)$$

Variables y and v refer to the displacement of the primary system and of the NES, respectively. A small mass of the NES and weak damping are assured by requiring that $\epsilon \ll 1$. All other variables are treated as $O(1)$ quantities.

In Refs. 7 and 8, it was shown that the structure and bifurcations of the periodic orbits of the undamped and unforced system enable one to understand the energy transfers in the weakly damped and impulsively loaded system. Therefore, all of the computed periodic orbits were gathered in a frequency-energy plot, represented in Fig. 2 for $M=k=C=1$, $\epsilon=0.05$. It is composed of several branches, each branch being a collection of periodic solutions with the same characteristics. The backbone of the frequency-energy plot is formed by two branches, namely $S11-$ and $S11+$, the latter being continued by $S13+$, $S13-$, $S15-$, $S15+$, etc. The other branches (e.g., $S21$, $U43$, $S14$) are referred to as tongues; each tongue is composed of two very close branches (e.g., $S12-$ and $S12+$) that emanate from the backbone branch but also coalesce into this branch. The following notations are adopted.

(i) Letters S and U refer to symmetric and unsymmetric solutions of the nonlinear boundary value problems that were solved in the calculation of the periodic orbits, respectively. The main qualitative difference between the periodic orbits on S and U branches is that they are represented by lines and Lissajous curves in the configuration space, respectively.

(ii) The two indices indicated for the S and U branches refer to how fast the NES is vibrating relative to the linear oscillator. For example, on branches $S11+$ and $S11-$ the NES engages in a 1:1 resonance capture with the primary system, whereas the NES is vibrating four times slower than the linear oscillator along $S14$.

(iii) The + and − signs in the notations of the branches indicate whether the two oscillators are in phase or out of phase during the periodic motion, respectively.

Due to the essential nonlinearity, the NES has no preferential resonant frequency. As a consequence, it may engage in an $i:j$ internal resonance with the linear oscillator, i and j being arbitrary relatively prime integers. A countable infinity of tongues is thus expected in the frequency-energy plot, each tongue being a realization of a different $i:j$ internal resonance between the primary system and the NES.

A close-up of the $S11+$ branch is presented in Fig. 3, where some representative periodic orbits are also superposed. The convention adopted in this paper is that the horizontal and vertical axes in the configuration space plots depict the displacement of the NES and primary system,

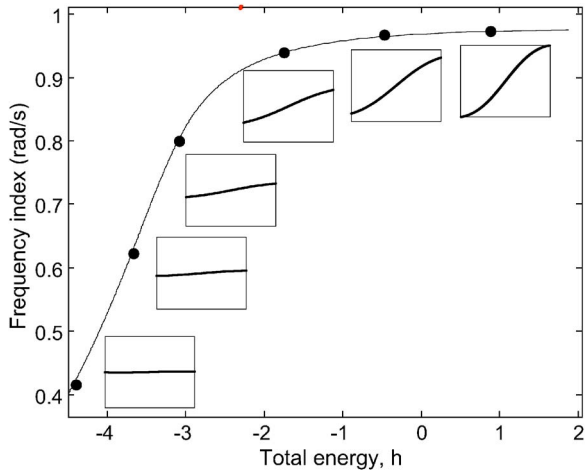


FIG. 3. Close-up of the $S11+$branch.

respectively. Furthermore, the aspect ratio is set so that increments on the horizontal and vertical axes are equal in size, enabling one to directly deduce whether the motion is localized in the linear or the nonlinear oscillator, respectively. The plot of Fig. 3 illustrates how the periodic orbits evolve with energy and frequency and how an irreversible and complete energy transfer to the NES is possible. As the energy decreases, the frequency of the motion diverges from the natural frequency of the linear oscillator (i.e., 1 rad/s in the present case), and the periodic orbits become more and more localized in the NES. Hence, the mode shape of the free nonlinear periodic motion (nonlinear normal mode) changes with energy variation; this feature is not encountered in linear normal modes. By decreasing the total energy, viscous dissipation therefore facilitates targeted energy transfer as the motion localizes from the linear to the nonlinear oscillator. The underlying dynamical phenomenon is a transient resonance capture on a 1:1 resonant manifold because the two oscillators vibrate with the same frequency, but this frequency varies in time with the amount of energy transferred. In the absence of damping, irreversible energy transfer cannot occur; the energy flows back and forth between the NES and the primary system, and a nonlinear beating occurs. We also note that the role of damping has been studied carefully by Gendelman[13] through the computation of damped nonlinear normal modes.

The resonance manifold cannot be reached immediately after the application of an impulsive excitation to the primary system because the shapes of the periodic orbits on the $S11+$ branch are not compatible with the NES being initially at rest. A transient bridging orbit, referred to as a special orbit, is then necessary in order to bring the motion into the domain of attraction of this manifold. This issue is not considered further herein, but a detailed discussion is available in Refs. 7 and 8. Specifically, it was demonstrated that the majority of tongues (e.g., $S31$, $U12$) in the frequency-energy plot carries at least one special orbit that through nonlinear beats triggers targeted energy transfer (energy pumping) from the linear to the nonlinear oscillator.

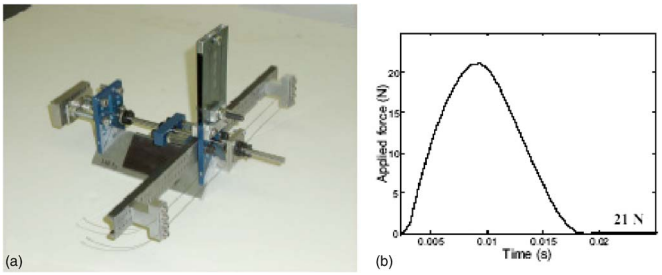


FIG. 4. Experimental setup. (a) General configuration; (b) experimental impulsive force (21 N).

A final remark is that the NES cannot absorb any frequency with equal effectiveness across the spectrum. As shown in Ref. 8, there seems to be a well-defined critical threshold of energy that separates high- from low-frequency special orbits, i.e., those that localize or not to the NES, respectively. As a result, the transfer of a significant amount of energy from the linear oscillator to the NES is only possible above this threshold.

III. EXPERIMENTAL SETUP

A. Description of the experimental fixture

The experimental fixture built to examine the energy transfers in the two degrees of freedom system described by Eqs. (1) is depicted in Fig. 4(a). A schematic of the system is provided in Fig. 5, detailing major components.

The primary system of mass M, grounded by means of a linear leaf spring k, consisted of a car made of aluminum angle stock that was supported on a straight air track, depicted in Fig. 5(a). The NES of mass ϵ, highlighted in Fig. 5(b), consisted of a shaft supported by two bearings. A dashpot was connected to one end of the shaft, allowing adjustable viscous damping between the primary system and the NES, while steel plates clamped two steel wires at the other end. The wires were configured with almost no pretension, realizing the essential nonlinearity C. They were connected

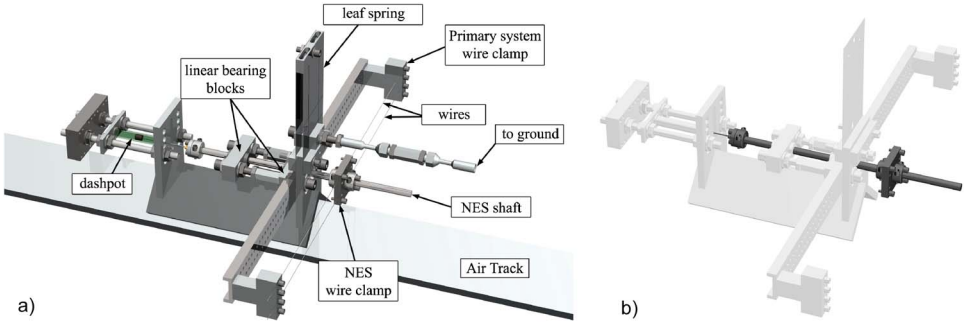


FIG. 5. (a) Schematic of the experimental fixture; (b) mass repartition (grey: primary system; black: NES.

TABLE I. System parameters identified using modal analysis and the restoring force surface method.

Parameter	Value
M	1.266 kg
ϵ	0.140 kg
k	1143 N/m
$\epsilon\lambda_1$	0.155 N s/m
$\epsilon\lambda_2$	0.4 N s/m
C	0.185×10^7 N/m$^{2.8}$
α	2.8

to the primary system using another clamp, the position of which could be modified. For this experiment, the wires' span was adjusted to 12 in. For further details about the construction of the essential nonlinearity, the interested reader can refer to Ref. 5

A long-stroke shaker provided a controlled (and repeatable) short impulse to the primary system. A representative input (broadband) force is given in Fig. 4(b).

The response of both oscillators was measured using accelerometers. An estimate of the corresponding velocity and displacement was obtained by integrating the measured acceleration. The resulting signals were then high-pass filtered to remove the spurious components introduced by the integration procedure.

B. System identification

The goal of system identification is to exploit input and output measurements performed on the structure using vibration sensing devices in order to estimate all the parameters governing the equation of motion (1). It should be noted that, prior to system identification, the primary system and the NES were weighed, and their masses were found to be equal to $M=1.266$ kg and $\epsilon=0.140$ kg, respectively, which implies a mass ratio ϵ/M equal to 0.11. This represents a smaller increase of the total mass of the system compared to previous experimental measurements performed on another NES configuration[5] for which the mass ratio was 0.47.

System identification was carried out in two separate steps. First, the primary system was disconnected from the NES, and modal analysis was performed on the disconnected primary system using the stochastic subspace identification method.[14] The natural frequency and the critical viscous damping ratio were estimated to be 4.78 Hz and 0.2%, respectively. Because the mass of the primary system was known, the stiffness and the damping parameters were easily deduced from this modal analysis; their values are listed in Table I.

In the second step, the primary system was clamped, and an impulsive force was applied to the NES using an instrumented hammer. The NES acceleration and the applied force were measured. The restoring force surface method[15] was then used to estimate the nonlinearity C and the damping coefficient $\epsilon\lambda_2$. In essence, Newton's second law was applied,

$$f_{\mathrm{NL}}(v,\dot{v},y,\dot{y}) = p - \epsilon\ddot{v}, \tag{2}$$

where $f_{\mathrm{NL}}(v,\dot{v},y,\dot{y})$ was the restoring force and p the external force [for simplicity, the temporal dependence is omitted]. Equation (2) shows that the time history of the restoring force can be calculated directly from the measurement of the acceleration and the external force and from the knowledge of the mass. This is illustrated for the 21N force level in Fig. 6(a). The representation of the restoring force in terms of the relative displacement $v-y$ in Fig. 6(b) demonstrates that the linear component of the nonlinear stiffness was negligible; in other words, an essential nonlinearity was realized. The model

$$f_{\mathrm{NL}}(v,\dot{v},y,\dot{y}) = \epsilon\lambda_2(\dot{v}-\dot{y}) + C(v-y)^3 \tag{3}$$

could then be fitted to the measured estimate of the restoring force, and least-squares parameter estimation could be used to obtain the values of coefficients C and $\epsilon\lambda_2$. For greater flexibility, the functional form of the nonlinear stiffness was relaxed to

$$f_{\mathrm{NL}}(v,\dot{v},y,\dot{y}) = \epsilon\lambda_2(\dot{v}-\dot{y}) + C|v-y|^{\alpha}\,\mathrm{sign}(v-y). \tag{4}$$

The three unknown parameters, namely the nonlinear coefficient, the exponent of the nonlinearity and the dashpot constant, were estimated by following the same procedure as in Ref. 16; i.e., one looks for the minimum of the normalized mean-square error between the measured and predicted restoring forces as a function of the exponent of the nonlinearity. The resulting parameters are listed in Table I. The best results have been obtained using an exponent equal to 2.8 which is not far from the theoretical value of 3.

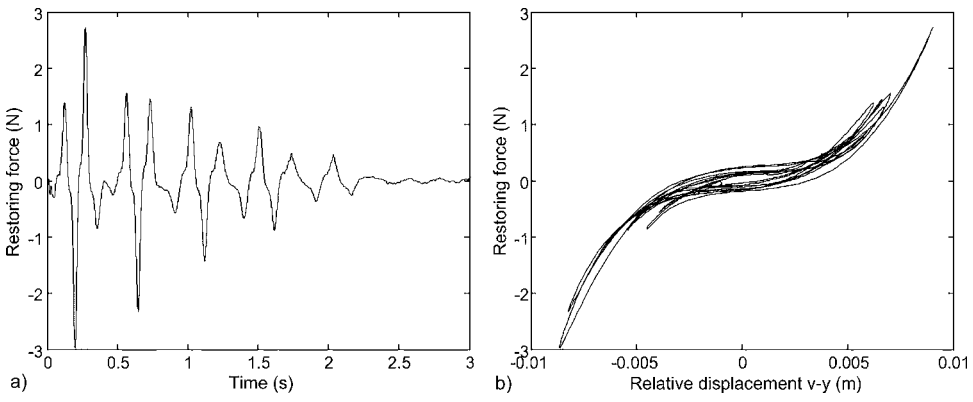


FIG. 6. Measured restoring force represented as a function of (a) time and (b) relative displacement $v-y$.

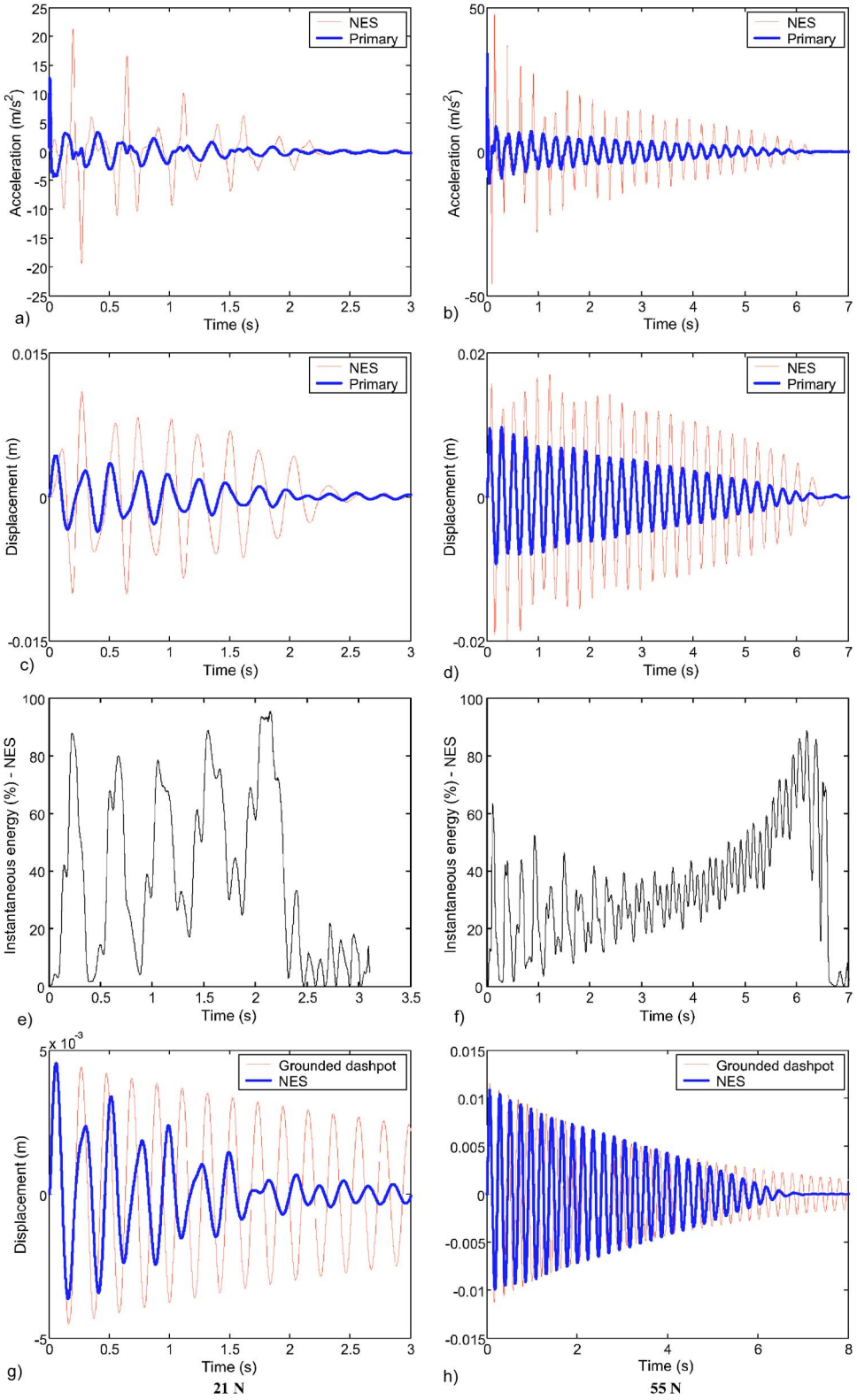


FIG. 7. Experimental results for low damping (left column: 21 N; right column: 55 N; note differing durations): (a), (b) measured accelerations; (c), (d) measured displacements; (e), (f) percentage of instantaneous total energy in the NES; (g), (h) displacement of the primary system (NES versus grounded dashpot).

IV. INTERPRETATION OF THE EXPERIMENTAL RESULTS

Two series of experimental tests were conducted in which the primary system was impulsively loaded. In the first series of tests, the damping in the NES was kept relatively low in order to highlight the different mechanisms for targeted energy transfers. Additional tests were performed to investigate whether the energy transfers to an NES can also take place with increased damping.

A. Low-damping case

In the low-damping case, several force levels ranging from 21 N to 55 N were considered, but for conciseness, only the results for the lowest and highest force levels are displayed in Fig. 7.

At the 21 N level, the acceleration and displacement of the NES are higher than those of the primary system, which indicates that the NES participates in the system dynamics to

TABLE II. Nonlinear beating phenomenon: energy transferred to the NES and transfer time.

Excitation level (N)	Energy transferred to the NES (%)	Transfer time (s)
21	88	0.23
29	72	0.20
34	67	0.19
45	64	0.13
55	63	0.12

a large extent. Figure 7(e), showing the percentage of instantaneous total energy carried by the NES, illustrates that vigorous energy exchanges take place between the two oscillators. However, it can also be observed that the channeling of energy to the NES is not irreversible. After 0.23 s, as much as 88% of the total energy is present in the NES, but this number drops down to 1.5% immediately thereafter. Hence, in this case, energy quickly flows back and forth between the two oscillators, which is characteristic of a nonlinear beating phenomenon. Another indication that a nonlinear beating occurs is that the envelope of the NES response undergoes large modulations.

At the 55 N level, the nonlinear beating phenomenon still dominates the early regime of the motion. A less vigorous but faster energy exchange is now observed as 63% of the total energy is transferred to the NES after 0.12 s. These quantities also hold for the intermediate force levels listed in Table II. It should be noted that these observations are in close agreement with the analytical and numerical studies reported in Refs. 7 and 8; indeed, in this case, the special orbits are such that they transfer smaller amounts of energy to the NES, but in a faster fashion when the force level is increased.

The main qualitative difference from the case of the lowest force level is that now there exists a second regime of motion. After approximately 2.5 s the motion is captured in the domain of attraction of the 1:1 resonance manifold, as clearly evidenced in Fig. 7(f). This graph also demonstrates the irreversibility of this energy transfer, at least until escape from resonance capture occurs around $t=6.2$ s. Another manifestation of the resonance capture is that the envelope of the displacement and acceleration signals decreases almost monotonically in this regime; no modulation is observed. The system is capable of sustaining the resonance capture during a large part of the motion (i.e., from $t=2.5$ s to $t=6.2$ s).

A qualitative means of assessing the energy dissipation by the NES is to compare the response of the primary system in the following two cases: (a) when the NES is attached to the primary system, which corresponds to the present results; (b) when the NES is disconnected, but its dashpot is installed between the primary system and ground; this corresponds to a single degree of freedom linear oscillator with added damping. Case (b) was not realized in the laboratory, but the system response was computed using numerical simulation. Figures 7(g) and 7(h) compare the corresponding displacements of the linear oscillator in the aforementioned two different system configurations. It can be observed that the NES performs much better than the grounded dashpot for the 21 N level, but this is less obvious for the 55 N level. This might mean that, when the nonlinear beating phenomenon is capable of transferring a significant portion of the total energy to the NES, it might be a more useful mechanism for energy dissipation.

B. High-damping case

Several force levels ranging from 31 N to 75 N were considered, but only the results for the 31 N level are presented herein. The damping constant was identified to be 1.48 Ns/m, which means that damping can no longer be considered to be of order ϵ. The increase in damping is also reflected in the measured restoring force in Fig. 8(f).

The system response shown in Figs. 8(a) and 8(b) is almost entirely damped out after 5 to 6 periods. The NES acceleration and displacement are still higher than the corresponding responses of the primary system, which means that targeted energy transfers may also occur in the presence of higher damping. The percentage of instantaneous total energy in the NES never reaches values close to 100% as in the previous case, but we conjecture that this is due to the increased damping value; as soon as energy is transferred to the NES, it is almost immediately dissipated by the dashpot.

A comparison of the performance of the NES and the grounded dashpot is given in Fig. 8(e). A quantitative measure of energy dissipation is available through the computation of the energy dissipated in the NES normalized by the total input energy

$$E_{\mathrm{diss}}(t)=\frac{\epsilon\lambda_2\int_0^t \dot{v}(\tau)^2\,d\tau}{\int_0^{t_{\max}} p(\tau)\dot{y}(\tau)d\tau}. \tag{5}$$

Experimental and simulated estimates of this quantity are depicted in Fig. 8(d). This demonstrates that as much as 96% of the total input energy is dissipated in the NES. There is very good agreement between predictions and measurements, validating the mathematical model developed in Sec. III.

C. Further results

The wavelet transform (WT) is a relevant technique for time-frequency analysis. In contrast to the fast Fourier transform (FFT) that assumes signal stationarity, the WT involves a windowing technique with variable-sized regions. Small time intervals are considered for high frequency components, whereas the size of the interval is increased for lower-frequency components. As a result, the WT offers a means of computing the temporal evolution of the frequency components of a vibration signal; it is usually represented in a time-frequency plane.

In the present study, the WT is represented in a energy-frequency plane by substituting the instantaneous total energy in the system for time. This enables one to superpose the WT and the frequency-energy plot. In Fig. 9, the backbone curve of the frequency energy plot of the experimental

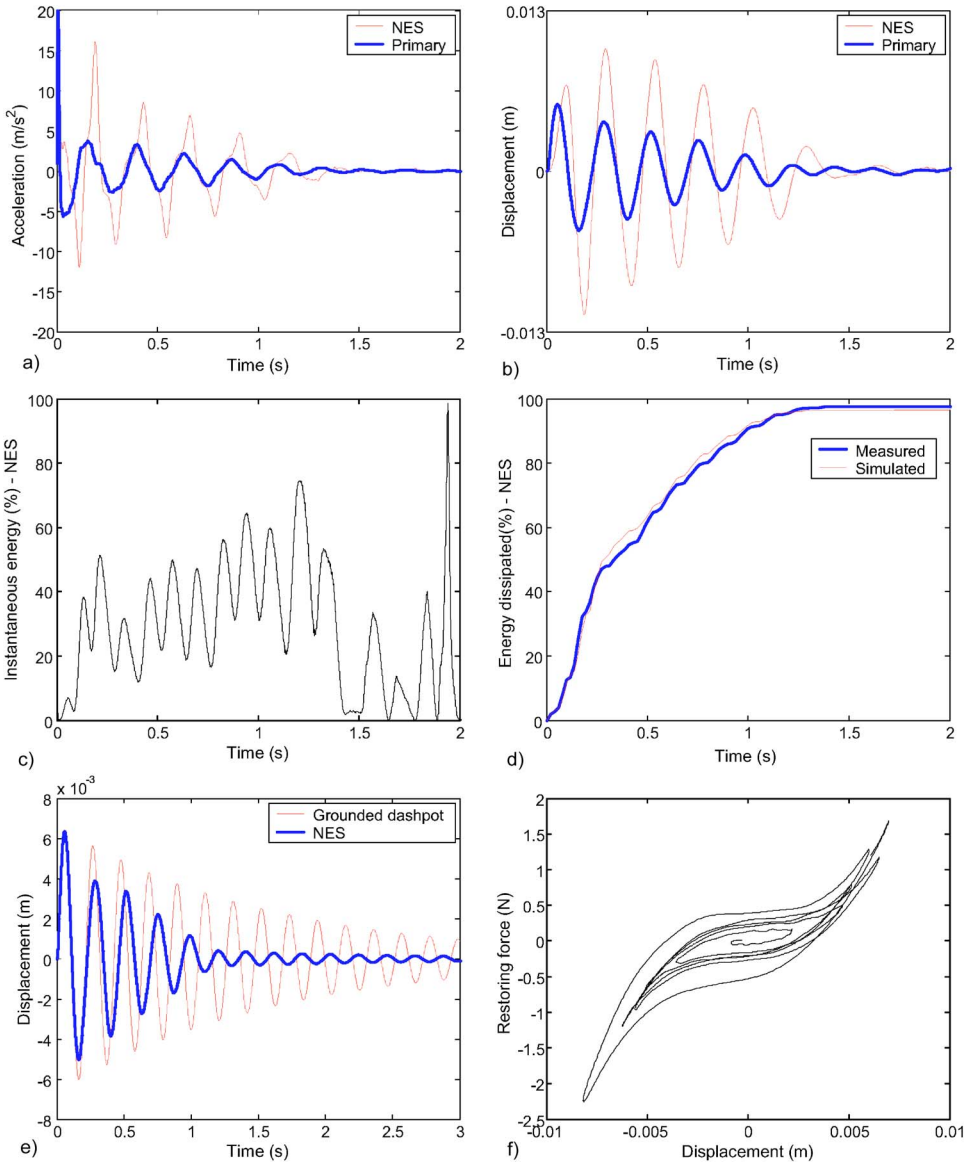


FIG. 8. Experimental results for high damping (31 N): (a) measured accelerations; (b) measured displacements; (c) percentage of instantaneous total energy in the NES; (d) measured and simulated energy dissipated in the NES; (e) displacement of the primary system (NES versus grounded dashpot); (f) restoring force.

system, represented by a solid line, is superposed on the WT of the relative displacement $v-y$. Shaded areas correspond to regions where the amplitude of the WT is high, whereas lightly shaded regions correspond to low amplitudes. This plot is a schematic representation because it superposes damped (the WT) and undamped (the frequency-energy plot) responses and is used for descriptive purposes only. However, it represents a useful tool for the interpretation of the dynamics. It indicates the following.

(i) The dynamics of the system is indeed nonlinear, as the predominant frequency component of the NES varies with energy.

(ii) There are strong harmonic components developing during the nonlinear beating phenomenon. Once these harmonic components disappear, the NES engages in

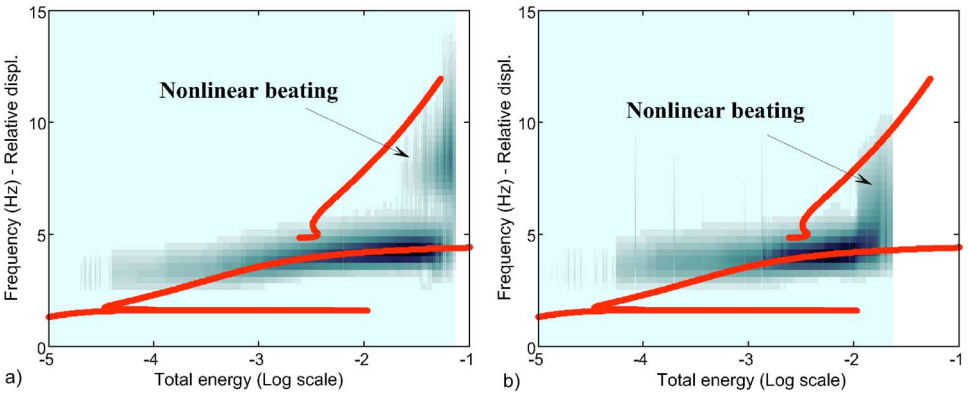


FIG. 9. Superposition of the wavelet transform of the relative displacement across the nonlinearity and the frequency-energy plot. (a) 55 N, low damping; (b) 31 N, high damping.

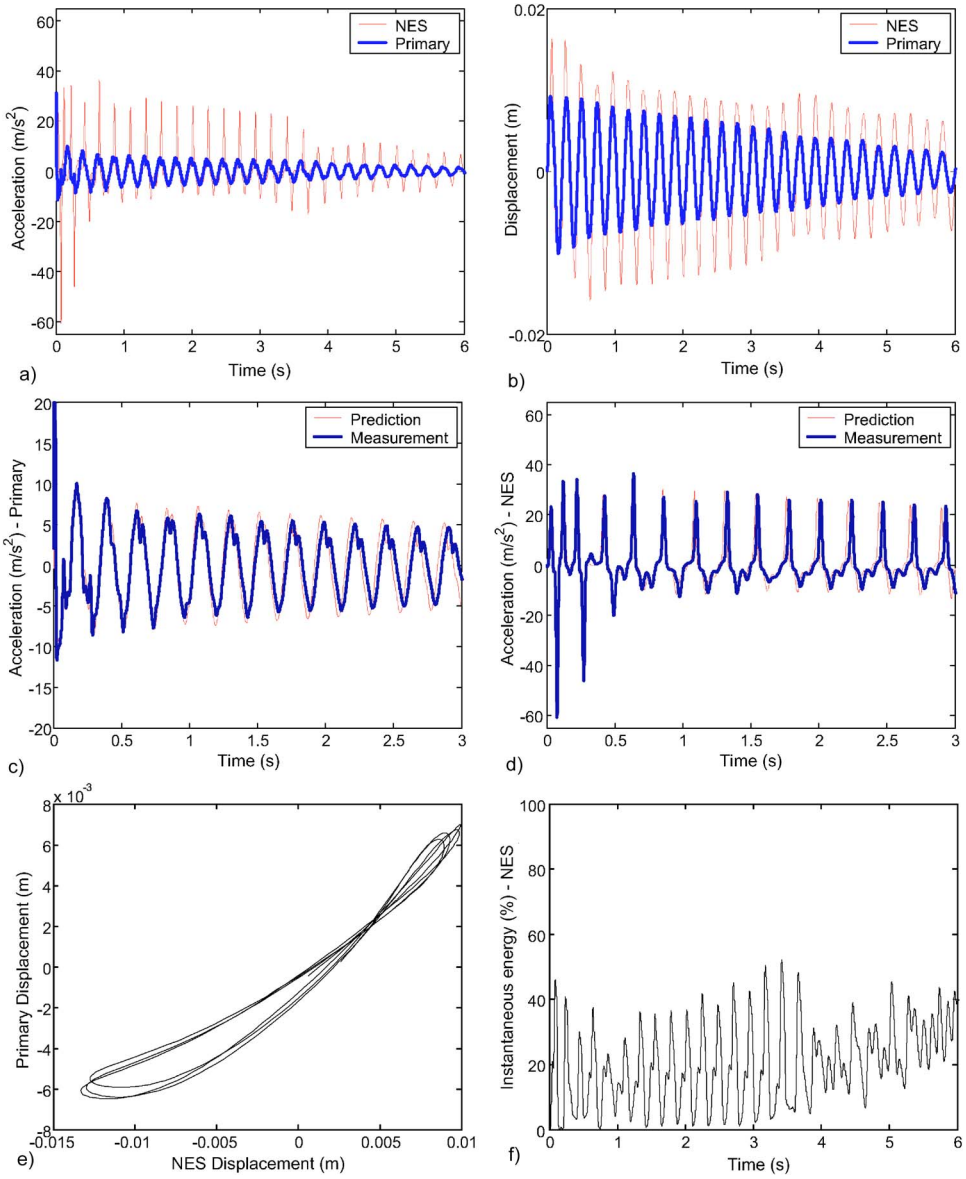


FIG. 10. Case of increased nonlinear coefficient: (a) measured accelerations; (b) measured displacements; a comparison between predicted and measured accelerations: (c) primary system; and (d) NES; (e) motion in the configuration space; (f) percentage of instantaneous total energy in the NES.

a 1:1 resonance capture with the linear oscillator at a frequency approximately equal to the natural frequency of the disconnected linear oscillator.

(iii) The predominant frequency component of the NES follows the backbone branch for most of the signal. This validates our conjecture that the weakly damped, transient dynamics can be interpreted mainly in terms of the periodic orbits of the underlying Hamiltonian system.

Additional measurements were performed using a stiffer nonlinearity by shortening the span of the wire from 12 to 10 in. and by increasing the wire diameter from 0.010 in. to 0.020 in., which results in a nonlinear coefficient of 1.65 10^7 N/m^3. An inspection of the accelerations and displacements shown in Figs. 10(a) and 10(b) reveals that the NES is no longer vibrating symmetrically with respect to its equilibrium position, particularly between $t=1$ s and $t=3$ s. Interestingly enough, there is almost a pointwise agreement between the experimental accelerations and those predicted by the identified numerical model in Figs. 10(c) and 10(d). A better understanding of this particular regime of motion can be gleaned from a snapshot of the configuration space [see Fig. 10(e)]. Apparently, the motion might be captured in the domain of attraction of a tongue on which the characteristic motion is not symmetric with respect to the origin of the configuration space. It turns out that the motion takes the form of a closed loop, which might mean that a U branch is excited. However, due to the existence of a countable infinity of tongues, and due to the presence of damping, it is difficult to ascertain with certainty what tongue is reached. Finally, Fig. 10(f) illustrates that a nonlinear beating occurs during this particular regime, but the energy exchanges are not so vigorous, as approximately 40% is transferred to the NES.

V. CONCLUDING REMARKS

Our purpose in this study is the experimental investigation of targeted energy transfers to a NES. By facilitating these energy transfers, one can promote energy dissipation of a major portion of externally induced energy in the NES.

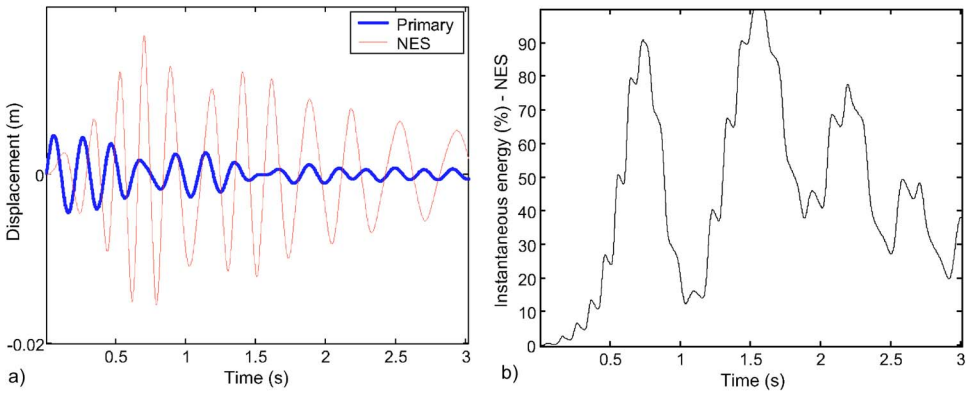


FIG. 11. Targeted energy transfer for a mass ratio of 0.04. (a) Displacements; (b) percentage of instantaneous total energy in the NES.

Two basic mechanisms governing targeted energy transfer have been highlighted, namely the excitation of a transient bridging orbit—resulting in a nonlinear beating phenomenon—and resonance capture into a 1:1 resonance manifold—resulting in irreversible energy flow from the primary system to the NES—with the former mechanism triggering the latter. As a result, this seemingly simple system may possess very complicated dynamics. However, it should be noted that satisfactory agreement was obtained between analytical and computational predictions and experimental measurements throughout this study, this, in spite of the transient and strongly nonlinear (nonlinearizable) nature of the NES dynamics.

The mass ratio between the NES and the primary system considered herein is equal to 11%, but it is worth inquiring whether vigorous energy exchanges can still occur for smaller ratios. Figure 11 confirms that this is indeed possible for a mass ratio of 4% (we note that the nonlinear coefficient had to be modified because decreasing the NES mass shifts the backbone branch toward lower energies in the frequency-energy plot, but this effect can be compensated for by decreasing the nonlinear coefficient). Actually, the limiting mass ratio for this particular setup is 2%. Below this threshold, the special orbits are no longer capable of transferring a sufficient amount of energy to the NES, and targeted energy transfer does not take place. Specifically, it was proven in Ref. 8 that the energy transferred to the NES during the beating tends to zero as the NES mass tends to zero.

Finally, it should be noted that essentially nonlinear attachments are promising for structures with multiple degrees of freedom. Due to the absence of a preferential resonant frequency, a NES has the potential to resonate with (and extract energy from) virtually any mode of the structure. This will be investigated in further detail in subsequent studies.

ACKNOWLEDGMENTS

This work was funded in part by AFOSR Contracts No. F49620-01-1-0208 and No. 00-AF-B/V-0813. One of the authors (GK) is supported by a grant from the Belgian National Fund for Scientific Research (FNRS) which is gratefully acknowledged. The support of the Fulbright and Duesberg Foundations which made GK's visit to the University of Illinois possible is also gratefully acknowledged.

[1]S. Aubry, G Kopidakis, A. M. Morgante, and G. P. Tsironis, "Analytic conditions for targeted energy transfer between nonlinear oscillators or discrete breathers," Physica B **296**, 222–236 (2001).
[2]G. Kopidakis, S. Aubry, and G. P. Tsironis, "Targeted energy transfer through discrete breathers in nonlinear systems," Phys. Rev. Lett. **87**, 165501 (2001).
[3]O. V. Gendelman, "Transition of energy to nonlinear localized mode in highly asymmetric system of nonlinear oscillators," Nonlinear Dyn. **25**, 237–253 (2001).
[4]A. F. Vakakis, "Inducing passive nonlinear energy sinks in vibrating systems," J. Vibr. Acoust. **123**, 324–332 (2001).
[5]D. M. McFarland, L. A. Bergman, and A. F. Vakakis, "Experimental study of nonlinear energy pumping occurring at a single fast frequency," Int. J. Non-Linear Mech. **40**, 891–899 (2005).
[6]A. F. Vakakis, D. M. McFarland, L. A. Bergman, L. I. Manevitch, and O. Gendelman, "Isolated resonance captures and resonance capture cascades leading to single- or multi-mode passive energy pumping in damped coupled oscillators," J. Vibr. Acoust. **126**, 235–244 (2004).
[7]Y. S. Lee, G. Kerschen, A. F. Vakakis, P. N. Panagopoulos, L. A. Bergman, and D. M. McFarland, "Complicated dynamics of a linear oscillator with an essentially nonlinear local attachment," Physica D **204**, 41–69 (2005).
[8]G. Kerschen, Y. S. Lee, A. F. Vakakis, D. M. McFarland, and L. A. Bergman, "Irreversible passive energy transfer in coupled oscillators with essential nonlinearity," SIAM J. Appl. Math..
[9]V. I. Arnold, *Dynamical Systems III*, Encyclopedia of Mathematical Sciences (Springer-Verlag, Berlin, 1988).
[10]D. Quinn, R. Rand, and J. Bridge, "The dynamics of resonance capture," Nonlinear Dyn. **8**, 1–20 (1995).
[11]R. Haberman, R. Rand, and T. Yuster, "Resonant capture and separatrix crossing in dual-spin spacecraft, nonlinear dynamics," Nonlinear Dyn. **18**, 159–171 (1999).
[12]D. L. Vainchtein, E. V. Rovinsky, L. M. Zelenyi, and A. I. Neishtadt, "Resonances and particle stochastization in nonhomogeneous electromagnetic fields," Journal of Nonlinear Science **14**, 173–205 (2004).
[13]O. V. Gendelman, "Bifurcations of nonlinear normal modes of linear oscillator with strongly nonlinear damped attachment," Nonlinear Dyn. **37**, 117–125 (2004).
[14]P. Van Overschee and B. DeMoor, *Subspace Identification For linear Systems: Theory, Implementation, Applications* (Kluwer Academic, Boston, 1996).
[15]S. F. Masri and T. K. Caughey, "A nonparametric identification technique for nonlinear dynamic systems," J. Appl. Mech. **46**, 433–441 (1979).
[16]G. Kerschen, V. Lenaerts, S. Marchesiello, and A. Fasana, "A frequency domain vs. a time domain identification technique for nonlinear parameters applied to wire rope isolators," J. Dyn. Syst., Meas., Control **123**, 645–650 (2001).

A state-space coupling method for fluid-structure interaction analysis of plates

Sheng Li[a)]
Department of Naval Architecture, Dalian University of Technology, Dalian 116024, People's Republic of China

(Received 29 April 2004; revised 25 February 2005; accepted 3 May 2005)

A state-space coupling method is presented for the direct solution of fluid-loaded natural frequencies and mode shapes of plates. This method expands the frequency-dependent term in the Rayleigh integral in a power series on circular frequency. After factoring out the frequency term from the integrands, integration involved in setting up the acoustic impedance coefficient matrices is confined to a frequency-independent part. The acoustic impedance coefficient matrices are therefore frequency-independent and can be coupled directly with the structural matrix into a canonical state-space form to yield the fluid-loaded modes via the direct eigenvalue analysis. A fluid-loaded stiffened plate is involved to demonstrate the method. Numerical results of the fluid-loaded natural frequencies, mode shapes, and modal damping ratios are given to show the efficacy of the state-space coupling method. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1940449]

PACS number(s): 43.40.Rj [MO] Pages: 800–805

I. INTRODUCTION

Coupled fluid-structure problems frequently arise in various engineering applications. Among them the problem of determining the natural frequencies and mode shapes of the fluid-loaded structure is of significant practical interest. Giordano and Koopmann[1] introduced a state-space method for the direct solution of fluid-loaded resonances and mode shapes. Their method is based on well-established boundary element method and finite element method which are used to discretize the acoustic and structural domains, respectively. The state-space method uses a polynomial fit to approximate the individual elements of the acoustic impedance matrix to remove its frequency dependence and couples the acoustic and structural matrices in a canonical state space which is also in the form of a standard complex eigenvalue problem. Where the state-space method introduced by Giordano and Koopmann used an acoustic impedance representation based on surface velocity for least-squares fit, Cunefare and De Rosa[2] demonstrated a simple modification to their method by using an impedance definition in terms of surface displacement. This simple modification reduces the order of the system of equations and improves the computational efficiency and the numerical performance. Cunefare and De Rosa[2] also showed that the results obtained through the state-space method are capable of producing both correct and incorrect eigenvalues and the quality of the results depends on the quality of the underlying fit to the acoustic impedance. They further suggested that the global and local correlation measures should be used to ensure the accuracy of the results.

This paper describes a state-space coupling method for the direct solution of fluid-loaded resonances and mode shapes of plates. The key topic of the proposed change to the state-space method by Giordano and Koopmann[1] and Cunefare and De Rosa[2] is the removal of the frequency dependence of the acoustic impedance matrix, which could avoid the step of the least-squares fit to approximate the individual elements of the acoustic impedance matrix with previously calculated impedance matrices. In the present work, the implicit frequency dependence of the acoustic model is made explicit by expanding the frequency-dependent term in the Rayleigh integral in a power series on circular frequency. After factoring out the frequency term from the integrands, numerical integration involved in setting up the acoustic impedance coefficient matrices is confined to a frequency-independent part. The acoustic impedance coefficient matrices are therefore frequency-independent and can be coupled directly with the structural matrix into a canonical state-space form to yield the fluid-loaded modes via the direct eigenvalue analysis.

II. ACOUSTIC MODEL

If a planar surface extends over an infinite half-space, the acoustic pressure at any field point P according to the Rayleigh integral can be described as follows:[3]

$$p(P) = i\omega\rho \int_S e^{-ikR} v_n(Q)/2\pi R dS, \tag{1}$$

where $p(P)$ is the acoustic pressure at the field point P and has a harmonic time dependency of $e^{i\omega t}$, $i=(-1)^{1/2}$, ω is the circular frequency of excitation, ρ is the density of the acoustic medium, $k=\omega/c$ is the wave number, c is the speed of sound, $v_n(Q)$ is the normal velocity of the vibrating surface at a point Q on the plate surface, $R=|Q-P|$, S is the plate surface.

A numerical solution to the Rayleigh integral equation (1) can be achieved by discretizing the plate surface S into a number of surface elements and nodes. For each position of P, the surface integral in Eq. (1) can be replaced by a sum of integrals over the surface elements (denoted by S_j, where $j=1,\ldots,$ number of elements). The coordinates, pressures, and

a)Electronic mail: shengli@dlut.edu.cn

normal velocities x_i, p, v_n at any points on a surface element are assumed to be related to the coordinates, pressures, and normal velocities x_i^l, p^l, v_n^l ($l=1,\ldots,L$, L is the number of nodes on the surface element) at nodal points on the element, by

$$x_i = \sum_{l=1}^{L} N_l(\xi,\eta)x_i^l, p = \sum_{l=1}^{L} N_l(\xi,\eta)p^l, v_n = \sum_{l=1}^{L} N_l(\xi,\eta)v_n^l, \tag{2}$$

where $N_l(\xi,\eta)$ are the shape functions of the local coordinates $-1 \leqslant \xi \leqslant 1$ and $-1 \leqslant \eta \leqslant 1$.

Then place P at each of nodal points on the surface successively. For each collocation point P, substitute Eq. (2) into Eq. (1) and integrate the equation over the entire surface. The integration is actually done on an element-by-element basis. Each collocation point P and element S_j combination produces an element coefficient vector,

$$z^l = \int_{S_j} N_l(\xi,\eta)\left(i\omega\rho\frac{e^{-ikR}}{2\pi R}\right)J(\xi,\eta)d\xi d\eta, \tag{3}$$

where $z^l = z_R^l + iz_I^l$ has real part z_R^l and imaginary part z_I^l, as

$$z_R^l = \int_{S_j} N_l(\xi,\eta)\left(\frac{-\omega\rho\sin(-kR)}{2\pi R}\right)J(\xi,\eta)d\xi d\eta, \tag{4}$$

$$z_I^l = \int_{S_j} N_l(\xi,\eta)\left(\frac{\omega\rho\cos(-kR)}{2\pi R}\right)J(\xi,\eta)d\xi d\eta, \tag{5}$$

where $J(\xi,\eta)$ is the Jacobian of the transformation Eq. (2).

Assemble the element coefficient vector z^l into a global matrix $[Z]$. This produces

$$\{p\} = [Z]\{v_n\}, \tag{6}$$

where $[Z]$ is the acoustic impedance matrix, $\{p\}$ and $\{v_n\}$ are the vectors consisting of the field values at the nodal locations of a grid defining the plate surface for the surface acoustic pressure and normal velocity. Thus, the link between the acoustic pressure and velocity on the plate surface at a single, specified frequency is formulated in Eq. (6).

III. STRUCTURAL MODEL

For a fluid-loaded structure, if the structure is modeled with finite elements, the resulting matrix equation of motion for the structural degrees of freedom (DOF) can be written as[4]

$$(-\omega^2[M] + i\omega[C] + [K])\{x\} = \{F\} - [G][A]\{p\}, \tag{7}$$

where $[M]$, $[C]$, and $[K]$ are the structural mass, damping, stiffness matrices, respectively, $\{x\}$ is the displacement vector for all structural DOF, $\{F\}$ is the external load vector, $[G]$ is the transformation matrix to transform a vector of normal forces to a vector of forces to all structural DOF, matrix $[A]=\int_S [N]^T[N]dS$, $[N]$ is the matrix of interpolation functions.

IV. STATE-SPACE COUPLING

The vector of normal velocity $\{v_n\}$ in Eq. (6) is related to the vector of the structural velocity $\{v\}$ and the vector of the structural displacement $\{x\}$ by the transformation matrix $[G]$

$$\{v_n\} = [G]^T\{v\} = i\omega[G]^T\{x\}. \tag{8}$$

The traditional approach to handling the structural equation (7) in coupling with the acoustic equation (6) is to eliminate the structural variables from Eqs. (6)–(8). By doing so, one obtains an equation which only contains the surface pressure as the unknown variable to describe the coupled structural acoustic system.[4] It should be noted that the resulting system equation of the coupled structural acoustic system is frequency dependent since the matrix $[Z]$ in Eq. (6) has to be constructed for each single frequency. Consequently, this modeling approach provides no ready means to identify the natural frequencies and modes shapes of the underlying systems.

Giordano and Koopmann[1] and Cunefare and De Rosa[2] have developed the state-space methods for the direct solution of the fluid-loaded resonances and corresponding mode shapes. More detail on the method may be found in their papers. The purpose of the least-squares fit in their state-space methods is to remove the implicit frequency dependence of the acoustic matrix. The present method obtains the frequency-independent acoustic matrix directly and avoids the step of the least-squares fit with previously calculated impedance matrices.

From Eqs. (4) and (5), it can be seen that the element coefficient vectors z_R^l and z_I^l for each P and S_j combination contain integration of $\sin(-kR)$ and $\cos(-kR)$, respectively. Therefore, the final global matrix $[Z]$ in Eq. (6) is implicitly frequency dependent. To make the implicit frequency dependence of the acoustic impedance matrix explicit, the frequency factor ω should be factored out from the integrands in Eqs. (4) and (5). To do this, first use algebraic polynomials

$$\begin{aligned} P_n(-kR) &= P_n\left(-\frac{\omega R}{c}\right) \\ &= a_0 + a_1\left(-\frac{\omega R}{c}\right) + \cdots + a_n\left(-\frac{\omega R}{c}\right)^n \\ &= b_0 + b_1\omega + \cdots + b_n\omega^n = \sum_{i=0}^{n} b_i\omega^i \end{aligned} \tag{9}$$

to approximate $\sin(-kR)$ and $\cos(-kR)$ in Eqs. (4) and (5), respectively, where n is a non-negative integer and $a_0,\ldots,a_n$ are real constants. Then, substitute an algebraic polynomial $\Sigma_{i=0}^n b_{Si}\omega^i$ for $\sin(-kR)$ and an algebraic polynomial $\Sigma_{i=0}^n b_{Ci}\omega^i$ for $\cos(-kR)$ into Eqs. (4) and (5), respectively, to yield

$$z_R^l = \int_{S_j} N_l(\xi,\eta)\left(\frac{-\omega\rho\sin(-kR)}{2\pi R}\right)J(\xi,\eta)d\xi d\eta,$$
$$\approx \int_{S_j} N_l(\xi,\eta)\left(\frac{-\omega\rho}{2\pi R}\sum_{i=0}^{n} b_{Si}\omega^i\right)J(\xi,\eta)d\xi d\eta$$
$$\approx \int_{S_j} N_l(\xi,\eta)\left(\sum_{i=1}^{n+1} c_{Si}(R)\omega^i\right)J(\xi,\eta)d\xi d\eta$$
$$\approx \sum_{i=1}^{n+1}\omega^i\int_{S_j} c_{Si}(R)N_l(\xi,\eta)J(\xi,\eta)d\xi d\eta$$
$$\approx \sum_{i=1}^{n+1}\omega^i z_{Ri}^l, \tag{10}$$

$$z_I^l \approx \sum_{i=1}^{n+1}\omega^i z_{Ii}^l. \tag{11}$$

It is apparent that the element coefficient vectors z_{Ri}^l and z_{Ii}^l are independent of the frequency ω. Assembling the element coefficient vectors z_{Ri}^l into global matrices $[Z_{Ri}]$, and z_{Ii}^l into global matrices $[Z_{Ii}]$, the matrices $[Z_{Ri}]$ and $[Z_{Ii}]$ are also independent of frequency ω. The final global coefficient matrices $[Z]$ can be expressed as

$$[Z] = \sum_{i=1}^{n+1}\omega^i([Z_{Ri}] + i[Z_{Ii}]). \tag{12}$$

The implicit frequency dependence of the acoustic impedance coefficient matrices is now made explicit.

Next, consider the problem of finding a polynomial of a specific degree to approximate the functions $\sin x$ and $\cos x$. Naturally, the power series of $\sin x$ and $\cos x$ can be used as the approximating algebraic polynomials, which is,

$$\sin x = \frac{x}{1} - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots, \tag{13}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots. \tag{14}$$

The truncated error $R_m(x)$ referring to the error involved in using a truncated summation of the first m terms of the above-mentioned infinite series is bounded by

$$|R_m(x)| \leqslant \frac{x^{2m+1}}{(2m+1)!} \quad \text{for } \sin x, \tag{15}$$

$$|R_m(x)| \leqslant \frac{x^{2m}}{(2m)!} \quad \text{for } \cos x. \tag{16}$$

Using a truncated summation of the first three terms of Eqs. (13) and (14), the real and imaginary parts of the acoustic impedance matrix $[Z]$ can be expressed as

$$\mathrm{Re}([Z]) = \omega^2[Z_{R2}] + \omega^4[Z_{R4}] + \omega^6[Z_{R6}], \tag{17}$$

$$\mathrm{Im}([Z]) = \omega[Z_{I1}] + \omega^3[Z_{I3}] + \omega^5[Z_{I5}]. \tag{18}$$

Considering time harmonic motion and replacing the product of the different powers of ω and the structural velocity with the different order time derivative of the displacement vector, the acoustic load vector can be expressed as

$$[G][A]\{p\} = [D_1]\{\ddot{x}\} - [D_2]\{\dddot{x}\} - [D_3]\{\ddddot{x}\} + [D_4]\{\dddddot{x}\} + [D_5]\{\ddddddot{x}\} - [D_6]\{\dddddddot{x}\}, \tag{19}$$

where $[D_1]=[G][A][Z_{I1}][G]^T$, $[D_2]=[G][A][Z_{R2}][G]^T$, $[D_3]=[G][A][Z_{I3}][G]^T$, $[D_4]=[G][A][Z_{R4}][G]^T$, $[D_5]=[G][A][Z_{I5}][G]^T$, $[D_6]=[G][A][Z_{R6}][G]^T$ are real matrices and independent of frequency ω.

Substituting Eq. (19) into Eq. (7) and recasting the coupled system into a state-space form,

$$\begin{bmatrix} -D_6 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & I \end{bmatrix}\begin{Bmatrix} \dddddddot{x} \\ \ddddddot{x} \\ \dddddot{x} \\ \ddddot{x} \\ \dddot{x} \\ \ddot{x} \\ \dot{x} \end{Bmatrix} + \begin{bmatrix} D_5 & D_4 & -D_3 & -D_2 & D_1+M & C & K \\ -I & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -I & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -I & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -I & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -I & 0 \end{bmatrix}\begin{Bmatrix} \ddddddot{x} \\ \dddddot{x} \\ \ddddot{x} \\ \dddot{x} \\ \ddot{x} \\ \dot{x} \\ x \end{Bmatrix} = \begin{Bmatrix} F \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}. \tag{20}$$

It can be seen that $[D_i](i=1,2,3,4,5,6)$ matrices are real and therefore only a real-valued eigenvalue routine is required by setting the right-hand side of Eq. (20) to zero to determine the eigenvalues and eigenvectors. However, the eigenvalues and eigenvectors are complex and eigenvalues occur as complex conjugate pairs with the associated eigenvectors also being complex conjugate pairs.

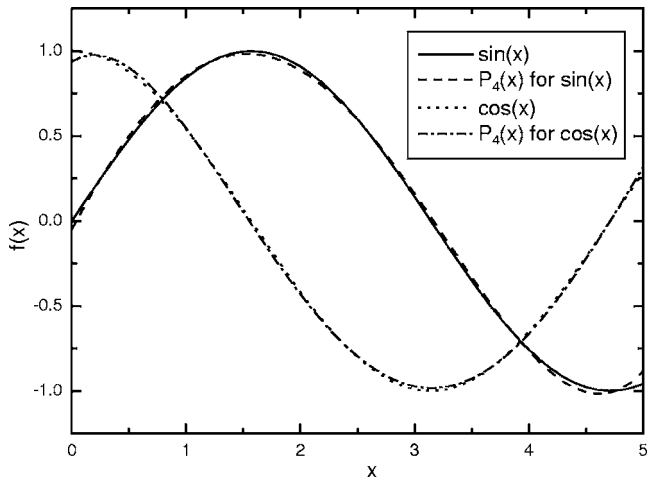


FIG. 1. Graphs of the least-squares approximating polynomials.

Alternatively, the least-squares approximating polynomial can be employed for $\sin x$ and $\cos x$ on an interval, for example, $kR \in [0,5]$. The least-squares approximating polynomial of degree four on the interval [0, 5] is

$$P_4(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4, \tag{21}$$

where $a_0 = -0.050\,901\,191\,344\,391\,56$, $a_1 = 1.266\,439\,732\,754\,480\,7$, $a_2 = -0.286\,743\,453\,443\,564\,8$, $a_3 = -0.093\,225\,327\,493\,989\,4$, and $a_4 = 0.018\,647\,117\,481\,992\,697$ for $\sin x$, and $a_0 = 0.939\,015\,875\,535\,418\,5$, $a_1 = 0.387\,333\,311\,633\,401\,8$, $a_2 = -1.074\,752\,208\,283\,267\,7$, $a_3 = 0.319\,276\,016\,231\,581\,64$, and $a_4 = -0.024\,961\,928\,915\,423\,31$ for $\cos x$, respectively. Figure 1 shows the least-squares approximating polynomials.

Using Eq. (21), the real and imaginary parts of the acoustic impedance matrix $[Z]$ can be expressed as

$$\mathrm{Re}([Z]) = \omega[Z_{R1}] + \omega^2[Z_{R2}] + \omega^3[Z_{R3}] + \omega^4[Z_{R4}] + \omega^5[Z_{R5}], \tag{22}$$

$$\mathrm{Im}([Z]) = \omega[Z_{I1}] + \omega^2[Z_{I2}] + \omega^3[Z_{I3}] + \omega^4[Z_{I4}] + \omega^5[Z_{I5}]. \tag{23}$$

The acoustic load vector can be written as

$$[G][A]\{p\} = -i[D_1]\{\ddot{x}\} - [D_2]\{\dddot{x}\} + i[D_3]\{\ddddot{x}\} + [D_4]\{\dddddot{x}\} - i[D_5]\{\ddddddot{x}\}, \tag{24}$$

where $[D_1]=[G][A]([Z_{R1}]+i[Z_{I1}])[G]^T$, $[D_2]=[G][A]([Z_{R2}]+i[Z_{I2}])[G]^T$, $[D_3]=[G][A]([Z_{R3}]+i[Z_{I3}])[G]^T$, $[D_4]=[G][A]([Z_{R4}]+i[Z_{I4}])[G]^T$, $[D_5]=[G][A]([Z_{R5}]+i[Z_{I5}])[G]^T$ are complex matrices and independent of frequency ω.

Substituting Eq. (24) into Eq. (7) and recasting the coupled system into a state-space form,

$$\begin{bmatrix} -iD_5 & 0 & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 & 0 \\ 0 & 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & 0 & I \end{bmatrix} \begin{Bmatrix} \dddddot{x} \\ \ddddot{x} \\ \dddot{x} \\ \ddot{x} \\ \dot{x} \\ x \end{Bmatrix} + \begin{bmatrix} D_4 & iD_3 & -D_2 & -iD_1+M & C & K \\ -I & 0 & 0 & 0 & 0 & 0 \\ 0 & -I & 0 & 0 & 0 & 0 \\ 0 & 0 & -I & 0 & 0 & 0 \\ 0 & 0 & 0 & -I & 0 & 0 \\ 0 & 0 & 0 & 0 & -I & 0 \end{bmatrix} \begin{Bmatrix} \dddddot{x} \\ \ddddot{x} \\ \dddot{x} \\ \ddot{x} \\ \dot{x} \\ x \end{Bmatrix} = \begin{Bmatrix} F \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{Bmatrix}. \tag{25}$$

It can be seen that $[D_i](i=1,2,3,4,5)$ matrices are complex and a complex-valued eigenvalue routine is required.

V. NUMERICAL RESULTS

Numerical example of a fluid-loaded stiffened plate is presented to demonstrate the method. The simply supported square plate (dimension $a=1$ m) of thickness $h=5$ cm, reinforced by four stiffeners of rectangular cross section, is shown in Fig. 2.[5] The depth and width of the stiffeners are $H=7.5$ cm and $W=5$ cm, respectively. The plate and stiffeners material is steel ($\rho_s=7850$ kg/m^3, $E=210.0$ GPa, $\nu=0.3$). The mass and stiffness proportional damping constants α and β are assumed to be 34.26 and 1.996×10^{-6} for the stiffened plate, respectively. The acoustic fluid is water with density $\rho=1000$ kg/m^3 and speed of sound $c=1500$ m/s. The plate is modeled by Mindlin plate elements and four stiffeners are modeled by Timoshenko beam elements and the eccentricity of the stiffeners is taken into account by a transformation that makes beam DOF "slave" to "master" DOF in the plate.[6] Although this eccentric beam modeling may introduce an

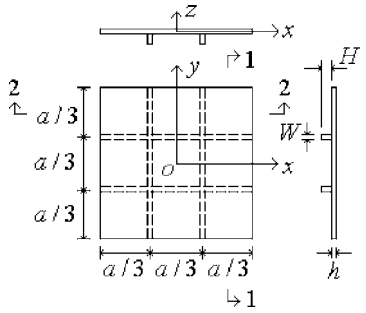


FIG. 2. Schematic of a stiffened plate.

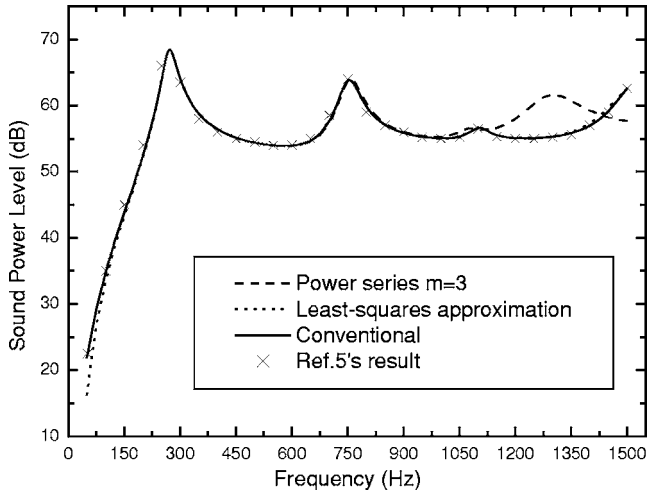


FIG. 3. Comparison of computed sound power level.

incompatibility and cause an error, Gupta and Ma[7] pointed out the error can be confined within an acceptable limit with a relatively few number of elements. The Rayleigh integral is made by discretizing the plate surface into the same mesh as plate finite element mesh.

To evaluate the goodness of the approximation of using the power series $m=3$ [Eqs. (17) and (18)] and the least-squares approximation [Eqs. (22) and (23) to express the conventional acoustic impedance matrix $[Z]$ in Eq. (6), the traditional coupling method[4] is used to compute the radiated sound power level (dB, $re:10^{-12}$ W). The excitation is a transverse point force of magnitude $F_0=1$ N located at $x_0=7.5$ cm, $y_0=7.5$ cm. The frequency is varied from 50 to 1500 Hz. The radiated power is calculated from the surface integral

$$W=\frac{1}{2}\int_S \mathrm{Re}(pv_n^*)dS, \tag{26}$$

where the asterisk $(*)$ denotes the complex conjugate. The computational result using a 12×12 mesh is presented and the comparison between the present and the Berry's results[5] is shown in Fig. 3. It can be seen that the least-squares approximation agrees well throughout the frequency range and the power series approximation looks good only for less than 1100 Hz.

Table I presents the complex eigenvalue λ_k of the stiffened plate with the given proportional damping in water using the present state-space method with the above 12×12 mesh. By analogy with the properties of a one-degree-of-freedom system, the natural frequency ω_k or f_k and modal damping ratio ζ_k are calculated as

$$\omega_k=2\pi f_k=|\lambda_k|,\quad \zeta_k=-\frac{\mathrm{Re}(\lambda_k)}{|\lambda_k|}. \tag{27}$$

The corresponding natural frequencies and modal damping ratios are also given in Table I. The water-loaded resonances in Table I are obtained using the conventional coupling method with the 12×12 mesh by a frequency sweep with increments of 1 Hz. It can be seen from Table I that the computed frequencies match the water-loaded resonances very well. To show what damping levels are being specified with the given Rayleigh damping parameters α and β for the stiffened plate, Table I lists the first four complex eigenvalues of the stiffened plate with the proportional damping *in vacuo* with the same 12×12 mesh. The *in vacuo* natural frequencies published in Ref. 5 also are shown in Table I.

TABLE I. Modal properties for the lowest four modes of the stiffened plate.

		Mode #k			
		1	2	3	4
Water-loaded (Power series $m=3$)	$\mathrm{Re}(\lambda_k)$	−115	−174	−179	−262
	$\mathrm{Im}(\lambda_k)$	1695	4729	4741	6757
	f_k (Hz)	271	753	755	1076
	ζ_k	0.0676	0.0368	0.0377	0.0387
Water-loaded (Least-squares approximation)	$\mathrm{Re}(\lambda_k)$	−113	−182	−182	−205
	$\mathrm{Im}(\lambda_k)$	1696	4715	4715	6903
	f_k (Hz)	271	751	751	1099
	ζ_k	0.0664	0.0385	0.0385	0.0297
Water-loaded resonances	f_k (Hz)	272	754	754	1102
Damped *in vacuo*	$\mathrm{Re}(\lambda_k)$	−22	−47	−47	−78
	$\mathrm{Im}(\lambda_k)$	2198	5509	5509	7814
	f_k (Hz)	350	877	877	1244
	ζ_k	0.0100	0.0086	0.0086	0.0100
In vacuo[a]	f_k (Hz)	342	883	883	1271

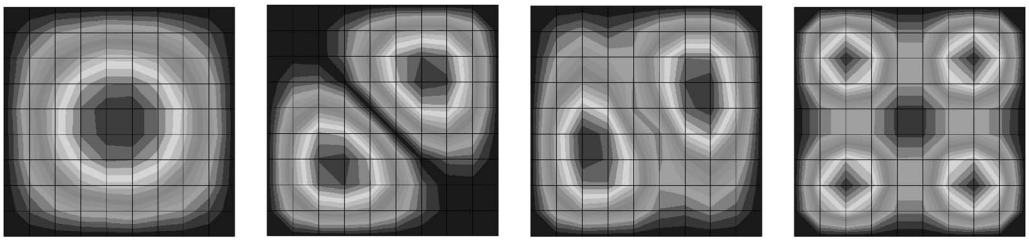
[a]Reference 5.

To show what damping levels are being caused by the radiation loading, the first four complex eigenvalues of the stiffened plate without structural damping in water is calculated with the 12×12 mesh using the least-squares approximation and the computed modal damping ratios are shown in Table II.

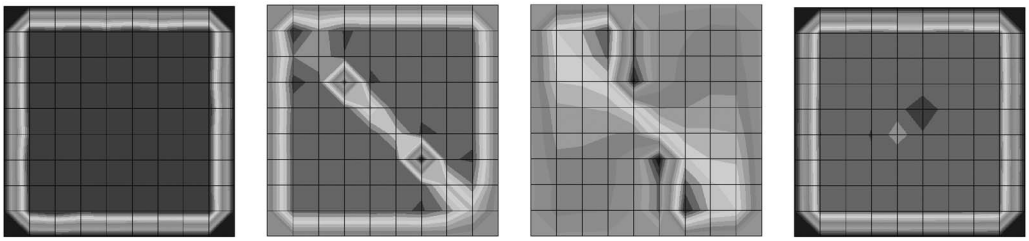
Figure 4 illustrates the water-loaded mode shapes for the first four structural modes of the stiffened plate using a 9×9 mesh with the least-squares approximation. There are surface contour plots using vibration magnitude and phase angle for the water-loaded complex modes.

TABLE II. The first four modal damping ratios caused by the radiation loading of the water-loaded stiffened plate.

		Mode #k			
		1	2	3	4
Water-loaded (Least-squares Approximation)	$\mathrm{Re}(\lambda_k)$	−100	−148	−148	−147
	$\mathrm{Im}(\lambda_k)$	1699	4720	4720	6910
	f_k (Hz)	271	752	752	1100
	ζ_k	0.0587	0.0314	0.0314	0.0213



(a) The contour plot using vibration magnitudes of the water-loaded complex modes



(b) The contour plot using phase angles of the water-loaded complex modes

FIG. 4. The mode shapes of the first four modes of the water-loaded stiffened plate.

VI. CONCLUSIONS

A state-space coupling method has been developed for the direct solution of fluid-loaded natural frequencies and mode shapes of plates. This method expands the frequency-dependent term in the Rayleigh integral in a power series on circular frequency. After factoring out the frequency term from the integrands, numerical integration involved in setting up the acoustic impedance coefficient matrices is confined to a frequency-independent part. The acoustic impedance coefficient matrices are therefore frequency-independent and can be coupled directly with the structural matrix into a canonical state-space form to yield the fluid-loaded modes via the direct eigenvalue analysis. Numerical results for a fluid-loaded stiffened plate are presented to demonstrate the method. The fluid-loaded natural frequencies, mode shapes, and modal damping ratios are given to show the efficacy of the state-space coupling method.

ACKNOWLEDGMENT

The author is grateful for the support of the National Natural Science Foundation of China. (No. 10402004).

[1]J. A. Giordano and G. H. Koopmann, "State-space boundary element-finite element coupling for fluid-structure interaction analysis," J. Acoust. Soc. Am. **98**, 363–372 (1995).
[2]K. A. Cunefare and S. De Rosa, "An improved state-space method for coupled fluid-structure interaction analysis," J. Acoust. Soc. Am. **105**, 206–210 (1999).
[3]F. Fahy, *Sound and Structural Vibration: Radiation, Transmission and Response* (Academic, London, 1985).
[4]G. C. Everstine and F. M. Henderson, "Coupled finite element/boundary element approach for fluid-structure interaction," J. Acoust. Soc. Am. **87**, 1938–1947 (1990).
[5]A. Berry and C. Locqueteau, "Vibration and sound radiation of fluid-loaded stiffened plates with consideration of in-plane deformation," J. Acoust. Soc. Am. **100**, 312–319 (1996).
[6]R. D. Cook, S. S. Malkus, and M. E. Plesha, *Concepts and Applications of Finite Element Analysis* (Wiley, New York, 1989).
[7]A. K. Gupta and P. S. Ma, "Error in eccentric beam formulation," Int. J. Numer. Methods Eng. **11**, 1473–1477 (1977).

Time domain computational modeling of viscothermal acoustic propagation in catalytic converter substrates with porous walls

N. S. Dickey and A. Selamet[a)]
Department of Mechanical Engineering, The Ohio State University, Columbus, Ohio 43212

K. D. Miazgowicz, K. V. Tallio, and S. J. Parks
Ford Motor Company, Dearborn, Michigan 48124

(Received 4 August 2004; revised 26 May 2005; accepted 29 May 2005)

Models for viscothermal effects in catalytic converter substrates are developed for time domain computational methods. The models are suitable for use in one-dimensional approaches for the prediction of exhaust system performance (engine tuning characteristics) and radiated sound levels. Starting with the "low reduced frequency" equations for viscothermal acoustic propagation in capillary tubes, time domain submodels are developed for the frequency-dependent wall friction, frequency-dependent wall heat transfer, and porous wall effects exhibited by catalytic converter substrates. Results from a time domain computational approach employing these submodels are compared to available analytical solutions for the low reduced frequency equations. The computational results are shown to agree well with the analytical solutions for capillary geometries representative of automotive catalytic converter substrates. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1974759]

PACS number(s): 43.50.Gf, 43.20.Mv [ANN] Pages: 806–817

I. INTRODUCTION

Catalytic converters (CCs) are a necessary component for reducing pollutant emissions of spark-ignited automotive engines. Significant effort goes into the design and placement strategy of the CC to ensure fast light off, adequate pollutant conversion, and long life of the substrate. While the functional purpose of the CC is to reduce emissions, its presence inevitably affects the breathing characteristics of the exhaust system. A well-known consequence of CCs is an increase in backpressure that increases pumping losses and reduces peak power. In addition to the backpressure increase, which is largely related to the steady flow behavior of the CC, the unsteady flow characteristics of the CC can also be important. The reflection and dissipation of pressure waves by the CC contribute to radiated sound as well as the wave dynamics and tuning characteristics of the exhaust system. The importance of the wave dynamics of the CC has increased with the use of variable valve timing (VVT) and close-coupled catalytic converters. Variable valve timing allows engine designers to take greater advantage of exhaust system tuning (which can adversely affect idle and low-speed behavior without VVT). This, and the use of close-coupled (or light-off) CCs that are located close to the exhaust port mean that the CC can significantly influence engine performance by affecting the tuning characteristics of the exhaust system.

Since the CCs can influence the flow performance, tuning characteristics, and acoustic behavior of an exhaust system, representative CC models should be included in predictive tools for engine performance (wave dynamics) and radiated sound levels. The most commonly used predictive tools that are capable of predicting both engine performance and radiated sound levels are one-dimensional time domain approaches. For these models, the wave dynamics and acoustics at low to moderate frequencies (up to about 1500 Hz) are of primary interest. Also, since model complexity increases computational costs, and optimization studies typically require a large number of simulations, it is desirable to eliminate as many details as possible (chemical kinetics, for example). The goal is thus to employ the simplest model that can adequately describe the unsteady flow characteristics of the CC.

A typical CC uses a ceramic catalyst substrate monolith that has from 300 to 900 cells/in.^{2}, with each cell representing a parallel tube, or capillary path, for fluid flow (Fig. 1). Though a variety of cell cross sections are possible, square cells are common. The ceramic substrate is covered with a washcoat, upon which the catalytically active metals reside. To allow the flow to more effectively use the substrate cross-sectional area, and also reduce flow losses, the outer housing has transition sections at each end of the monolith (upstream diffuser and downstream reducer). The substrate is mounted in its housing with a swelling mat or wire mesh around its periphery to provide holding pressure. If a wire mesh is used, a parallel acoustic path may exist around the substrate (depending on the density of the mesh). For CCs with a swelling mat, however, a previous study suggests that for the frequencies of interest, the mat may reasonably be considered as rigid and impermeable.[1] In this case, the acoustic behavior of the CC is determined by the housing geometry and the substrate behavior.

[a)]Electronic mail: selamet.1@osu.edu

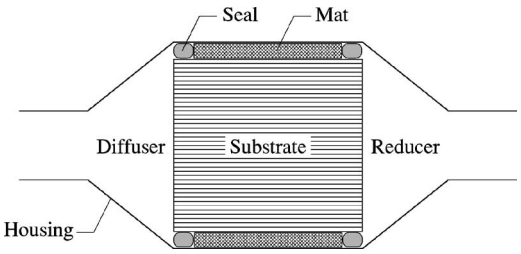


FIG. 1. Basic geometry of catalytic converter assembly.

The substrate in a CC assembly may be analyzed through a linear frequency domain analysis of laminar compressible flow in a bundle of narrow tubes or capillaries. Early works by Kirchhoff[2] and Rayleigh[3] considered the propagation of sound waves in cylindrical tubes with zero mean flow. Numerous subsequent workers have developed approximate analytical and numerical solutions for the same fundamental equations (see, for example, the summaries by Tijdeman[4] and Beltman[5]). More recently, the interest in practical applications of CCs and thermoacoustic devices has led to works considering more realistic (and complex) conditions with mean flow, axial temperature gradients, and porosity of the cell walls in the substrate.[6–11] These studies have shown that viscothermal effects decrease the speed of propagation of a pressure disturbance (phase speed) and increase the amount that the disturbance is attenuated as it propagates (attenuation rate). Moreover, the phase speed and attenuation rate are frequency dependent, so the different frequency components of a complex, multifrequency wave will propagate and decay at different rates.

While a number of frequency-domain approaches are available, relatively little work has been done on time domain approaches suitable for one-dimensional engine simulation models. One reasonably simple approach is to employ a frequency-dependent wall friction model. This approach is common in the analysis of pressure pulsations in liquid-filled pipes. Trikha[12] has presented an approach that efficiently applies the frequency-dependent friction term from the earlier analytical work of Zielke.[13] Other approaches have been presented as well, though none appears to be clearly superior for the frequency range and flow conditions of interest in this study.[14,15] Since the approaches used in waterhammer studies are based on incompressible flow, the viscous boundary layer is accounted for, but the thermal boundary layer is neglected. In addition, the work of Arnott *et al.*[6] indicates that small pores in the ceramic substrate material (porous walls of the capillaries) may be important.

The objective of this study is to develop improved methods to model the acoustics and wave dynamics of CCs with one-dimensional time domain engine simulation tools. The linear "low reduced frequency" equations for capillary acoustics are employed to develop submodels for the frequency-dependent viscothermal and porous wall effects that are suitable for inclusion in a time domain computational model. The current viscothermal submodels are developed for capillaries of circular cross section, though the general approach may be applied to noncircular capillaries. Results from the time domain computational model are obtained for small-amplitude wave propagation in circular capillaries with zero mean flow. The computational results are compared to available analytical solutions of the low reduced frequency equations for capillaries with circular and square cross sections.

II. ACOUSTIC PROPAGATION IN A CAPILLARY TUBE

In the following, equations will be presented for acoustic propagation in capillaries of constant cross section. The basic assumptions applied are that perturbations to the flow are small in comparison with the mean, and there is no mean flow in the system. For the capillary sizes and frequencies of interest, the fundamental equations simplify to the low reduced frequency equations.[4] To investigate the accuracy of the time domain submodels, and to illustrate the relative importance of the factors contributing to the acoustic behavior of CCs, the computational results will be compared to analytical solutions for both circular and rectangular capillary cross sections under the following conditions: (1) nonporous adiabatic walls, (2) nonporous isothermal walls, and (3) porous isothermal walls. The analytical solutions for the low reduced frequency equations considering each of these factors (or combinations thereof) are available in existing works.[2–7,16,17] For convenience, and consistency of notation, solutions for the conditions of interest are presented here without derivation.

A. The low reduced frequency equations

For the propagation of small disturbances, the linearized forms of the continuity, momentum (Navier-Stokes), and energy equations apply. With zero mean flow, and using the ideal gas equation of state to couple the thermodynamic variables, the linearized equations may be expressed as[2,3]

$$\frac{\partial \rho_t}{\partial t} + \rho_0 \nabla \cdot \boldsymbol{V} = 0, \tag{1}$$

$$\rho_0 \frac{\partial \boldsymbol{V}}{\partial t} + \nabla P - \frac{4\mu}{3} \nabla (\nabla \cdot \boldsymbol{V}) + \mu \nabla \times (\nabla \times \boldsymbol{V}) = 0, \tag{2}$$

$$\rho_0 c_p \frac{\partial T_t}{\partial t} - \lambda_T \nabla^2 T_t - \frac{\partial P}{\partial t} = 0, \tag{3}$$

$$P = \rho_t R_0 T_t, \tag{4}$$

where ρ is density, V is velocity, P is pressure, T is temperature, μ is absolute viscosity, c_p is specific heat at constant pressure, λ_T is thermal conductivity, R_0 is the gas constant, and a variable subscripted 0 denotes a mean or reference value. For the density and temperature, the subscript t is used to indicate a "total" value (mean plus acoustic perturbation).

Equations (1)–(4) have been solved for small harmonic perturbations in circular capillaries by Kirchhoff.[2] For a particular fluid and capillary geometry, the viscothermal wave propagation may be obtained in terms of the reduced frequency $k=\omega R/c_0$ and a dimensionless shear wave number $s=R\sqrt{\omega/\nu}$, where $\omega=2\pi f$ is the angular frequency, f is the frequency, R is the capillary radius (or transverse length

scale) and $\nu=\mu/\rho_0$ is the kinematic viscosity. While the full Kirchhoff solution results in a transcendental equation which must be solved numerically, analytical solutions are available for certain limiting cases of practical interest.[2–4,16] For the conditions in the channels of a catalytic converter substrate, it can be assumed that $k \ll 1$ and $k/s \ll 1$, and the expressions simplify to the low reduced frequency equations initially presented by Zwikker and Kosten[18] and subsequently used in a number of analyses (for more complete discussions of the low reduced frequency equations see, for example, the works of Tijdeman,[4] Stinson,[17] Arnott *et al.*,[6] and Beltman[5]). Noting that the geometry and physics of the problem suggest separate length scales, the vector variables and operations for the axial and cross directions are treated separately. The acoustic variables are introduced as perturbations from the mean using

$$\begin{aligned} P &= p_0 + p, \\ \boldsymbol{V} &= \boldsymbol{v}_0 + \boldsymbol{v} = \boldsymbol{v} = u + \boldsymbol{v}_c, \\ \rho_t &= \rho_0 + \rho, \\ T_t &= T_0 + T, \end{aligned} \tag{5}$$

where u is the axial, or x-direction velocity and the subscript c denotes directions in the plane of the tube cross section. Equations (1)–(4) can then be simplified and rearranged to yield the low reduced frequency equations as[6,7,17] (see Appendix A for details of the coordinate systems and vector operations).

$$\frac{\partial \rho}{\partial t} + \rho_0 \frac{\partial u}{\partial x} + \rho_0 \nabla_c \cdot \boldsymbol{v}_c = 0, \tag{6}$$

$$\nabla_c p = 0, \tag{7}$$

$$\rho_0 \frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} - \mu \nabla_c^2 u = 0, \tag{8}$$

$$\rho_0 c_p \frac{\partial T}{\partial t} - \lambda_T \nabla_c^2 T - \frac{\partial p}{\partial t} = 0, \tag{9}$$

$$\frac{p}{p_0} = \frac{\rho}{\rho_0} + \frac{T}{T_0}. \tag{10}$$

From Eq. (7), it can be seen that the pressure is constant over the capillary cross section. The other acoustic variables may vary over the cross section (depending on the wall boundary conditions), but are readily formulated in terms of area-averaged mean values for consistency with plane wave analyses.

B. Analytical harmonic solution

With the assumption of harmonic ($e^{i\omega t}$) time dependence in Eqs. (6)–(9), the low reduced frequency equations for mass, axial momentum and energy conservation may be expressed as

$$i\omega\rho + \rho_0 \frac{\partial u}{\partial x} + \rho_0 \nabla_c \cdot \boldsymbol{v}_c = 0, \tag{11}$$

$$i\omega\rho_0 u + \frac{\partial p}{\partial x} - \mu \nabla_c^2 u = 0, \tag{12}$$

$$i\omega\rho_0 c_p T - \lambda_T \nabla_c^2 T - i\omega p = 0. \tag{13}$$

The solution of Eqs. (11)–(13) for various geometries and wall boundary conditions has been discussed in previous works and will not be elaborated fully here. Additional details and summaries of previous works are available in the literature.[4–7,17]

The pressure and mean velocity (u_m) solutions for the low reduced frequency equations can be expressed as

$$p = C_+ e^{-\Gamma k_0 x} + C_- e^{\Gamma k_0 x}, \tag{14}$$

$$u_m(x) = i \frac{\Gamma \mathcal{F}(s)}{\rho_0 c_0} [C_+ e^{-\Gamma k_0 x} - C_- e^{\Gamma k_0 x}], \tag{15}$$

where

$$\Gamma = \left[\frac{[\gamma + (\gamma - 1)\mathcal{F}(s)]}{\mathcal{G}(s\sigma)} - i \frac{\mathcal{P}}{k Z_w A_c} \frac{1}{\mathcal{G}(s\sigma)} \right]^{1/2} \tag{16}$$

is the propagation constant, C denotes an amplitude constant, the subscripts + and − represent disturbances moving in the positive and negative directions, respectively, $k_0=\omega/c_0$ is the acoustic wave number, γ is the ratio of specific heats, $\sigma = \sqrt{\mu c_p/\lambda_T}$ is the square root of the Prandtl number, $\mathcal{P}$ is the capillary perimeter, A_c is the capillary cross-sectional area, and $Z_w = p/\rho_0 c_0 \boldsymbol{v}_{c,\text{wall}} \cdot \boldsymbol{n}_{\text{wall}}$ is the normalized wall impedance of the capillary walls. The functions $\mathcal{F}(s)$ and $\mathcal{G}(s\sigma)$ result from the variation of velocity and temperature over a capillary cross section, respectively. Therefore, they depend on the cross-sectional capillary geometry and the velocity and temperature boundary conditions imposed at the capillary wall. Combination of Eqs. (14) and (15) allows a normalized characteristic impedance of the capillary to be computed as

$$Z = \frac{p_+}{\rho_0 c_0 u_{+,m}} = -\frac{i}{\Gamma \mathcal{F}(s)}. \tag{17}$$

The solution given by Eqs. (14)–(16) depends on fluid properties, capillary geometry, wall boundary conditions, and the capillary wall impedance. The present study considers analytical solutions for air at atmospheric conditions in capillaries with circular and square cross section. The wall boundary condition for the axial velocity is given by the no-slip condition. For the temperature boundary condition, the gas temperature can be assumed equal to the wall temperature (isothermal walls), since the heat capacity of the gas is much less than that of the substrate material. While the assumption of isothermal walls is typically appropriate, adiabatic (insulated) walls are also considered to investigate the relative importance of wall heat transfer. The impedance of porous capillary walls is specified following the work of Arnott *et al.*[6] where the individual pores in the substrate walls are treated as small, rigidly terminated capillaries having av-

TABLE I. Analytical solutions for circular and square cross sections.

	Circular cross section	Rectangular cross section
$\mathcal{F}(s)$	$J_2(i^{3/2}s)/J_0(i^{3/2}s)$	$(-4is^2/\psi^2)\Sigma_{m=0}^{\infty}\Sigma_{n=0}^{\infty}[1/\alpha_m^2\beta_n^2(\alpha_m^2+\beta_n^2+is^2)]$
$\mathcal{G}(s\sigma)$ (isothermal walls)	$\mathcal{F}(s\sigma)$	$\mathcal{F}(s\sigma)$
$\mathcal{G}(s\sigma)$ (adiabatic walls)	-1	-1

erage pore radius R_w and length l_w. For wall porosity (open area ratio) Ω_w, the impedance of the wall pores can be approximated as[6]

$$Z_w = -\frac{i}{\Omega_w \gamma k_0 l_w}. \tag{18}$$

Equation (18) represents a limiting case of viscothermal propagation for small k, s, and l_w, and corresponds to isothermal compression of a lumped volume of fluid in each wall pore.

Solutions for the functions $\mathcal{F}(s)$ and $\mathcal{G}(s\sigma)$ for circular and rectangular cross sections are summarized in Table I where, for the rectangular cross section, ψ is the aspect ratio of the capillary, $\alpha_m=(2m+1)\pi/2$, and $\beta_n=(2n+1)\pi/2\psi$. With the additional specification of wall impedance, the expressions in Table I allow the capillary propagation constant and characteristic impedance to be determined from Eqs. (16) and (17), respectively.

III. TIME DOMAIN COMPUTATIONS

The time domain computational techniques of interest in this study are based on the nonlinear equations for one-dimensional flow in ducts of variable cross section. Since the present study investigates model accuracy by comparing to linear analytical solutions, the nonlinear terms could be omitted here without consequence. However, the model is expected to be useful in engine simulation studies where these terms are not negligible, so the full forms of the equations are retained. The balance equations for mass, momentum, and energy may be expressed, respectively, as

$$\frac{\partial}{\partial t}(\rho_t A) + \frac{\partial}{\partial x}(\rho_t(AU + A_p U_p)) = 0, \tag{19}$$

$$\frac{\partial}{\partial t}(\rho_t AU) + \frac{\partial}{\partial x}(\rho_t AU^2) + \frac{\partial}{\partial x}(PA) - \tau_w \mathcal{P} = 0, \tag{20}$$

$$\frac{\partial}{\partial t}(\rho_t e) + \frac{\partial}{\partial x}(\rho_t e(AU + A_p U_p)) + P\frac{\partial}{\partial x}(UA + U_p A_p) - \tau_w \mathcal{P} U + q_w \mathcal{P} = 0, \tag{21}$$

where U is the velocity, τ_w is the wall shear stress, $\mathcal{P}$ is the perimeter, e is the internal energy, q is the wall heat transfer rate, and the subscript p denotes wall perforations (typically employed with respect to perforated tube silencers). The ideal gas equation of state,

$$P = (\gamma - 1)\rho_t e, \tag{22}$$

is used to relate the thermodynamic variables and close the system of equations. Note that since the fluid motion is assumed to be one-dimensional, all flow variables in Eqs. (19)–(22) represent mean values over a cross section.

Typically, the wall shear stress and heat transfer are obtained from steady flow relationships. For laminar flow, the steady wall shear and heat transfer may be calculated as

$$\tau_w = \frac{4\mu}{R} U, \tag{23}$$

$$q_w = 1.83\frac{\lambda_T}{R}(T_t - T_{t,\text{wall}}). \tag{24}$$

Since these relationships are based on steady flow, they will only apply in the limit as frequency approaches zero. In the following, expressions will be developed to implement frequency-dependent wall friction and heat transfer in the one-dimensional treatment. The approach is analogous to the method used by Zielke,[13] who analyzed oscillating incompressible flow in a conduit to obtain a frequency-dependent friction relationship for waterhammer analyses. In the present study, the low reduced frequency equations are employed to develop expressions for frequency-dependent wall shear and heat transfer. The study considers cylindrical capillaries only, although the basic approach can be applied to other geometries as well. After the expressions are developed, an approximation technique presented by Trikha[12] is employed to reduce computational requirements.

In addition to frequency-dependent wall shear and heat transfer, an approach is developed to account for the effects of wall porosity. While a direct method could be used to seek modified versions of the conservation equations for porous walls, this might require substantial modifications to an existing code or numerical scheme. However, as indicated by Eqs. (19)–(21), most engine simulation codes already include a structure to account for the porous walls of perforated tube silencers.[19,20] To take advantage of this fact, an approach is developed that allows porous wall substrates to be modeled with slight modifications to the typical treatment of an analogous perforated tube silencer.

A. Frequency-dependent wall friction

For a cylindrical capillary, the low reduced frequency momentum equation [Eq. (8)] can be expressed as

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} - \frac{1}{\nu}\frac{\partial u}{\partial t} - \frac{1}{\mu}\frac{\partial p}{\partial x} = 0. \tag{25}$$

With the assumption that the fluid is initially at rest, taking the Laplace transform of Eq. (25) gives

$$\frac{\partial^2 \hat{u}}{\partial r^2} + \frac{1}{r}\frac{\partial \hat{u}}{\partial r} - \frac{w}{\nu}\hat{u} - \frac{1}{\mu}\frac{\partial \hat{p}}{\partial x} = 0, \tag{26}$$

where a caret denotes a Laplace transformed variable as $\hat{g}(w) = \mathcal{L}(g(t))$. Applying boundary conditions of $u=0$ at the tube wall ($r=R$) and boundedness at $r=0$ yields the solution to Eq. (26) as

$$\hat{u}(r,w) = \frac{1}{\rho_0 w}\frac{\partial \hat{p}}{\partial x}\left[\frac{J_0\left(i\sqrt{\frac{w}{\nu}}r\right)}{J_0\left(i\sqrt{\frac{w}{\nu}}R\right)} - 1\right]. \tag{27}$$

Integrating Eq. (27) over the cross section of the tube allows the transform of the mean duct velocity to be determined as

$$\hat{u}_m(w) = \frac{1}{\pi R^2}\int_0^R 2\pi r \hat{u} dr = \frac{1}{\rho_0 w}\frac{\partial \hat{p}}{\partial x}\left[\frac{J_2\left(i\sqrt{\frac{w}{\nu}}R\right)}{J_0\left(i\sqrt{\frac{w}{\nu}}R\right)}\right]. \tag{28}$$

The transform of the wall shear stress is then

$$\hat{\tau}_w = -\mu\left.\frac{\partial \hat{u}}{\partial r}\right|_{r=R} = \frac{R}{2}\frac{\partial \hat{p}}{\partial x}\left[\frac{J_2\left(i\sqrt{\frac{w}{\nu}}R\right)}{J_0\left(i\sqrt{\frac{w}{\nu}}R\right)} + 1\right]. \tag{29}$$

From Eqs. (28) and (29), the wall shear stress can be determined from the mean velocity as

$$\hat{\tau}_w(w) = \frac{\rho_0 R w}{2}\left[\frac{J_0\left(i\sqrt{\frac{w}{\nu}}R\right)}{J_2\left(i\sqrt{\frac{w}{\nu}}R\right)} + 1\right]\hat{u}_m(w). \tag{30}$$

Since Eq. (30) does not approach zero as w approaches infinity, the expression is unsuitable for inversion. This difficulty may be avoided by expressing $\hat{\tau}_w$ in terms of the transform of the acceleration (rate of change of the *mean* velocity). The acceleration is given by

$$\frac{\partial \hat{u}_m}{\partial t}(w) = w\hat{u}_m(w) = \frac{1}{\rho_0}\frac{\partial \hat{p}}{\partial x}\left[\frac{J_2\left(i\sqrt{\frac{w}{\nu}}R\right)}{J_0\left(i\sqrt{\frac{w}{\nu}}R\right)}\right] \tag{31}$$

and the transform of the wall shear stress can be expressed as

$$\hat{\tau}_w(w) = \frac{\rho_0 R}{2}\left[\frac{J_0\left(i\sqrt{\frac{w}{\nu}}R\right)}{J_2\left(i\sqrt{\frac{w}{\nu}}R\right)} + 1\right]\frac{\partial \hat{u}_m}{\partial t}(w). \tag{32}$$

The inverse transformation of Eq. (32) gives (see Ref. 13)

$$\tau_w(t) = \frac{4\mu}{R}\left[U(t) + \frac{1}{2}\int_0^t \frac{\partial U(\phi)}{\partial t}W(t-\phi)d\phi\right], \tag{33}$$

where $U=u_m$ and $W(t)$ is a function of the dimensionless time variable $\tilde{t} = \nu t/R^2$ given by

$$W(t) = W'(\tilde{t}) = e^{-26.3744\tilde{t}} + e^{-70.8493\tilde{t}} + e^{-135.0198\tilde{t}} + e^{-218.9216\tilde{t}} + e^{-322.5544\tilde{t}} + \cdots. \tag{34}$$

For small values of $\tilde{t}$, Eq. (34) converges very slowly, so an alternate expression is used as[13]

$$W(t) = W'(\tilde{t}) = 0.282\,095\tilde{t}^{-1/2} - 1.250\,000 + 1.057\,855\tilde{t}^{1/2} + 0.937\,500\tilde{t} + 0.396\,696\tilde{t}^{3/2} - 0.351\,563\tilde{t}^2\ldots, \quad \tilde{t} < 0.02. \tag{35}$$

Inspection of Eq. (33) shows that the instantaneous wall shear stress is equal to the steady state value plus a term that is dependent on the past history of the velocity changes. This equation is suitable for inclusion in a time domain modeling approach. However, the form of the integral in Eq. (33) would require storage of and computations on a large number of time intervals in the past. To reduce the computational effort, an approximate approach is used that greatly reduces the required storage and number of calculations.

The frequency-dependent wall shear is approximated using the approach of Trikha.[12] The basis of the method is to approximate $W'(\tilde{t})$ with a curve-fit expression as

$$W'(\tilde{t}) \cong W'_{\text{app}}(\tilde{t}) = \sum_{i=1}^{n} W'_i(\tilde{t}), \tag{36}$$

where the functions $W'_i(\tilde{t})$ are specified as

$$W'_i(\tilde{t}) = a_i e^{-b_i\tilde{t}}, \tag{37}$$

where a_i and b_i are constants determined in the curve fit of $W'_{\text{app}}(\tilde{t})$ to Eqs. (34) and (35). With the introduction of

$$y_i(t) = \int_0^t W_i(t-\phi)\frac{\partial U(\phi)}{\partial t}d\phi, \tag{38}$$

where $W_i(t) = W'_i(\tilde{t})$, Eq. (33) can be expressed as

$$\tau_w(t) \cong \frac{4\mu}{R}\left[U(t) + \frac{1}{2}\sum_{i=1}^{N} y_i(t)\right]. \tag{39}$$

For a small computational time step Δt, the approximation of $W'(\tilde{t})$ by a summation of exponential functions allows $y_i(t+\Delta t)$ to be determined from $y_i(t)$ as (see Appendix B)

$$y_i(t+\Delta t) = y_i(t)e^{-b_i(\nu/R^2)\Delta t} + a_i(U(t+\Delta t) - U(t)). \tag{40}$$

Equation (40) provides a relationship between $y_i(t+\Delta t)$ and $y_i(t)$ which, combined with Eq. (39), is suitable for use in a time domain computational scheme. Note that the number of additional variables that need to be stored at each computational cell is equal to N, the number of terms in the curve fit of Eq. (36). Moreover, relatively simple expressions for the y_i values have replaced time-consuming numerical integrations of stored values.

B. Frequency-dependent wall heat transfer

Development of the time domain approach for frequency-dependent wall heat transfer is similar to the approach for frequency-dependent wall shear stress. Note that the adiabatic wall boundary condition is equivalent to simply setting $q_w=0$, so only the case of isothermal walls requires consideration.

The low reduced frequency energy equation for a cylindrical capillary may be expressed as

$$\frac{\partial^2 T}{\partial r^2}+\frac{1}{r}\frac{\partial T}{\partial r}-\frac{\rho_0 c_p}{\lambda_T}\frac{\partial T}{\partial t}-\frac{1}{\lambda_T}\frac{\partial p}{\partial t}=0. \tag{41}$$

Taking the Laplace transform of Eq. (41) and solving for isothermal wall boundary conditions yields

$$\hat{T}(r,w)=\frac{1}{\rho_0 c_p}\hat{p}\left[\frac{J_0\left(i\sigma\sqrt{\frac{w}{\nu}}r\right)}{J_0\left(i\sigma\sqrt{\frac{w}{\nu}}R\right)}-1\right], \tag{42}$$

and the mean duct temperature is

$$\hat{T}_m(w)=\frac{1}{\pi R^2}\int_0^R 2\pi r\hat{T}dr=\frac{1}{\rho_0 c_p}\hat{p}\left[\frac{J_2\left(i\sigma\sqrt{\frac{w}{\nu}}R\right)}{J_0\left(i\sigma\sqrt{\frac{w}{\nu}}R\right)}\right]. \tag{43}$$

The wall heat transfer can then be related to the rate of change of the mean temperature as

$$\hat{q}_w=-\lambda_T\left.\frac{\partial\hat{T}}{\partial r}\right|_{r=R}=\frac{\rho_0 c_p R}{2}\left[\frac{J_0\left(i\sigma\sqrt{\frac{w}{\nu}}R\right)}{J_2\left(i\sqrt{\frac{w}{\nu}}R\right)}+1\right]\frac{\partial\hat{T}_m}{\partial t}(w). \tag{44}$$

The inverse transformation of Eq. (44) gives

$$q_w(t)=\frac{4\lambda_T}{R}\left[T(t)+\frac{1}{2}\int_0^t\frac{\partial T}{\partial t}W(t-\phi)d\phi\right], \tag{45}$$

where $W(t)$ is the same function given by Eqs. (34) and (35) but with the the dimensionless time variable now given by $\tilde{t}=\lambda_T t/\rho_0 c_p R^2$. The approximate approach to numerically integrate Eq. (45) is identical to that for the wall shear stress.

C. Wall porosity

The porous wall effects are incorporated in a manner similar to a common approach for the treatment of perforated tube silencers. As depicted in Fig. 2, the substrate is represented by two ducts that communicate through a perforated interface. The main duct represents the open area of the capillary channels, while the smaller secondary duct represents the open area of the porous walls. The computational treatment of the main duct is identical to that of a perforated tube silencer (with the additional inclusion of frequency-dependent wall shear and heat transfer). Simple modifications are made to the secondary duct treatment which, for the frequency range of interest, yields a wall impedance equivalent to Eq. (18).

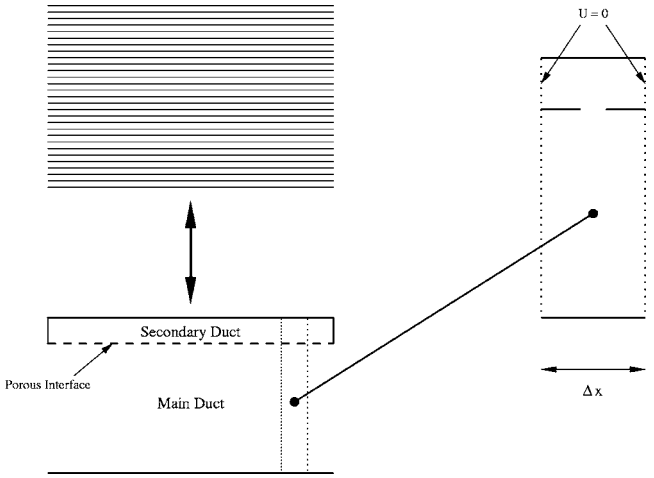


FIG. 2. Representation of substrate with porous walls.

In the computational approach, all wall pores in a computational cell of length Δx are treated as a single lumped volume (see Fig. 2). The cross-sectional area of the secondary duct is determined from the total pore volume as

$$N_c\Omega_w\mathcal{P}\Delta x l_w=A_s\Delta x, \tag{46}$$

where N_c is the total number of capillaries in the substrate, A_s is the cross-sectional area of the secondary duct, and Δx is the computational cell size. Since each wall pore is assumed to be isolated, there is no axial fluid motion in the substrate and the the momentum equation for the secondary duct is simply replaced with

$$U=0. \tag{47}$$

Axial conduction is neglected in Eqs. (19)–(21), so the effect of Eq. (47) is to completely eliminate axial mass and energy transfer in the secondary duct. Inspection of Eq. (18) reveals that the wall impedance is purely capacitive and corresponds to isothermal compression of the fluid in the wall pores. The energy equation for the secondary duct is therefore replaced with

$$\frac{\partial e}{\partial t}=0. \tag{48}$$

Consistent with the approach for perforated tube silencers, communication between the main and secondary ducts is treated using an "interface" momentum equation given by[20,21]

$$\rho_t l_{\mathrm{eq}}\frac{dU_p}{dt}+\mathcal{R}U_p-(p_i-p_0)=0, \tag{49}$$

where, for a perforated interface, l_{eq} is an "equivalent length," $\mathcal{R}$ is the resistance, and the subscripts i and 0 denote the main and secondary duct, respectively. For the porous wall interface, l_{eq} is specified as $0.85R_w$ (the Rayleigh end correction) and losses are neglected with $\mathcal{R}=0$. The computational approach requires the interface equation in order to couple the primary and secondary ducts.

However, in comparison to the analytical approach, the interface equation introduces an additional impedance in series with the lumped volume of the secondary duct. In essence, at each computational cell, the interface/volume combination is a small Helmholtz resonator attached to the main duct. For harmonic motion, the wall impedance due to this combination is found as

$$Z_w = -\frac{i}{\gamma k_0 l_w}(1 - \gamma k_0^2 l_w l_{\mathrm{eq}}) \tag{50}$$

which, since $k_0 l_w \ll 1$ and $k_0 l_{\mathrm{eq}} \ll 1$, reduces to the wall impedance given by Eq. (17).

D. Numerical implementation

The foregoing frequency-dependent friction, frequency-dependent heat transfer, and porous wall models have been incorporated in the numerical approach of Chapman *et al.*[22] In the computational approach, the frequency-dependent models use four terms in the curve fit expression for $W'_{\mathrm{app}}(\tilde{t})$. The curve fit expression is given by

$$W'_{\mathrm{app}}(\tilde{t}) = 27.271e^{-8960.2\tilde{t}} + 10.883e^{-1540.3\tilde{t}} + 4.5983e^{-266.33\tilde{t}} + 2.0e^{-39.871\tilde{t}}, \tag{51}$$

or, in terms of the coefficients,

$$a_1 = 27.271, \quad a_2 = 10.883, \quad a_3 = 4.5983, \quad a_4 = 2.0,$$

$$b_1 = 8960.2, \quad b_2 = 1540.3, \quad b_3 = 266.33,$$

$$b_4 = 39.871.$$

IV. RESULTS

In the following, results from the computational approach are compared to the analytical solutions of the low reduced frequency equations. The computational results (based on the circular capillary formulations of Sec. III) are compared to analytical solutions for both circular and square capillaries. Results are first presented for basic capillary propagation characteristics (attenuation rate, phase speed, and impedance). To more closely represent the behavior of a CC substrate, comparisons are then made for the acoustic reflection and transmission characteristics of a representative substrate placed within an anechoically terminated duct. For the results presented here, the dimensions of the capillaries are taken from a substrate with square capillaries having 400 cells/in.^{2}. Wall porosities and characteristic dimensions of the pores within the wall are based on the substrate measured by Arnott *et al.*, though the length of the cell pores has been reduced somewhat to account for smaller wall thickness. The parameters for the substrate and fluid (air) are listed in Table II.

The general propagation characteristics of the substrate capillaries are presented in terms of the attenuation rate, normalized phase speed, and normalized impedance of a propagating wave. The attenuation rate is determined from the real part of the propagation constant as

TABLE II. Substrate dimensions and characteristic parameters.

Capillary cross section	Circular, square
Capillary radius/semiwidth	R=0.55 mm
Substrate axial length	L_s=152.4 mm
Substrate open area ratio	Ω=0.757
Wall porosity	Ω_w=0, 0.490
Wall pore diameter	d_w=0.1 mm
Wall pore depth	l_w=0.081 mm
Speed of sound	c_0=344 m/s
Fluid density	ρ_0=1.2 kg/m^3
Viscosity	μ=18.5$\times 10^{-6}$ N s/m^2
Specific heat ratio	γ=1.4
Prandtl number	Pr=0.707

$$\text{Attenuation Rate (dB/axial unit)} = -20\log_{10}e^{-\Re(\Gamma k_0)}, \tag{52}$$

where the "axial unit" is determined by the units of k_0. The normalized phase speed (c_P) is obtained from the imaginary part of the propagation constant as

$$\frac{c_P}{c_0} = \frac{1}{\Im(\Gamma)}, \tag{53}$$

and the normalized characteristic impedance of the fluid is given by Eq. (17). Note that in Eqs. (53) and (17), the phase speed and characteristic impedance are normalized by the limiting case for inviscid, adiabatic wave propagation. Therefore, as viscothermal effects become negligible ($s \to \infty$), the normalized values will approach 1.

The comparisons for a substrate in a duct were performed to more closely represent the overall acoustic characteristics of an automotive CC substrate. The acoustic behavior of the substrate is characterized in terms of the reflection coefficient

$$\mathrm{RC} = \frac{p_{\mathrm{re}}}{p_{\mathrm{inc}}}, \tag{54}$$

and the transmission coefficient

$$\mathrm{TC} = \frac{p_{\mathrm{tr}}}{p_{\mathrm{inc}}}, \tag{55}$$

where the subscripts re, inc, and tr denote reflected, incident, and transmitted components, respectively. The transmission and reflection coefficients can be determined from the substrate transfer matrix, which relates the acoustic variable across the substrate as

$$\begin{Bmatrix} p \\ \rho_0 c_0 u_m \end{Bmatrix}_u = [T] \begin{Bmatrix} p \\ \rho_0 c_0 u_m \end{Bmatrix}_d, \tag{56}$$

where the subscripts u and d denote upstream and downstream locations and

$$[T] = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} \tag{57}$$

is the transfer matrix. The reflection and transmission coefficients can be determined from the transfer matrix components as

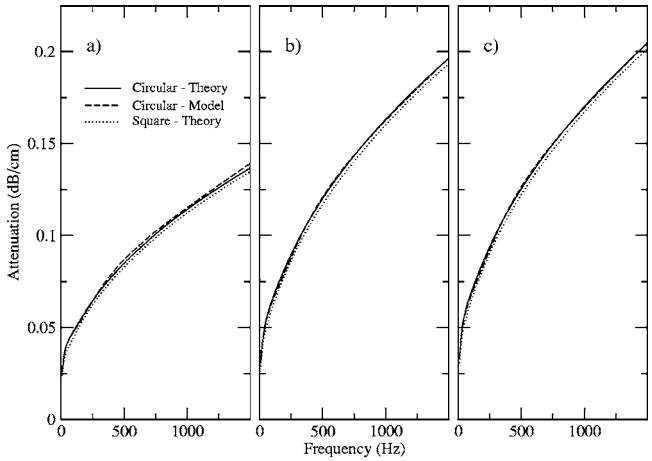


FIG. 3. Attenuation rate for acoustic propagation in capillary: (a) adiabatic, solid walls; (b) isothermal, solid walls; (c) isothermal, porous walls.

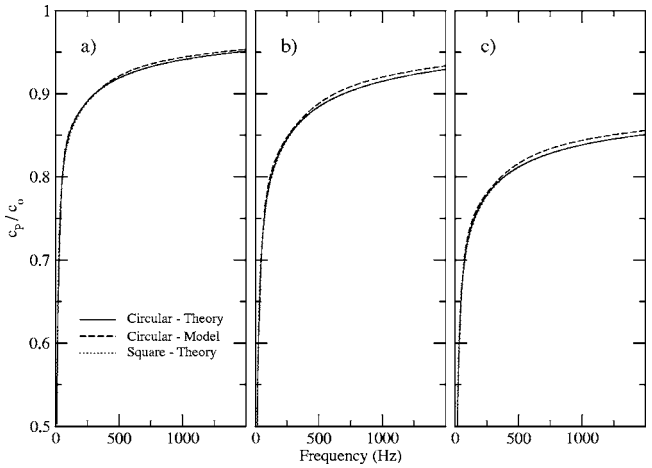


FIG. 4. Normalized phase speed for acoustic propagation in capillary: (a) adiabatic, solid walls; (b) isothermal, solid walls; (c) isothermal, porous walls.

$$\mathrm{RC} = \frac{(T_{11} - T_{21})(1 + R_d) + (T_{12} - T_{22})(1 - R_d)}{(T_{11} + T_{21})(1 + R_d) + (T_{12} + T_{22})(1 - R_d)}, \tag{58}$$

$$\mathrm{TC} = \frac{2}{(T_{11} + T_{21})(1 + R_d) + (T_{12} + T_{22})(1 - R_d)}, \tag{59}$$

where the downstream reflection coefficient R_d is determined by the termination characteristics. For an anechoic termination, $R_d = 0$.

In addition to viscothermal propagation, the substrate will affect acoustic propagation due to the changes in cross-sectional area at its inlet and outlet. The overall transfer matrix may be expressed as

$$[T] = [T_{\mathrm{contraction}}][T_c][T_{\mathrm{expansion}}], \tag{60}$$

where the contraction and expansion transfer matrices account for the area changes at the upstream and downstream ends of the substrate, respectively. The overall transfer matrix components can then be found as

$$T_{11} = \cosh(\Gamma k_0 L_s), \tag{61}$$

$$T_{12} = \frac{Z}{\Omega}\sinh(\Gamma k_0 L_s), \tag{62}$$

$$T_{21} = \frac{\Omega}{Z}\sinh(\Gamma k_0 L_s), \tag{63}$$

$$T_{22} = \cosh(\Gamma k_0 L_s), \tag{64}$$

where L_s is the length of the substrate and $\Omega = A_{\mathrm{open}}/A_{\mathrm{total}}$ is the open area ratio of the substrate.

A. Capillary propagation characteristics

In this section, the computational results are compared to circular and square cross section analytical solutions for cases of: (a) nonporous adiabatic walls, (b) nonporous isothermal walls, and (c) porous isothermal walls. The substrate and fluid parameters are included in Table II.

The attenuation rates for the different cases are shown in Fig. 3. For all cases, the well-known behavior of dissipative attenuation becoming more significant as frequency increases is seen. All cases show a significant difference from the ideal case of adiabatic and inviscid propagation with attenuation rate of zero. As might be expected, differences in attenuation rate between the two cross-sectional shapes can be distinguished, but are relatively insignificant. The differences between the adiabatic (a) and isothermal (b) results show that changes due to wall heat transfer are significant. Differences in attenuation rate due to wall porosity [compare (b) and (c)] are less significant. The computational model closely matches the circular capillary analytical solution (upon which the model is based) in all three cases. In case (a), slight deviations between the computations and the circular theory are noticeable, and are most likely due to the approximate method to determine $W'(\tilde{t})$.

For the phase speed comparisons depicted in Fig. 4, differences due to cross-sectional shape are nearly indistinguishable (the circular and square capillary analytical solutions generally overlay each other). In contrast to the attenuation rate, however, the effect of wall porosity is larger than that of wall heat transfer. Note that wall heat transfer

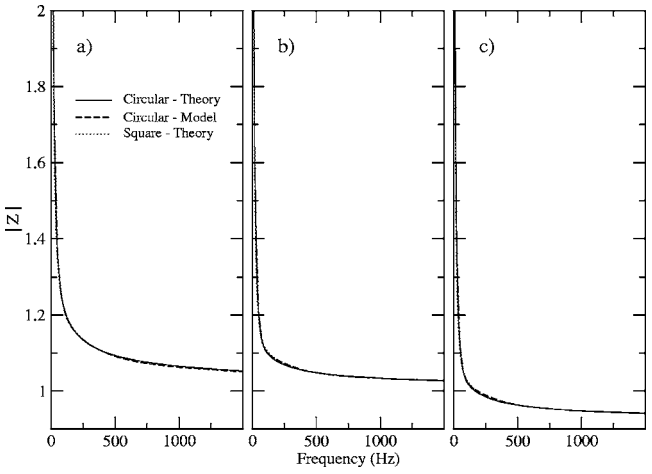


FIG. 5. Normalized impedance magnitude for acoustic propagation in capillary: (a) adiabatic, solid walls; (b) isothermal, solid walls; (c) isothermal, porous walls.

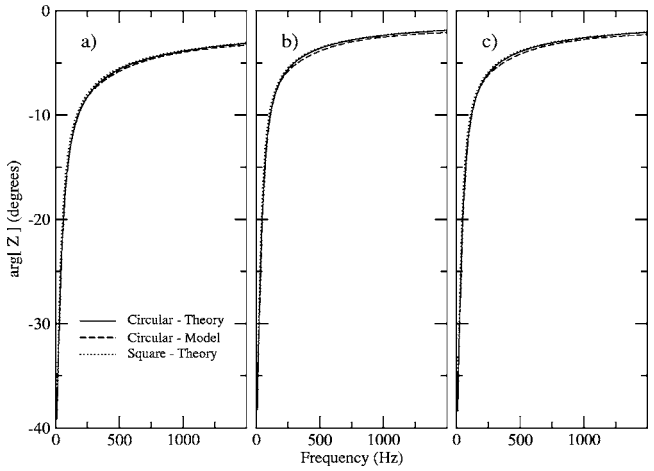


FIG. 6. Argument of impedance magnitude for acoustic propagation in capillary: (a) adiabatic, solid walls; (b) isothermal, solid walls; (c) isothermal, porous walls.

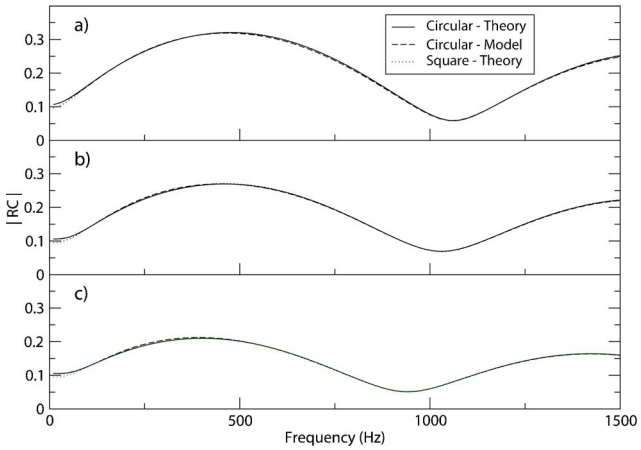


FIG. 8. (Color online) Magnitude of reflection coefficient for catalyst substrate: (a) adiabatic, solid walls; (b) isothermal, solid walls; (c) isothermal, porous walls.

and wall porosity both decrease the phase speed, so errors due to neglecting these two factors will compound. The computational results match the theoretical solutions well, though there is a small, but consistent overprediction of the phase speed at frequencies above approximately 500 Hz for all cases.

Figures 5 and 6 include the magnitude and argument of the normalized capillary impedance, respectively. Both the magnitude and argument of Z are relatively insensitive to cross-sectional shape. In general, findings for the magnitude of Z are similar to those for the phase speed: Porosity decreases $|Z|$, while neglecting heat transfer increases it and, except at the lowest frequencies, the effects of porosity are somewhat greater than those of heat transfer. Findings for the argument of Z are similar to those for the attenuation rate. The effect of wall porosity is of small consequence. Changes due to wall heat transfer are significantly larger than those for wall porosity, but still moderate. For all cases considered in Figs. 5 and 6, the computational results closely match the theory.

B. Substrate transmission/reflection characteristics

In this section, analytical and computational results are compared for the transmission and reflection characteristics of a substrate (refer to Table II for dimensions) placed in an anechoically terminated duct (see Fig. 7). This geometry is somewhat more representative of the actual behavior of a catalyst substrate since, in addition to the viscothermal wave propagation in the capillaries, the effects of reflections due to changes in cross-sectional area at the inlet and outlet are accounted for. Computational and analytical results are compared for the same cases considered in the preceding section.

Figure 8 depicts the magnitude of the substrate reflection coefficient RC. Although viscothermal attenuation and wave reflections from the outlet of the substrate contribute to the reflection coefficient, its magnitude is mainly determined by the changes in cross-sectional area and characteristic impedance at the substrate inlet. Differences in $|\mathrm{RC}|$ between the different cases therefore correlate with changes in the characteristic impedance from Fig. 5. The larger impedance of the adiabatic case (a) causes greater reflection at the inlet, while the reduction in impedance due to wall porosity (c) decreases the reflection coefficient. Differences between the model and two analytical solutions are nearly indistinguishable. However, at the lowest frequencies, the theory for the square capillary is slightly lower than the circular results (both model and theory) for all three cases. For the argument of the reflection coefficient (Fig. 9), the adiabatic (a) and isothermal (b) cases are quite similar at frequencies below approximately 750 Hz. Above this frequency, phase differences are more significant. The introduction of porous walls (c) causes larger phase changes that can be seen for the entire

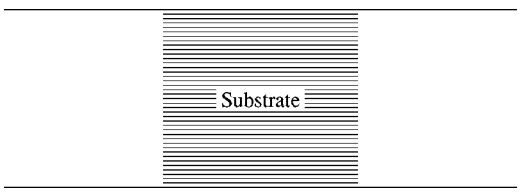


FIG. 7. Geometry used in substrate analysis.

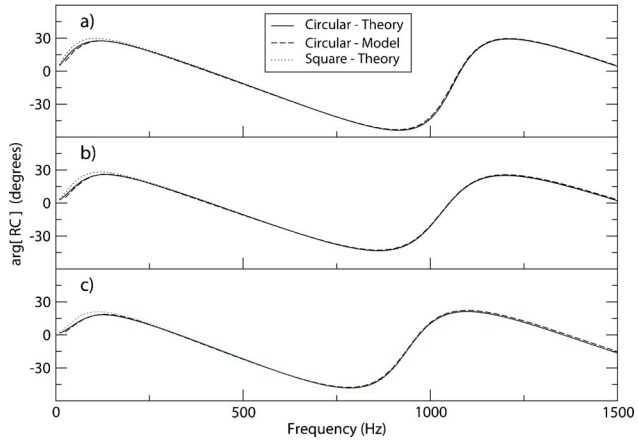


FIG. 9. (Color online) Argument of reflection coefficient for catalyst substrate: (a) adiabatic, solid walls; (b) isothermal, solid walls; (c) isothermal, porous walls.

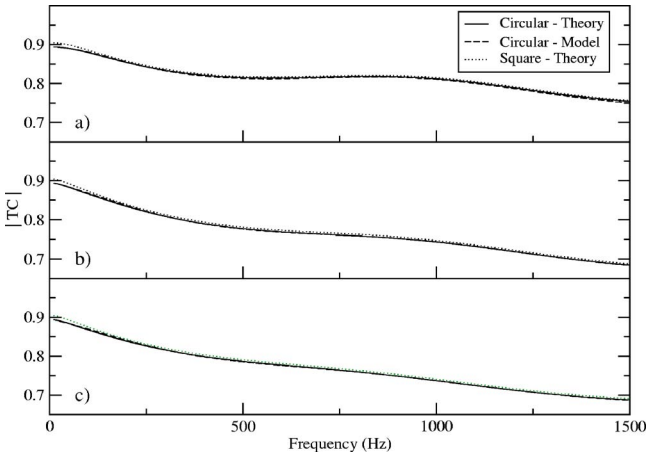


FIG. 10. (Color online) Magnitude of transmission coefficient for catalyst substrate: (a) adiabatic, solid walls; (b) isothermal, solid walls; (c) isothermal, porous walls.

frequency range. Differences between the numerical approach and circular capillary theory are nearly indistinguishable, but small differences are again seen between the circular and square results.

Results for the magnitude of the transmission coefficient (TC) are given in Fig. 10. For the most part, the results for |TC| reflect the findings for the attenuation rate in Fig. 3. Porosity has a rather small effect on |TC|, while the effects of heat transfer are substantial, particularly at the higher frequencies. For all cases, the computational approach matches the circular theory for |TC| very well. The theory for the square capillary is slightly higher than the circular results for the entire frequency range, with this difference increasing slightly at the lowest frequencies. For the argument of the transmission coefficient (Fig. 11), the effects of heat transfer are small, and the adiabatic and isothermal theory are rather close over nearly the entire frequency range. The porosity changes the phase of TC significantly, with the phase difference increasing with frequency. Differences between the numerical results and two theoretical solutions are, for the most part, indistinguishable over the entire frequency range (apparent differences near the 360° phase transition are due to lower frequency resolution of the computational results).

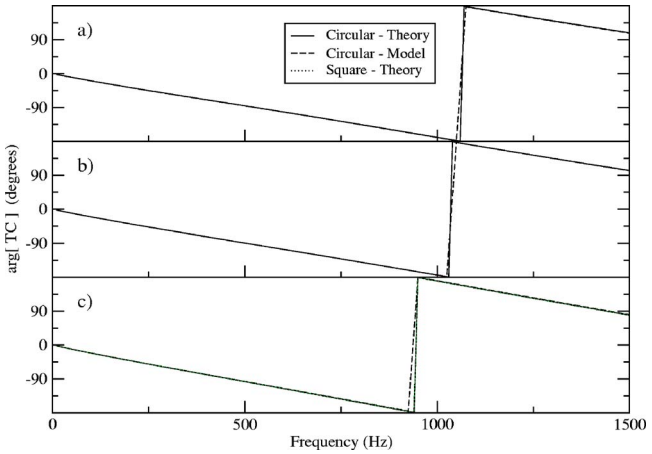


FIG. 11. (Color online) Argument of transmission coefficient for catalyst substrate: (a) adiabatic, solid walls; (b) isothermal, solid walls; (c) isothermal, porous walls.

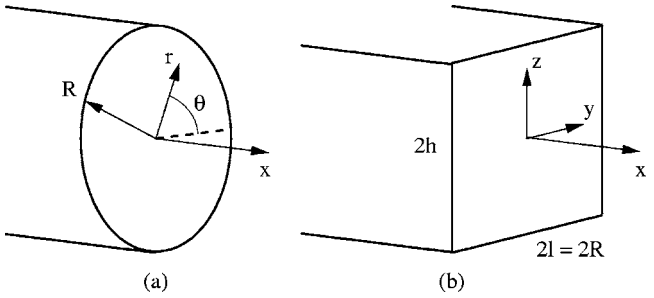


FIG. 12. Geometry and coordinate systems for (a) circular and (b) rectangular capillaries.

V. CONCLUSIONS

Models for frequency-dependent friction, frequency-dependent heat transfer, and substrate wall porosity have been developed that are suitable for modeling catalytic converter substrates in one-dimensional time domain approaches for automotive exhaust systems. The computational models have been shown to correlate well with theoretical solutions for the propagation of sound in capillary tubes. The method employed in this study is based on circular capillaries, but may be applied to develop similar submodels for different capillary cross sections. However, comparisons between square and circular capillaries suggest that differences for most practical geometries will be slight. The comparisons also illustrate that effects due to viscous wall shear, wall heat transfer, and wall porosity are all potentially important. The relative importance of the viscothermal effects may be determined from theoretical considerations. The expected importance of wall porosity, however, is somewhat more uncertain, as data for the porosity parameters do not appear to be readily available. Finally, while the actual conditions in an automotive catalytic converter may deviate from the linear assumptions used to develop the models, it is expected that the models will be a substantial improvement from the typical wall shear and heat transfer coefficients based on steady laminar flow. More work is necessary, however, to address the actual conditions in an automotive exhaust system with mean flow, high-amplitude disturbances, large temperature gradients, and substrate temperatures that depend on numerous engine operating parameters.

APPENDIX A: COORDINATE SYSTEMS

The coordinate systems used for circular and rectangular capillaries are included in Fig. 12. For both geometries, the axial coordinate is taken as x. For the circular capillary, the fluid motion is assumed to be axisymmetric. Velocities and vector operators for the circular geometry are therefore

$$\boldsymbol{v} = e_x u + e_r v,$$

$$\nabla_a = e_x \frac{\partial}{\partial x},$$

$$\nabla_c = e_r \frac{\partial}{\partial r},$$

$$\nabla_a^2 = \frac{\partial^2}{\partial x^2},$$

$$\nabla_c^2 = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r}. \tag{A1}$$

For the rectangular geometry, the velocities and vector operators are

$$\boldsymbol{v} = e_x u + e_y v + e_z w,$$

$$\nabla_a = e_x \frac{\partial}{\partial x},$$

$$\nabla_c = e_y \frac{\partial}{\partial y} + e_z \frac{\partial}{\partial z},$$

$$\nabla_a^2 = \frac{\partial^2}{\partial x^2},$$

$$\nabla_c^2 = \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}. \tag{A2}$$

APPENDIX B: APPROXIMATION OF $W'(\tilde{t})$

The numerical approach approximates $W'(\tilde{t})$ with a summation of exponential functions as

$$W'(\tilde{t}) \cong W'_{\text{app}}(\tilde{t}) = \sum_{i=1}^{n} W'_i(\tilde{t}) = \sum_{i=1}^{n} a_i e^{-b_i \tilde{t}}, \tag{B1}$$

where a_i and b_i are constants determined in the curve fit of $W'_{\text{app}}(\tilde{t})$ to Eqs. (34) and (35). With the introduction of

$$y_i(t) = \int_0^t W_i(t-\phi)\frac{\partial U(\phi)}{\partial t} d\phi, \tag{B2}$$

where $W_i(t) = W'_i(\tilde{t})$, Eq. (33) can be expressed as

$$\tau_w(t) \cong \frac{4\mu}{R}\left[U(t) + \frac{1}{2}\sum_{i=1}^{N} y_i(t) \right]. \tag{B3}$$

The reduction in storage for a computational approach is obtained by relating y_i at time $t+\Delta t$ to its value at time t, rather than requiring a numerical integration over numerous previous time steps. From Eq. (B2)

$$y_i(t+\Delta t) = \int_0^{t+\Delta t} W_i(t+\Delta t-\phi)\frac{\partial U(\phi)}{\partial t} d\phi, \tag{B4}$$

which, by splitting the integration limits and using

$$W_i(t+\Delta t-\phi) = W_i(t+\Delta t-\phi) - W_i(t-\phi) + W_i(t-\phi), \tag{B5}$$

can be rearranged as

$$y_i(t+\Delta t) = \int_t^{t+\Delta t} W_i(t+\Delta t-\phi)\frac{\partial U(\phi)}{\partial t} d\phi + \int_0^t [W_i(t+\Delta t-\phi) - W_i(t-\phi)]\frac{\partial U(\phi)}{\partial t} d\phi + \int_0^t W_i(t-\phi)\frac{\partial U(\phi)}{\partial t}\, \partial\, \phi. \tag{B6}$$

Using Eq. (B1) in the first term on the right-hand side of Eq. (B6) gives

$$\int_t^{t+\Delta t} W_i(t+\Delta t-\phi)\frac{\partial U(\phi)}{\partial t} d\phi = \int_t^{t+\Delta t} a_i e^{b_i(\nu/R^2)(t+\Delta t-\phi)}\frac{\partial U(\phi)}{\partial t} d\phi, \tag{B7}$$

which, for small Δt, can be expressed by a difference equation as (see Ref. 12)

$$\int_t^{t+\Delta t} W_i(t+\Delta t-\phi)\frac{\partial U(\phi)}{\partial t} d\phi = a_i(U(t+\Delta t) - U(t)) \quad \text{for } \Delta t \to 0. \tag{B8}$$

Inserting Eq. (B1) into the second term on the right-hand side of Eq. (B6) gives, after some rearrangement,

$$\int_0^t [W_i(t+\Delta t-\phi) - W_i(t-\phi)]\frac{\partial U(\phi)}{\partial t} d\phi = (e^{-b_i(\nu/R^2)\Delta t} - 1)\int_0^t a_i e^{-b_i(\nu/R^2)(t-\phi)}\frac{\partial U(\phi)}{\partial t} d\phi, \tag{B9}$$

or

$$\int_0^t [W_i(t+\Delta t-\phi) - W_i(t-\phi)]\frac{\partial U(\phi)}{\partial t} d\phi = (e^{-b_i(\nu/R^2)\Delta t} - 1) y_i(t). \tag{B10}$$

Finally, the third term on the right-hand side of Eq. (B6) gives simply

$$\int_0^t W_i(t-\phi)\frac{\partial U(\phi)}{\partial t} d\phi = y_i(t). \tag{B11}$$

Inserting Eqs. (B8)–(B11) into Eq. (B6) gives

$$y_i(t+\Delta t) = y_i(t) e^{-b_i(\nu/R^2)\Delta t} + a_i(U(t+\Delta t) - U(t)). \tag{B12}$$

Equation (B12) provides the desired relationship between $y_i(t+\Delta t)$ and $y_i(t)$.

[1] N. S. Dickey, A. Selamet, S. J. Parks, K. V. Tallio, K. D. Miazgowicz, and P. M. Radavich, "Acoustic characteristics of automotive catalytic converter assemblies," SAE 2004-01-1002, 2004.

[2] G. Kirchhoff, "Über den einfluß der wärmeleitung in einem gase auf die schallbewegung" ("On the influence of heat conduction in a gas on sound propagation"), Poggendorfer Annalen **134**, 177–193 (1868).

[3] J. W. S. Rayleigh, *The Theory of Sound* (Dover, New York, 1945), Vol. II.

[4] H. Tijdeman, "On the propagation of sound waves in cylindrical tubes," J. Sound Vib. **39**, 1–33 (1975).

[5] W. M. Beltman, "Viscothermal wave propagation including acousto-elastic interaction," Ph.D. thesis, University of Twente, Enschede, The

Netherlands, 1998.
[6]W. P. Arnott, J. M. Sabatier, and R. Raspet, "Sound propagation in capillary-tube-type porous media with small pores in the capillary walls," J. Acoust. Soc. Am. **90**, 3299–3306 (1991).
[7]W. P. Arnott, H. E. Bass, and R. Raspet, "General formulation of thermoacoustics for stacks having arbitrarily shaped pore cross-sections," J. Acoust. Soc. Am. **90**, 3228–3237 (1991).
[8]R. J. Astley and A. Cummings, "Wave propagation in catalytic converters: Formulation of the problem and finite element solution scheme," J. Sound Vib. **188**, 635–657 (1995).
[9]K. W. Jeong and J. G. Ih, "A numerical study on the propagation of sound through capillary tubes with mean flow," J. Sound Vib. **198**, 67–79 (1996).
[10]K. S. Peat, "Convected acoustic wave motion along a capillary duct with an axial temperature gradient," J. Sound Vib. **203**, 855–866 (1997).
[11]E. Dokumaci, "On transmission of sound in circular and rectangular narrow pipes with superimposed mean flow," J. Sound Vib. **210**, 375–389 (1998).
[12]A. K. Trikha, "An efficient method for simulating frequency-dependent friction in transient liquid flow," J. Fluid Mech. **97**, 97–105 (1975).
[13]W. Zielke, "Frequency-dependent friction in transient pipe flow," J. Basic Eng. **90**, 109–115 (1968).
[14]A. Bergant, A. R. Simpson, and J. Vítkovský, "Developments in unsteady pipe flow friction modelling," J. Hydraul. Res. **39**, 249–257 (2001).
[15]K. Suzuki, T. Taketomi, and S. Sato, "Improving Zielke's method of simulating frequency-dependent friction in laminar pipe flow," J. Fluids Eng. **113**, 569–573 (1991).
[16]D. E. Weston, "The theory of the propagation of plane sound waves in tubes," Proc. Phys. Soc. London, Sect. B **66**, 695–709 (1953).
[17]M. R. Stinson, "The propagation of plane sound waves in narrow and wide circular tubes, and generalization to uniform tubes of arbitrary cross-sectional shape," J. Acoust. Soc. Am. **89**, 550–558 (1991).
[18]C. Zwikker and C. W. Kosten, *Sound Absorbing Materials* (Elsevier, Amsterdam, 1949).
[19]A. Onorati, "Nonlinear fluid dynamic modeling of reactive silencers involving extended inlet/outlet and perforated ducts," Noise Control Eng. J. **45**, 35–51 (1997).
[20]N. S. Dickey, A. Selamet, and J. M. Novak, "Multi-pass perforated tube silencers: A computational approach," J. Sound Vib. **211**, 435–448 (1998).
[21]J. W. Sullivan and M. J. Crocker, "Analysis of concentric-tube resonators having unpartitioned cavities," J. Acoust. Soc. Am. **64**, 207–215 (1978).
[22]M. Chapman, J. M. Novak, and R. A. Stein, "Numerical modeling of inlet and exhaust flows in multi-cylinder internal combustion engines," in *Flows in Internal Combustion Engines*, edited by T. Uzkan (ASME WAM, Austin, TX, 1982).

Reflection of a spherical wave by acoustically hard, concave cylindrical walls based on the tangential plane approximation

Yoshinari Yamada[a] and Takayuki Hidaka
Takenaka R & D Institute, 1-5-1, Otsuka, Inzai, Chiba, 270-1395, Japan

(Received 11 June 2004; revised 12 December 2004; accepted 9 May 2005)

The tangential plane approximation (TPA) is introduced to investigate the spherical wave reflection from smooth concave cylindrical walls, and as a practical calculation scheme, asymptotic expression of the reflected sound from that surface with a relatively large dimension is derived. The physical condition under which TPA holds is derived for the spherical wave incidence; moreover, the essential properties of the reflected sound from the curved walls, which cannot be treated by the geometrical acoustics, are discussed through the numerical calculation for infinitely long cylinders. A formula of the reflection factor of the two-dimensional curved surface is obtained for the purpose of the room acoustical design. This formula coincides with that based on the geometrical acoustics when the frequency is infinitely high. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1944527]

PACS number(s): 43.55.Br, 43.55.Ka [MK] Pages: 818–831

I. INTRODUCTION

It has been known that the sound waves reflected by smooth concave walls concentrate at certain locations in a room (Rayleigh, 1926). This phenomenon can be predicted by means of the ray tracing if the geometry of the curved wall is mathematically given. When the geometrical acoustics is introduced to estimate the reflected sound field, some difficulties arise, such as no frequency dependency taken into consideration or the inability to obtain substantial solutions near the caustics or in the shadow zones (Babič, 1991). However, there are only a few papers on the wave-theoretical treatment on the sound field generated by the curved wall (Wahlström, 1985; Kuttruff, 1992) despite the fact that the practical guidelines for the room acoustical design are described in many references (Cremer and Muller, 1982).

The purpose of this paper is to investigate the sound reflection from various types of the concave wall through the wave-theoretical analysis aiming at cylindrical surfaces with a relatively high likeliness in application to architectural designs, thereby contributing to room acoustical designs. The tangential plane approximation (Bass and Fuks, 1979) is applied to comparatively large-sized cylindrical walls. This replaces the surface field on a curved wall with that on the tangential plane; as a result, the reflected field is described by a simple integral representation. Many applications are found in theoretical studies on wave scattering from rough surfaces (McCammon and McDaniel, 1986; Wirgin, 1989; Voronovich, 1996). However, since the plane wave incidence was considered in the majority of the existing studies, there are few reported cases on the spherical wave incidence, mainly relating to room acoustical applications.

In Sec. II, the principle of the tangential plane approximation is introduced and the physical conditions are identified, under which this approximation is established with respect to the reflected field generated by a point source and a curved wall. The subsequent results of the reflection analyses would be ensured from the theoretical aspect by using the derived conditions to limit the scope of examination.

In Sec. III, the sound reflection from finite cylinders is formulated based on the tangential plane approximation. By evaluating the integral along the polar axis of cylinder by means of the uniform stationary phase method (Borovikov, 1994), a practical method to calculate the reflected sound field is developed. The numerical accuracy of the method is then verified.

In Sec. IV, the essential properties of the reflected sound field due to the finite curvature of walls are numerically investigated by analyzing the sound reflection from infinitely long cylinders. The comparison with the related theories based on the geometrical acoustics (Kravtsov *et al.*, 1999; Kuttruff, 2000) is also given.

II. TANGENTIAL PLANE APPROXIMATION

A. Basic formulas for sound reflection by a curved wall

Part of the boundaries enclosing the room Ω is composed of a smooth concave wall Γ that is acoustically hard. When a point source at $\mathbf{R_s} \in \Omega$ is radiating the sinusoidal waves with the angular frequency ω, the sound field at $\mathbf{R} \in \Omega$ due to the source and the curved wall Γ is represented by the Helmholz formula for the Neumann problem (Voronovich, 1994),

$$p(k,\mathbf{R}) = p_d(k,\mathbf{R},\mathbf{R_S}) - \int_{\mathbf{r}\in\Gamma} p(k,\mathbf{r}) \frac{\partial}{\partial N(\mathbf{r})} \frac{\exp(ik|\mathbf{r}-\mathbf{R}|)}{4\pi|\mathbf{r}-\mathbf{R}|} dS(\mathbf{r}), \quad (1)$$

where $p_d = \exp(ik|\mathbf{R}-\mathbf{R_s}|)|\mathbf{R}-\mathbf{R_s}|^{-1}$ is the direct sound from the source [$\exp(-i\omega t)$ time dependence is omitted]; $k = \omega/c = 2\pi/\lambda$ is the wave number (c is the sound speed and λ is the wavelength); $dS(\mathbf{r})$ is the surface element of the

[a]Electronic mail: yamada.yoshinari@takenaka.co.jp

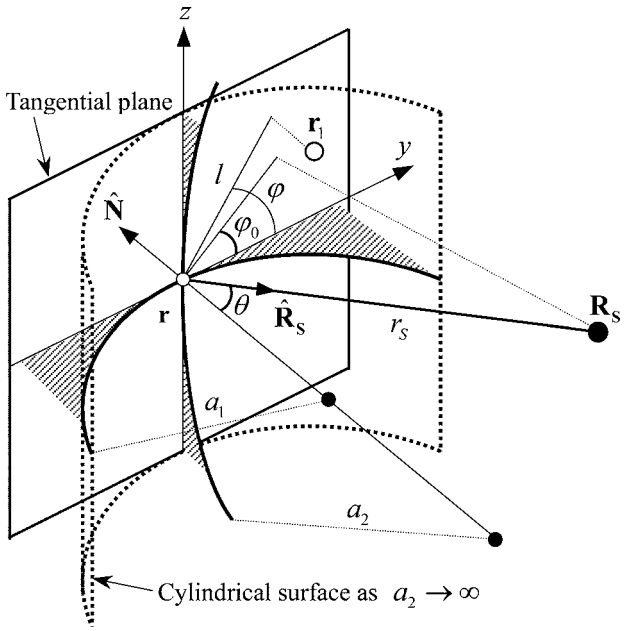


FIG. 1. Small area on a concave boundary approximated by the ellipsoidal paraboloid $x=y^2/(2a_1)+z^2/(2a_2)$. The limit $a_2 \to \infty$ is assumed for cylindrical surfaces.

curved wall Γ; and $\partial/\partial N(\mathbf{r})$ denotes the differentiation in the direction of the outward normal vector $\hat{\mathbf{N}}$. The surface field $p(k,\mathbf{r})$ on the curved wall Γ is generally unknown and given as solutions of the integral equation

$$p(k,\mathbf{r}) = 2p_d(k,\mathbf{r},\mathbf{R_S}) - 2\int_{\mathbf{r}'\in\Gamma} p(k,\mathbf{r}')\frac{\partial}{\partial N(\mathbf{r}')}\frac{\exp(ik|\mathbf{r}'-\mathbf{r}|)}{4\pi|\mathbf{r}'-\mathbf{r}|}dS(\mathbf{r}'), \tag{2}$$

which is obtained by operating $\mathbf{R}\to\mathbf{r}$ to Eq. (1) (Filippi *et al.*, 1999).

The tangential plane approximation (TPA), or the Kirchhoff approximation, replaces the surface field in Eq. (2) with that on the tangential plane $2p_d(k,\mathbf{r},\mathbf{R_s})$ (a sum of the incident and reflected fields). Hereby, the reflected sound wave from the curved wall Γ, which corresponds to the integral in Eq. (1), is simplified to

$$p_r(k,\mathbf{R}) \approx -2\int_{\mathbf{r}\in\Gamma} p_d(k,\mathbf{r},\mathbf{R_S}) \times \frac{\partial}{\partial N(\mathbf{r})}\frac{\exp(ik|\mathbf{r}-\mathbf{R}|)}{4\pi|\mathbf{r}-\mathbf{R}|}dS(\mathbf{r}). \tag{3}$$

B. Applicable condition of TPA for the spherical wave incidence

Equation (2) can be expanded to the Neumann series as follows:

$$p(k,\mathbf{r}) = \sum_{n=0} p_n(k,\mathbf{r}),$$

$$p_0(k,\mathbf{r}) = 2p_d(k,\mathbf{r},\mathbf{R_S}), \tag{4}$$

$$p_1(k,\mathbf{r}) = -2\int_{\mathbf{r}_1\in\Gamma} p_0(k,\mathbf{r}_1)\frac{\partial}{\partial N(\mathbf{r}_1)} \times \frac{\exp(ik|\mathbf{r}_1-\mathbf{r}|)}{4\pi|\mathbf{r}_1-\mathbf{r}|}dS(\mathbf{r}_1), \cdots .$$

It is found that TPA corresponds to the first term $p_0(k,\mathbf{r})$ and is valid if the higher terms $p_n(k,\mathbf{r})$ $(n\geqslant 1)$ are negligible. In this study, the asymptotic solution of Eq. (4) is used to examine the applicability of TPA. Similar approaches have been developed in the theoretical studies on plane waves scattering at rough surfaces (Lynch, 1970; Liszka *et al.*, 1982; Fuks and Voronovich, 1999). Referring to them, the higher terms in Eq. (4) can be neglected when

(1) multiple reflections along the concave surface do not exist, and
(2) local interactions with the surface field in the vicinity are sufficiently small.

The existence of the multiple reflections that relates with the stationary phase solution can be predicted by means of the ray tracing. On the other hand, the local interaction should be reformulated to match the spherical wave incidence.

For this purpose, the second term $p_1(k,\mathbf{r})$ in Eq. (4) is evaluated under the condition $k\to\infty$ assuming that the elliptic paraboloid $x=(2a_1)^{-1}y^2+(2a_2)^{-1}z^2$ approximates a small neighborhood of the incident point $\mathbf{r}$ as shown in Fig. 1 (Belobrov and Fuks, 1985). Expressing the source position as $\mathbf{R_s}=r_s\hat{\mathbf{R}}_\mathbf{s}+\mathbf{r}$ $[\hat{\mathbf{R}}_\mathbf{s}=(\cos\theta,\sin\theta\cos\varphi_0,\sin\theta\sin\varphi_0)]$ and transforming the vector $\mathbf{r}_1-\mathbf{r}=(x,y,z)$ to $(l^2\kappa(\varphi)/2, l\cos\varphi, l\sin\varphi)$ $[\kappa(\varphi)=a_1^{-1}\cos^2\varphi+a_2^{-1}\sin^2\varphi]$ by the polar coordinates (l,φ) defined on the tangential plane, the following approximations hold in the vicinity of $\mathbf{r}$:

$$|\mathbf{r}_1-\mathbf{r}| \approx l, \quad \hat{\mathbf{N}}(\mathbf{r}_1)\cdot(\mathbf{r}_1-\mathbf{r}) \approx l^2\kappa(\varphi)/2,$$

$$|\mathbf{r}_1-\mathbf{R_S}| \approx r_S\sqrt{1-2\sin\theta\cos(\varphi-\varphi_0)(lr_S^{-1})+\{1-\kappa(\varphi)r_S\cos\theta\}(lr_S^{-1})^2} \triangleq f(l,\varphi). \tag{5}$$

Hereby, $p_1(k,\mathbf{r})$ is rewritten as

$$p_1(k,\mathbf{r}) \approx -\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} \kappa(\varphi)(ikl-1)f(l,\varphi)^{-1} \times \exp[ik\{l+f(l,\varphi)\}]dl\,d\varphi, \tag{6}$$

where the relation $\partial/\partial N(\mathbf{r}_1)=\hat{\mathbf{N}}(\mathbf{r}_1)\cdot(\mathbf{r}_1-\mathbf{r})|\mathbf{r}_1-\mathbf{r}|^{-1}\partial/\partial|\mathbf{r}_1-\mathbf{r}|$ was taken into account. This integral can be evaluated by the asymptotic method described in Appendix A if the stationary point is not near the integration boundaries. As a result, the solution of Eq. (6) is given by

$$p_1(k,\mathbf{r}) = (D_R + iD_I)\frac{\exp(ikr_S)}{r_S} + O(k^{-3}), \tag{7}$$

where

$$D_R = \frac{6}{kr_S\cos^2\theta}\frac{C_1}{2k\cos^3\theta} - \frac{8}{kr_S}\frac{B_+}{2k\cos^3\theta} - \frac{C_2}{8k^2\cos^6\theta}, \quad D_I = \frac{C_1}{2k\cos^3\theta}, \tag{8}$$

and

$$C_1 = B_+(1+\cos^2\theta) + B_-\cos 2\varphi_0\sin^2\theta,$$

$$C_2 = \sum_{m=0}^{2} T_{2m}L_{2m}(\theta)\cos 2m\varphi_0\tan^{2m}(\theta/2),$$

$$B_+ = (1/a_1 + 1/a_2), \quad B_- = (1/a_1 - 1/a_2), \tag{9}$$

$$T_0 = B_+^2 + B_-^2/2, \quad T_2 = 2B_+B_-, \quad T_4 = B_-^2/2,$$

$$L_m(\theta) = Z_m^3 + Z_m^2(3+\cos^2\theta) + 3Z_m(1+\sin^2\theta) + \sin^2\theta(5+\cos^2\theta), \quad Z_m = 1 + m\cos\theta.$$

In Eq. (8), the third term $C_2(k\cos^3\theta)^{-2}$ of D_R and $D_I = C_1(k\cos^3\theta)^{-1}$ coincide with the coefficients for the plane wave incidence by Belobrov and Fuks (1985). Other two terms depending on the source distance r_s are additive ones for the spherical wave incidence. Taking into account that they are multiplied by D_I, the local interaction can be neglected ($D_R+iD_I\approx 0$) if the following inequalities are satisfied at the same time:

$$D_I \triangleq C_1(k\cos^3\theta)^{-1} \ll 1, \quad K \triangleq 6(kr_S\cos^2\theta)^{-1} \leqslant 1. \tag{10}$$

Thus, the applicability of TPA depends on the local parameters at the incident point: the principal radii a_1 and a_2 of the wall; the source distance r_s and the directional angle φ_0; the angle of incidence θ. Qualitatively, the radius of curvature and the source distance have to be much larger than the wavelength, and the angle of incidence has to be comparatively small.

C. Application to cylindrical walls

In order to apply Eq. (10) to cylindrical walls, the limit $a_2\to\infty$ is calculated (see Fig. 1). Then, $D_I\to(1-\sin^2\varphi_0\sin^2\theta)(ka_1\cos^3\theta)^{-1}$ $(a_2\to\infty)$ is obtained.

When the source and the incident point are on the same plane perpendicular to the polar axis of cylinder ($\varphi_0=0$), writing the source distance and the angle of incidence as r_0 and θ_0, respectively, D_I and K in Eq. (10) are specified as

$$D_I = (ka_1\cos^3\theta_0)^{-1} \triangleq D_{10}, \quad K_0 = (kr_0\cos^2\theta_0)^{-1} \triangleq K_0. \tag{11}$$

When the incident point moves away from this sectional plane by z_L in the direction parallel to the polar axis, the geometrical relations $r_s=(r_0^2+z_L^2)^{1/2}$, $\sin^2\varphi_0 = z_L^2(r_0^2\sin^2\theta_0 + z_L^2)^{-1}$, and $\cos\theta = r_0r_s^{-1}\cos\theta_0$ hold. Accordingly, the applicable condition of TPA is given by

$$D_I = r_0^{-1}\sqrt{r_0^2+z_L^2}D_{10} \ll 1, \quad K = r_0^{-1}\sqrt{r_0^2+z_L^2}K_0 \leqslant 1. \tag{12}$$

It is apparent from Eq. (12) that the length of cylinder (the dimension along the polar axis) has to be finite. When the incident point is considerably distant from the sectional plane including the source ($z_L\to\infty$ and $\theta\to\pi/2$), the stationary point of Eq. (6) is located near the integration boundary (the incident point or infinity), that is, the arrival of the reflected wave along the cylinder surface is suggested.

If one requires $D_I \leqslant g_l$ ($g_l \ll 1$) at every point on a cylindrical wall, the allowable distance between a curved edge and the sectional plane including the source (the upper limit of z_L) is derived from Eq. (12) as

$$z_L \leqslant \min[r_0(g_l^2D_{10}^{-2}-1)^{1/2},\ r_0(K_0^{-2}-1)^{1/2}], \tag{13}$$

where min is the minimum value while r_0, G_{10}, and K_0 change along the lateral surface. In this study, $g_l=0.1$ is assumed. Substituting $D_R=D_I=0.1$ into Eq. (7) ($K\leqslant 1$ ensures $D_R\leqslant D_I$) and comparing with $p_0(k,\mathbf{r})$ in Eq. (4), the error of the surface field by TPA is less than 0.5 dB in this case and would be allowable for the practical purposes in room acoustics.

III. REFLECTED SOUND FIELD FROM FINITE CYLINDER SECTORS

A. Formulation based on TPA

The canonical two-dimensional cylinders (the circular, parabolic, elliptic, and hyperbolic cylinders) are expressed in the Cartesian coordinate system as follows:

Circular cylinder $\quad x^2+y^2=a^2\ (a>0),$ (14a)

Parabolic cylinder $\quad y^2=4b(x+b)\ (b>0),$ (14b)

Elliptic cylinder $\quad x^2/a_e^2+y^2/b_e^2=1\ (a_e>b_e>0),$ (14c)

Hyperbolic cylinder $\quad x^2/a_h^2-y^2/b_h^2=1\ (a_h>0,\ b_h>0).$ (14d)

Here, the polar axis is parallel to the z axis. Although Eqs. (14a)–(14d) are useful to understand the geometries of each cylinder as shown in Table I, the co-focal, orthogonal curvilinear coordinates in Table II, i.e., the polar coordinates, the parabolic coordinates, and two kinds of the elliptic coordinates (Morse and Feshback, 1953) are used to formulate the sound reflections. In the following introducing the general-

TABLE I. Geometrical properties of two-dimensional cylinders.

	Circular	Parabolic	Elliptic	Hyperbolic
Equation	(14a)	(14b)	(14c)	(14d)
Geometrical focus F	(0,0)	(0,0)	$(\pm d_e,0)$ $d_e=\sqrt{a_e^2-b_e^2}$	$(\pm d_h,0)$ $d_h=\sqrt{a_h^2+b_h^2}$
Eccentricity ε	0	1	$0<d_e/a_e<1$	$d_h/a_h>1$
Radius of curvature A	a	$\frac{2(x+2b)^{3/2}}{\sqrt{b}}$	$(a_e b_e)^2\left(\frac{x^2}{a_e^4}+\frac{y^2}{b_e^4}\right)^{3/2}$	$(a_h b_h)^2\left(\frac{x^2}{a_h^4}+\frac{y^2}{b_h^4}\right)^{3/2}$

ized coordinates (ξ,η) in order to treat the four coordinate systems together, the reflection surfaces are defined as $\xi=\xi_b=constant$, $\eta_1\leqslant\eta\leqslant\eta_2$, and $z_1\leqslant z\leqslant z_2$. When a source is at $\mathbf{R_s}=(x_0,y_0,0)$ on the xy plane, the reflected sound field at $\mathbf{R}=(x_r,y_r,z_r)$ can be rewritten as follows (Appendix B) when TPA is applied:

$$p_r(k,\mathbf{R})=-\frac{ik}{2\pi}\int_{\eta_1}^{\eta_2}\int_{z_1}^{z_2}f(\eta,z)\exp[ik\varphi(\eta,z)]dz\,d\eta, \quad (15)$$

where η is the integration variable in the direction perpendicular to the z axis, and

$$f(\eta,z)=\frac{1}{2U(\eta,z)V(\xi_b,\eta,z)^2}\left[\frac{\partial\nu(\xi,\eta)^2}{\partial\xi}\right]_{\xi=\xi_b}\frac{h_\eta(\xi_b,\eta)}{h_\xi(\xi_b,\eta)}, \quad (16a)$$

$$\varphi(\eta,z)=U(\eta,z)+V(\xi_b,\eta,z). \quad (16b)$$

Here,

$$h_\xi(\xi,\eta)=\sqrt{\left\{\frac{\partial x(\xi,\eta)}{\partial\xi}\right\}^2+\left\{\frac{\partial y(\xi,\eta)}{\partial\xi}\right\}^2},$$
$$h_\eta(\xi,\eta)=\sqrt{\left\{\frac{\partial x(\xi,\eta)}{\partial\eta}\right\}^2+\left\{\frac{\partial y(\xi,\eta)}{\partial\eta}\right\}^2}, \quad (17a)$$

$$U(\eta,z)=|\mathbf{r}-\mathbf{R_s}|=\sqrt{u(\eta)^2+z^2},$$
$$V(\xi,\eta,z)=|\mathbf{r}-\mathbf{R}|=\sqrt{\nu(\xi,\eta)^2+(z-z_r)^2}, \quad (17b)$$

TABLE II. The co-focal, orthogonal curvilinear coordinate system (ξ,η) as a generalized representation of the polar coordinates (r,ϕ), the parabolic coordinates (μ,γ), and two kinds of the elliptic coordinates (ρ,φ) and (φ,ρ). The curve $\xi=\xi_b=constant$ is used as a reflection surface. $x(\xi,\eta)$ and $y(\xi,\eta)$ represent the transformations from (ξ,η) to the Cartesian coordinates (x,y). $h_\xi(\xi,\eta)$ and $h_\eta(\xi,\eta)$ are the scale factors defined by Eq. (17a).

	Circular	Parabolic	Elliptic	Hyperbolic
(ξ,η)	(r,ϕ)	(μ,γ)	(ρ,φ)	$(\varphi,\rho)(0\leqslant\varphi\leqslant\pi)$
$\xi=\xi_b=constant$	$r_b\equiv a$	$\mu_b\equiv\sqrt{2b}$	$\rho_b\equiv\sinh^{-1}(b_e/d_e)$	$\varphi_b\equiv\sin^{-1}(b_h/d_h)$
$x(\xi,\eta)$	$r\cos\phi$	$(\gamma^2-\mu^2)/2$	$d_e\cosh\rho\cos\varphi$	$d_h\cos\varphi\cosh\rho$
$y(\xi,\eta)$	$r\sin\phi$	$\mu\gamma$	$d_e\sinh\rho\sin\varphi$	$d_h\sin\varphi\sinh\rho$
$h_\xi(\xi,\eta)$	1	$\sqrt{\gamma^2+\mu^2}$	$d_e\sqrt{\cosh^2\rho-\cos^2\varphi}$	$d_h\sqrt{\cosh^2\rho-\cos^2\varphi}$
$h_\eta(\xi,\eta)$	r			

$$u(\eta)=\sqrt{\{x(\xi_b,\eta)-x_0\}^2+\{y(\xi_b,\eta)-y_0\}^2},$$
$$\nu(\xi,\eta)=\sqrt{\{x(\xi,\eta)-x_r\}^2+\{y(\xi,\eta)-y_r\}^2}, \quad (17c)$$

where h_ξ and h_η are the scale factors with respect to ξ and η, respectively [refer to Eq. (B1)]. The specific expressions of the scale factors, the transformation formulas $x=x(\xi,\eta)$ and $y=y(\xi,\eta)$ for each coordinate system are summarized in Table II.

The integrand in Eq. (15) oscillates significantly, when the wall dimension is larger than the wavelength. Accordingly, the integral with respect to z is decomposed into

$$I=\int_{-\infty}^{z_2}f(\eta,z)\exp[ik\varphi(\eta,z)]dz-\int_{-\infty}^{z_1}f(\eta,z)\exp[ik\varphi(\eta,z)]dz, \quad (18)$$

and evaluated by the uniform stationary phase method (Borovikov, 1994). Differentiating Eq. (16b), the stationary point z_s is obtained as

$$z_s(\eta)=u(\eta)\{u(\eta)+\nu(\xi_b,\eta)\}^{-1}z_r. \quad (19)$$

The short wave asymptotic solution of Eq. (18) is given by the following functions' combinations in Table III, whose manners depend on the location of the stationary point relative to the integration boundaries z_j $(j=1,2)$ as shown in Fig. 2:

TABLE III. Combination of the functions Eqs. (20a)–(20c) to obtain asymptotic solutions of the integral Eq. (18). Regions I–IV are defined in Fig. 2.

	Regions where the phase stationary point z_s is located			
Integral	I	II or III	IV	Edge z_j
$\int_{-\infty}^{z_1}dz$	I_e	I_u+I_e	I_s+I_e	$I_s/2$
$\int_{-\infty}^{z_2}dz$	I_s+I_e	I_u+I_e	I_e	$I_s/2$

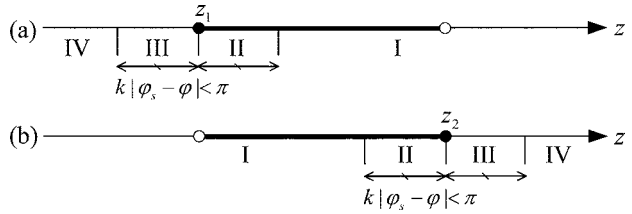


FIG. 2. Regions I–IV defining locations of the stationary point z_s relative to (a) the lower integration boundary z_1 and (b) the upper integration boundary z_2 (see also Table III).

$$I_s(\eta) = (1+i)\sqrt{\pi k^{-1}}\, g_s(\eta)\exp[ik\varphi_s(\eta)],$$

$$I_e(\eta, z_j) = -ik^{-1} g_e(\eta, z_j)\exp[ik\varphi(\eta, z_j)],$$

$$I_u(\eta, z_k) = F[\mathrm{sign}(z_j - z_s) d_k(\eta, z_j)^{1/2}] I_s(\eta) + \frac{i}{\sqrt{2k}} \frac{g_s(\eta)\exp[ik\varphi(\eta, z_j)]}{\mathrm{sign}(z_j - z_s) d_k(\eta, z_j)^{1/2}}, \tag{20}$$

where $F(x) = (\pi i)^{-1/2} \int_{-\infty}^{x} \exp(is^2) ds$ is the Fresnel integral, and

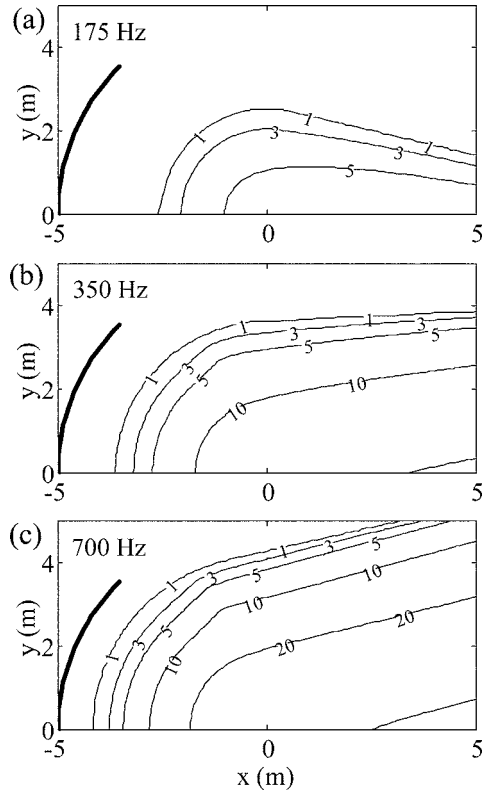


FIG. 3. Allowable distance from the plane $z=0$ to a curved edge, for the circular cylinder with the radius 5 m and directional angles of 135° and 225°, as contour lines in the coordinates representing source positions. The frequency is (a) 175, (b) 350, and (c) 700 Hz.

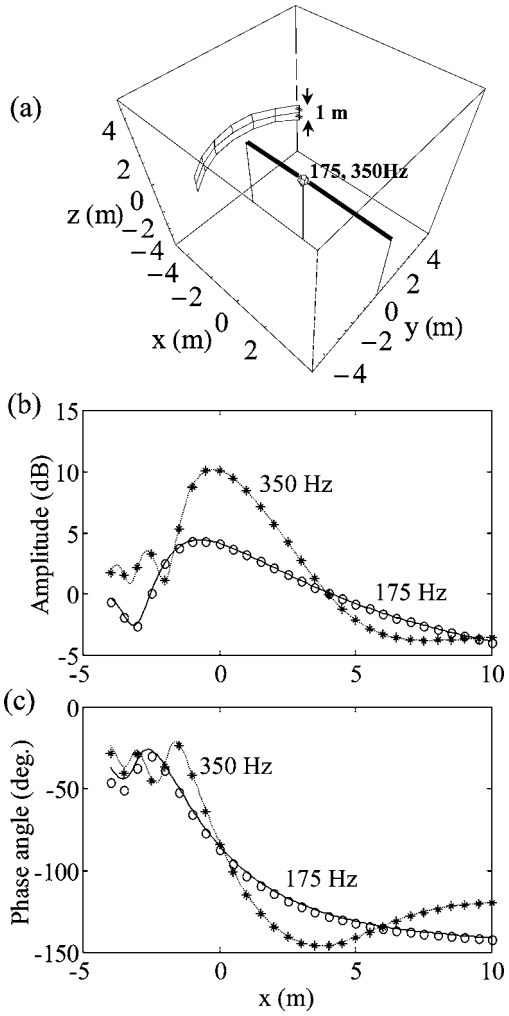


FIG. 4. Reflected sound from finite sectors of the circular cylinder with 5 m radius and directional angles of 135° and 225°: (a) geometrical condition including a source (dodecahedron) and receivers (thick solid line); (b), (c) amplitude and phase angle (relative values) of the reflected sound by Eq. (20) (lines) and the double integral Eq. (15) (marks), parameter is the frequency.

$$g_s(\eta) = \frac{u(\eta) + \nu(\xi_b, \eta)}{\sqrt{u(\eta)\nu(\xi_b, \eta)}\,\varphi_s(\eta)^{(3/2)}} \frac{[\partial\nu(\xi, \eta)^2/\partial\xi]_{\xi=\xi_b}}{2\nu(\xi_b, \eta)} \frac{h_\eta(\xi_b, \eta)}{h_\xi(\xi_b, \eta)},$$

$$g_e(\eta, z_j) = \frac{1}{U(\eta, z_j)(z_j - z_r) + V(\xi_b, \eta, z_j) z_j} \times \frac{[\partial\nu(\xi, \eta)^2/\partial\xi]_{\xi=\xi_b}}{2V(\xi_b, \eta, z_j)} \frac{h_\eta(\xi_b, \eta)}{h_\xi(\xi_b, \eta)}, \tag{21}$$

$$\varphi_s(\eta) = \sqrt{(u(\eta) + \nu(\xi_b, \eta))^2 + z_r^2},$$

$$d_k(\eta, z_j) = k|\varphi_s(\eta) - \varphi(\eta, z_j)|.$$

In the traditional (nonuniform) stationary phase method, the solutions of Eq. (18) are given only by I_s and I_e. Another function I_u is introduced to avoid that I_e diverges when the stationary point approaches the integration boundary ($z_s \to z_j$). The transition regions (Regions II and III in Fig. 2), in which I_u should be used, are defined by

$$d_k(\eta, z_j) < \pi. \tag{22}$$

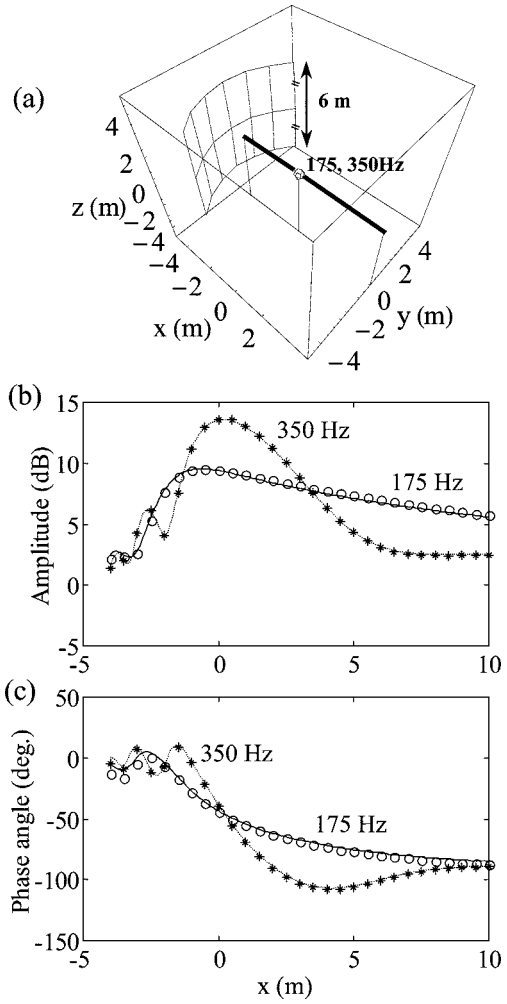


FIG. 5. Same as Fig. 4 but for different condition.

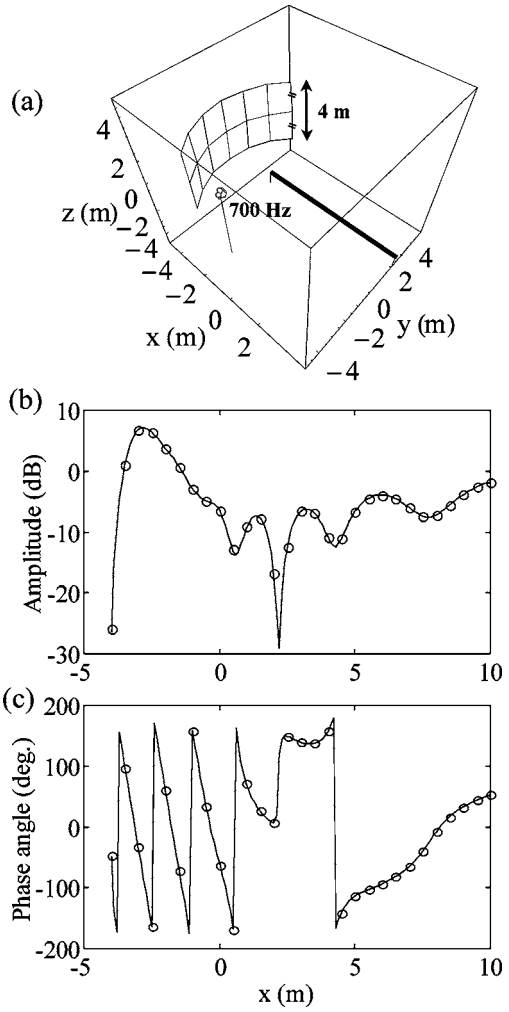


FIG. 6. Same as Figs. 4 and 5 but for different condition.

Equation (15) resulted in the single integral with respect to η, which is obviously the suitable form to the numerical integration.

B. Numerical accuracy of the asymptotic solution using Eq. (20)

In order to examine the numerical accuracy of the asymptotic solution using Eq. (20), the reflected sound from the circular cylinder was calculated and compared with that obtained by the double integral Eq. (15), which is the most exact solution under the assumption of TPA. Provided that the radius is 5 m and the directional angles are 135° and 225°, the length of cylinder, the source and receiver positions, and the frequency were changed variously.

Figure 3 shows the allowable distance between the sectional plane $z=0$ and a curved edge as contour lines in the coordinates representing the source position, which is calculated by Eq. (13) for the frequency 175, 350, and 700 Hz. Using this figure, the variables excepting the receiver position were limited. It is noted that TPA is inapplicable for frequencies lower than 109 Hz because D_{10} in Eq. (11) is greater than 0.1.

Figures 4 and 5 are some calculation results of the reflected sound field. In the top drawing the geometrical condition is shown; in the mid and bottom graphs the relative values of the amplitude and the phase angle are plotted. The location of the stationary point is fixed in Regions II and I under the conditions in Figs. 4 and 5, respectively. When the frequency is 175 Hz, the maximum errors of the amplitude and the phase angle are about 0.4 dB and 8°, respectively (solid line versus circle). Reductions of errors in the calculation results for the frequency 350 Hz (broken line versus asterisk) follow the general property of asymptotic solutions (the larger the error, the shorter the wavelength is). On the other hand, Fig. 6 shows the calculation result under a more complicated condition where the location of the stationary point changes over all regions—Regions I–IV with the moving of the receiver. The asymptotic solution achieves good accuracy as well in this extreme case (solid line versus circle).

The examined cylinder sizes are typical as the boundaries in actual rooms, such as the ceilings, surfaces around the stage, and the rear walls (Beranek, 2004). Besides, the similarity law is established between the wall dimension and the wavelength. For example, if the wall dimension is doubled, one can reduce the frequency to half without the increase of error because the precision of the asymptotic solution depends on the radius to wavelength ratio under the

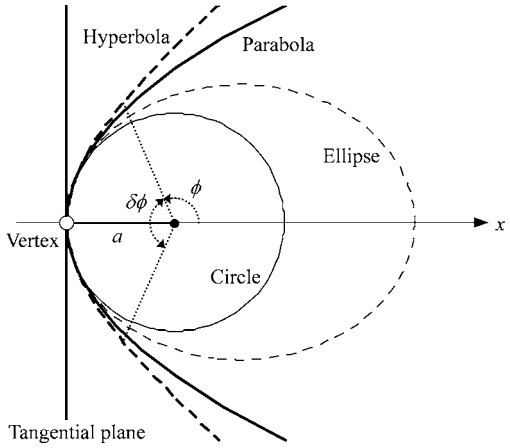


FIG. 7. Definition of reflection surfaces on infinitely long cylinders.

similarity law. All of the derived formulas can be applied to convex walls by multiplying Eq. (15) with −1 and ensure the same accuracy as in concave cases. However, utilized to thin panel reflectors, it should be understood that contributions of the rear face and the edges are neglected.

IV. ESSENTIAL PROPERTIES OF SOUND REFLECTION BY CYLINDRICAL WALLS

A. Definition of the reflection surface

In order to investigate the essential properties of sound reflections due to the finite curvatures of walls, infinitely long cylinders are considered. In addition, the receiver is prepared on the sectional plane $z=0$ where the source exists. In this case, since the stationary point z_s is always in region I and the function I_e vanishes, the reflected sound is given by

$$p_r = \frac{1-i}{2}\sqrt{\frac{k}{\pi}}\int_{\eta_1}^{\eta_2}\frac{\exp[ik(u+\nu)]}{\sqrt{u+\nu}}\frac{[\partial\nu^2/\partial\xi]_{\xi=\xi_b}}{2\nu\sqrt{u\nu}}\frac{h_\eta}{h_\xi}d\eta. \tag{23}$$

Equation (23) approximates the reflected sound from finite cylinders fairly if its length is sufficiently long that I_e is neglected. For example, I_e composing the calculation results in Fig. 5 is smaller than I_s by at least 10 dB and is not almost contributive.

Of the four cylinders with two dimensions, the curvature is not constant for the parabolic, elliptic, and hyperbolic cylinders and changes along the lateral surface. Therefore, assuming that the sound wave from a source impinges on the area around the vertex where the variation of the curvature is outstanding, the sound reflections are compared under the distinguishable condition. Thus, the symmetric portions around $(x,y)=(-b,0)$ of the parabolic cylinder, $(-a_e,0)$ on the major axis of the elliptic cylinder, and $(a_h,0)$ on the branch $x>0$ of the hyperbolic cylinder are chosen as the reflection surfaces. Also, supposed that the four cylinders always have the same curvature at the vertices, the lateral sizes are measured by the radius a and the included angle $\delta\phi=2(\pi-\phi)$ (ϕ is the directional angle) of the circular cylinder (Fig. 7). In this case, the specific parameters of the parabolic, elliptic, and hyperbolic cylinders are given as $b=a/2$, $a_e=a(1-\varepsilon^2)^{-1}$, $b_e=a(1-\varepsilon^2)^{-1/2}$, $a_h=a(\varepsilon^2-1)^{-1}$, and $b_h=a(\varepsilon^2-1)^{-1/2}$, where ε is the eccentricity (refer to Table I). Moreover, the integration variable in Eq. (23) is related with the directional angle ϕ as follows:

$$\text{Parabolic cylinder}\quad \eta \triangleq \gamma = \sqrt{2b}\{\cos\phi + (1+\sin^2\phi)^{1/2}\}\sin^{-1}\phi, \tag{24a}$$

$$\text{Elliptic cylinder}\quad \eta \triangleq \varphi = \sin^{-1}\left[\frac{b_e\sin\phi}{a_e}\frac{d_e^2\cos\phi + a_e(a_e^2+d_e^2\sin^2\phi)^{1/2}}{a_e^2\sin^2\phi + b_e^2\cos^2\phi}\right], \tag{24b}$$

$$\text{Hyperbolic cylinder}\quad \eta \triangleq \rho = \sinh^{-1}\left[\frac{b_h\sin\phi}{a_h}\frac{d_h^2\cos\phi + a_h(a_h^2+d_h^2\sin^2\phi)^{1/2}}{a_h^2\sin^2\phi - b_h^2\cos^2\phi}\right]. \tag{24c}$$

B. Reflected sound field at the geometrical focus under the total focusing

If a point source is at the geometrical focus of the circular and the elliptic cylinder, all of the reflected rays concentrate on their geometrical focuses (refer to Table I). Under those conditions, i.e., the total focusing, the integral in Eq. (23) can be solved in closed forms because $u+\nu$ in the phase term is constant. Reflected rays are also collected perfectly in the geometrical focus when a source is at a considerably large distance on the principal axis of the parabolic cylinder, that is, the plane wave incidence can be assumed. The reflected sound in this case is obtained by setting $p_d = \exp(-ikx)$ and using $\int_{-\infty}^{\infty}\exp(ikV)(\pi V)^{-1}dz = iH_0(k\nu)$ in Eqs. (B3) and (B4), where H_0 is the Hankel function of zeroth order. As a result, the reflected sound field at the geometrical focus under the total focusing is given as

$$\text{Circular cylinder}\quad p_r = \sqrt{\frac{ka}{i\pi}}(\phi_2-\phi_1)\times\frac{e^{i2ka}}{2a}, \tag{25a}$$

$$\text{Parabolic cylinder}\quad p_r = \sqrt{\frac{2kb}{i\pi}}\{\sinh^{-1}[(2b)^{-1/2}\gamma_2] - \sinh^{-1}[(2b)^{-1/2}\gamma_1]\}\times e^{i2kb}, \tag{25b}$$

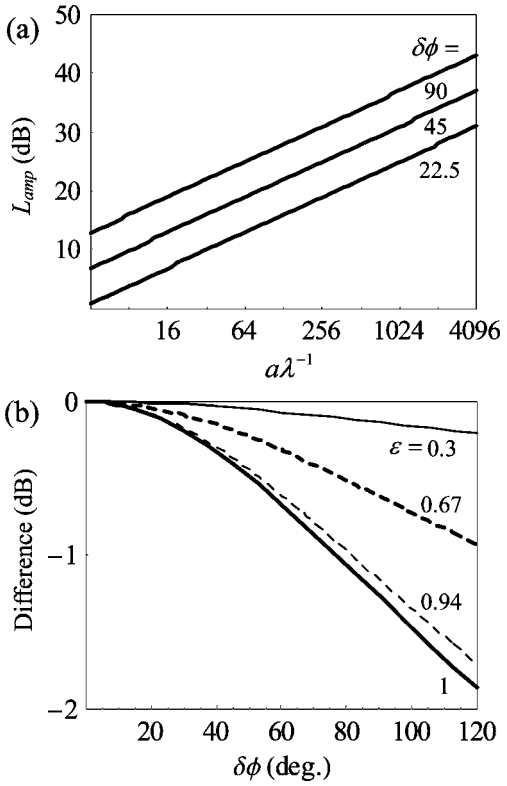


FIG. 8. Focusing effect of circular, parabolic, and elliptic cylinders: (a) L_{amp} of circular cylinder as a function of the radius to wavelength ratio $a\lambda^{-1}$, the parameter $\delta\phi$ is the included angle; (b) difference with L_{amp} of parabolic and elliptic cylinders as a function of included angle, the parameter ε is the eccentricity ($\varepsilon=1$ for parabolic cylinders).

$$\text{Elliptic cylinder} \quad p_r = \sqrt{\frac{ka_e}{i\pi}}\{E_1(\varphi_2, 1-a_e^2 b_e^{-2}) - E_1(\varphi_1, 1-a_e^2 b_e^{-2})\} \times \frac{e^{i2ka_e}}{2a_e}, \tag{25c}$$

where $E_1(\varphi,x)=\int_0^{\varphi}(1-x\sin^2 t)^{-1/2}dt$ is the elliptic integral of the first kind.

One of the most reasonable ways to investigate the sound reflection from the curved wall is to compare with that from an infinite plane surface. Then, the normalized amplitude by the reflected sound from the tangential plane at the vertex L_{amp} is introduced. In Eqs. (25a)–(25c), the terms following the operator $\times$ represent the reflected sound from that plane. Substituting $b=a/2$ and $a_e=a(1-\varepsilon_e^2)^{-1}$ into Eqs. (25b) and (25c), it is found that the normalized amplitude L_{amp} increases at a rate of 3 dB as ka is doubled. This result follows that cylindrical walls curve in a single direction, compared with the rate 6 dB for the paraboloid of revolution obtained by Wahlström (1985). On the other hand, the dependence on the sectional shape would be different in each cylinder.

In order to compare the focusing effects of the three cylinders precisely, L_{amp} was calculated by Eqs. (25a)–(25c). In the upper graph of Fig. 8 L_{amp} of the circular cylinders are plotted as functions of the radius to wavelength ratio $a\lambda^{-1}$; in the lower graph differences with L_{amp} of the parabolic cylinder ($\varepsilon\equiv 1$) and the elliptic cylinders ($\varepsilon=0.3$, 0.67, and 0.94) are shown. It is seen that the dependence on the sectional shape is small within the range $\delta\phi\leqslant 75°$ since the difference of L_{amp} is less than 1 dB. That is, the focused sound energy is roughly proportional to the size of the inscribed circle, $a\,\delta\phi$.

C. Spatial distribution of the reflected sound

As a next examination, the spatial distribution of the reflected sound was calculated by the numerical integration of Eq. (23) in order to investigate the reflected sound fields under rather general conditions where envelopes of the reflected rays, i.e., the caustics, are formed or concentrations of the reflected rays do not appear obviously.

As an example, Figs. 9–11 show the calculation results when the radius and the included angle are 10 m and 90° and the eccentricities of the elliptic and hyperbolic cylinders are 0.67 and 1.25, respectively. The average values of the normalized amplitude L_{amp} over one octave bandwidth for the midfrequency 500 Hz are presented as the contour lines in each graph (a)–(d), together with the caustics (solid line) and the shadow boundaries (broken line) obtained by the method described in Appendix C. Although the spatial distributions follow the shapes of the caustic and the shadow boundary, some common properties are found. For every cylinder, strong reflections are observed over a wide area in the illuminated zones. Considerably intense reflections arrive near the caustics, whose amplitudes take the maximum values around the caustic cusps. From the precise investigation on the band averaged L_{amp} by using magnified figures, it is found that the maximum values for the circular cylinder are about 18, 18, and 16 dB in Figs. 9–11, respectively, and equal to or greater than those for other cylinders by 1 dB. This result almost coincides with the focused sound energy in Fig. 8 [the point at $a\lambda^{-1}=14.5$ and $\delta\phi=90$ in (a), and the curves for $\varepsilon=0.67$ and 1 in (b)].

Figure 12 shows the calculation result when the included angle and the midfrequency are 60° and 250 Hz but other conditions are same as in Fig. 9. The spatial distributions for each cylinder are quite similar in contrast with those in Fig. 9, although the caustics or the shadow boundaries are still distinguishable. The maximum values of the band averaged L_{amp} are about 12 dB for every cylinder and equal to the focused sound energy in Fig. 8 as well.

D. Similarity of reflected sound fields

In this section, the physical condition under which the reflected sound fields from the four cylinders are substantially identical is derived.

Hereafter, the directional angle is redefined by $\chi=\pi-\phi$ measured from the negative side of the principal axis. Writing the functions in Eq. (23) for the circular cylinder as

$$u \triangleq u_c(\chi) = \sqrt{(a\cos\chi + x_0)^2 + (a\sin\chi - y_0)^2}, \tag{26a}$$

$$\nu \triangleq \nu_c(\chi) = \sqrt{(a\cos\chi + x_r)^2 + (a\sin\chi - y_r)^2}, \tag{26b}$$

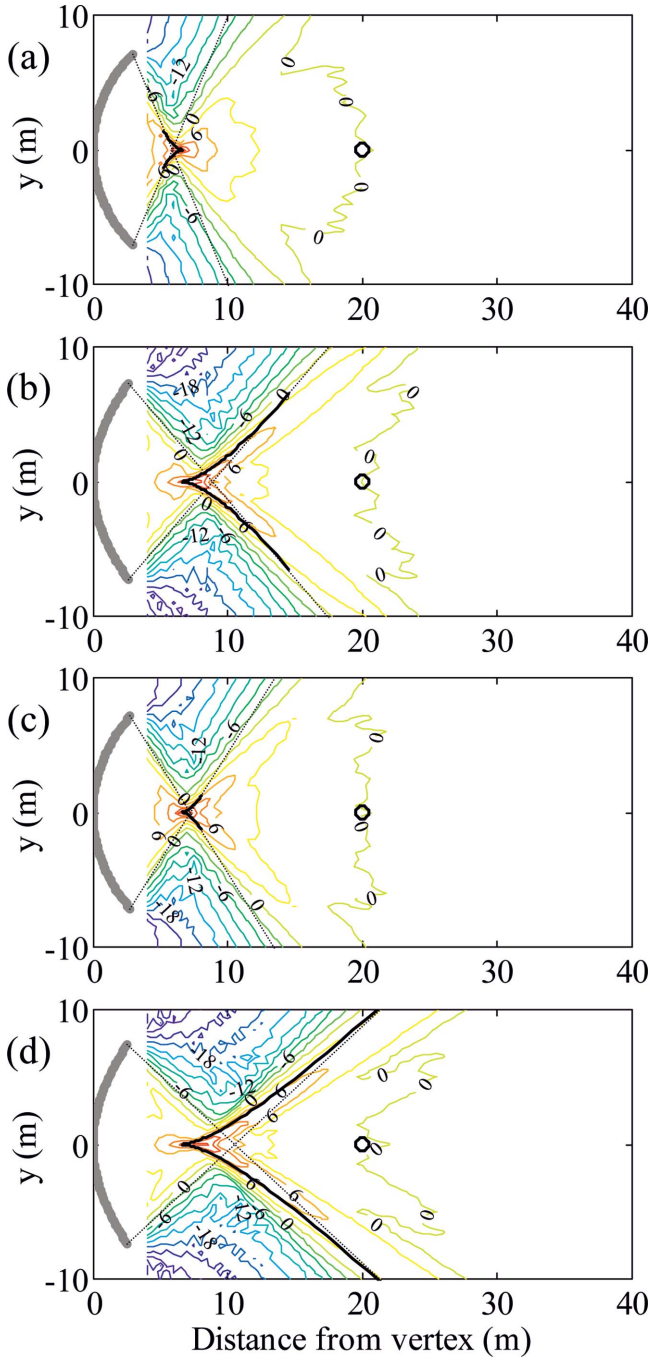


FIG. 9. Spatial distribution of reflected sound from four cylinders with 10 m radius and included angle of 90°: (a) circular, (b) parabolic, (c) elliptic (eccentricity 0.67), and (d) hyperbolic cylinder (eccentricity 1.25). Average value of L_{amp} over one octave bandwidth for midfrequency of 500 Hz is presented as contour lines together with source position (circle), the shadow boundaries (broken line), and the caustics (solid line).

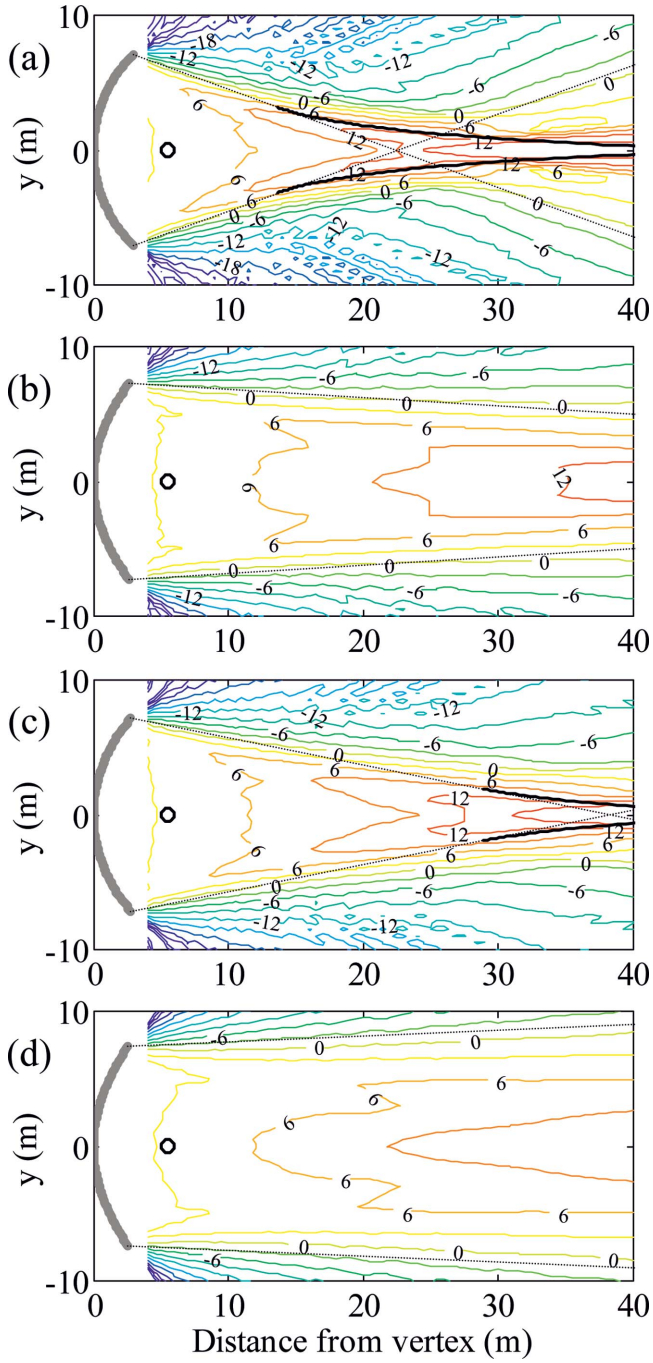


FIG. 10. Same as Fig. 9 but for different source position.

$$2^{-1}[\partial\nu^2/\partial r]_{r=a}h_\chi h_r^{-1}d\chi \triangleq t_c(\chi) = a(a + x_r\cos\chi - y_r\sin\chi)d\chi, \tag{26c}$$

the relations with the following functions for the parabolic cylinder are investigated:

$$u \triangleq u_p(\gamma) = \sqrt{\{(\gamma^2 - a)/2 - (x_0 + a/2)\}^2 + (\sqrt{a}\gamma - y_0)^2}, \tag{27a}$$

$$\nu \triangleq \nu_p(\gamma) = \sqrt{\{(\gamma^2 - a)/2 - (x_r + a/2)\}^2 + (\sqrt{a}\gamma - y_r)^2}, \tag{27b}$$

$$2^{-1}[\partial\nu^2/\partial\mu]_{\mu=\sqrt{a}}h_\gamma h_\mu^{-1}d\gamma \triangleq t_p(\gamma) = [\sqrt{a}\{(\gamma^2 + a)/2 + (x_r + a/2)\} - y_r\gamma]d\gamma. \tag{27c}$$

The variable γ is expanded into the Taylor series around $\chi = 0$ as

$$\gamma(\chi) = \sqrt{a}\{\chi - \chi^3/6 + 2\chi^5/15 + o(\chi^7)\}. \tag{28}$$

Also, the distance between the two cylinders is given as

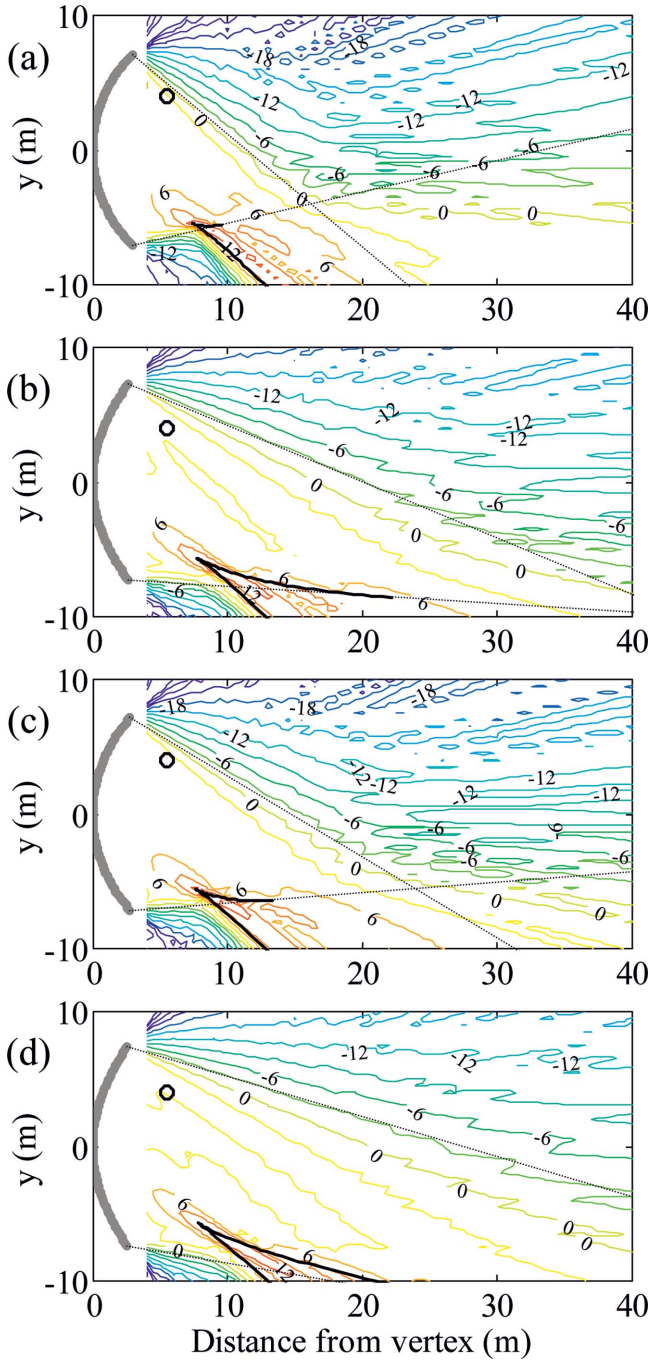


FIG. 11. Same as Figs. 9 and 10 but for different source position.

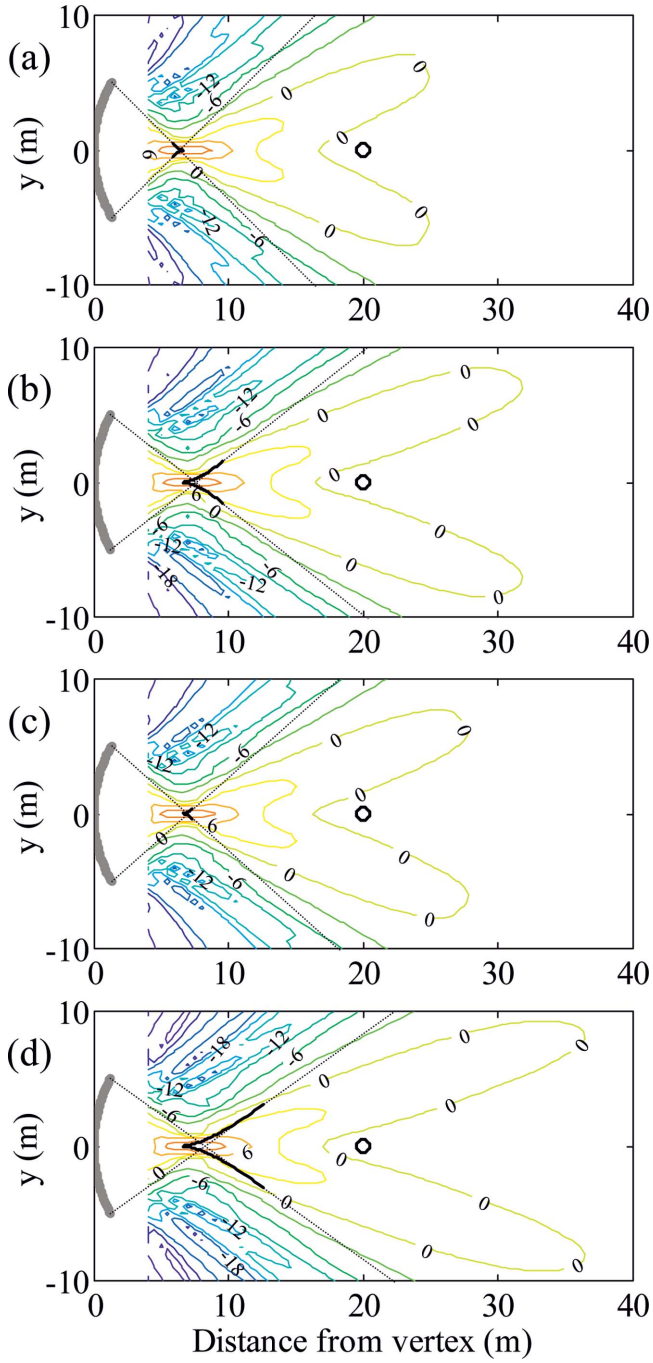


FIG. 12. Same as Fig. 9 but for included angle of 60° and the midfrequency 250 Hz.

$$d_{cp}(\chi) = a\{\chi^4/8 + o(\chi^6)\}. \tag{29}$$

The substitution of Eq. (28) into Eqs. (27a)–(27c) leads to

$$u_p(\chi) \approx u_c(\chi)\sqrt{1 + (a\chi^4/4)(a + \chi_0 - y_0\chi)u_c(\chi)^{-2}}, \tag{30a}$$

$$v_p(\chi) \approx v_c(\chi)\sqrt{1 + (a\chi^4/4)(a + x_r - y_r\chi)v_c(\chi)^{-2}}, \tag{30b}$$

$$t_p(\chi) \approx t_c(\chi) + (a\chi^3/8)\{4y_r + (2a + 5x_r)\chi\}d\chi. \tag{30c}$$

The square roots in Eqs. (30a) and (30b) can be approximated in the order of $o(u_c^{-4})$ and $o(v_c^{-4})$ because the second term is always much smaller than unity. As a result, the integrand of the parabolic cylinder g_p and that of the circular cylinder g_c are related by

$$g_p \approx \exp\left[ik\frac{a\chi^4}{8}\left\{\frac{a + x_0 - y_0\chi}{u_c(\chi)} + \frac{a + x_0 - y_0\chi}{v_c(\chi)}\right\}\right]g_c. \tag{31}$$

Accordingly, the reflected sound fields from the two cylinders with the directional angles $\chi = \pm\chi_c$ ($\chi_c > 0$) are identical when the following inequality is fulfilled:

$$G_{cp}(\pm\chi_c) \triangleq k\frac{a\chi_c^4}{8}\left|\frac{a + x_0 \mp y_0\chi_c}{u_c(\pm\chi_c)} + \frac{a + x_r \mp \gamma_r\chi_c}{v_c(\pm\chi_c)}\right| \ll 1. \tag{32}$$

It is assumed that both the source and the receiver are on the principal axis ($y_0 = y_r = 0$) as shown in the top drawing of

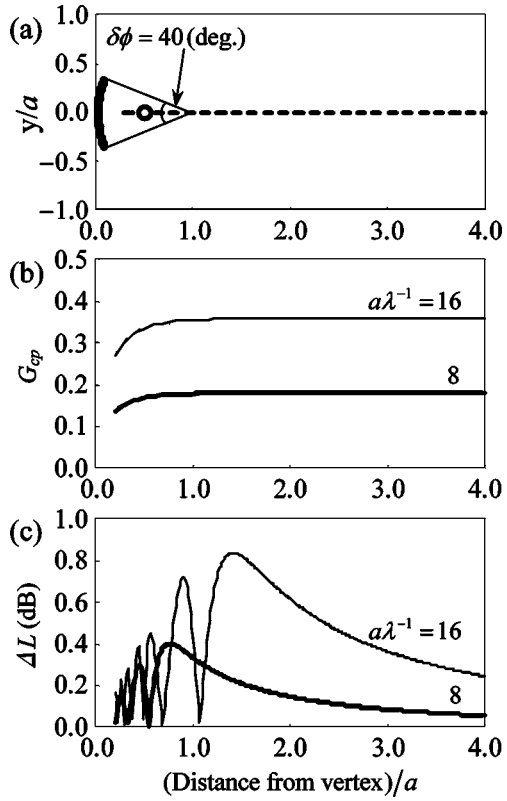


FIG. 13. Numerical examples of G_{cp} and ΔL for reflected sound field from circular and parabolic cylinder when source and receiver are on the principal axis, as functions of receiver distance from the vertex normalized by the radius a: (a) geometrical condition including a source (circle) and receivers (broken line); (b), (c) curves of G_{cp} and ΔL, parameter is the radius to wavelength ratio $a\lambda^{-1}$.

Fig. 13. The mid and bottom graphs in Fig. 13 are numerical examples of $G_{cp}(\chi_c)$ and the difference of the reflected sound energy ΔL between the circular and parabolic cylinders as functions of the receiver distance normalized by the radius a, for the included angle $\delta\phi = 2\chi_c = 40°$ and the radius to wavelength ratio $a\lambda^{-1} = 8$ and 16. One can presume a linear relationship between G_{cp} and the maximum value of ΔL, $\max(\Delta L)$. From further examinations for $\delta\phi \leqslant 40°$ (Fig. 14), it was found that the empirical relation $\max(\Delta L) \approx 2.4 G_{cp}$ holds, in addition, the curves of G_{cp} are flat at a certain distance from the cylinder and almost independent on the source position. Accordingly, the substitution of $x_0 = x_r = \infty$ into Eq. (32) yields a simple measure for the similarity as follows:

$$G_{cp} \to 2kd_{cp}(\chi_c) \approx a\lambda^{-1}\delta\phi^4/10 \ll 1, \tag{33a}$$

$$\max(\Delta L) \to 4.8kd_{cp}(\chi_c) \approx a\lambda^{-1}\delta\phi^4/4. \tag{33b}$$

Equation (33a) means that one can regard the two cylinders as acoustically equivalent walls when the distance between them d_{cp} is much smaller than the wavelength. Thus, its significance is still maintained even if the source or the receiver is off the principal axis. For the elliptic or the hyperbolic cylinder, Eqs. (33a) and (33b) can be corrected by multiplying the eccentricity ε because the distance from the circular cylinder d is given by $d \approx \varepsilon d_{cp}$.

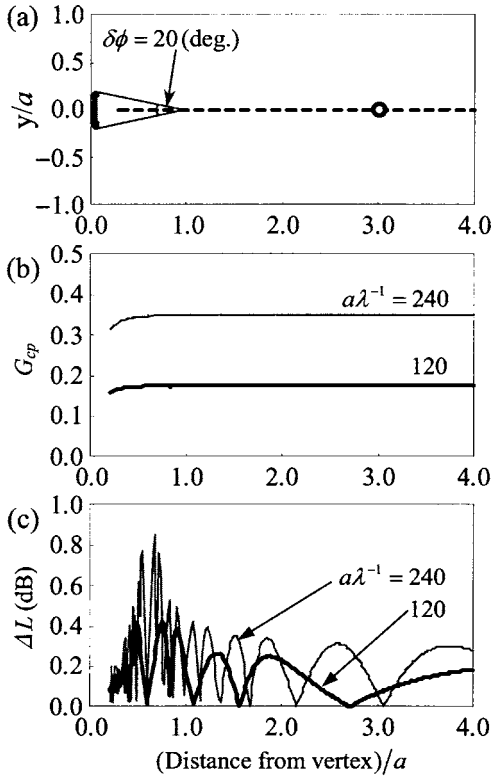


FIG. 14. Same as Fig. 13 but for different condition.

E. Reflected sound field from the cylindrical wall with a small included angle

When the included angle is so small that $\cos\chi \approx 1$ is satisfied, the reflected sound from the two-dimensional cylinder would be identical in a wide range of the radius to wavelength ratio $a\lambda^{-1}$, excepting the hyperbolic cylinder with extremely large eccentricity. The accuracy of 0.4% is ensured in this approximation of the cosine function if $|\chi|$ is smaller than 6°. In this case, it is confirmed by manipulating the caustic equation in Appendix C that the caustics of each cylinder shorten and concentrate to the specified point,

$$\begin{aligned} s_F &= \frac{as_0}{2s_0 - a}\frac{1}{1 + 2y_0^2 s_0^{-1}(2s_0 - a)^{-1}} + o(\chi), \\ y_F &= \frac{-ay_0}{(2s_0 - a)}\frac{1}{1 + 2y_0^2 s_0^{-1}(2s_0 - a)^{-1}} + o(\chi), \end{aligned} \tag{34}$$

where (s_F, y_F) is the relative position to the vertex, and s_0 is the source distance along the principal axis. Based on the geometrical acoustics, Kuttruff (2000, p. 107) proposed a simple formula for the normalized amplitude L_{amp} of the curved wall with a small included-angle when the source and the receiver are on the principal axis, and his formula is expressed as

$$L_{\text{amp}} = 10\log_{10}\left|\frac{a(s_0 + s_r)}{2s_0 s_r - a(s_0 + s_r)}\right|^n, \tag{35}$$

where s_r is the receiver distance, and $n = 1$ for cylindrical walls. L_{amp} in Eq. (35) can be regarded as the reflection factor of the curved wall for the normal incidence. In the

remaining part, a solution of the reflected sound field under that condition is derived to revise Eq. (35).

Using $\sin\chi=\chi+o(\chi^3)$ and $\cos\chi=1-\chi^2/2+o(\chi^4)$, Eqs. (26a) and (26b) are rewritten as

$$u_c(\chi)\approx(a+x_0)\sqrt{1-ax_0\chi^2(a+x_0)^{-2}},$$
$$\nu_c(\chi)\approx(a+x_r)\sqrt{1-ax_r\chi^2(a+x_r)^{-2}}, \tag{36a}$$

$$t_c(\chi)\approx(a+x_r)[1-x_r\chi^2(a+x_r)^{-1}/2]. \tag{36b}$$

The root functions in Eq. (36a) can be approximated in the order of $o(\chi^{-4})$ if $a|x_0|\chi^2(a+x_0)^{-2}\ll 1$ and $a|x_r|\chi^2(a+x_r)^{-2}\ll 1$ are satisfied. These assumptions are valid so long as the source and the receiver are not near the cylinder. Substituting the obtained functions into Eq. (23) and integrating within $\chi=[-\chi_c,\chi_c]$, the following formula is obtained:

$$p_r\approx\sqrt{\frac{k}{2i}}\frac{a\sqrt{x_0+x_r+2a}}{\sqrt{(x_0+a)(x_r+a)}}\frac{\mathrm{erf}(\sqrt{ikaC}\chi_c)}{\sqrt{ikaC}}\times\frac{\exp[ik(x_0+x_r+2a)]}{x_0+x_r+2a}, \tag{37}$$

where

$$C\equiv\{x_0(a+x_0)^{-1}+x_r(a+x_r)^{-1}\}/2, \tag{38}$$

and $\mathrm{erf}(x)=2\pi^{-1/2}\int_0^x\exp(-t^2)dt$ is the error function. Since the reflected sound from the tangential plane at the vertex is given by the term following the operator $\times$ in Eq. (37), one obtains L_{amp} of the cylindrical wall with the small included-angle $\delta\phi=2\chi_c$,

$$L_{\mathrm{amp}}=10\log_{10}\left|\frac{a(s_0+s_r)}{2s_0s_r-a(s_0+s_r)}\right|+20\log_{10}\left|\mathrm{erf}\left[\sqrt{\frac{ika}{2}\frac{|2s_0s_r-a(s_0+s_r)|}{s_0s_r}}\chi_c\right]\right|, \tag{39}$$

where the source position and the receiver position are redefined in the same way as Eq. (35). The first term in Eq. (39) coincides with the geometrical acoustics solution Eq. (35) and diverges when the receiver approaches the focal point $s_F=as_0(2s_0-a)^{-1}$ in Eq. (34). On the other hand, the second term is the correction based on the wave theory and L_{amp} remains finite by virtue of this term excepting the condition $k\to\infty$ where the geometrical acoustics is rigorous.

A definite measure to estimate the validity of the geometrical acoustics can be derived by evaluating the error function in Eq. (39). If one requires the precision about 1 dB, i.e., $|\mathrm{erf}(x)|<10^{1/20}$, the following restriction is obtained:

$$G_{\mathrm{GA}}\triangleq a\lambda^{-1}(\delta\phi/2)^2|1-(as_0^{-1}+as_r^{-1})/2|>1. \tag{40}$$

When s_r changes, the term $|1-(as_0^{-1}+as_r^{-1})/2|$ in Eq. (40) varies monotonously between three specific values: infinity on the cylinder; zero at the focal point; the asymptotic value $|1-(s_0/a)^{-1}/2|$ at the infinite distance from the cylinder. This suggests that the applicability of the geometrical acoustics depends on the source position and the receiver position.

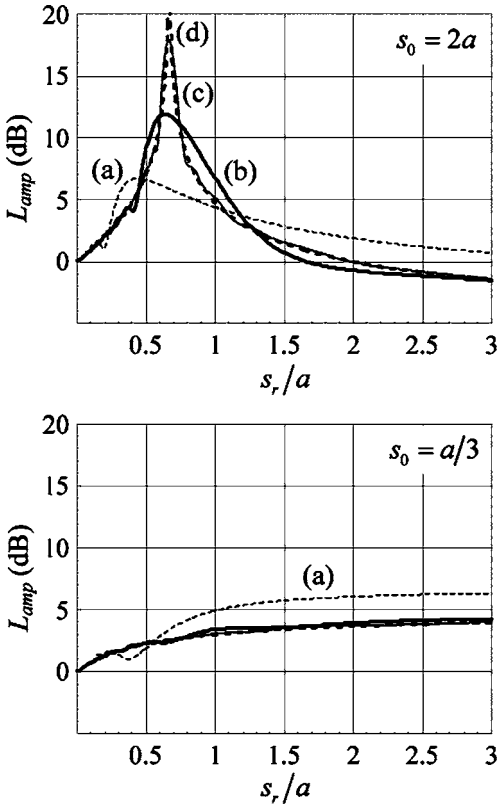


FIG. 15. Plot of L_{amp} averaged over one octave bandwidth vs the receiver distance to radius ratio s_r/a, for circular cylinder with the included angle 10°. Upper: the source distance is $s_0=2a$; lower: the source distance is $s_0=a/3$. The radius to wavelength ratio $a\lambda^{-1}$ is (a) 60, (b) 235, (c) 941, and (d) ∞.

Figure 15 shows the calculation results of L_{amp} for the cylinders with the included angle $\delta\phi=10°$ as functions of the receiver distance to radius ratio s_ra^{-1}, where two source positions $s_0=2a$ (the upper graph) and $a/3$ (the lower graph) are considered. The curves (a) to (d) correspond to the averaged values of L_{amp} over one octave bandwidth for the radius to wavelength ratio, $a\lambda^{-1}=60$, 235, 941, and ∞, respectively. The curve (d) coincides with the geometrical acoustics solution. In the upper graph where the focal point is at $s_r/a=0.67$, it is interesting that the maximum point of L_{amp} is not always identical with the focal point. However, the difference between L_{amp} values at the maximum and the focal point is less than 1 dB. Most important is that the peak of L_{amp} shifts toward the focal point while growing at a rate of 3 dB per octave with the increase of $a\lambda^{-1}$, and L_{amp} at locations distant from the focal point gradually converges to the geometrical acoustics solution. In the lower graph where no focal point exists, the calculation result based on the geometrical acoustics coincides with the wave solutions for sufficiently large $a\lambda^{-1}$ ($a\lambda^{-1}=235$ or 941 in this case) over the observation area.

That is to say, the geometrical acoustics is valid when the focal point (or the caustic) is not formed and the wavelength is sufficiently shorter than the wall dimension $(a\lambda^{-1}(\delta\phi/2)^2\gg 1)$.

V. CONCLUSION

By introducing the tangential plane approximation to the spherical wave reflection problem by the smooth, concave

cylindrical surface with larger dimension compared to the wavelength, the asymptotic expression of the reflected sound field was derived. First, the physical condition under which the expression is valid was obtained, and it was shown that this condition holds for practical room acoustical applications, such as sound reflections by the stage enclosure, the rear wall, the balcony front, and so on. Next, the reflected sound from the infinitely long cylinder was examined by means of the numerical calculation since the contribution of the edge [I_e in Eq. (20)] is not very significant in practical cases although the realistic cylindrical walls in rooms have finite length. As a result, following essential features were obtained:

(a) When a cylindrical wall is modeled by four cylinders: the circular, the parabolic, the elliptic, and the hyperbolic cylinder (Table I), the corresponding reflected sound field is significantly different when $\varepsilon a\lambda^{-1}\delta\phi^4/10 \ll 1$ does not hold [Eq. (33a)]. Here, ε, a, $\delta\phi$, and λ are the eccentricity of the cylinder, the radius and the included angle of the inscribed circle, and the wavelength, respectively.

(b) Under the total focusing, the amplitude of the reflected sound at the geometrical focus increases by the ratio of 3 dB per octave and are approximately proportional to the size of the inscribed circle, $a\delta\phi$. In addition, the reflected sound field around the caustic cusp has similar characteristics to those.

(c) The geometrical acoustics is valid unless the focal point or the caustic is formed. However, the correction by the wave theory is required if $a\lambda^{-1}(\delta\phi/2)^2 \gg 1$ is not satisfied [Eq. (39)].

Finally, the reflection factor of the cylindrical wall was given [Eq. (35)], which corresponds to the geometrical acoustics formula by Kuttruff (2000, p. 107) under the condition $k \to \infty$.

APPENDIX A: EVALUATION OF THE INTEGRAL EQ. (6)

The integration with respect to the variable l in Eq. (6) is written in the general form,

$$I(k,\varphi) = \int_{l_1}^{l_2} g(l,\varphi)\exp[ikh(l,\varphi)]dl. \tag{A1}$$

If the stationary point [the solution of $h'(l,\varphi) \triangleq \partial h(l,\varphi)/\partial l = 0$] is not near the integration boundaries l_1 and l_2, Eq. (A1) can be evaluated under the condition $k \to \infty$ by (Borovikov, 1994)

$$I(k,\varphi) = \frac{\exp[ikh(l_1,\varphi)]}{h'(l_1,\varphi)}\sum_{n=1}^{\infty}(i/k)^n g_{n-1}(l_1,\varphi) - \frac{\exp[ikh(l_2,\varphi)]}{h'(l_2,\varphi)}\sum_{n=1}^{\infty}(i/k)^n g_{n-1}(l_2,\varphi), \tag{A2}$$

$$g_{n+1}(l,\varphi) = \{g_n(l,\varphi)/h'(l,\varphi)\}', \quad g_0(l,\varphi) = g(l,\varphi). \tag{A3}$$

Substituting the integrands of Eq. (6) into Eq. (A2) and ending the operation at $n \leqslant 3$, the following solution is obtained:

$$I(k,\varphi) = -ik^{-1}I_1(\varphi) + k^{-2}\{I_3(\varphi)\cos\theta - 4I_2(\varphi)r_s^{-1}\} + O(k^{-3}). \tag{A4}$$

Here,

$$I_1 = \frac{\kappa(\varphi)}{\{1-\sin\theta\cos(\varphi-\varphi_0)\}} + \frac{\kappa(\varphi)}{\{1-\sin\theta\cos(\varphi-\varphi_0)\}^2},$$

$$I_2 = \frac{\kappa(\varphi)}{\{1-\sin\theta\cos(\varphi-\varphi_0)\}^3}, \tag{A5}$$

$$I_3 = \frac{\kappa(\varphi)^2}{\{1-\sin\theta\cos(\varphi-\varphi_0)\}^3} + \frac{3\kappa(\varphi)^2}{\{1-\sin\theta\cos(\varphi-\varphi_0)\}^4}.$$

The integral with respect to the directional angle φ is calculated, substituting Eqs. (A4) and (A5) into Eq. (6). After applying the transformation $\sin\theta\cos(\varphi-\varphi_0) = s\cos\varphi + t\sin\varphi$ ($s \equiv \sin\theta\cos\varphi_0, t \equiv \sin\theta\sin\varphi_0$) to the denominators in Eq. (A5), the following integration formulas and their derivatives lead to Eqs. (7)–(9):

$$\int_0^{2\pi}\frac{d\varphi}{x - s\cos\varphi - t\sin\varphi} = \frac{2\pi}{\sqrt{x^2-s^2-t^2}},$$

$$\int_0^{2\pi}\frac{\cos\varphi d\varphi}{x - s\cos\varphi - t\sin\varphi} = \frac{2\pi}{x^2-s^2-t^2+x\sqrt{x^2-s^2-t^2}}, \tag{A6}$$

$$\int_0^{2\pi}\frac{\cos^2\varphi d\varphi}{x - s\cos\varphi - t\sin\varphi} = 2\pi\frac{x(-s^2+t^2)\sqrt{x^2-s^2-t^2} + t^2(t^2-x^2) + s^2(t^2+x^2)}{(s^2+t^2)^2\sqrt{x^2-s^2-t^2}}.$$

APPENDIX B: REPRESENTATION OF REFLECTED FIELDS BY THE CURVILINEAR COORDINATES

When expressing the arclengths corresponding to the variations $d\eta$ along $\xi = constant$ and $d\xi$ along $\eta = constant$ as ds_ξ and ds_η, respectively, they have the following relations:

$$ds_\xi = h_\eta(\xi,\eta)d\eta, \quad ds_\eta = h_\xi(\xi,\eta)d\xi, \tag{B1}$$

where h_ξ and h_η are the scale factors defined by Eq. (17a). Accordingly, the surface element $dS(\mathbf{r})$ at $\mathbf{r} = (x(\xi_b,\eta), y(\xi_b,\eta), z)$ on the cylinder $\xi_b = constant$ ($\eta_1 \leqslant \eta \leqslant \eta_2, z_1 \leqslant z \leqslant z_2$) is

$$dS(\mathbf{r}) = ds_\xi\, dz = h_\eta(\xi_b,\eta)d\eta\, dz. \tag{B2}$$

Writing the distances $|\mathbf{r}-\mathbf{R_s}|$ and $|\mathbf{r}-\mathbf{R}|$ as in Eq. (17b), the functions in Eq. (3) are represented as follows:

$$p_d(k,\mathbf{r},\mathbf{R_s}) = \frac{\exp[ikU(\eta,z)]}{U(\eta,z)}, \tag{B3}$$

$$\frac{\partial}{\partial N}\frac{\exp[ikV(\xi_0,\eta,z)]}{4\pi V(\xi_b,\eta,z)}$$
$$=\frac{h_\xi^{-1}}{4\pi}\left[\frac{\partial V}{\partial \xi}\right]_{\xi=\xi_b}\frac{\partial}{\partial V}\frac{\exp(ikV)}{V}$$
$$=\frac{h_\xi^{-1}}{4\pi}\frac{1}{2V}\left[\frac{\partial \nu^2}{\partial \xi}\right]_{\xi=\xi_b}\frac{(ikV-1)\exp(ikV)}{V^2}. \quad \text{(B4)}$$

Assuming that the receiver is far from the cylinder as compared with the wavelength ($kV \gg 1$) and substituting Eqs. (B2)–(B4) into Eq. (3), Eq. (15) is obtained.

APPENDIX C: THE CAUSTICS FORMED BY CYLINDERS

When a source is at $\mathbf{R_s}=(x_0,y_0,0)$, and point $\mathbf{p}=(x_p,y_p,z_p)$ on a ray reflected from $\mathbf{r}=(x(\xi_b,\eta),y(\xi_b,\eta),z)$ on the cylinder $\xi=\xi_b=constant$ is given by

$$x_p = x(\eta)+\{u_x(\eta)-2n_x(\eta)^2u_x(\eta) - 2n_x(\eta)n_y(\eta)u_y(\eta)\}U(\eta,z)^{-1}t,$$
$$y_p = y(\eta)+\{u_y(\eta)-2n_y(\eta)^2u_y(\eta) - 2n_x(\eta)n_y(\eta)u_x(\eta)\}U(\eta,z)^{-1}t,$$
$$z_p = \{1+U(\eta,z)^{-1}\}zt \quad (t>0), \quad \text{(C1)}$$

where the vector element of $\mathbf{U}=\mathbf{r}-\mathbf{R_s}$ and the normal $\hat{\mathbf{N}}$ were written as $(u_x(\eta),u_y(\eta),z)$ and $(n_x(\eta),n_y(\eta),0)$, respectively, and $U=|\mathbf{U}|$ is defined by Eq. (17b).

The following equation is established on the caustics $\mathbf{c}=(x_c,y_c,z_c)$ (Kravtsov and Orlov, 1999).

$$J=\begin{vmatrix} \partial x_p/\partial\eta & \partial y_p/\partial\eta & \partial z_p/\partial\eta \\ \partial x_p/\partial z & \partial y_p/\partial z & \partial z_p/\partial z \\ \partial x_p/\partial t & \partial y_p/\partial t & \partial z_p/\partial t \end{vmatrix}=0, \quad \text{(C2)}$$

which has two real roots t_1 and t_2,

$$t_1=-U,$$
$$t_2=\frac{f_1u_x+f_2u_y}{2f_3u_x^2+2f_4u_y^2+4f_5u_xu_y+f_6(u_xu_y'-u_x'u_y)}U, \quad \text{(C3)}$$

where the operator $'$ means $\partial/\partial\eta$, and the following abbreviations are used:

$$f_1=2n_xn_yx'-(2n_x^2-1)y', \quad f_2=(2n_y^2-1)x'-2n_xn_yy',$$
$$f_3=-(2n_x^2-1)n_xn_y'+(2n_x'+1)n_x'n_y,$$
$$f_4=-(2n_y^2+1)n_xn_y'+(2n_y^2-1)n_x'n_y,$$
$$f_5=(2n_y^2-1)n_xn_x'-(2n_x^2-1)n_y'n_y, \quad f_6=(2n_x^2-1)+2n_y^2. \quad \text{(C4)}$$

Substituting $t=t_2$ into Eq. (C1), the caustic equation is obtained. It is obvious that the caustics form a cylindrical surface whose polar axis is parallel to that of the reflection surface.

Babič, V. M. and Buldyrev, N. Y. (**1991**). *Short-Wavelength Diffraction Theory* (Springer, Berlin), Chap. 2.

Bass, F. G. and Fuks, I. M. (**1979**). *Wave Scattering from Statistically Rough Surfaces* (Pergamon, Oxford), Chap. 7, pp. 220–229.

Belobrov, A. V. and Fuks, I. M. (**1985**). "Short-wave asymptotic analysis of the problem of acoustic wave diffraction by a rough surface," Sov. Phys. Acoust. **31**, 442–445.

Beranek, L. L. (**2004**). *Concert Halls and Opera House* (Springer, New York).

Borovikov, V. A. (**1994**). *Uniform Stationary Phase Method* (The Institution of Electrical Engineers, London), Chap. 2.

Cremer, L. and Muller, H. (**1982**). *Principal and Applications of Room Acoustics* (Applied Science, London/New York), Vol. **1**, Chap. 1.3.

Filippi, P., Habault, D., Lefebvre, J. P., and Bergassoli, A. (**1999**). *Acoustics: Basic Physics, Theory and Methods* (Academic, London), Chap. 3.

Fuks, I. M. and Voronovich, A. G. (**1999**). "Wave diffraction by a concave statistically rough surface," Waves Random Media **9**, 501–520.

Kravtsov, Yu. A. and Orlov, Yu. I. (**1999**). *Caustics, Catastrophes and Wave Fields* Springer, Berlin), Chaps. 2–3.

Kuttruff, H. (**1992**). "Some remarks on the simulation of sound reflection from curved walls," Acustica **77**, 176–182.

Kuttruff, H. (**2000**). *Room Acoustics* (Spon, London), Chap. 4, p. 107.

Liszka, E. G. and McCoy, J. J. (**1982**). "Scattering at a rough boundary—Extensions of the Kirchhoff approximation," J. Acoust. Soc. Am. **71**, 1093–1100.

Lynch, P. J. (**1970**). "Curvature corrections to rough-surface scattering at high frequencies, " J. Acoust. Soc. Am. **47**, 804–815.

McCammon, D. F. and McDaniel, S. T. (**1986**). "Surface reflection: On the convergence of a series solution to a modified Helmholz integral equation and the validity of the Kirchhoff approximation," J. Acoust. Soc. Am. **79**, 64–70.

Morse, P. M. and Feshbach, H. (**1953**). *Methods of Theoretical Physics, Part I* (McGraw-Hill, New York), Chap. 5.

Rayleigh, L. (**1926**). *The Theory of Sound* (Macmillan, London), Vol. **II**, pp. 124–128.

Voronovich, A. G. (**1994**). *Wave Scattering from Rough Surfaces* (Springer, Berlin), Chap. 5.

Voronovich, A. G. (**1996**). "On the theory of electromagnetic waves scattering from the sea surface at low grazing angles," Radio Sci. **31**, 1519–1530.

Wahlström, S. (**1985**). "The parabolic reflector as an acoustical amplifier," J. Audio Eng. Soc. **33**, 418–429.

Wirgin, A. (**1989**). "Scattering from hard and soft corrugated surfaces: Iterative corrections to the Kirchhoff approximation through the extinction theorem," J. Acoust. Soc. Am. **85**, 670–679.

On the measurement of the Young's modulus of small samples by acoustic interferometry

F. Simonetti[a] and P. Cawley
Department of Mechanical Engineering, Imperial College, London, SW7 2AZ, United Kingdom

A. Demčenko
Ultrasound Institute, Kaunas University of Technology, LT-51368 Kaunas, Lithuania

(Received 5 January 2005; revised 4 May 2005; accepted 5 May 2005)

This paper describes and validates an interferometric technique for the measurement of the Young's modulus of limited size samples. The technique is a generalization of the torsional guided wave interferometry introduced in a previous paper [Simonetti and Cawley, J. Acoust. Soc. Am. **115**, 157–164 (2004)] for the characterization of the shear properties of soft materials. The method is based on the transmission of the fundamental longitudinal guided mode through a sample clamped between two buffer rods. The main difference with other interferometric techniques is the use of a delay line placed between the sample and one of the buffer rods. While with conventional interferometry the sample length needs to be larger than half of the shortest wavelength which can be propagated through the sample, the introduction of the delay line removes this limitation enabling the characterization of very small samples (compared to the wavelength). In addition, this method is suitable for testing highly attenuative materials which can be cut into thin specimens so reducing the energy absorption as the acoustic signal propagates through the sample. The principle of the method is discussed and measurements on a variety of specimens are presented. Results for long material specimens tested without delay line and short samples of the same material tested with the delay line agree within 1%. Moreover, measurements are in good agreement with results obtained from conventional methods and literature. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1942387]

PACS number(s): 43.58.Dj, 43.35.Mr [YHB] Pages: 832–840

I. INTRODUCTION

This paper was motivated by the need to monitor degradation of graphite in nuclear reactors where only very small samples can be extracted. The degradation process affects the tensile strength of the graphite and can be assessed by estimating the real part of the dynamic Young's modulus at low frequency (where material dispersion can be neglected). Although several techniques which correlate strength and stiffness versus deformation have been developed and implemented in international standards such as the ASTM (see, for instance, methods for plastic characterization[1]) they become inapplicable when the specimen volume is limited to a few millimeter cube. Under these conditions, one possibility is to use ultrasonic techniques which probe the sample by means of mechanical waves.[2] However, these methods require the dimensions of the sample to be larger than half of the propagated wavelength, λ, limiting the possibility of testing small (compared to λ) samples. In this paper, this will be referred to as the "*limited size problem*." Since λ decreases as the frequency increases, it could be argued that a sample can be tested regardless of its dimensions provided that the testing frequency is high enough. However, the high frequency properties of a material can be different from the static or low frequency ones due to hysteresis phenomena such as viscoelasticity. Moreover, the attenuation of an acoustic wave with propagation distance increases with frequency making it difficult or even impossible to transmit energy through the sample (note that even when the sample is thin the attenuation can be very large since the testing frequency needs to be high in order to separate multiple echoes).

Recently, the authors have proposed an interferometric technique for the measurement of the shear properties of viscoelastic materials at frequencies below 100 kHz which partially overcomes the limited size problem.[3] The technique consists of clamping a small cylindrical sample between two metallic buffer rods as shown in Fig. 1(a), the diameter of the sample being the same as that of the rods. The fundamental torsional guided mode,[4] $T(0,1)$, is sent from the free end of one rod and detected at the free ends of each rod. The measured transmission coefficient spectrum of $T(0,1)$ through the sample exhibits peaks which correspond to the through thickness resonances of the sample if it were free, i.e.,

$$f_n = \frac{c_s n}{2d}, \tag{1}$$

where f_n is the frequency at which the nth transmission peak occurs and c_s and d are the shear sound velocity and length of the sample, respectively. As a consequence, by measuring the frequency interval between two consecutive peaks, c_s can be obtained by inverting Eq. (1). If the density, ρ, of the sample is known, the shear modulus, G, is given by $c_s^2\rho$. The technique can provide the shear properties of the sample regardless of its transversal dimension

[a]Electronic mail: f.simonetti@imperial.ac.uk

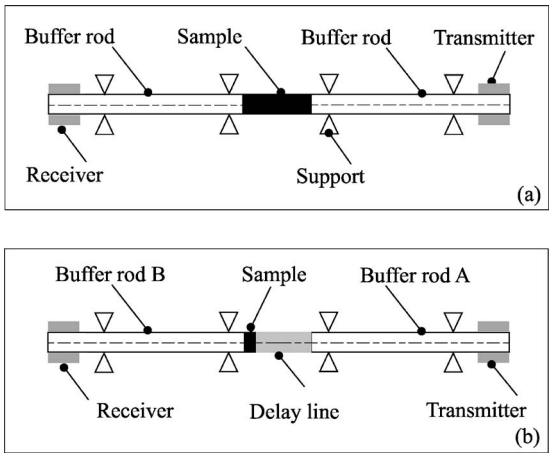


FIG. 1. (a) Diagram of the apparatus used for shear property measurements; (b) addition of the delay line.

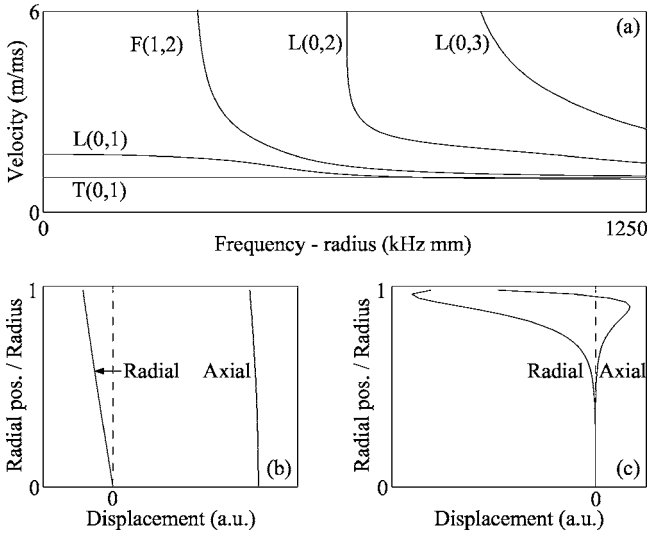


FIG. 2. (a) Nylon rod dispersion curves; material properties are given in Table I; (b) $L(0,1)$ radial and axial components of the displacement at 150 kHz mm; (c) $L(0,1)$ mode shape at 1250 kHz mm.

(sample radius) whereas the sample length needs to be thicker (larger) than half of the shortest wavelength which can be propagated through the sample in order to produce the interference peaks.

In a similar fashion, the Young's modulus could be obtained by considering the interference of the fundamental extensional mode, $L(0,1)$. In this case, it is expected that interference peaks would still occur at the frequencies given by Eq. (1), where c_s is now replaced with the phase velocity of $L(0,1)$. However, the interference of extensional guided waves is a much more complicated phenomenon than that of torsional modes. In particular, the interference of $T(0,1)$ is the exact cylindrical analogue of the scattering of a shear horizontal plane wave which impinges a layer separating two infinite half spaces at normal incidence.[3] Therefore, the expression for the transmission coefficient of $T(0,1)$ is the same as that of plane shear waves and the mode is transmitted without being converted into other modes[3] (neither propagating nor nonpropagating[5]). On the other hand, there is not such a close correspondence between extensional guided waves and longitudinal plane waves. The first important difference is in the frequency dependence of the phase velocity of $L(0,1)$. Figure 2(a) shows the phase velocity versus frequency-radius product for the $T(0,1)$ mode and for the first three extensional symmetric modes, $L(0,1)$, $L(0,2)$, and $L(0,3)$, propagating in a nylon rod whose properties are summarized in Table I. The second flexural mode of the first family, $F(1,2)$, is also represented, its significance being discussed in Sec. III. The dispersion curves have been calculated by neglecting energy absorption in the rod and by assuming that the acoustic properties are frequency independent. While the $T(0,1)$ mode is non-dispersive (its phase velocity is the same as the material shear velocity), the phase velocity spectrum of $L(0,1)$ exhibits a plateau region up to 300 kHz mm but then suddenly decreases and approaches the Rayleigh velocity as the frequency-radius product increases [Fig. 2(a)]. The phase velocity dispersion of $L(0,1)$ is accompanied by substantial variations of the displacement fields through the rod cross section (called mode shapes). As an example, Fig. 2(b) shows a typical mode shape corresponding to a frequency-radius of 150 kHz mm, which lies on the plateau region of $L(0,1)$, whereas Fig. 2(c) is the mode shape of $L(0,1)$ for large values of the frequency-radius product (1250 kHz mm). As the frequency-radius product tends to zero the radial displacement of $L(0,1)$ vanishes and the mode behaves as a longitudinal plane wave which propagates at the speed[4] $\sqrt{E/\rho}$, E being the Young's modulus. On the other hand, for large values of the frequency-radius product, $L(0,1)$ becomes a surface wave.

The complexity of the $L(0,1)$ mode shape has a major effect on the transmission mechanism of $L(0,1)$ through a sample clamped between two buffer rods. For the $T(0,1)$ mode the displacement field, which is tangential, varies linearly with the radius and perfect mode shape matching between $T(0,1)$ propagating in the buffer rods and $T(0,1)$ propagating in the sample is always possible.[3] On the other hand, the mode matching for $L(0,1)$ is never perfect. As a consequence, the transmission of $L(0,1)$ is always accompanied by the generation of other modes which can be propagating or nonpropagating[5] [note that in Fig. 2(a) propagating modes are shown only]. This suggests that the transmission of $L(0,1)$ will not follow Eq. (1) which is derived under the hypothesis of plane wave propagation. It is important to observe that these effects can also occur with other methods such as the resonant bar technique originally developed by

TABLE I. Material properties used to calculate the phase velocity dispersion curves. The shear velocities were measured by using the technique introduced in Ref. 3 around 50 kHz, whereas the longitudinal velocities are those given in Ref. 2, with the exception of the longitudinal velocity of nylon which was measured at 500 kHz.

	c_s (m/s)	c_L (m/s)	ρ (kg/m^3)
Nylon 6	1040	2343	1140
Lucite	1330	2680	1180
Teflon	400	1340	2160

Norris and Young[6–8] and subsequently modified by Adams and Coppendale for measuring the complex moduli of thin samples.[9]

Nevertheless, if there are interference peaks occurring in the plateau region of $L(0,1)$, Eq. (1) holds with a degree of accuracy which increases as the peaks move toward lower frequency-radius product values since at low frequency $L(0,1)$ behaves as a plane wave. Unless the sample dimensions are limited and the material absorption is not too large, these values could be tailored to the very low frequency-radius regime by choosing very long samples [see Eq. (1)].

This paper demonstrates how the limited size problem can be addressed by adding a delay line between the sample to be tested and one of the buffer rods as shown in Fig. 1(b). The delay line is made of a known material (Lucite™ in this paper) and its length is chosen in order to produce interference peaks at very low frequency-radius product values [where $L(0,1)$ behaves as a plane wave] when the delay line is directly clamped between the two buffer rods (without sample). This implies that when the sample is inserted between one of the buffers and the delay line, it will produce a shift of the transmission peaks toward lower frequency-radius values. This is trivial when the sample is made of the same material as the delay line since the sample and delay line now behave as a monolithic sample whose length is the sum of the individual lengths [see Eq. (1)]. In other words, with the introduction of the delay line, the peaks of the transmission spectrum through the sample clamped between the two buffers (without the delay line), which would occur at very large frequency-radius values when the sample is short, are shifted toward the low frequency-radius regime where $L(0,1)$ behaves as a longitudinal plane wave over all the transmission line (buffer rods, sample, and delay).

In Sec. II the analytical expression for the transmission coefficient of a longitudinal plane wave propagating through an infinitely wide bilayer separating two half spaces is derived. Sections III and IV describe the method and the experimental setup, respectively. In Sec. V the feasibility of interferometric measurements with the delay line is demonstrated by comparing the results obtained for long (tested without delay) and short (tested with the delay line) nylon samples. Moreover, the effects of dispersion are investigated experimentally by testing Teflon samples. The results are further supported by comparison with literature data and independent ultrasonic measurements. Finally, the method sensitivity is analyzed in Sec. VI.

II. PLANE WAVE TRANSMISSION COEFFICIENT

Let us consider a plane wave, I, incident on an infinitely wide bilayer separating two half spaces as shown in Fig. 3, each medium being assumed to be homogeneous and isotropic. Here, the half spaces represent the two buffer rods while layers "2" and "3" correspond to the delay line and the sample, respectively. The analytical expression of the transmission coefficient can be obtained by following the approach described by Chew.[10]

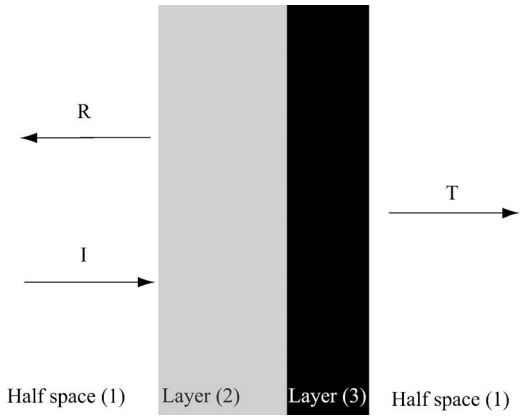


FIG. 3. Schematic diagram of the transmission and reflection through a bilayer separating tow half spaces. I is the incident wave, T and R are the transmitted and reflected waves.

Consider first the transmission and reflection coefficients, T' and R', of a plane wave impinging on the interface between a half space, "1," and an infinitely wide layer, "2," backed by a half space, "3," i.e.,

$$R' = R_{12} + \frac{T_{12}R_{23}T_{21}e^{2ik_2d_2}}{1 - R_{21}R_{23}e^{2ik_2d_2}}, \tag{2}$$

$$T' = \frac{T_{12}T_{23}e^{ik_2d_2}}{1 - R_{23}R_{21}e^{2ik_2d_2}}, \tag{3}$$

where R_{ij} and T_{ij} are the reflection and transmission coefficients at the interface between half spaces of the relevant materials ("i" and "j") when an incident wave travels from a half space of material "i" to a half space of material "j." By defining the complex impedance as

$$Z = \rho\frac{c}{1 + i\alpha c/\omega}, \tag{4}$$

where ρ is the density, ω the angular frequency, c the plane wave phase velocity, and α the attenuation in nepers per unit length, the reflection and transmission coefficients are

$$R_{ij} = \frac{Z_j - Z_i}{Z_i + Z_j}, \tag{5}$$

$$T_{ij} = \frac{2Z_j}{Z_i + Z_j}. \tag{6}$$

Moreover, k_2 is the wave number within the layer ($k_2 = \omega/c_2 + i\alpha_2$) and d_2 is the layer thickness.

Expression (2) and (3) can be employed to calculate the transmission coefficient, T, through the bilayer (Fig. 3). In particular, the half space "1" where the plane wave is incident and the layer "2" can be replaced with an artificial half space "0." As a consequence, according to Eq. (3) the transmission coefficient of a wave traveling from the half space "0" through the layer "3" and emerging in the half space "1" is

$$T = \frac{T_{03}T_{31}e^{ik_3d_3}}{1 - R_{31}R_{30}e^{2ik_3d_3}}. \tag{7}$$

R_{03} is the reflection coefficient of a wave which impinges on the half space "0" from a half space of material "3." Since the half space "0" is made of the half space "1" and the layer "2" from Eq. (2) it follows that

$$R_{30} = R_{32} + \frac{T_{32}R_{21}T_{23}e^{2ik_2d_2}}{1 - R_{23}R_{21}e^{2ik_2d_2}}, \tag{8}$$

similarly it can be shown that

$$T_{03} = \frac{T_{12}T_{23}e^{ik_2d_2}}{1 - R_{23}R_{21}e^{2ik_2d_2}}. \tag{9}$$

Equation (7) along with Eqs. (8) and (9) provides the analytical expression for the transmission coefficient. Moreover, implicit in the definition of Eqs. (5) and (6) is the hypothesis that no mode conversion occurs when a plane wave impinges the interface between two half spaces "*i*" and "*j*" at normal incidence.

III. METHOD

The technique proposed in this paper assumes that expression (7) is also valid for the scattering of $L(0,1)$ in the frequency range where its dispersion is negligible. For a given frequency range and radius, the dispersion of $L(0,1)$ depends on the material properties of the rod in which it propagates. For a homogeneous and isotropic rod, the transition of $L(0,1)$ from plane wave to surface wave [see Figs. 2(b) and 2(c)] is marked by the cutoff frequency of the second flexural mode of the first family, $F(1,2)$, shown in Fig. 2 (for a classification of flexural modes see, for instance, Rose[11]) which, to a first approximation, is given by $c_s/4R$, where R is the radius. Therefore, the lower the material shear velocity, the smaller the extent of the frequency range in which $L(0,1)$ is nondispersive. For the entire transmission line (buffer rods, sample, and delay line) this extent will depend on the material with the lowest shear velocity, which for instance, can be measured by torsional guided wave interferometry.[3]

After characterizing the frequency extent of the nondispersive region of $L(0,1)$ the material Young's modulus can be derived from the measured transmission coefficient spectrum. For this purpose it can be observed that for given buffers and delay line properties, expression (7) can be thought of as known function of the frequency and two unknown parameters: the phase velocity, c_{ph}, and the attenuation coefficient of $L(0,1)$ in the sample, $\bar{\alpha}$, which is defined in Ref. 3 and whose units are neper per wavelength (Np/wl). The unknown parameters can be obtained by a least-squares fit of Eq. (7) to the measured transmission coefficient spectrum. The modulus of the dynamic Young's modulus, E, is then derived from the phase velocity by observing that in the nondispersive region of $L(0,1)$ the following relationship holds:[12]

$$E = c_{\mathrm{ph}}^2\rho. \tag{10}$$

It should be emphasized that the frequency dependence of the material longitudinal and shear sound velocities has been neglected in the previous discussion, the dispersion of $L(0,1)$ being due to the geometry of the waveguide only. Material dispersion modifies the "geometrical dispersion" of $L(0,1)$ because the longitudinal and shear bulk velocities change with frequency. For instance, a strong dispersion of the shear velocity can produce a large variation of the $F(1,2)$ cutoff frequency so affecting the frequency extent of the nondispersive region of $L(0,1)$. Also a very low frequency transition between material rubbery and glassy behavior can introduce high dispersion at low frequency in the plateau region of $L(0,1)$. As an example, the transition temperature of Nylon 6 at 200 Hz is 350 K approximately.[13] As a result, at temperatures lower than 350 K, the material dispersion is high at frequencies below 200 Hz according to the time-temperature superposition principle.[14] For a 2.5 mm radius rod the transition would occur below 0.5 kHz mm which is a very small frequency-radius product on the scale of Fig. 2. Therefore, there will be a quasi-nondispersive region for $L(0,1)$ above a few kilohertz. This implies that in a sufficiently narrow frequency range the dispersion of the material properties of metals and many plastics can be neglected as long as the transition between material rubbery and glassy behavior occurs outside this range. Therefore, in the rest of this paper the effects of material property dispersion will not be considered. Moreover, the phase velocity appearing in Eq. (10) which corresponds to the phase velocity of $L(0,1)$ in the plateau region of the frequency interval where material dispersion is neglected will be referred to as the nondispersive phase velocity.

IV. EXPERIMENTS

A schematic diagram of the apparatus employed for the measurements is given in Fig. 1(b). Two stainless steel buffer rods 770 mm length, 5 mm diameter were employed. The alignment of the rods was ensured by cylindrical perspex supports which allowed the rods to slide axially with minimum friction, so reducing the possibility of $L(0,1)$ being reflected from the supports. In order to excite and detect $L(0,1)$, a pair of piezoelectric transducers were firmly clamped onto the lateral surface of each rod at one end (for more details on the transducers see Alleyne *et al.*[15]). For each rod the transducer polarization direction was oriented parallel to the rod axis and the two piezoelectric elements were connected in parallel in order to induce a uniform axial motion along the rod circumference. A custom-made wave-form generator-power amplifier excited the transducers of rod A [see Fig. 1(b)] by a one cycle Hanning windowed toneburst with a center frequency of 60 kHz. The transmission coefficient was measured by operating the transducers of rod A in pitch-catch mode with those of rod B (i.e., send from rod A and receive from rod B). The pitch-catch response was amplified and transmitted to a digital oscilloscope (LeCroy 9400) and subsequently stored in a PC. Acoustic coupling between the buffer rods, delay line, and sample was achieved by applying a moderate compressional axial load at the free ends of each buffer by means of a screw as in previous

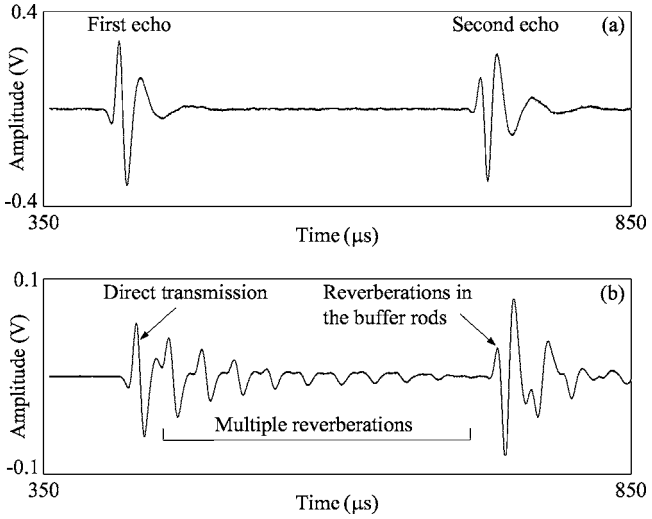


FIG. 4. (a) Typical pulse-echo response of the transducer pair of rod A when neither sample nor delay line are present; (b) signal transmitted through the delay line without sample.

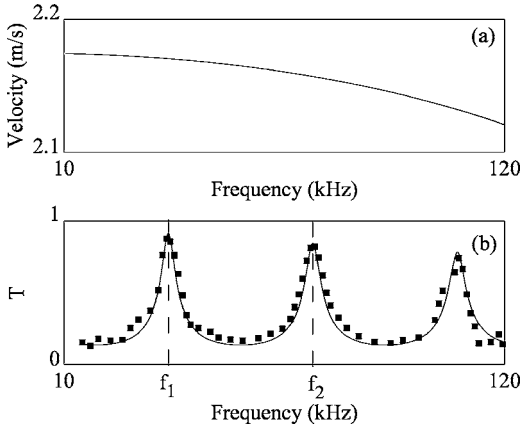


FIG. 5. (a) $L(0,1)$ phase velocity in a 5-mm-diam Lucite rod; (b) transmission coefficient spectrum for the Lucite delay clamped between the two buffers. (■) Experiments; (—) least-squares fit.

experiments with $T(0,1)$.[3] However, in this case the axial load was much lower than that used for torsional experiments because, by contrast with torsional waves, longitudinal waves can be transmitted through frictionless contact surfaces. It should be noticed that care has to be taken to avoid eccentricity of the axial load and buffer misalignment. These precautions are needed in order to avoid the generation of unwanted flexural modes which would be excited from nonuniform acoustic coupling at the contact interfaces. Most of the experiments were performed without coupling gel; however, when the finish of the sample contact surfaces was not smooth enough a very thin layer of couplant was employed.

In order to measure the transmission coefficients the apparatus was calibrated by measuring the reflection from the free end of one of the buffer rods, and the transmission and reflection at the interface between the two buffers when they are put in contact (without sample and delay line). By assuming that no energy is dissipated at the contact interface a calibration parameter which corrects for differences in the electronics of the transmit/receive lines and transducers and their coupling conditions was derived (for more details see Simonetti and Cawley[3]). The transmission coefficient through the sample and delay line was obtained by multiplying the transmitted signal spectrum by this parameter (the parameter does not compensate for nonideal contact between sample, delay line and buffers; good contact is ensured by the applied axial load). Note that this procedure does not require the use of calibration samples.

The delay line was a 30-mm-length 5-mm-diam Lucite (Methyl methacrylate) cylinder which exhibits low damping below 100 kHz and whose properties are given in the next section. All the experiments were carried out at room temperature (296 K).

Figure 4(a) is a typical time trace showing the first two consecutive reflections of $L(0,1)$ from the free end of rod A measured in pulse-echo mode when there is no contact with the delay line (Fig. 1). Figure 4(b) shows the signal transmitted from the transducer pair mounted on rod A and received by the pair on rod B, when the delay line is clamped between the two buffer rods (without sample). The first part of the signal (up to 700 μs) contains the pulse directly transmitted through the delay line followed by a series of multiple reverberations within the delay line. For a more detailed discussion on the origin of the multiple reverberations see Ref. 3. Similar reverberations are observed when the sample is inserted between the delay line and one of the buffers. The signal arriving after 700 μs corresponds to multiple reverberations within the buffer rods which are due to the finite length of the buffers. Since the method proposed in Sec. III assumes that the two buffers are infinitely long the signals arriving after 700 μs are gated out for the calculation of the transmission coefficient.

V. RESULTS AND DISCUSSION

In the following the feasibility of interferometric measurements with the delay line is demonstrated and validated experimentally by comparing the results obtained for long specimens tested without the delay line and thin specimens of the same material tested with the delay line. The absolute results are supported by a comparison with literature data and independent ultrasonic measurements. Section V A applies standard interferometry to the characterization of the delay line properties, which are needed to characterize thin nylon and Teflon™ samples in Secs. V B and V C.

A. Delay line

Figure 5(a) shows the phase velocity for the $L(0,1)$ mode propagating in a 5-mm-diam Lucite rod whose properties are given in Table I. Note that the shear velocities in Table I have been measured by using the technique introduced in Ref. 3, whereas the longitudinal velocities are those given in Ref. 2, with the exception of the longitudinal velocity of nylon, which was measured as explained in Sec. V B. The densities given in Table I are from Ref. 2. These values agreed with direct density measurements within 1–2 %, which corresponds to the level of uncertainty of the measurements.

TABLE II. Comparison of measured phase velocity and Young's modulus with literature data and other independent measurements.

	Measurements			Literature/other meas.	
	c_{ph} (m/s)	$\bar{\alpha}$ (Np/wl)	E (GPa)	c_{ph} (m/s)	E (GPa)
Lucite	2165	0.02	5.53	2170	5.56
Nylon 6 20 mm	1790	0.06	3.65	1730	3.41
Nylon 6 5.3 mm	1800	0.22	3.69	1730	3.41
Teflon 13.5 mm	653	0.19	0.92	680	1.00
Teflon 2.4 mm	656	0.62	0.93	680	1.00

It can be observed that up to 120 kHz the maximum phase velocity variation is within 2.3% of its value at low frequency; therefore, within this frequency range it can be assumed that $L(0,1)$ behaves as a plane wave. Note that all the dispersion curves shown in this paper have been calculated by using the software DISPERSE.[16] Figure 5(b) shows the measured transmission coefficient spectrum through the Lucite sample clamped between the two buffers (without sample), obtained by Fourier transforming the signal shown in Fig. 4(b).

The solid line in Fig. 5(b) is the least-squares fit of expression (7) to the experimental spectrum in the frequency range between 30 and 120 kHz. In this case it is assumed that the delay line length is zero and the sample is the Lucite cylinder. The best fit leads to the values of the phase velocity and attenuation coefficient given in Table II. From the value of the phase velocity, the Young's modulus can be calculated by using Eq. (10). For comparison, the $L(0,1)$ nondispersive phase velocity and the Young's modulus calculated from the Lucite properties given in (Table I) are shown in the last two columns of Table II. These values have been calculated by substituting the acoustic velocities, c_S and c_L, in the expression relating the Young's and shear moduli to the Poisson's ratio. Since the nondispersive phase velocity is given by Eq. (10), after some algebra one obtains

$$c_{ph} = \sqrt{\frac{3c_L^2c_s^2 - 4c_s^4}{c_L^2 - c_s^2}} \quad \text{or} \quad E = \rho\frac{3c_L^2c_s^2 - 4c_s^4}{c_L^2 - c_s^2}. \tag{11}$$

As these expressions contain the shear velocity to the power of four, it follows that the estimate of the Young's modulus and phase velocity are very sensitive to the shear velocity accuracy.

It can be observed that the phase velocity could also be derived from Eq. (1) by measuring the frequency interval between the first two peaks, i.e., $f_1 = 36$ kHz and $f_2 = 72$ kHz. This would lead to a phase velocity of 2136 m/s, which is slightly different from that given in Table II. However, the velocity derived from the best fit procedure, which involves all the data measured between 30 and 120 kHz, is more reliable since it is more robust to measurement errors.

B. Nylon samples

Long nylon specimens without delay line and thin samples with the delay line were tested, so as to assess the degree of accuracy of the thin sample measurements. Figure 6(a) shows the measured transmission coefficient for a 20 mm length nylon sample without the delay line. The solid line is the least-squares fit of Eq. (7) to these data over the frequency range between 30 and 120 kHz; the corresponding values of the phase velocity and attenuation are given in Table II. Figure 6(b) is the transmission coefficient for a 5.3 mm length nylon sample and the 30 mm length Lucite delay line. Note that in Fig. 6(b) the experimental data around the frequency F_1 are corrupted by the two pole 25 kHz highpass filter of the amplifier. This explains the unrealistically high value of the measured transmission coefficient (unity at 28 kHz) and is the reason why the least-squares fitting was performed between 30 and 120 kHz. The phase velocity and attenuation obtained from the least-squares fit of Eq. (7) to the measured transmission coefficient [solid line in Fig. 6(b)] are given in Table II (third row). There is excellent agreement between the Young's moduli obtained for the long and thin samples, the relative difference being around 1%.

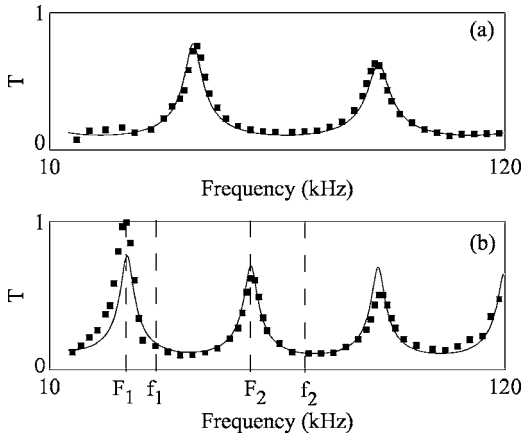


FIG. 6. $L(0,1)$ transmission coefficient spectra through: (a) 20-mm-length nylon sample without delay, (b) 5.3-mm-length Nylon sample and 30 mm Lucite delay line. (■) Experiments; (—) least-squares fit.

These results confirm that the technique proposed in this paper is capable of addressing the limited size problem. In the frequency range considered in these experiments, the wavelength propagated through the nylon sample ranges from a minimum of 15 mm at 120 kHz to a maximum of 60 mm at 30 kHz, which are much larger than the sample dimensions (5 mm diameter, 5.3 mm length).

It has to be mentioned that although the phase velocities obtained for the short and long sample agree very well, the absorption coefficient obtained for the short sample is much higher than that of the long specimen. Such a large difference is probably due to the presence of three contact interfaces: buffer-delay, delay-sample, and sample-buffer, which produce higher energy dissipation than in the case of the sample clamped between the two buffers directly. This makes it very difficult to estimate the actual absorption coefficient of the material; however, this does not affect the accuracy of the phase velocity measurements.

In order to appreciate the sensitivity of the method, the frequencies, f_1 and f_2, where the transmission peaks through the delay line occur [Fig. 5(b)], have been marked in Fig. 6(b). The presence of the sample results in a clear shift of the peak transmission frequencies which determines the method

sensitivity, the larger the frequency shift the higher the sensitivity of the entire transmission spectrum to the sample properties. Moreover, without the delay line the first two interference peaks for the 5.3 mm sample would occur at 168 kHz (420 kHz mm) and 336 kHz (840 kHz mm) at which $L(0,1)$ can no longer be considered as a plane wave (see Fig. 2). Instead, by using the delay line the transmission peaks occur at the frequencies $F_1=28$ kHz and $F_2=59$ kHz, which lie in the frequency regime where the dispersion of $L(0,1)$ is negligible (see Fig. 2).

The previous results have demonstrated that measurements with the delay line lead to the same results as standard interferometry. In order to further validate the results a comparison with the phase velocity and Young's modulus predicted from the acoustic properties of nylon was performed. The longitudinal sound velocity was obtained by measuring the time of flight of a longitudinal pulse sent through the thickness of a nylon plate 17 mm thick, the center frequency of the pulse being 500 kHz; note that this is only an approximation of the longitudinal velocity in the 30–120 kHz range needed in the rod measurements since there may be some velocity dispersion at higher frequencies and the plate and rod material properties may also be slightly different. The shear velocity was measured by using the technique proposed by Simonetti and Cawley[3] in the frequency range between 30 and 120 kHz. The results are given in Table I.

The nondispersive phase velocity and Young's modulus calculated with the measured bulk properties of nylon by using Eq. (11) (see the last two columns in Table II) agree well with the interferometric measurements (with and without the delay line), the average difference being around 3% (7% for the Young's modulus). This difference is likely to be due to the above-discussed dispersion and material variation effects.

C. Teflon samples

Teflon samples were tested in order to investigate the effects of the $L(0,1)$ mode dispersion on the transmission coefficient spectrum. For a Teflon rod, the $L(0,1)$ mode undergoes severe dispersion as the frequency increases up to 120 kHz, which leads to a relative phase velocity reduction of 40% as shown in Fig. 7(a). (Teflon properties are from Table I). Above approximately 70 kHz, $L(0,1)$ cannot be regarded as a plane wave as its mode shape varies along the radius rapidly, by contrast with the displacement field in the buffers which can still be considered uniform. As a result, the transmission of $L(0,1)$ through a Teflon sample clamped between the two buffers will be accompanied by strong mode conversion into nonpropagating and propagating [above the cutoff frequency of the second extensional mode, $L(0,2)$] modes which impedes the transmission of energy in $L(0,1)$. This is clearly shown in Fig. 7(b), which is the transmission coefficient for a 13.5-mm-length Teflon sample clamped between the buffers (without delay line). Above 70 kHz, the measured transmission coefficient decreases rapidly, and transmission peaks are no longer visible [see the square dots in Fig. 7(b)]. Moreover, energy dissipation is also responsible for the decay of the transmission coefficient with frequency, since the wave attenuation increases with frequency. Note that the experimental data around the first transmission peak, which occurs below 30 kHz, should not be considered as part of the general trend of the transmission coefficient with frequency due to the presence of the amplifier cutoff.

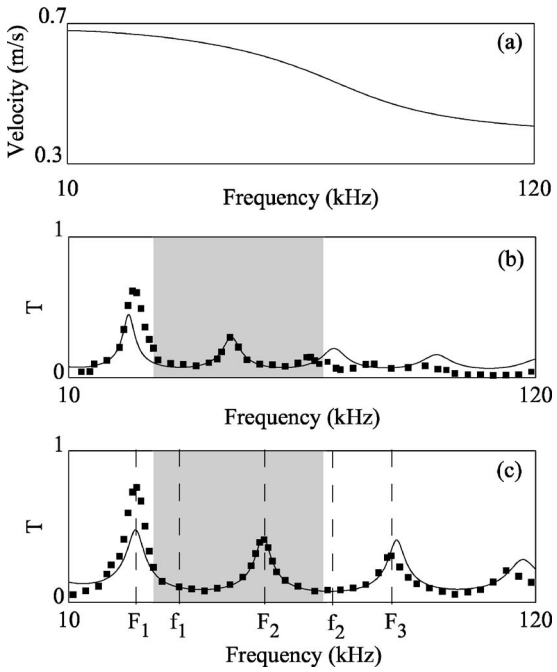


FIG. 7. (a) $L(0,1)$ phase velocity in a 5-mm-diam Teflon rod; (b) transmission coefficient spectrum for the 13.5-mm-length sample clamped between the two buffers; (c) transmission coefficient spectrum for the 2.4-mm-length sample and delay line; (■) experiments; (—) least-squares fit. The shaded area identifies the frequency range where the best fit is performed.

The solid line in Fig. 7(b) is the best fit of Eq. (7) to the experimental data, by setting the delay line length equal to zero. Although function (7) has been plotted from 10 to 120 kHz, the least-squares fit was performed in the frequency range between 30 and 70 kHz [see the gray shaded area in Fig. 7(b)]. The choice of the low frequency limit is dictated by the amplifier cutoff, whereas the upper limit removes the frequencies at which $L(0,1)$ no longer behaves as a plane wave. The results are given in Table II.

Figure 7(c) shows the transmission coefficient for a 2.4-mm-length Teflon sample with the Lucite delay line. Thanks to the presence of the delay line, the first two interference peaks occur at $F_1=26$ and $F_2=56$ kHz rather than 140 and 280 kHz, which would be expected in the case of transmission without delay line [note that in Fig. 7(c) f_1 and f_2 are the peak transmission frequencies of the delay line alone]. Also in this case the least-squares fit was carried out between 30 and 70 kHz [gray shaded area in Fig. 7(c)], the values of phase velocity and attenuation being given in Table II.

From Table II it can be seen that the two values of the Young's modulus obtained for the long and thin specimens agree very well, the relative difference being less than 0.5%. For a qualitative comparison, the last two columns of Table II provide the phase velocity and Young's modulus obtained from the acoustic properties given in Table II through Eq. (11). The relatively large difference, 7.5%, could be due to

the uncertainty on the value of the longitudinal bulk velocity which has not been measured directly (the value given in Table I is from the literature).

The occurrence of the third peak in Fig. 7(c) around $F_3=86$ kHz, seems to contradict the argument on the generation of other modes above 70 kHz and the subsequent reduction of the transmission coefficient. However, this apparent inconsistency can be explained by observing that in the presence of material absorption all the nonpropagating modes become propagating.[17–19] These modes are highly attenuated with distance and are completely absorbed within one wavelength propagation distance. Therefore, while they are not able to pass through the 13.5-mm-length sample, they can transmit energy through the shorter sample, leading to the presence of the third transmission peak.

VI. SENSITIVITY ANALYSIS

This section provides an analysis of the method sensitivity as a function of the sample and delay line properties. The sensitivity is assessed by considering the nondimensional ratio

$$S=(f_1-F_1)/f_1, \tag{12}$$

where f_1 and F_1 are the frequencies where the first transmission peak occurs for the delay line clamped between the buffers and for both sample and delay line clamped between the buffers, respectively. According to this definition, the sensitivity is high when the introduction of the sample causes a large frequency shift of the first transmission peak compared to that due to the delay line if it were clamped between the buffers directly.

The nonlinear nature of the method implies that S will depend on the properties of each component of the transmission line (buffers, delay line, and sample). By neglecting absorption effects, this would lead to eight parameters: the phase velocities of $L(0,1)$ propagating in each component, c_{sample}, c_{buffers}, and $c_{\text{delay line}}$, their densities, ρ_{sample}, ρ_{buffers}, $\rho_{\text{delay line}}$, and the sample and delay line lengths, d_{sample} and $d_{\text{delay line}}$. However, in order to characterize the dependence of S on the sample and delay line properties, such a number can be reduced to six by considering steel buffers only. Numerical calculations performed by using Eq. (7) have shown that, to a very high degree of accuracy, S depends on three nondimensional ratios obtained from the original six parameters, i.e.,

$$S=S\left(\frac{c_{\text{sample}}}{c_{\text{delay line}}},\frac{d_{\text{sample}}}{d_{\text{delay line}}},\frac{\rho_{\text{sample}}}{\rho_{\text{delay line}}}\right). \tag{13}$$

Figure 8 shows the dependence of the sensitivity on the velocity and length ratios, $c_{\text{sample}}/c_{\text{delay line}}$ and $d_{\text{sample}}/d_{\text{delay line}}$, for two values of the density ratio $\rho_{\text{sample}}/\rho_{\text{delay line}}$. In particular, the solid lines refer to a density ratio of 0.97, which corresponds to the combination nylon-Lucite, whereas the dotted lines are representative of the Teflon–Lucite combination ($\rho_{\text{sample}}/\rho_{\text{delay line}}=1.83$). The two families of curves have been calculated for the velocity ratio ranging between 0.1 and 1.2 (note that the velocity ratios labeled in Fig. 8 refer to the nylon–Lucite family, those for the Teflon–Lucite family, which are not shown for clarity, vary from 0.1 to 1.2 from top to bottom as for the nylon–Lucite family.)

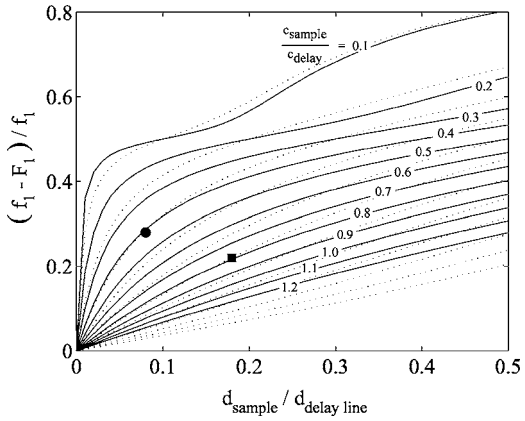


FIG. 8. Sensitivity curves as a function of the velocity and length ratios for: (—) $\rho_{\text{nylon}}/\rho_{\text{Lucite}}=0.97$; (···)$\rho_{\text{Teflon}}/\rho_{\text{Lucite}}=1.83$; (■) tested nylon sample; (●)tested Teflon specimen.

Let us consider the nylon–Lucite family. For a given sample to delay line length ratio, the sensitivity increases as the sample to delay line velocity ratio decreases. However, as the velocity ratio decreases the transmission coefficient also decreases due to the impedance mismatch. As a consequence, the delay line velocity should be chosen by considering the tradeoff between the frequency shift (f_1-F_1) and the transmission coefficient magnitude.

For a given sample to delay line velocity ratio, the sensitivity increases with the sample to delay line length ratio. On the other hand, for very thin samples ($d_{\text{sample}}/d_{\text{delay line}}<0.1$), the sensitivity decreases with the sample to delay line velocity ratio very rapidly (Fig. 8). This means that in order to test very thin samples, the delay line should be much "faster" than the sample. Note that in this case the amplitude of the transmission coefficient is affected by the velocity ratio only slightly; in the limit for $d_{\text{sample}}=0$ the peak transmission coefficient tends to unity regardless of the sample velocity.

The dependence of S on the density ratio is more complicated and does not exhibit a monotonic behavior. For low values of the velocity ratio, the sensitivity increases with the density ratio if the length ratio is large, while it decreases if the length ratio is low. For instance, when the velocity and length ratios are 0.2 and 0.5, respectively, the sensitivity for Teflon is higher than that for nylon. On the other hand, for a 0.1 length ratio and 0.2 velocity ratio, the sensitivity difference is reversed. In addition, for large velocity ratios, the sensitivity decreases as the density ratio increases (bottom curves in Fig. 8). In general, the density ratio does not have a dramatic effect on the sensitivity and its influence decreases with the length ratio (Fig. 8).

VII. CONCLUSIONS

In this paper it has been shown that the measurement of the Young's modulus of limited size specimens can be addressed by means of interferometric measurements. It has been demonstrated that the introduction of a delay line of

suitable length enables the generation of interference peaks in the low frequency regime where the fundamental extensional mode, $L(0,1)$, can be considered as a plane wave. This results in the possibility of probing samples with acoustic waves whose wavelengths are much larger than the sample dimensions. Therefore, the technique is ideal for testing very small samples (compared to the wavelength) and highly attenuative materials which can be cut into thin specimens, so reducing the energy absorption as the acoustic signal propagates through the sample.

The accuracy of the measurements obtained for thin material samples tested with the delay line has been assessed by comparing the results with the measurements made for long samples of the same material tested without the delay line, the agreement being within 1%.

Moreover, it has been shown that, due to the dispersion of $L(0,1)$, the frequency-radius product at which the experiments are performed can have a strong effect on the transmission coefficient spectrum, resulting in a drastic decay of the transmitted energy with frequency.

The technique is fast, minimum sample preparation is required and the use of couplant is not required unless the sample contact surfaces are too rough. The limit on the absolute minimum sample length that can be tested depends on the signal to noise level of the system, the length of the delay line, and the velocity and density mismatch between the sample and delay line. The main drawback of the technique compared to conventional static measurements of the Young's modulus is the need for the sample density to be known.

[1]ASTM D 638-01, Standard test method for tensile properties of plastics.
[2]P. McIntire, Ultrasonic Testing, Nondestructive Testing Handbook, Vol. 7 (American Society for Nondestructive Testing, Columbus, Ohio, 1991).
[3]F. Simonetti and P. Cawley, "Ultrasonic interferometry for the measurement of shear velocity and attenuation in viscoelastic solids," J. Acoust. Soc. Am. **115**, 157–164 (2004).
[4]K. F. Graff, *Wave Motion in Elastic Solids* (Clarendon, Oxford, 1975).
[5]B. A. Auld, *Acoustic Fields and Waves in Solids* (Krieger, Malabar, FL, 1990), Vol. 2.
[6]S. L. Garrett, "Resonant acoustic determination of elastic moduli," J. Acoust. Soc. Am. **88**, 210–221 (1990).
[7]J. L. Buchanan, "Numerical solution for the dynamic moduli of a viscoelastic bar," J. Acoust. Soc. Am. **81**, 1775–1786 (1987).
[8]D. M. Norris and W. C. Young, "Complex modulus measurement by longitudinal vibration testing," Exp. Mech. **10**, 93–96 (1970).
[9]R. D. Adams and J. Coppendale, "Measurement of the elastic moduli of structural adhesives by a resonant bar technique," J. Mech. Eng. Sci. **18**, 149–158 (1976).
[10]W. C. Chew, *Waves and Fields in Inhomogeneous Media* (IEEE Press, New York, 1995).
[11]J. L. Rose, *Ultrasonic Waves in Solid Media* (Cambridge University Press, Cambridge, UK, 1999).
[12]F. M. Guillot and D. H. Trivett, "A dynamic young's modulus measurement system for highly compliant polymers," J. Acoust. Soc. Am. **114**, 1334–1345 (2003).
[13]N. G. McCrum, B. E. Read, and G. Williams, *Anelastic and Dielectric Effects in Polymeric Solids* (Wiley, London, 1967).
[14]J. D. Ferry, *Viscoelastic Properties of Polymers* (Wiley, New York, 1980).
[15]D. N. Alleyne, B. Pavlakovic, M. J. S. Lowe, and P. Cawley, "Rapid, long range inspection of chemical plant pipework using guided waves," Insight **43**, 93–96 (2001).
[16]B. N. Pavlakovic, M. J. S. Lowe, D. N. Alleyne, and P. Cawley, "Disperse: A general purpose program for creating dispersion curves," in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. 16, pp. 185–192.
[17]F. Simonetti and P. Cawley, "On the nature of shear horizontal wave propagation in elastic plates coated with viscoelastic materials," Proc. R. Soc. London **460**, 2197–2221 (2004).
[18]F. Simonetti, "Lamb wave propagation in elastic plates coated with viscoelastic materials," J. Acoust. Soc. Am. **115**, 2041–2053 (2004).
[19]F. Simonetti and M. J. S. Lowe, "On the meaning of Lamb mode nonpropagating branches," J. Acoust. Soc. Am. (in press).

Acoustic time delay estimation and sensor network self-localization: Experimental results

Joshua N. Ash[a] and Randolph L. Moses[b]
Department of Electrical and Computer Engineering, Ohio State University, 205 Dreese Laboratory, Columbus, Ohio 43210

(Received 20 August 2004; revised 28 April 2005; accepted 25 May 2005)

Experimental results are presented on propagation, coherence, and time-delay estimation (TDE) from a microphone array in an outdoor aeroacoustic environment. The primary goal is to understand the achievable accuracy of acoustic TDE using low-cost, commercial off-the-shelf (COTS) speakers and microphones. In addition, through the use of modulated pseudo-noise sequences, the experiment seeks to provide an empirical understanding of the effects of center frequency, bandwidth, and signal duration on TDE effectiveness and compares this to the theoretical expectations established by the Weiss-Weinstein lower bound. Finally, sensor network self-localization is performed using a maximum likelihood estimator and the time-delay estimates. Experimental network localization error is presented as a function of the acoustic calibration signal parameters. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1953307]

PACS number(s): 43.60.Jn, 43.28.Tc, 43.28.Gq [DKW] Pages: 841–850

I. INTRODUCTION

Sensor networks are emerging in a large number of civilian and military applications to sense and process information about their surroundings.[1–4] To make use of this information the locations of the individual sensors need to be known. However, in many scenarios manual assignment of sensor locations is impossible or impractical due to the volume of sensors deployed or the placement method. Therefore, the problem of self-localizing sensor networks becomes increasingly important.

There are a number of techniques for self-localization of sensors based on measurements of time-of-arrival (TOA), time-difference-of-arrival (TDOA), direction-of-arrival (DOA), or received signal strength measurements.[5–8] Several techniques employ hybrid schemes that combine multiple types of measurements, such as ranging information from TOA estimates and angular information from DOA estimates. In this paper we consider TDOA measurements from acoustic sources to localize elements of a microphone sensor array. The resultant localization accuracy is closely tied to the accuracy with which we can estimate TDOA. In this paper we explore, through statistical bounds and outdoor experimentation, the limits on the mean square error of TDOA estimates for an acoustic sensor network.

Some previous experimental work in acoustic source localization includes the investigations by Kozick and Sadler[9,10] on received signal coherence and localization of a moving vehicle. Other authors have considered source localization by intersecting hyperboloids defined by relative arrival times[11] and by intersecting arrival angles.[12] Relative and absolute localization of the array elements themselves was performed by Dosso[13] using impulsive sources in an underwater setting.

The goals and contributions of this work are that we illustrate the effectiveness of acoustic source localization using uncalibrated commercial off-the-shelf (COTS) equipment and that we explore through experimentation the impact of acoustic source signal parameters on time-delay estimation (TDE) and source localization. The theory of time-delay estimation from propagating waves is relatively well understood,[14–17] however very little work exists to validate this theory in the aeroacoustic regime. In this paper we compare theoretical expectations of TDE performance to experimental observations in an aeroacoustic environment. In particular, we evaluate TDE performance with respect to SNR and the source signal parameters of bandwidth, center frequency, and signal duration.

The remainder of this paper is organized as follows. Section II describes our experimental procedure, including hardware setup and signal generation. In Sec. III we present empirical results characterizing the attenuation and coherence loss of the acoustic channel. Section IV contains the results of our empirical study of the effects of center frequency, bandwidth, and signal duration on time-delay estimation accuracy. In Sec. V we present the results of self-calibrating an outdoor sensor network using broadband pseudo-noise (PN) sequences. Finally, in Sec. VI we conclude.

II. EXPERIMENTAL PROCEDURE

A. Equipment and environment

As depicted in Fig. 1, a linear array of eight Knowles BL-1994 microphones, each separated by 7.7 m, was used in this experiment. All microphones were positioned at the same height (h_m=18 cm) above the grass surface. Each microphone was equipped with an 8-cm-radius spherical windscreen and was connected to a National Instruments data acquisition system by a coaxial cable. The data acquisition system consisted of an eight-channel analog signal condi-

[a]Electronic address: ashj@ece.osu.edu
[b]Electronic address: randy@ece.osu.edu

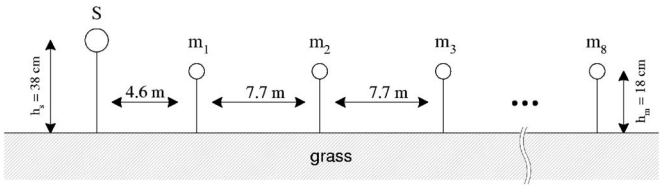


FIG. 1. Linear array used in field measurements (not to scale).

tioner (performing amplification and anti-alias filtering to 4 kHz) followed by an analog-to-digital converter. Each channel was sampled at 12 k samples per second with 12 bits per sample and stored to disk for subsequent analysis and processing.

The low-cost COTS sound source consisted of a portable stereo (Sony CFDG55) playing prerecorded signals from a compact disc. The portable stereo was located collinear with the array 4.6 m from the first microphone and with a height of h_s=38 cm as shown in Fig. 1.

The microphone array was set up in a flat, mowed grass field, located in Glacier Ridge Metro Park of Union County, Ohio, on the afternoon of 16 June 2003. The temperature was 21 °C, the relative humidity was approximately 83%, and the wind was light (<2 m/s) with no dominant direction clearly discernable. First, background noise was recorded to establish the noise spectrum of the microphone outputs. Then, a series of PN sequences was played and recorded to study empirical propagation and time-delay estimation. Finally, the array was reconfigured in a nonlinear fashion and a small subset of the PN sequences were played from various locations to serve as calibration signals for a self-localization experiment described in Sec. V.

B. Source signal generation

A set of source signals, $s(t)$, was generated spanning different time durations, center frequencies, and bandwidths. The source signals were based on PN sequences that were generated from maximum-length shift-register sequences using an m-stage shift register with linear feedback.[18] The binary PN sequences $\{b_i\}_{i=1}^{n}$ drawn from $\{-1, +1\}$ have length $n=2^m-1$ and were chosen for their nearly ideal periodic autocorrelation

$$R(k)=\begin{cases} n, & k=0 \\ -1, & 1\leqslant k\leqslant n-1. \end{cases} \tag{1}$$

The baseband source signal is built up from the sequence values

$$s_{bb}(t)=\sum_{i=1}^{n} b_i p(t-iT_c), \tag{2}$$

where $p(t)$ is a unit amplitude rectangular pulse of duration T_c, and then modulated to the desired center frequency, F_c, to obtain the desired source signal for transmission

$$s(t)=\sin(2\pi F_c t)s_{bb}(t). \tag{3}$$

The bandwidth of $s(t)$ is controlled through T_c, while the signal duration is equal to nT_c. Figure 2 illustrates a typical PN-based source signal at baseband and passband in the time and autocorrelation domains.

In our experiment we chose center frequencies that ranged from 100 to 2000 Hz, bandwidths that ranged from 3 to 3200 Hz, and signal durations that ranged from 0.2 to 10 s. In total, 319 PN source signals were evaluated. The source signals were generated in MATLAB at a sampling rate of 44.1 kHz, exported as audio (.wav) files, and then copied to an audio CD that was played by the portable stereo system during the experiments.

C. Postprocessing

Figure 3 illustrates the normalized power spectral density (PSD) of a $\{F_c$=200 Hz, B=63 Hz, T=2 s$\}$ source signal as it was designed, $G_s(f)$, and as it was received at the first microphone, $G_{r1,r1}(f)$. In this paper, $G_{ri,ri}(f)$ denotes the PSD of the signal received at microphone i, and $G_{ri,rj}(f)$ denotes the cross-spectral density of signals received at microphones i and j. The power spectra in Fig. 3 and elsewhere are estimated using Welch's method of averaging periodograms obtained from windowed segments of time series data.[14,19] A 20-ms Hanning window and 50% window overlap are used throughout this paper.

Harmonic components in the portable stereo are clearly visible in the received signal of Fig. 3. The additional bandwidth provided by these components would aid in our subsequent attempts of time-delay estimation, however they interfere with our attempt to characterize TDE performance based on bandwidth. To provide a fair comparison of the

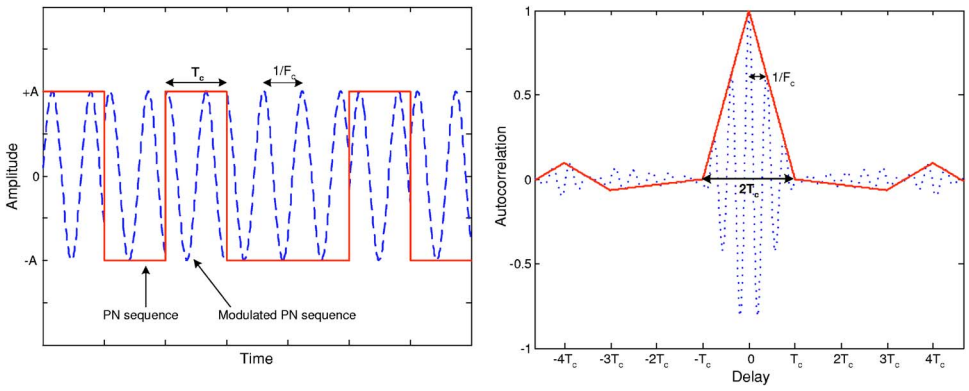


FIG. 2. Example of a baseband PN source signal, $s_{bb}(t)$, and the modulated version, $s(t)$, for transmission (left). Example normalized autocorrelation functions for the baseband (———) and passband (:) signals (right).

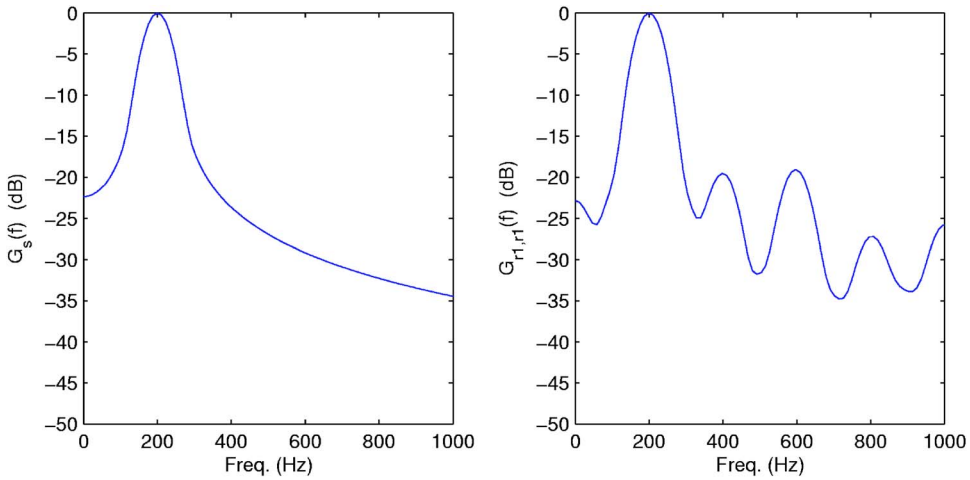


FIG. 3. Power spectral density, $G_s(f)$, of a $\{F_c=200\ \text{Hz}, B=63\ \text{Hz}, T=2\ \text{s}\}$ source signal as designed (left) and as received at the first microphone $G_{r1,r1}(f)$ (right). Harmonic components, likely caused by nonlinearities in the sound system, are clearly visible in the received signal.

different source signals, we include a bandpass filter in the postprocessing phase to eliminate the harmonics from the portable stereo. The bandpass filter was centered at F_c with bandwidth set to the designed bandwidth of the signal, which was taken as the null-to-null bandwidth given by $B=2T_c$.

We also note that postexperiment analysis of the data indicated that measurements from microphones 2 and 8 were corrupt due to hardware failures in those channels. Thus, results from these two microphone signals have been omitted from the empirical results below.

III. EMPIRICAL PROPAGATION

In this section we present the results of experiments that were designed to give an empirical understanding of the acoustic channel over short distances (<75 m) and moderate bandwidths. The aim of this section is not to develop a precise model of acoustic propagation, but rather to empirically study the trends of parameters that are important to time-delay estimation. In particular, we evaluate signal attenuation and signal coherence as a function of distance.

A. Noise and attenuation

Figure 4 presents the observed power spectral density (PSD) of the background noise observed at microphones 1, 3, and 7 (chosen arbitrarily for illustration). The close agreement of the noise PSDs indicates similar noise levels at each position and similar frequency responses from each microphone. The drop at 4000 Hz is due to the antialiasing filter in the signal conditioning hardware.

In Fig. 5 we present the observed PSDs for received signals from a high bandwidth PN sequence $\{F_c=2000\ \text{Hz}, B=3200\ \text{Hz}, T=10\ \text{s}\}$. Because the sound source was uncalibrated, the exact PSD of the transmitted signal was unknown and the attenuation to the first microphone could not be determined. There are, however, several trends that can be noted from the figure. As expected, the signal experiences greater attenuation as the source-receiver distance is increased. These losses also exceed those expected from spherical spreading and atmospheric absorption alone. For example, in going from microphone 1 to 3, the distance from the source approximately quadruples and a loss of 12.7 dB would be expected from spherical spreading;[20] moreover, the expected atmospheric absorption is less than 1 dB for the distances and frequencies considered.[21] However, for most frequencies the actual power loss between $G_{r1,r1}(f)$ and $G_{r3,r3}(f)$ exceeds 20 dB, implying the presence of other losses. Also apparent in the figure is a sequence of nulls whose position in frequency decreases with distance. This trend is the opposite of what would be expected from simple destructive interference from a single ground reflection. Although the details of turbulence models are outside the scope of this paper, we comment that basic models imposing random phase combining of direct and reflected rays at each microphone can explain this phenomenon. The interested reader is referred to Ref. 22 and Ref. 20, Appendix K.

By examining the difference in signal power and the average noise power in Fig. 5 we can investigate the SNR as a function of frequency and distance. For example, at 2000 Hz, microphone signal 7 (which is 50.8 m from the sound source) has an SNR of approximately 18 dB in this experiment. All of the SNRs are observed to be greater than 10 dB except near the null frequency for microphone 7.

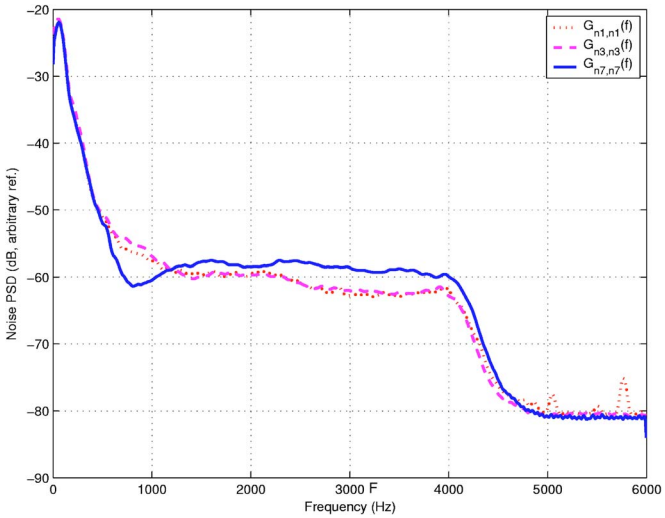


FIG. 4. Power spectral density of background noise as observed at microphones 1, 3, and 7. The close agreement indicates similar noise levels and microphone responses.

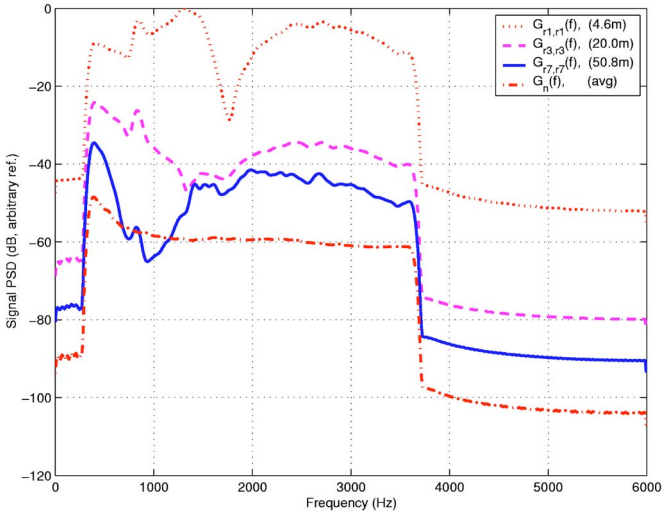


FIG. 5. Received PSDs of a high bandwidth modulated PN sequence $\{F_c = 2000\ \text{Hz}, B = 3200\ \text{Hz}, T = 10\ \text{s}\}$ at microphones 1, 3, and 7. The sharp cutoffs at 400 and 3600 Hz are due to the postprocessing bandpass filter (see Sec. II C). For comparison, the average noise floor after bandpass filtering, $G_n(f)$, is also plotted.

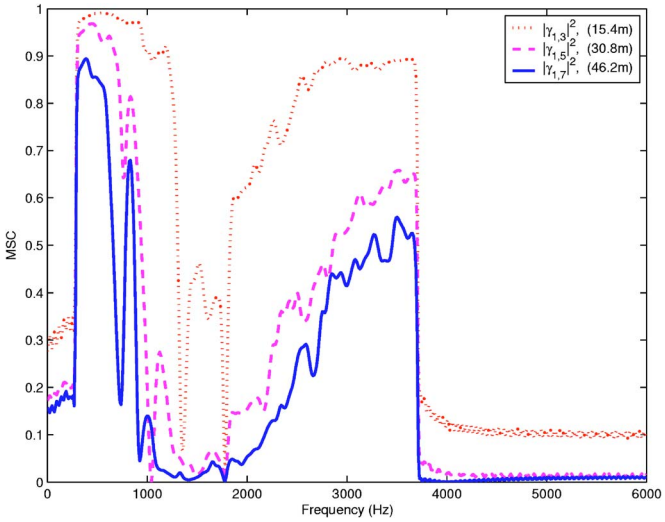


FIG. 6. Magnitude-squared coherence (MSC) as a function of frequency for microphone separations of 15.4, 30.8, and 46.2 m.

B. Coherence

Signal coherence, $\gamma_{ri,rj}(f)$, describes the degree of correlation between like frequency components in two signals $r_i(t)$ and $r_j(t)$. As we describe in Sec. IV, signal coherence plays an important role in time-delay estimation because it affects the quality of the cross correlation function that is used in estimating time delay. In this section we present empirical observations of signal coherence as a function of microphone separation and frequency. The coherence between signals received at microphones i and j is calculated from the PSD estimates as

$$\gamma_{ri,rj}(f) = \frac{G_{ri,rj}(f)}{\sqrt{G_{ri,ri}(f)G_{rj,rj}(f)}}. \tag{4}$$

In the subsequent figures we plot the magnitude-squared coherence (MSC): $|\gamma_{ri,rj}(f)|^2$ - a quantity bounded by 0 and 1, with endpoints representing no correlation and perfect correlation between like frequency components, respectively.[14]

Figure 6 illustrates the coherence for the same 3200-Hz broadband signal whose PSD is shown in Fig. 5. The spatial coherences for pairs of received signals separated by 15.4, 30.8, and 46.2 m are presented from the evaluation of $\gamma_{r1,r3}(f)$, $\gamma_{r1,r5}(f)$, and $\gamma_{r1,r7}(f)$, respectively. This figure provides a general understanding of the coherence effects in the channel during this experiment. The region of severe coherence loss between 1000 and 2000 Hz is attributed to the power loss at the same frequencies. In this case, the background noise is a more significant portion of the received signal and the coherence is reduced accordingly. We observe this by noting that the points of lowest signal power correspond to minima in coherence as well. For example, microphones 1 and 3 experience power minima at 1300 and 1800 Hz, respectively (see Fig. 5), hence $|\gamma_{r1,r3}(f)|^2$ is significantly reduced at these frequencies in Fig. 6. The same holds true for the other microphone pairs. Also apparent in Fig. 6 is the general loss of coherence as distance increases. In addition to power loss, we believe coherence loss is partially a result of distortions in the signal wavefront due to atmospheric turbulence.[22] Coherence versus distance is further investigated below.

In Fig. 7 we explicitly plot the coherence as a function of distance for six low bandwidth signals with different center frequencies. Each line is a plot of $\{|\gamma_{r1,ri}(F_c)|^2\}_{i=3}^{7}$, and the six lines are generated from center frequencies $F_c \in \{100, 200, 400, 800, 1600, 2000\}$. Here, $B = 30$ Hz and

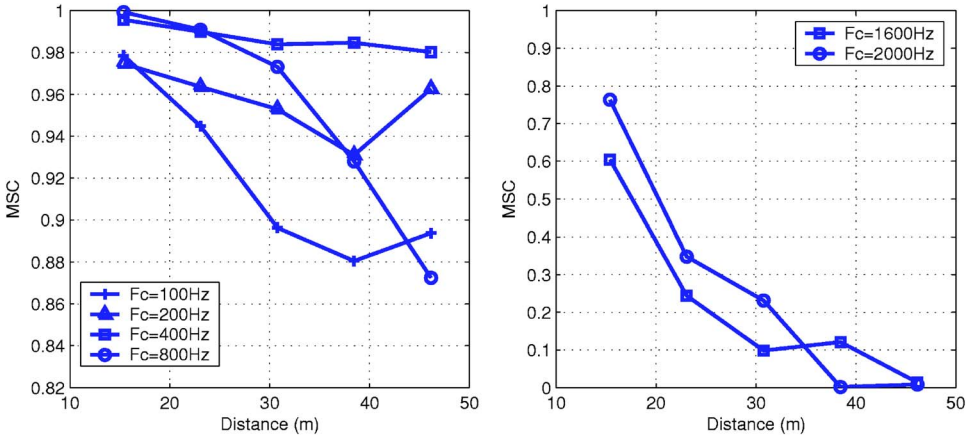


FIG. 7. Magnitude-squared coherence (MSC) versus distance for different center frequencies.

$T=1$ s for all six signals. From Fig. 7, we observe that the 400-Hz signal maintains coherence near unity over the entire distance range, while the 100-, 200-, and 800-Hz signals exhibit slowly decreasing coherence with distance. In contrast, the 1600- and 2000-Hz signals experience severe coherence loss as the distance is increased. For both of these higher center frequencies, the MSC falls below 0.2 when the distance exceeds 32 m.

The impact of coherence loss on time-delay estimation is studied in the following section.

IV. TIME-DELAY ESTIMATION

A. Background

The time-delay estimation (TDE) problem is to estimate the time-difference-of-arrival between received signals at two distant receivers. For a transmitted signal $s(t)$ we have the following received signal model for signals received at sensors i and j:

$$r_i(t) = h_i(t) * s(t) + n_i(t),$$
$$r_j(t) = h_j(t) * s(t-\tau) + n_j(t), \tag{5}$$

where $*$ denotes convolution, τ is the delay to be estimated, $h_i(t)$ is the deterministic channel impulse response from source to node i, and the $n_i(t)$ are independent and identically distributed (i.i.d.) additive noise signals that are assumed to be uncorrelated with $s(t)$ and with each other. As the distance from source to receiver is increased, $h_i(t)$ is expected to apply greater attenuation to $s(t)$.

We use the simple cross-correlator (SCC) to estimate the delay. The SCC delay estimate is the position of the peak in the sample cross correlation between two received signals,[17]

$$\hat{\tau} = \arg\max_{\tau} R_{ri,rj}(\tau). \tag{6}$$

It is recognized that the generalized cross correlator (GCC) is the maximum likelihood estimator in this problem,[17] however we chose the SCC because of its ease in implementation and for its robustness. The GCC requires knowledge of the signal and noise power spectra and can give poor performance if the estimated spectra are mismatched from the true spectra (see, e.g., Ref. 14). Moreover, the loss of statistical efficiency incurred with the SCC is modest for the signal and noise spectra observed and the bandwidths encountered in these experiments. The experiments described in this paper were performed in an open field to minimize the effects of multipath from buildings and other obstacles. For a consideration of identifying first-arriving signal components in a multipath environment see Ref. 23.

Figure 8 presents the results of a computer simulation that illustrates key aspects of the TDE problem. In this example a $\{F_c=200$ Hz$, B=30$ Hz$, T=1$ s$\}$ point source signal is subjected to geometrical spreading, and the resulting SNR and TDE effects are examined. From the lower portion of Fig. 8 we observe that as the sensor separation increases the SNR decreases 6 dB per doubling of distance, as expected for spherical expansion and constant background noise. The

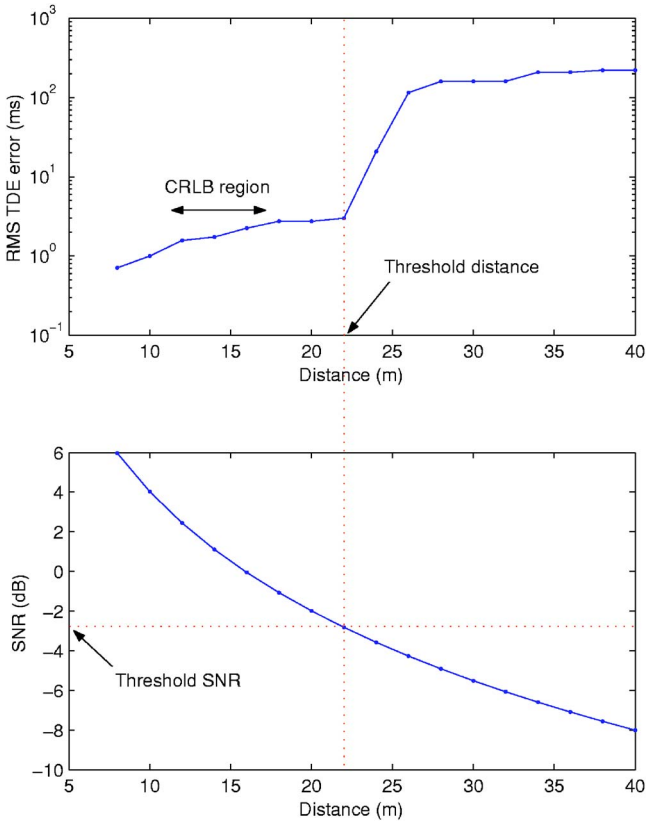


FIG. 8. Simulation result illustrating the threshold phenomenon in time-delay estimation. As the intersensor distance increases, the SNR falls below the threshold SNR and the estimation error rapidly increases. This is due to the peak ambiguity problem that arises when attempting to identify the maximum in the received signals' cross correlation.

rms time-delay estimation error is given in the upper half of the figure for the same distances. At a certain distance, about 22 m in Fig. 8, the SNR drops below a *threshold* signal-to-noise ratio, denoted SNR_{th}, and the error dramatically increases. This is due to the well-known threshold effect in time-delay estimation[15,24] and is caused when the estimator can no longer reliably identify the peak in the cross- correlation because of excess noise or signal decorrelation in the channel. This peak ambiguity results in high estimation error and occurs whenever $\mathrm{SNR} < \mathrm{SNR}_{th}$.

Next we consider means of statistically bounding the TDE error for certain signal and noise parameters. The Cramér-Rao lower bound[25] (CRLB) is the usual statistical tool of choice, however the CRLB is a local bound that is only tight under high SNR conditions. An alternative bound that is tighter over all SNRs is the Weiss-Weinstein lower bound[26] (WWLB). The WWLB is derived for the time-delay estimation problem[16] when the unknown delay is assumed to have a uniform prior distribution over $[-D/2, D/2]$. The variance of the time-delay estimate is then bounded by

$$\sigma_\tau^2 \geq \max_{0<h<D} J(h), \tag{7}$$

where $J(h)$ is given by

$$J(h)=\begin{cases}\dfrac{\frac{1}{2}h^2(1-h/D)^2e^{-\Gamma(1-R_s(h))/2}}{1-h/D-(1-2h/D)e^{-\Gamma(1-R_s(2h))/4}}, & 0\leqslant h<D/2,\\ \frac{1}{2}h^2(1-h/D)e^{-\Gamma(1-R_s(h))/2}, & D/2\leqslant h<D,\end{cases} \tag{8}$$

and where $R_s(h)$ is the source signal's autocorrelation function and $\Gamma=2E/N_0$ is the so-called postintegration SNR, with E being the signal energy and N_0 the single-sided noise power. When $\mathrm{SNR}>\mathrm{SNR}_{th}$ there is no peak ambiguity in the signals' cross- correlation and the bound in Eq. (7) reduces to the CRLB:[15]

$$\sigma_\tau^2 \geqslant \frac{1}{8\pi^2 BTF_c^2\mathrm{SNR}} \quad (\mathrm{SNR}>\mathrm{SNR}_{th}), \tag{9}$$

where B is the signal bandwidth in Hz, and T is the signal duration in seconds. From Ref. 15, the value of SNR_{th} can be estimated from the source's time-bandwidth product BT and bandwidth to center frequency ratio B/F_c:

$$\mathrm{SNR}_{th}=\frac{6}{\pi^2(BT)}\left(\frac{F_c}{B}\right)^2\left[\phi^{-1}\left(\frac{B^2}{24F_c^2}\right)\right]^2, \tag{10}$$

where $\phi(y)=1/\sqrt{2\pi}\int_y^\infty e^{-t^2/2}dt$.

Although the bound in Eq. (7) accounts for peak ambiguities under low SNR, it is still an optimistic bound because it assumes the received signals are fully coherent except for additive noise. In practice, the signals are not fully coherent as we observed in Sec. III. As such, this bound underestimates the error because the assumption of a linear, time-invariant, convolutional channel in Eq. (5) is violated in the acoustic setting. Under these conditions, the CRLB would also be overly optimistic even for high SNRs. To account for this, Kozick and Sadler[9] propose an *effective SNR* that treats the coherent portion of a received waveform as "signal" and treats the incoherent portion as additional noise. With this designation, the effective SNRs of a signal received at microphones i and j are, respectively,

$$\mathrm{SNR}_i=\frac{|\gamma_{si,sj}(F_c)|G_{si,si}(F_c)}{G_n(F_c)+(1-|\gamma_{si,sj}(F_c)|)G_{si,si}(F_c)},$$
$$\mathrm{SNR}_j=\frac{|\gamma_{si,sj}(F_c)|G_{sj,sj}(F_c)}{G_n(F_c)+(1-|\gamma_{si,sj}(F_c)|)G_{sj,sj}(F_c)}, \tag{11}$$

where $G_n(f)$ is the noise PSD that is assumed equal at each microphone, $\gamma_{si,sj}$ is the coherence between the signal portion of the received waveforms (neglecting background noise) at microphones i and j, and similarly $G_{si,si}$ is the PSD of only the signal portion at microphone i. Using these modified SNR expressions, the joint SNR as given in Ref. 15 is modified as

$$\mathrm{SNR}=\frac{\mathrm{SNR}_i\mathrm{SNR}_j}{1+\mathrm{SNR}_i+\mathrm{SNR}_j}, \tag{12}$$

which can be compactly expressed in terms of the MSC of the received waveforms as

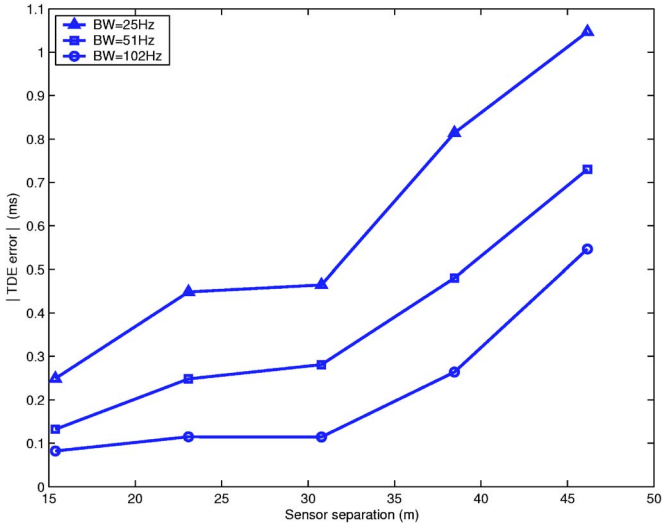


FIG. 9. Observed time-delay estimation error as a function of sensor separation and signal bandwidth. $F_c=100$ Hz, $T=10$ s in all cases.

$$\mathrm{SNR}=\frac{|\gamma_{ri,rj}(F_c)|^2}{1-|\gamma_{ri,rj}(F_c)|^2}. \tag{13}$$

With this formulation, we can compare the SNR of Eq. (13) to the threshold SNR of Eq. (10) for any level of signal coherence.

B. TDE accuracy: Experimental results

In this section we present the results of field experiments designed to study the effects of source signal parameters on time-delay estimation accuracy. A set of 319 different source signals formed from modulated PN sequences was generated with varying durations, bandwidths, and center frequencies as described in Sec. II. Each source waveform was then transmitted from an endfire position as shown in Fig. 1. The time delays were estimated using the SCC in Eq. (6) and compared to the true time differences obtained from the known geometry of the array.

In computing the true time delays from the known geometry we used an estimate of the speed of sound derived from a least-squares fit to a subset of our measurements. This subset was carefully selected to only contain signals with high SNR and from adjacent microphones. The speed of sound so obtained was 343.6 m/s.

1. Error versus bandwidth and signal length

Figure 9 illustrates the observed TDE error of three different source signals. The center frequency and signal duration were held constant at 100 Hz and 10 s, while only the bandwidth was varied. The time-delay estimate as a function of distance was empirically determined by cross correlating the received signal at microphone 1 with the received signals at microphones 3–7. The absolute error, when compared to the true delays, is plotted on the vertical axis. As expected, the higher bandwidth signals had better delay estimation performance. The low TDE errors suggest that all the points are above the threshold SNR and that we are in the CRLB region. From the CRLB in Eq. (9), we expect that doubling the

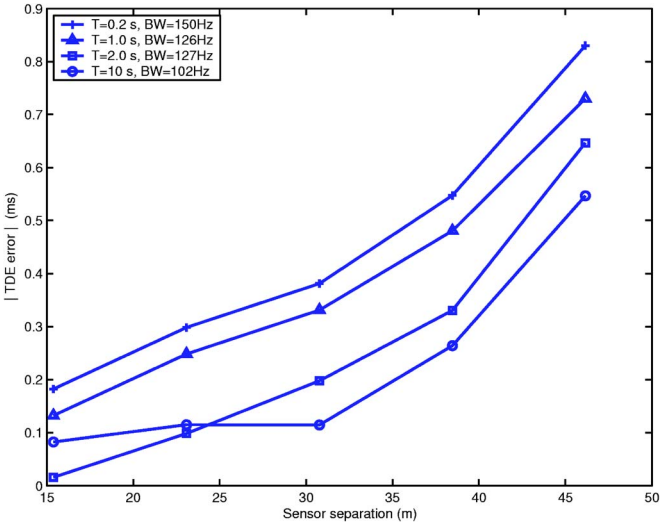


FIG. 10. Observed time-delay estimation error as a function of sensor separation and signal duration. F_c=100 Hz in all cases.

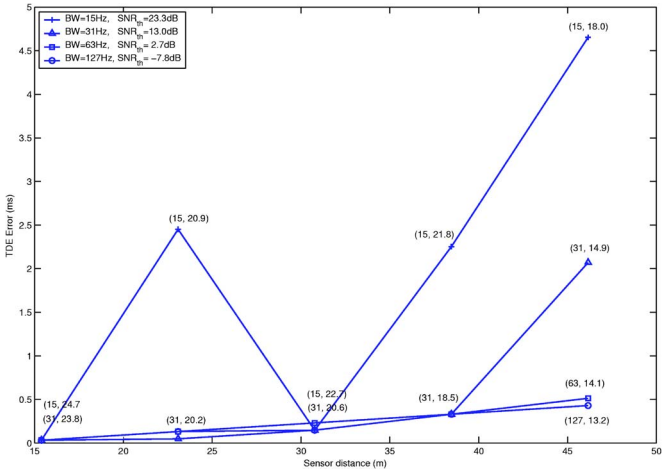


FIG. 11. TDE errors for signals with low bandwidths. As intersensor distance is increased, the two low-bandwidth signals fall below their SNR$_{th}$ and the TDE error dramatically increases. Selected points are marked as B (in Hz) and SNR (in dB). F_c=400 Hz, T=2 s.

bandwidth, and thus the time-bandwidth product, will decrease the TDE error by a factor of $(1-\sqrt{1/2})=29\%$. This expectation is generally confirmed in Fig. 9 where the error reduction, for doubling of bandwith, ranges from 25% to 59% with a median value of 43%.

Figure 10 was produced the same way as Fig. 9 except that signal duration was varied while center frequency and bandwidth were held constant. In this case the added signal length decreases the error as expected, but less than the amount predicted by the CRLB. For example, the time-bandwidth product for the $\{T=1.0\text{ s}, B=126\text{ Hz}\}$ case is 4.2 times the time-bandwidth product in the $\{T=0.2\text{ s}, B=150\text{ Hz}\}$ case, so the expected decrease in TDE error from Eq. (9) is $(1-\sqrt{1/4.2})=51\%$. The actual decrease ranges from 27% at 15.4 m to only 12% at 46.2 m. The reason for this discrepancy is unknown, although it was verified that the longer signals did not exhibit any additional coherence loss. It should also be noted that the errors reported in Figs. 9 and 10 are from a single realization of the experiment, whereas the CRLB represents the squared error averaged over multiple realizations.

Figure 11 illustrates TDE errors for a case where received signal SNRs fall below the threshold SNR in some cases. Figure 11 was produced the same way as Fig. 9 except the signal length has been reduced to T=2 s and the center frequency has been raised to F_c=400 Hz. From Eq. (10), both of these changes necessarily raise the value of SNR$_{th}$ which is given in the legend of the figure for the four different bandwidths. The coherence-corrected SNR values from Eq. (13) are also presented for selected points. In Fig. 11 we observe a rapid increase in error for the two lowest bandwidth signals as the distance increases. The increase in error is quantized to multiples of approximately 2.5 ms. This corresponds to the spacing $(1/F_c)$ between peaks of the correlation function and indicates that the estimator has selected the incorrect peak of the cross-correlation function (see Fig. 2).

The two highest bandwidth signals in Fig. 11 maintain low TDE error over the entire range of distances, and their SNRs are well above the predicted threshold. For example, at 46.2 m the 127-Hz signal is still 21.0 dB above its threshold. In contrast, the two lower bandwidth signals sometimes fall below their threshold SNR, giving rise to a dramatic increase in TDE error. The 31-Hz signal has a threshold SNR of 13.0 dB; measurements at 38.5 m and below are above this threshold, and low TDE errors are seen for these distances. At 46.2 m the estimator has clearly identified an incorrect peak in the cross correlation and the measured SNR (14.9 dB) is very close to the predicted threshold. Finally, for the 15-Hz signal, we observe that three of the five measurements exhibit large TDE errors and that these points all have SNR below the predicted 23.3 dB threshold. Thus, we see that the SNR threshold is a good predictor of whether the TDE error will be low or high in these experiments. As in Figs. 9 and 10, the experimental results in Fig. 11 are for single realizations, whereas the SNR threshold predicts average performance.

2. Error versus center frequency

Figure 12 illustrates the observed increase in TDE error as the center frequency of the PN sequence is increased from 100 to 2000 Hz for different PN sequence bandwidths. There are two major effects contributing to the increase in error. First, for a fixed bandwidth signal, the percent bandwidth, B/F_c, decreases with increasing center frequency and the threshold SNR increases according to Eq. (10). The second cause for the increase in TDE error is the loss of signal coherence as the center frequency increases—which from Eqs. (11) and (13) implies a loss in effective SNR. This connection is illustrated by comparing the observed coherences in Fig. 7 to the TDE errors in Fig. 12. Center frequencies of 1600 and 2000 Hz exhibit the lowest coherence and these have the highest TDE errors. Nearly perfect coherence was observed at F_c=400 Hz and this has the smallest TDE error over the range of bandwidths considered.

Figure 13 illustrates how the SNR threshold changes with center frequency and bandwidth. The top row of plots illustrates the observed TDE error as a function of bandwidth

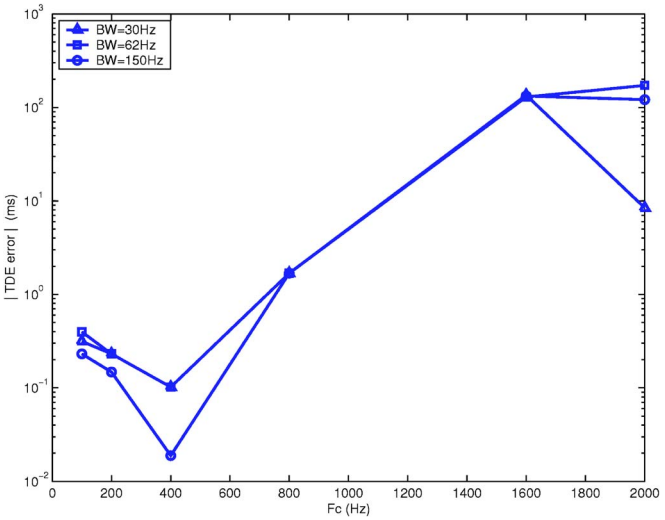


FIG. 12. Experimental time-delay estimation error versus center frequency. $T=1$ s, distance=30.8 m.

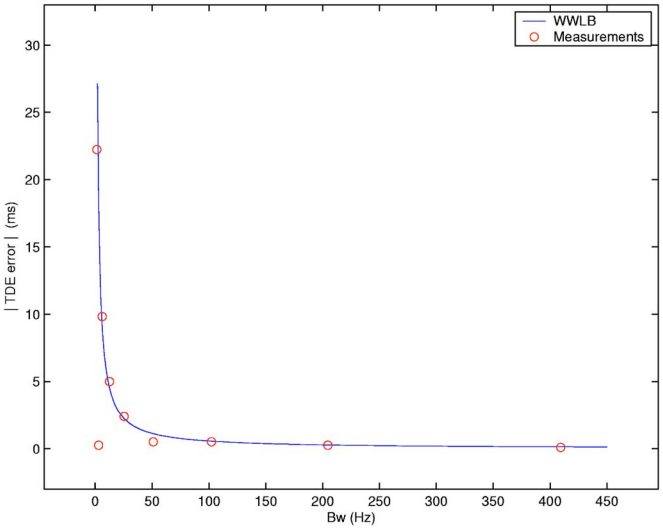


FIG. 14. Time-delay estimation error as a function of signal bandwidth as predicated by the WWLB (-) and for empirical measurements (○). $T=10$ s, $F_c=400$ Hz, distance=46.2 m.

for center frequencies of 200, 400, and 800 Hz. Each plot in the second row corresponds to the one above it and gives the measured total SNR [from Eq. (13)] for each signal. The solid lines in the second row of plots are the threshold SNRs as a function of bandwidth as predicted by Eq. (10). When the observed SNR is above SNR_{th}, we expect low TDE error; this is seen in the top row of plots. The match between the observed and theoretical bandwidth thresholds is not perfect, but the trends are evident.

In Fig. 14 we plot the TDE error bound predicted by the WWLB in Eq. (7) and compare it to empirical errors measured for the $F_c=400$ Hz case. The postintegration SNR for the plot was taken as $\Gamma=16$ dB by fitting Eq. (7) to the TDE measurements. The measured Γ was approximately 35 dB. The tight agreement between the bound and the estimates verifies the validity of the WWLB, however the effective SNR (16 dB) is less than the observed SNR (35 dB).

V. LOCALIZATION EXPERIMENT

In this section we present an experiment in which we estimate both sensor and signal source locations from TDOA measurements obtained from time-delay estimates. The self-localization scenario consists of a number of signal sources placed in a field of sensors with unknown locations. The sources, which also have unknown positions, each transmit a calibration source signal that is detected by a subset of the sensors and used to compute the TDOAs. It is assumed that the emission times of the sources are unknown, but that the sensors all have a common time base. The time measurements are then passed to a localization algorithm to determine the locations of the sensors. Because TDOA measurements only provide information about the relative configuration of sensors, we only evaluate the performance of relative localization. Absolute localization requires prior

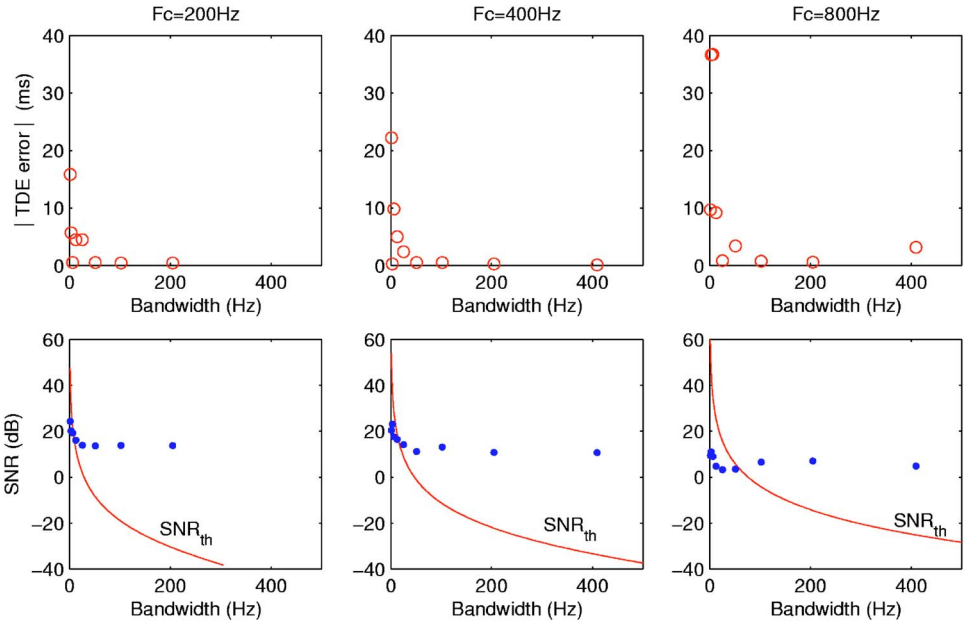


FIG. 13. Illustration of the threshold SNR at different center frequencies. The top row shows empirical TDE error versus bandwidth for three different PN signal center frequencies. Below each of these is a plot illustrating the empirical coherence-corrected SNR for each signal. The solid curves in the bottom row of figures show the threshold SNR values predicated by Eq. (10). The plots confirm low TDE error when the signal SNR is above the threshold and high TDE error when the SNR is below the threshold. $T=10$ s, intersensor distance=46.2 m.

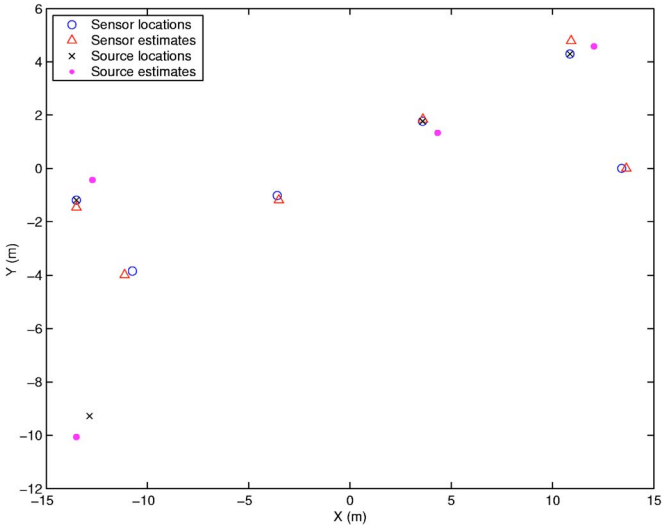


FIG. 15. Sensor network used in self-localization. Shown estimates correspond to calibration experiment 4 of Table I.

location information to determine the translation, rotation, and reflection of the relative solution. This prior information may be obtained by strategic placement of certain nodes or through a GPS receiver embedded in a subset of the sensors. For the purpose of this experiment, relative scene estimates are prescribed a common orientation and translated such that their center of mass is at the origin. In this section we present results of self-localizing an acoustic sensor network using a subset of the PN sequences described earlier as the calibration source signals.

The self-localization experiment used the same equipment described in Sec. II, but altered the acoustic array into the nonlinear configuration depicted in Fig. 15. The six sensors (microphones) are represented by the circles, and the four signal sources are represented by the ×'s. We emulated the four calibration sources by moving the portable stereo to these different positions. Neither the source nor sensor locations are known to the localization algorithm; in addition, while three source locations are co-located with sensors, this information was not provided to the localization algorithm. Because the signal emission times were unknown, the only information available to the localization algorithm was the TDOAs obtained from cross correlations of the received PN sequences as described in Sec. IV. With these time estimates, self-localization was then performed using the maximum likelihood algorithm described in Ref. 27 which simultaneously estimates the x and y locations of the sensors along with the locations and signal emission times of the sources. The unknown source locations and unknown emission times are considered nuisance parameters in the sensor localization problem. The results from a $T=2$ s calibration signal with $F_c=200$ Hz and $BW=127$ Hz are presented in Fig. 15. Sensor location estimates are shown by triangles, while estimates of the source locations are given by solid dots. The average sensor localization error of the scene estimate was calculated as

TABLE I. Source signals used in network self-calibration. Position estimates were made from time-delay estimates of these signals and errors were calculated from known true positions.

Experiment no.	F_c (Hz)	BW (Hz)	length (s)	Average localization error (m)
1	200	14	1	2.95
2	200	127	2	0.22
3	400	30	1	0.98
4	400	127	2	0.28
5	800	127	2	5.63
6	800	254	1	1.57
7	1600	1023	2	31.69

$$\frac{1}{N}\sum_{n=1}^{N}\sqrt{(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2}, \tag{14}$$

where $N=6$ is the number of sensors, x_i and $\hat{x}_i$ are the true and estimated x coordinates of the ith sensor, respectively, and y_i and $\hat{y}_i$ are the true and estimated y coordinates of the ith sensor. For this source signal, the average localization error was 0.28 m.

The preceding experiment was repeated for several different PN-based source signals. The source signal parameters and their resulting average localization errors are given in Table I. The estimates in Fig. 15 correspond to experiment 4 in this table. For most of the signals, we see good agreement between estimated and actual sensor locations. The exception is experiment 7, which has the worst performance even though it has the greatest bandwidth. The poor localization performance is due to poor time-delay estimates which are caused by the degraded signal coherence at $F_c=1600$ Hz. Similarly, the smallest average localization errors correspond to low-frequency source signals that were previously observed to have high signal coherence.

VI. CONCLUSIONS

We have presented experimental results from an outdoor field experiment designed to study the effects of source signal bandwidth, center frequency, and duration on both signal coherence and time-delay estimation error. One goal was to understand the achievable accuracy of acoustic TDE using low-cost commercial equipment and simple estimation algorithms.

Modulated pseudo-noise signals were used to estimate both signal coherence and TDE accuracy as functions of bandwidth, signal duration, and center frequency. A set of 319 different modulated PN source signals spanning these three signal parameters was transmitted toward a linear array of length 46.2 m, and received signals were used to assess signal coherence and TDE performance. A simple cross-correlation estimator was employed to examine time-delay estimation accuracy. In general, we found that the modulated PN source signals and the SCC estimator were effective in performing time-delay estimation using our uncalibrated COTS equipment.

Both signal coherence and TDE accuracy followed trends predicted by theory to within the limits of the low-cost hardware. As expected, signal coherence was generally higher at lower frequencies, but measured signal coherence dropped off slightly for frequencies below 400 Hz, as a result of greater noise levels and the reduced signal transmission power of the equipment at low frequencies. For center frequencies below 800 Hz, we obtained TDE errors on the order of 0.5 ms for a sensor separation of 46.2 m.

We also considered how well mathematical theory predicted TDE behavior in the aeroacoustic environment. The CRLB and WWLB provide a theoretical expectation that the mean-squared error of time-delay estimates will vary inversely with the source signal's time-bandwidth product and the coherence-corrected effective SNR; however, higher bandwidths require higher center frequencies, and coherence decreases sharply with distance at higher center frequencies. Our experiments have confirmed these theoretically predicted trends in the aeroacoustic regime; however, in many cases the actual performance was lower than theoretical predictions based on measured SNR values. Besides predicting the TDE error, detecting highly erroneous time-delay estimates is an important element in network self-localization. TDE theory predicts an SNR threshold below which estimates are subject to peak ambiguities resulting in relatively high error. Using Eqs. (10) and (13) we were successful in calculating a coherence-corrected SNR that, when compared to the theoretical threshold, correctly predicted whether the estimate was in the peak ambiguous region or not.

Finally, we presented results from an outdoor self-localization experiment where we used the modulated PN sequences to obtain time-difference-of-arrival estimates for six randomly placed sensors. Our primary goal was to experimentally assess localization performance as a function of the center frequency, bandwidth, and duration of calibration signals used to obtain TDE estimates that form the basis for localization estimates. Using four calibration sources we obtained an average localization error of 0.22 m in the best case.

ACKNOWLEDGMENTS

This work was prepared through collaborative participation in the Advanced Sensors Consortium sponsored by the U. S. Army Research Laboratory under the Collaborative Technology Alliance Program, Cooperative Agreement DAAD19-01-02-0008. The authors gratefully acknowledge the assistance of D. Keith Wilson in the evaluation of turbulence effects.

[1]S. Kumar, F. Zhao, and D. Shepherd, "Collaborative signal and information processing in microsensor networks," IEEE Signal Process. Mag. **19**(2), 13–14 (2002).

[2]N. Bulusu, J. Heidemann, and D. Estrin, "GPS-less low-cost outdoor localization for very small devices," IEEE Personal Commun. Mag. **7**, 28–34 (2000).

[3]G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," Commun. ACM **43**(5), 51–58 (2000).

[4]D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," IEEE International Conference on Acoustics, Speech, and Signal Processing, May 2001, Vol. 4, pp. 2033-2036.

[5]R. L. Moses, D. Krishnamurthy, and R. Patterson, "A self-localization method for wireless sensor networks," Eurasip J. Appl. Signal Process., Special Issue on Sensor Networks **2003**, 348–358 (2003).

[6]R. Moses and R. Patterson, "Self-calibration of sensor networks, Unattended ground sensor technologies and applications IV," Proc. SPIE **4743**, 108–119 (2002).

[7]N. Patwari, A. Hero III, M. Perkins, N. Correal, and R. O'Dea, "Relative location estimation in wireless sensor networks," IEEE Trans. Signal Process. **51**(8), 2137–2148 (Aug 2003).

[8]D. Niculescu and B. Nath, "*Ad hoc* positioning systems (APS) using AOA," Proceedings IEEE INFOCOM '03, April 2003.

[9]R. Kozick and B. Sadler, "Algorithms for localization and tracking of acoustic sources with widely separated sensors," 2000 Meeting of the MSS Specialty Group on Battlefield Acoustic and Seismic Sensing, 17–19 October 2000.

[10]R. Kozick and B. Sadler, "Source localization with distributed sensor arrays and partial spatial coherence," IEEE Trans. Signal Process. **52**(3), 601–616 (2004).

[11]J. L. Spiesberger, "Locating animals from their sounds and tomography of the atmosphere: Experimental demonstration," J. Acoust. Soc. Am. **106**(2), 837–846 (1999).

[12]B. G. Ferguson, L. G. Criswick, and K. W. Lo, "Locating far-field impulsive sound sources in air by triangulation," J. Acoust. Soc. Am. **111**, 104–116 (2002).

[13]S. E. Dosso, N. E. B. Collison, G. J. Heard, and R. I. Verrall, "Experimental validation of regularized array element localization," J. Acoust. Soc. Am. **115**, 2129–2137 (2004).

[14]G. Carter, *Coherence and Time Delay Estimation* (IEEE, Piscataway, NJ, 1993).

[15]A. J. Weiss and E. Weinstein, "Fundamental limitations in passive time delay estimation, part I: Narrow-band systems," IEEE Trans. Acoust., Speech, Signal Process. **31**(2), 472–485 (1983).

[16]A. J. Weiss, "Composite bound on arrival time estimation errors," IEEE Trans. Aerosp. Electron. Syst. **22**(6), 751–756 (1986).

[17]C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. Acoust., Speech, Signal Process. **24**(4), 320–327 (1976).

[18]J. Proakis, *Digital Communications* (McGraw-Hill, Boston, MA, 1995).

[19]P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," IEEE Trans. Audio Electroacoust. **AU-15**, 70–73 (1967).

[20]E. M. Salomons, *Computational Atmospheric Acoustics* (Kluwer Academic, Dordrecht, 2001).

[21]H. E. Bass, L. N. Bolen, R. Raspet, W. McBride, and J. Noble, "Atmospheric absorption of sound: further developments," J. Acoust. Soc. Am. **97**, 680–683 (1995).

[22]V. E. Ostashev and D. K. Wilson, "Coherence function and mean field of plane and spherical sound waves propagating through inhomogeneous anisotropic turbulence," J. Acoust. Soc. Am. **115**, 497–506 (2004).

[23]J. L. Spiesberger, "Linking auto- and cross-correlation functions with correlation equations: Application to estimating the relative travel times and amplitudes of multipath," J. Acoust. Soc. Am. **104**, 300–312 (1998).

[24]J. P. Ianniello, "Time delay estimation via cross-correlation in the presence of large estimation errors," IEEE Trans. Acoust., Speech, Signal Process. **30**(6), 998–1003 (1982).

[25]H. V. Poor, *An Introduction to Signal Detection and Estimation*, 2nd ed. (Springer, New York, 1994).

[26]A. J. Weiss and E. Weinstein, "Lower bounds on the mean square error in random parameter estimation," IEEE Trans. Inf. Theory **IT-31**(5), 680–682 (1985).

[27]R. L. Moses, D. Krishnamurthy, and R. Patterson, "An auto-calibration method for unattended ground sensors," IEEE International Conference on Acoustics, Speech, and Signal Processing, 13–17 May 2002, Vol. **3**(4), III-2941–III-2944.

Wavelet preprocessing for lessening truncation effects in nearfield acoustical holography

Jean-Hugh Thomas[a] and Jean-Claude Pascal[b]
Laboratoire d'Acoustique de l'Université du Maine, LAUM UMR-CNRS 6613, Avenue Olivier Messiaen 72085 Le Mans Cedex 09, France

(Received 22 July 2004; revised 31 March 2005; accepted 11 May 2005)

The goal of planar nearfield acoustical holography (NAH) is to recover the sound field at the sound source from pressure measurements made close to the source plane. The theory requires the pressure to be measured over a complete plane. Because experimentation consists of acquiring only a finite measurement aperture of the pressure field, it naturally causes erroneous values in the reconstructed field. Wavelet preprocessing applied to the pressure measurements in the nearfield provides a solution to lessen the effects due to the truncation of the hologram. The approach is based on a multiresolution analysis of the field from different wave number bands followed by selective spatial filtering of effects highlighted by the first analysis. Experimental results show the relevance of the method by comparison to standard NAH involving exponential filtering in the wave number domain. The computation of objective indicators based on distance measurements between wave number spectra and comparisons between patterns composed of relevant features drawn from experimental data are proposed to give objective criteria to prove the viability of the method. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1945469]

PACS number(s): 43.60.Sx, 43.60.Hj, 43.60.Pt [EGW] Pages: 851–860

I. INTRODUCTION

The aim of planar nearfield acoustical holography (NAH) is to reconstruct the sound field at the sound source from pressure measurements made close to the source plane. These measurements are carried out using an array of microphones or a single microphone moving across a finite size grid when the acoustic noise is stationary. Thus holography is applied to a truncated acoustic measurement field, which causes erroneous values in the reconstructed pressure field. Indeed, the discontinuity in the acquired pressure field increases the distortions of the usual reconstruction procedure due to the exponential amplification of the evanescent waves in the wave number spectrum. The goal of our study is to lessen the effects of the truncation of the hologram in planar NAH. A wavelet method is used for this purpose. Full details of the approach are given in the first part of the paper. In the second part, some indicators are proposed to compare results obtained by standard NAH and by NAH coupled with the wavelet preprocessing. The aim of this part is to prove the relevance of the method discussed, from objective criteria computed from the wave number spectra involved.

Several authors have already worked on the lessening of truncation effects. A weighting window is proposed[1] but it has the disadvantage of adding distortions when the gradient of the rebuilt pressure field is to be used to calculate the acoustical and structural intensity. Some authors have applied regularization approaches to NAH problems, based on the singular value decomposition (SVD) of the matrix which relates the normal velocity of the vibrating surface to the hologram pressure on the measurement plane.[2–4] Williams, in particular, presented a comparative study of regularization methods such as the Tikhonov procedure, the Landweber iteration, and the conjugate gradient approach.[5] In his application, he underlined the fact that the Veronesi filter,[6] or exponential filter, provided the best results, even in comparison with a modified Tikhonov approach, improving the results given by the three tested methods. That is why the approach using wavelet processing, presented in the paper, is compared to NAH with the use of the Veronesi filter as a regularization tool. The advantages of these regularization methods are that they process automatically and can be used for planar, cylindrical, and spherical geometries. Tikhonov regularizations in conjunction with a parameter-choice method [L-curve criterion or the generalized cross-validation method (GCV)] are also successfully used for reconstructing acoustic sources using the inverse boundary element method (IBEM)[7] or, more recently, hybrid NAH based on a modified Helmholtz equation least-squares (HELS) method.[8] Another approach consisted in reconstructing the pressure field of the source plane over an area larger than the small region, or "patch," where the pressure was measured.[9,10] The starting point of the method is the measurement plane, i.e., the hologram, extended by zero-padding to the size needed for the pressure field of the source plane. The idea is to reconstruct the pressure field on the source plane and then to propagate the latter to the measurement plane. The central part, whose dimension is the same as that of the patch, is replaced by the hologram and the process reiterates until the field propagated is similar to the hologram. The method seems to give interesting results even though it needs many iterations. Other techniques which accurately calculate the spatial derivatives in the wave number domain[11,12] cannot be employed in backward propagation because they limit the width or modify the shape of the k spectrum.

[a]Electronic mail: jean-hugh.thomas@univ-lemans.fr
[b]Electronic mail: jean-claude.pascal@univ-lemans.fr

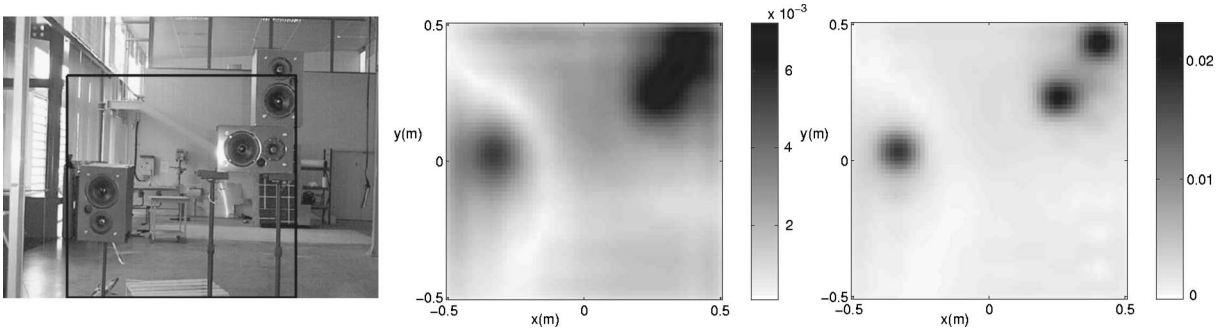


FIG. 1. The experiment with three loudspeakers in a vertical plane and a square frame is shown left. The 16×16 pressure field A_0 (center), measured from this frame, 10 cm from the source plane, is the starting point of the study. The pressure field directly acquired 1 cm from the source plane on the right is the reference field intended to be reconstructed from the image in the center using NAH with the wavelet preprocessing.

The approach discussed in the paper does not belong to regularization-based methods but involves multiresolution decomposition of the acoustical field. The aim of this analysis is to highlight truncation effects that appear at different resolutions. Once the effects have been located, they are selectively filtered. To our knowledge, El Khoury and Nouals were the first to propose this two-step method.[13] The stages of the method are as follows: multiresolution analysis and selective spatial filtering are described in Sec. III. The selective spatial filtering we proposed is, in particular, different from the Hamming window used by El Khoury and Nouals.[13] The wavelet processing is also presented from a wave number point of view: each decomposition of the acoustic field at one resolution involves specific wave number filtering, highlighted in Sec. III. Section IV describes the results obtained using the proposed wavelet method to experimentally reconstruct acoustic fields very near the source plane. The experimental configuration tested here involves a three-monopole point source plane that has already been tested successfully on simulated cases.[14] Several figures, illustrating back-propagated acoustic fields, are also displayed to compare results obtained from standard NAH and from NAH coupled with wavelet preprocessing. Because it is sometimes difficult to make visual comparisons of processing results in acoustic imaging, we propose in Sec. V to use objective indicators to prove the relevance of the wavelet method for NAH. First the wave number spectra resulting from standard NAH and from wavelet processing are compared to a reference spectrum through the computation of a similarity indicator between the spectra, which is based on distance measurements. The smallest distance gives the most relevant spectrum. Second, a pattern recognition approach is used: several features are extracted from the wave number spectra to create a pattern which may be then represented by a point in the feature space. Each method gives a pattern whose location in the feature space is compared to that of a reference pattern. The closer the pattern is to the reference, the more relevant the source reconstruction. As the steps of the proposed method are illustrated all along the paper by images of the experimental acoustic fields, the paper starts in Sec. II with a description of the experiment and explains the principle of the study.

II. DESCRIPTION OF THE STUDY AND THE EXPERIMENT

The experimental configuration for the holographic measurements consisted of three stationary loudspeakers excited in a vertical plane (Fig. 1). Two of them were close to each other. The pressure was measured 10 cm from the source plane on a square grid of 28×28 points. The scan was conducted in an automated point-by-point fashion using a single microphone. The complex pressure was retrieved from the phase relations between the acquired acoustic signals and the excitation signal. The step size in both x and y directions was $\Delta L=6.7$ cm, providing an overall scan dimension of 1.8×1.8 m^2. The study focuses on a smaller surface of 16×16 points located at the center of this area (see the black square frame in the picture in Fig. 1). This region, which measured 1.0×1.0 m^2, was chosen so that the highest loudspeaker was very near its border, resulting in unsatisfactory values in the reconstruction processing. The wavelength and the wave number of the stationary sources being studied are respectively $\lambda=0.85$ m and $k_0=2\pi/\lambda=7.4$ rad m^{-1} at the frequency of 400 Hz. The aim of our experiment is to reconstruct an acoustic field 1 cm from the source plane, from the acquired acoustic field 10 cm from the loudspeaker plane. From the experiment shown on the left in Fig. 1, the 16×16-point pressure field in the center, acquired 10 cm from the sources, is back-propagated 1 cm from the source plane using NAH. The pressure field resulting from the method discussed in the paper is expected to be similar to the reference one on the right directly acquired 1 cm from the source plane. These fields, as throughout the paper, are interpolated on a 64×64 grid using the Shannon interpolation from the computed fields on a 16×16 grid, giving sharper images.

The wavelet method,[13–15] which is both space and frequency selective, involves modifying the acquired acoustic pressure field before solving the holography inverse problem (NAH) as shown by the synopsis in Fig. 2. The method differs from the standard one in that it does not apply any filtering in the wave number spectrum after the Fourier transform. However, there is no denying that the wavelet preprocessing itself filters the data in the wave number domain. The aim of the next section is to explain how the acquired pressure field is modified by the wavelet method and in particular how its wave number spectrum is affected.

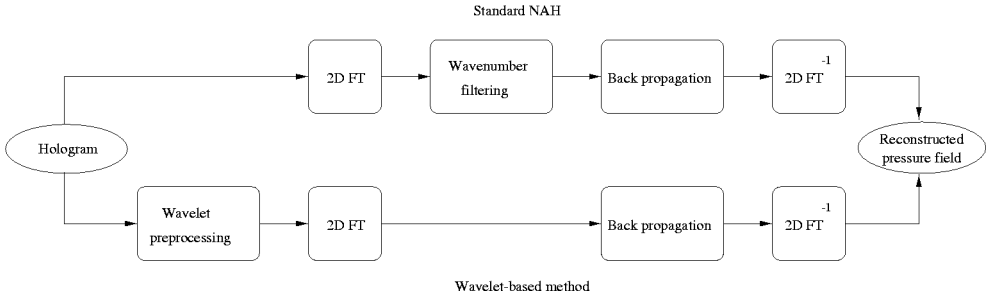


FIG. 2. Synopsis of the wavelet-based method and the standard holography. 2D FT denotes the two-dimensional Fourier transform and 2D FT^{-1} its inverse.

III. A WAVELET PREPROCESSING METHOD

A. Principle

The first stage is based on a multiresolution analysis[16] of the acquired pressure field 10 cm from the source plane. It provides subimages highlighting different characteristics of the pressure field according to a given resolution, in particular the edges caused by the finite size of the acoustic field. The main advantage of decomposition (which is used in data compression with MPEG video compression standards) is that the image studied may be reconstructed from the subimages, using a simple algorithm involving zero insertion and summation.

The second stage of the method is based on selective spatial filtering of the edges highlighted by the multiresolution analysis. The spatial filter is a Π-modified function which was devoted to pattern recognition applications[17] as a model of class membership function. Here it is applied to the subimages. Then the resulting subimages are added, yielding a modified pressure field to be used as an input of the holography algorithm.

B. Multiresolution analysis

In this part, multiresolution analysis is essentially presented in a physical sense as a filter bank analysis without the theory of wavelet bases which is reported in detail in the literature.[16,18] The method consists in analyzing a signal or an image at several resolutions. First it focuses on details, then it gives increasingly coarse approximations of the signal studied.

The decomposition of 1-D signals at one resolution involves a high-pass filter and a low-pass filter to separate the input signals into two components: an approximate and a detailed signal. At the first resolution, for 1-D signals sampled at frequency f_e, the high-pass filter extracts details in the $[f_e/4, f_e/2]$ band and the low-pass filter gives an approximation in the $[0, f_e/4]$ band. At the second resolution and so on, this two-step processing is reiterated on the approximation given from the previous resolution. At resolution j, the high-pass filter extracts details in the $[f_e/2^{j+1}, f_e/2^j]$ band and the low-pass filter gives an approximation in the $[0, f_e/2^{j+1}]$ band.

For images, the approach is similar except that the filters operate on the lines and the columns, involving four filtering operations instead of two. In this way, the first resolution wavelet analysis of A_0, the acquired pressure field 10 cm from the source plane, yields the four subimages A_1, H_1, V_1, and D_1 of Fig. 3 such that

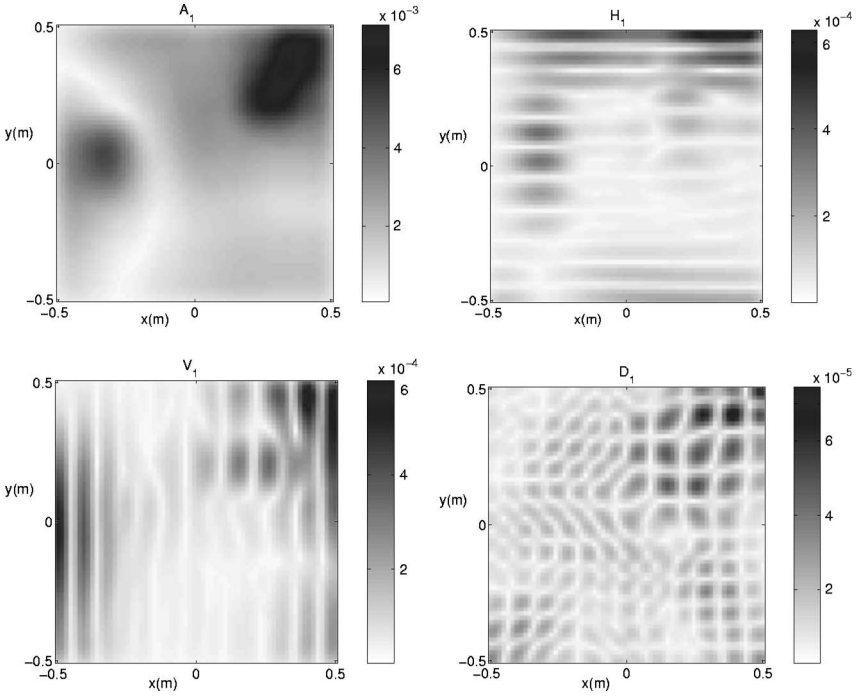


FIG. 3. Resulting acoustic fields (magnitude) without detail top left (A_1), with the horizontal edges top right (H_1), the vertical edges bottom left (V_1), and the corners bottom right (D_1) from a one-level multiresolution analysis of the acquired radiated field (A_0) in Fig. 1. The analysis is based on a Daubechies wavelet with eight vanishing moments.

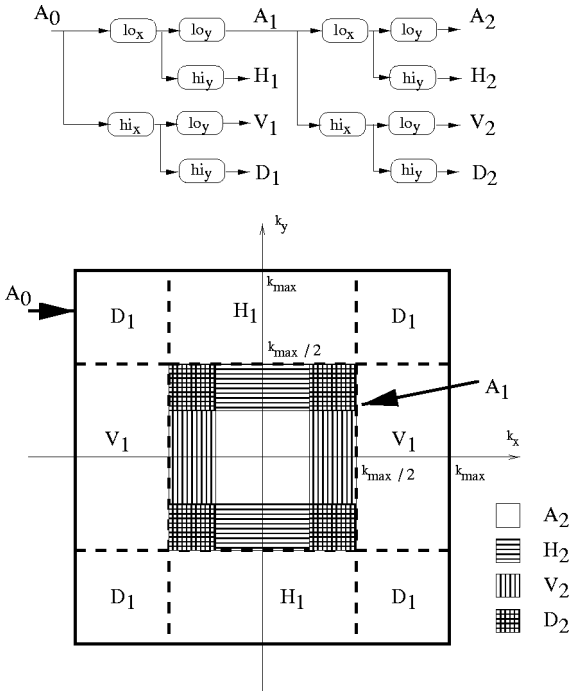


FIG. 4. Multiresolution analysis at level 2 from a wave number spectrum point of view: the original image A_0 is decomposed into seven subimages H_1, V_1, D_1, A_2, H_2, V_2, and D_2 using high-pass filtering (hi) and low-pass filtering (lo) horizontally (subscript x) or vertically (subscript y). The whole wave number domain of the acoustic field A_0, represented by a square of size k_{max}, is divided into several areas, each of them corresponding to a specific wave number band.

$$A_0 = A_1 + H_1 + V_1 + D_1. \tag{1}$$

The analysis highlights different components of the acoustic field as shown by analyzing wavelet processing in the wave number domain. Indeed, a j-resolution analysis of an acoustic field leads to filtering in the wave number space known as k space (see bottom part in Fig. 4):

(i) in the $[0, k_{max}/2^j]$ band in both directions k_x and k_y to obtain the approximation A_j (k_x and k_y are respectively wave numbers in x and y directions),

(ii) in the $[0, k_{max}/2^j]$ band in the k_x direction and in the $[k_{max}/2^j, k_{max}/2^{j-1}]$ band in the k_y direction to obtain image H_j highlighting the horizontal edges due to high-pass filtering on the vertical axis,

(iii) in the $[k_{max}/2^j, k_{max}/2^{j-1}]$ band in the k_x direction and in the $[0, k_{max}/2^j]$ band in the k_y direction to obtain image V_j highlighting the vertical edges due to high-pass filtering on the horizontal axis,

(iv) in the $[k_{max}/2^j, k_{max}/2^{j-1}]$ band in both directions k_x and k_y to obtain image D_j providing the details,

with $k_{max} = \pi/\Delta L$, the wave number limit due to space sampling ΔL.

In this approach, filtering is done on the lines, then on the columns or both on the lines and the columns (see top part in Fig. 4) while standard holography uses a unique circular filter that retains only wave numbers smaller than a k space cutoff k_c.

Figure 3 shows that features due to hologram truncation are spatially localized near the boundaries of the radiated field (see in particular the dark horizontal lines on image H_1 and the dark vertical lines on image V_1). Note that multiresolution analysis is performed using Daubechies wavelets computed from finite impulse responses of filters with 16 coefficients given in Ref. 16.

C. Selective spatial filtering

The aim is to remove the features due to the finite aperture of the antenna, emphasized by multiresolution analysis. Several standard windows (Hamming, Kaiser,…) can be used at this stage, but they generally have the disadvantage of reducing the area of the acquired acoustic field. Thus a two-dimensional Π-modified function Π_m shown in Fig. 5 and defined in 1-D by Eq. (2) is used: its shape is easily adjustable by changing two parameters λ and β. β sets the length of the top of the function and λ its slope:

$$\Pi_m(x,\lambda,\beta) = \begin{cases} 2\left(1 - \dfrac{x^2-\beta}{\lambda}\right)^2 & \text{if } \dfrac{\lambda}{2} + \beta \leqslant x^2 \leqslant \lambda + \beta, \\ 1 - 2\dfrac{[x^2-\beta]^2}{\lambda^2} & \text{if } \beta \leqslant x^2 \leqslant \dfrac{\lambda}{2} + \beta, \\ 1 & \text{if } 0 \leqslant x^2 \leqslant \beta, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The Π-modified functions of Fig. 5 are applied to the subimages provided by multiresolution analysis to filter the horizontal edges in the y direction, the vertical edges in the x direction, and the corners in both directions with a polar shape. Only the approximate subimage remains unchanged.

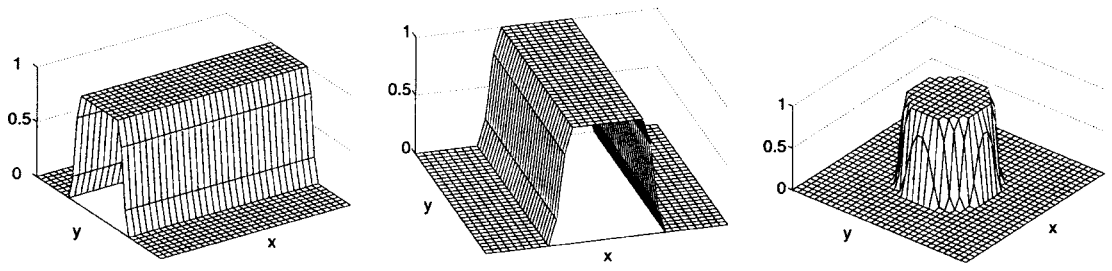


FIG. 5. Two-dimensional Π-modified functions (β=0.3, λ=0.2) to selectively filter horizontal edges in the y direction on the left, vertical edges in the x direction in the middle, and corners in both directions with a polar form on the right. The filters are respectively applied to acoustic fields H_1, V_1, and D_1 of Fig. 3 yielding the fields H_{1f}, V_{1f}, and D_{1f} of Fig. 6.

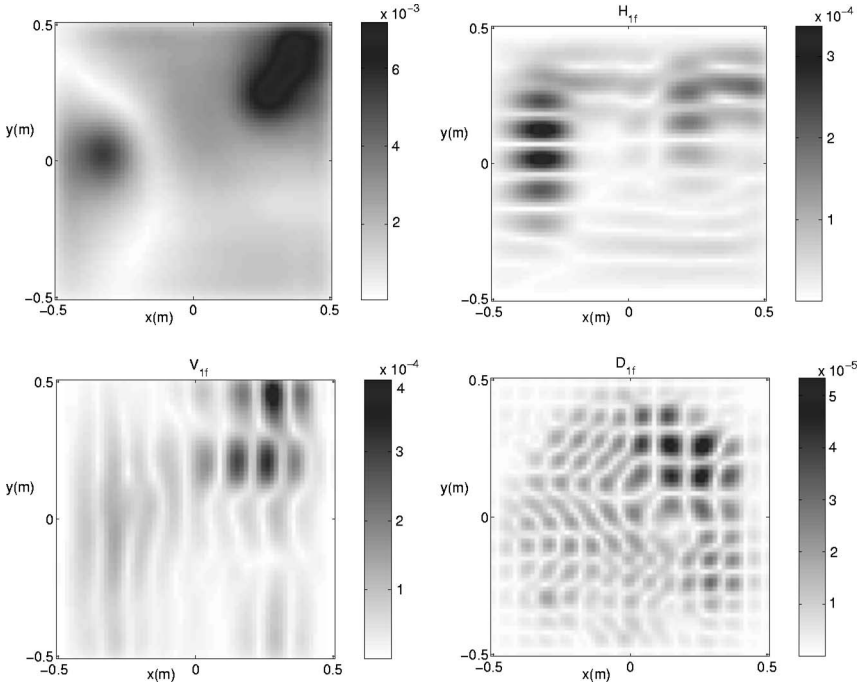


FIG. 6. Results (magnitude) of the filtering of the three detailed fields (H_1, V_1, D_1) in Fig. 3 by the Π-modified functions in Fig. 5 for horizontal edges top right (H_{1f}), vertical edges bottom left (V_{1f}), and corners bottom right (D_{1f}). The modulus of the modified acoustic field resulting from the addition of A_1, H_{1f}, V_{1f}, and D_{1f} is top left.

The three acoustic fields H_{1f}, V_{1f}, and D_{1f}, which result from selective spatial filtering near the borders of the antenna, are shown in Fig. 6. The strong edges have been removed, highlighting components that were masked before the selective filtering.

These three selectively filtered acoustic fields are then added to the approximation top left in Fig. 3 to give a new acoustic field $\hat{A}_1$ intended to be used as an input of the holography process (top left in Fig. 6):

$$\hat{A}_1 = A_1 + H_{1f} + V_{1f} + D_{1f}. \tag{3}$$

Figure 7 shows the wave number representation of the acquired pressure field 10 cm from the source plane (A_0) and the modified field resulting from the wavelet processing $(\hat{A}_1)$ to be presented to the NAH procedure. The modified acoustic field $\hat{A}_1$ results from a single resolution analysis. If wavelet preprocessing is performed at resolution j, the modified acoustic field $\hat{A}_j$ involves $3j+1$ subimages, $3j$ of which are spatially filtered, such that

$$\hat{A}_j = A_j + \sum_{i=1}^{j} (H_{if} + V_{if} + D_{if}). \tag{4}$$

IV. RESULTS FROM RECONSTRUCTED ACOUSTIC FIELDS

The back-propagation algorithm of standard NAH is run on either the acquired acoustic field 10 cm from the source plane or the modified pressure field of Eq. (4) resulting from the wavelet preprocessing. Figure 8 shows the reconstructed spatial acoustic fields 1 cm from the source plane from different configurations. Standard holography was investigated through the use of two filters in the wave number domain: the one proposed by Veronesi and Maynard[6] (also mentioned by Williams[5]), whose taper follows an exponential shape, and the one given by Li *et al.*,[19] derived from Veronesi's filter. The former is designed to eliminate too many propagative wave number components in the case of a weak slope at the break point. The two filters are tested with two cutoff wave numbers. The wavelet processing was performed for three resolutions, 1, 2, and 3. The influence of the size of the acquisition grid is also investigated. Whereas wavelet preprocessing always involves 16×16 pressure fields, standard holography is experimented on both 16×16 and 28×28 acoustic fields. Each calculation is made on a 32×32 grid. Indeed, the acquired fields are extended to 32×32 points by

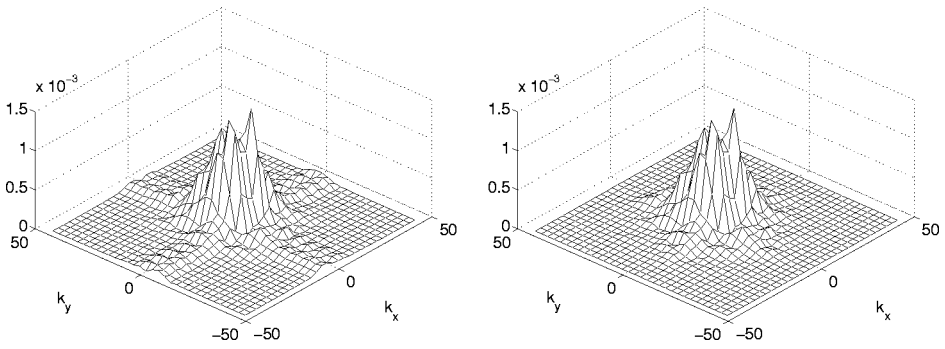


FIG. 7. K space representation (magnitude) of the acquired pressure field 10 cm from the source plane on the left and the modified field resulting from the wavelet processing on the right.

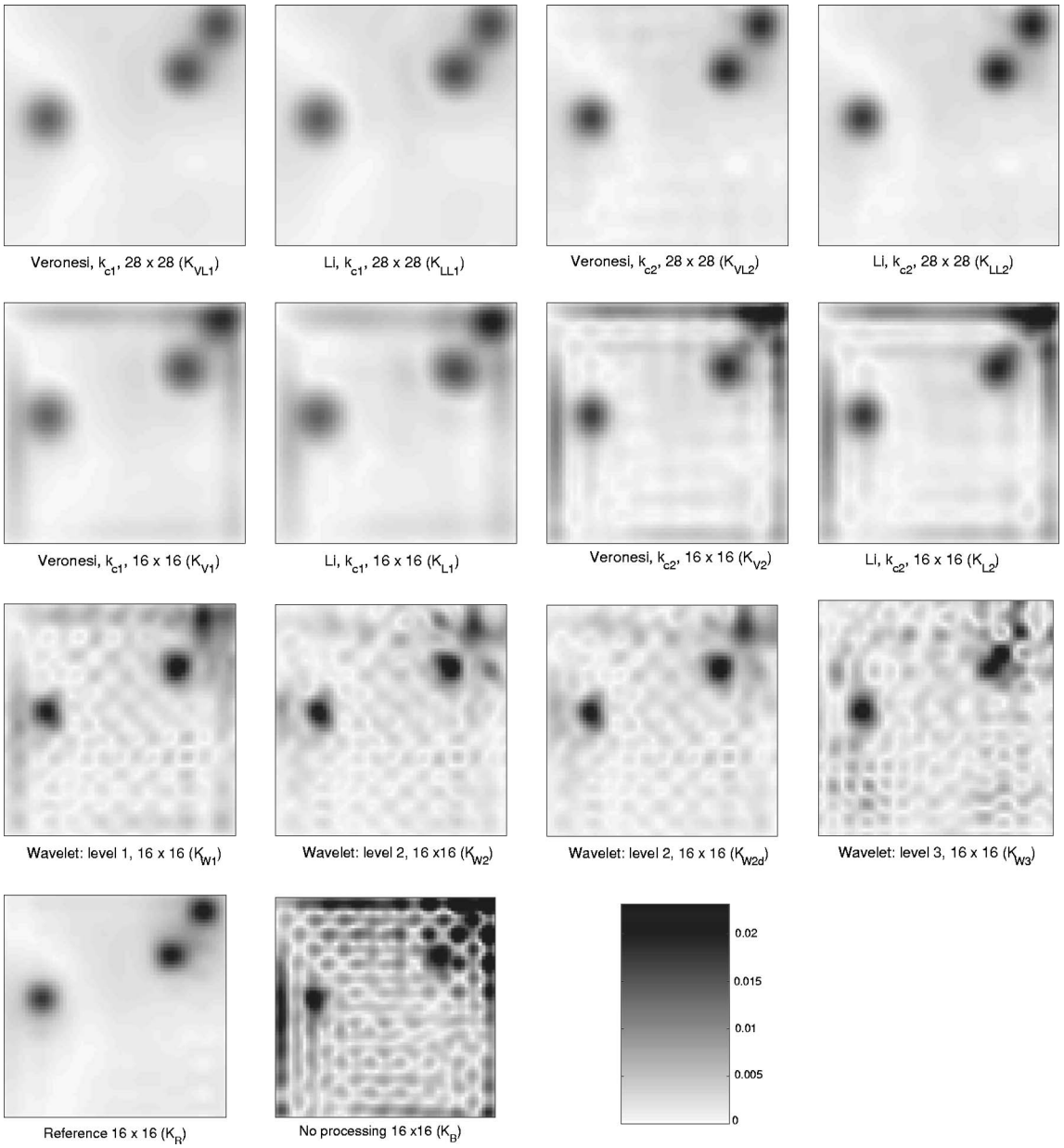


FIG. 8. Magnitudes of the reconstructed acoustic fields 1 cm from the source plane obtained from the acquired pressure field 10 cm from the loudspeaker plane. Each acoustic field obtained from standard NAH or wavelet processing can be compared to the reference one, bottom left, directly acquired 1 cm from the source plane. Each acoustic field is 1.0×1.0 m^2.

zero-padding corresponding to a 2.07×2.07 m^2 zone. However, the area of interest remains the 16×16 grid of dimension 1.0×1.0 m^2 shown in Fig. 8.

Finally, 14 reconstructed acoustic fields 1 cm from the source plane are shown:

(i) the reference acoustic field directly acquired 1 cm from the source plane,
(ii) the basic spatial field reconstructed from the acquired pressure field with no processing in k space,
(iii) eight spatial fields computed from the acquired pressure field 10 cm from the source plane and lessened by Veronesi or Li filtering with two different cutoff wave numbers $k_{c_1}=0.6k_{max}$ or $k_{c_2}=0.8k_{max}$, and
(iv) four spatial fields computed from the modified acquired pressure with a one-level, two-level, two-level with no spatial filtering of details, and three-level wavelet decomposition.

Every field has to be compared to the reference one at the bottom left of Fig. 8: the basic field obtained with no processing demonstrates the necessity to operate filtering in the wave number spectrum. The effects of Veronesi and Li filtering with the same cutoff wave number appear fairly similar: this is a known fact when the slope at the break point is steep. Acoustic fields computed from an extended area (28×28) logically do not exhibit distortion effects even if the sources are not as dark as those of the reference. By contrast, standard holography applied to a 16×16 grid leads to truncation effects at the edges, the sharpness of which

depends on the cutoff wave number. The cutoff wave number k_{c_1} is more satisfactory than k_{c_2}. Except for the third level decomposition, the fields resulting from the wavelet preprocessing give the best results: indeed, the three acoustic sources appear quite clearly and they are more confined and intense. The distortions due to the truncation of the antenna are smaller at level 2, compared to level 1. However, two dark spots can be seen on an equidistant line from the two up sources (see wavelet level two K_{W_2} in Fig. 8). These two points are due to overfiltering in the $[k_{max}/4, k_{max}/2]$ band. Indeed, the acoustic field back-propagated 1 cm from the sources, resulting from the method presented, when no detail filtering at level 2 is applied, does not reveal the spots (see wavelet level 2 $K_{W_{2d}}$ in Fig. 8).

V. EVALUATION OF THE VIABILITY OF THE METHOD

A. Introduction

Since the previous discussion was based on visual comparisons and since the acoustic fields resulting from the wavelet approach seem to show some background noise in Fig. 8, the conclusions may be arguable. In order to evaluate the viability of the wavelet method objectively, some indicators are necessary. We decided to compute the indicators on the wave number spectra (k spectra) taken before backward propagation which lead to the resulting spatial fields in Fig. 8. Indeed, the backpropagation algorithm is computed in the wave number domain for both standard holography and the wavelet approach (see Fig. 2). Furthermore, the k spectrum provides a powerful representation of the physics underlying the acoustic radiation of the sources of noise: the k spectrum separates the evanescent waves and the plane waves and highlights the directions of the plane waves radiated on the measurement plane.[20] The k spectrum is also used to compute acoustical velocity and, then, sound intensity.

That is why each k spectrum computed 10 cm from the source plane [yielding the reconstructed spatial acoustic fields 1 cm from the source plane (reported in Fig. 8)] is compared to the reference k spectrum K_R. This k spectrum is the two-dimensional Fourier transform of the pressure field directly acquired 1 cm from the source plane on a 16×16 grid, propagated by calculation at 10 cm. However, the visual comparison of the wave number spectra to the reference is not very clear, justifying why the spectra are not reported in the paper. Table I summarizes the notation and the characteristics of the wave number spectra K's. For the standard method, the first subscript denotes the type of the filter (V for Veronesi, L for Li) and the number denotes the cutoff wave number 1 for k_{c_1}, 2 for k_{c_2}. A capital L may appear when acquisition is made with a 28×28 "large" grid. For instance, K_{VL_1} denotes Veronesi filtering with the cutoff wave number k_{c_1} from a 28×28 grid extended to 32×32 by zero-padding. For the wavelet processing, the first subscript is W and the number indicates the level of the decomposition.

Two indicators are proposed for comparing these k spectra objectively: the first is based on similarity measurements between two wave number spectra, the second provides a relevant representation plane for the spectra where they are easy to compare.

B. Criteria for characterization

1. Similarity measurements

Thirteen wave number spectra K_P with P in $\{B, L_1, LL_1, L_2, LL_2, V_1, VL_1, V_2, VL_2, W_1, W_2, W_{2d}, W_3\}$ before backward propagation are compared to the reference K_R using several similarity measurements in the wave number domain $\mathcal{A}$, corresponding to the area covered by the array of measurement points. It is assumed that the minimum distance is obtained by the most relevant wave number spectrum. The similarity measurements are computed using the following distances, which Davy *et al.* used[21] for comparisons between time-frequency representations:

Manhattan distance

$$d_1(K_P,K_R)=\iint_{\mathcal{A}}|K_P(k_x,k_y)-K_R(k_x,k_y)|dk_x\,dk_y, \tag{5}$$

Euclidean distance

$$d_2(K_P,K_R)=\left[\iint_{\mathcal{A}}|K_P(k_x,k_y)-K_R(k_x,k_y)|^2\,dk_x\,dk_y\right]^{1/2}, \tag{6}$$

Correlation distance

$$d_{cor}(K_P,K_R)=\frac{[d_2(K_P,K_R)]^2}{\iint_{\mathcal{A}}|K_P(k_x,k_y)|^2\,dk_x\,dk_y+\iint_{\mathcal{A}}|K_P(k_x,k_y)|^2\,dk_x\,dk_y}, \tag{7}$$

Kolmogorov distance

$$d_{kol}(K_P^N,K_R^N)=\iint_{\mathcal{A}}|K_P^N(k_x,k_y)-K_R^N(k_x,k_y)|dk_x\,dk_y, \tag{8}$$

Küllback distance

$$d_{kul}(K_P^N,K_R^N)=\iint_{\mathcal{A}}[K_P^N(k_x,k_y)-K_R^N(k_x,k_y)]\times\log\frac{K_P^N(k_x,k_y)}{K_R^N(k_x,k_y)}\,dk_x\,dk_y, \tag{9}$$

Matusita distance

$$d_{mat}(K_P^N,K_R^N)=\sqrt{\iint_{\mathcal{A}}|\sqrt{K_P^N(k_x,k_y)}-\sqrt{K_R^N(k_x,k_y)}|^2\,dk_x\,dk_y}, \tag{10}$$

where the notation $K_P^N(k_x,k_y)$ emphasizes the normalization of the wave number spectrum:

$$K_P^N(k_x,k_y)=\frac{|K_P(k_x,k_y)|}{\iint_{\mathcal{A}}|K_P(k_x,k_y)|dk_x\,dk_y}. \tag{11}$$

Figure 9 shows the results obtained from the six different distance measurements defined above, between the reference and the studied wave number spectra. Full corresponding numerical results have already been reported.[22] Four distances (d_1, d_{kol}, d_{kul}, and d_{mat}) give almost the same positions and seem to validate the visual inspection from Fig. 8. Wave number spectra computed from a 28×28 grid are in general more accurate than those computed from a 16×16

TABLE I. Review of the wave number spectra to compare. MA_i means multiresolution analysis of level i. $K_{W_{2d}}$ is different from K_{W_2} by the fact that the modified pressure field at resolution 2 is computed with no filtering of the corners.

Wave number spectrum		Grid		Processing
Reference K_R		16×16		no
K_B		16×16		no
K_{L_1}	K_{LL_1}	16×16	28×28	Li (0.6 k_{max})
K_{L_2}	K_{LL_2}	16×16	28×28	Li (0.8 k_{max})
K_{V_1}	K_{VL_1}	16×16	28×28	Veronesi (0.6 k_{max})
K_{V_2}	K_{VL_2}	16×16	28×28	Veronesi (0.8 k_{max})
K_{W_1}	K_{W_3}	16×16		MA_1 MA_3
K_{W_2}	$K_{W_{2d}}$	16×16		MA_2

grid. The spectra obtained by the wavelet preprocessing at resolution 2 are the closest to the reference, whatever the distance used. However, these distance measurements have the disadvantage of comparing two images pixel by pixel. Therefore they are very sensitive to the position of the pressure field maxima.

2. A representative space for comparing k spectra

The aim of this approach is to associate a representative pattern with each of the 14 k spectra in Table I. In pattern recognition, a pattern is composed of d real numerical components and may then be represented by a point in the feature space $\mathbb{R}^d$. Two patterns are similar if the distance between them is short in the feature space. For the purposes of the study, the patterns are composed of the marginal wave number distribution, extracted from a k spectrum, which gives information about the relative positions of the wave numbers.

The marginal wave number distributions $E_x(k_x)$ and $E_y(k_y)$, respectively along k_x and k_y directions of the wavenumbers $K_P(k_x,k_y)$, are defined as

$$E_x(k_x) = \frac{\int_{\mathcal{A}} |K_P(k_x,k_y)|\,dk_y}{\iint_{\mathcal{A}} |K_P(k_x,k_y)|\,dk_x\,dk_y}, \tag{12}$$

$$E_y(k_y) = \frac{\int_{\mathcal{A}} |K_P(k_x,k_y)|\,dk_x}{\iint_{\mathcal{A}} |K_P(k_x,k_y)|\,dk_x\,dk_y}. \tag{13}$$

Figure 10 shows the marginal wave number distribution along k_x for several k spectra including the reference K_R. In comparison with K_R, K_{W_2} behaves well in [5, 20], as does K_{V_1} in [32, 44]. Wave numbers seem to be overfiltered in [5, 10] with wavelet processing at level 3, and underfiltered in [10, 30] at level 1. These effects are consistent with multiresolution analysis properties reported in Sec. III B. The impression of accuracy given by the reconstructed acoustic field 1 cm from the sources from a second level multiresolution analysis (Fig. 8) may be explained by the filtering operated in the [5, 20] band.

In the study, the patterns are composed of 64 components, given by both distributions E_x and E_y (32 for each one). Therefore the feature space is $\mathbb{R}^{64}$. It is obviously not appropriate to represent the 14 patterns in the feature space. Thus a principal component analysis (PCA) is performed on the pattern set to reduce the dimension of the representation space. It allows us to project the patterns on an orthogonal basis with two axes obtained by linear combination of the input features. This representation gives information about the similarities between the wave number spectra studied. The results are reported in Fig. 11. The projection made by the PCA led to a loss of information of 22%, which is sufficient to validate the representation.

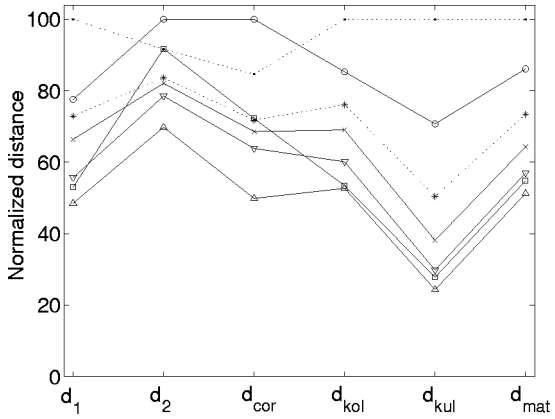


FIG. 9. Six distance measurements (Manhattan d_1, Euclidean d_2, Correlation d_{cor}, Kolmogorov d_{kol}, Küllback d_{kul}, Matusita d_{mat}) between the reference wave number spectrum K_R and the spectra K_B (dotted .), K_{W_3} (○), K_{V_2} (dotted ∗), K_{W_1} (×), K_{V_1} (▽), K_{VL_1} (□), and $K_{W_{2d}}$ (△). For each distance measurement, the longest distance is set to value 100. The distance measurements involving K_{L_1}, K_{L_2}, and K_{W_2} are not represented because they are respectively close to those involving K_{V_1}, K_{V_2}, and $K_{W_{2d}}$. It is the same for the distance measurements involving K_{LL_1}, K_{LL_2}, and K_{VL_2} that are similar to the measurement concerning K_{VL_1}.

Considering the PCA plane in Fig. 11, the marginal distribution seems to be a relevant feature. K_{W_2} and $K_{W_{2d}}$ are seen to behave very well: they are located near the reference. The similarities between Veronesi and Li filtering and the distance from K_B and the other points may also be noticed. The spectra computed from a large grid are the closest to the reference. In conclusion, an examination of the projection in the plane of the 14 patterns (see Fig. 11), provided by the wave number spectra studied, seems to give the same impression as the visual analysis of the reconstructed acoustic fields of Fig. 8. It confirms that the results obtained from the wavelet preprocessing for resolution 2 appeared more accurate than those given by standard holography with Veronesi or Li filtering. Furthermore, the marginal wave number distribution is established as a relevant feature for characterizing wave number spectra.

VI. CONCLUSIONS

We have demonstrated the relevance of wavelet preprocessing applied to the acquired pressure field in the nearfield of stationary sources before back-propagation by standard NAH. The method may also be used in a nonstationary context. In this context, the constraint is based particularly on the fact that the microphones of the array must operate simultaneously.

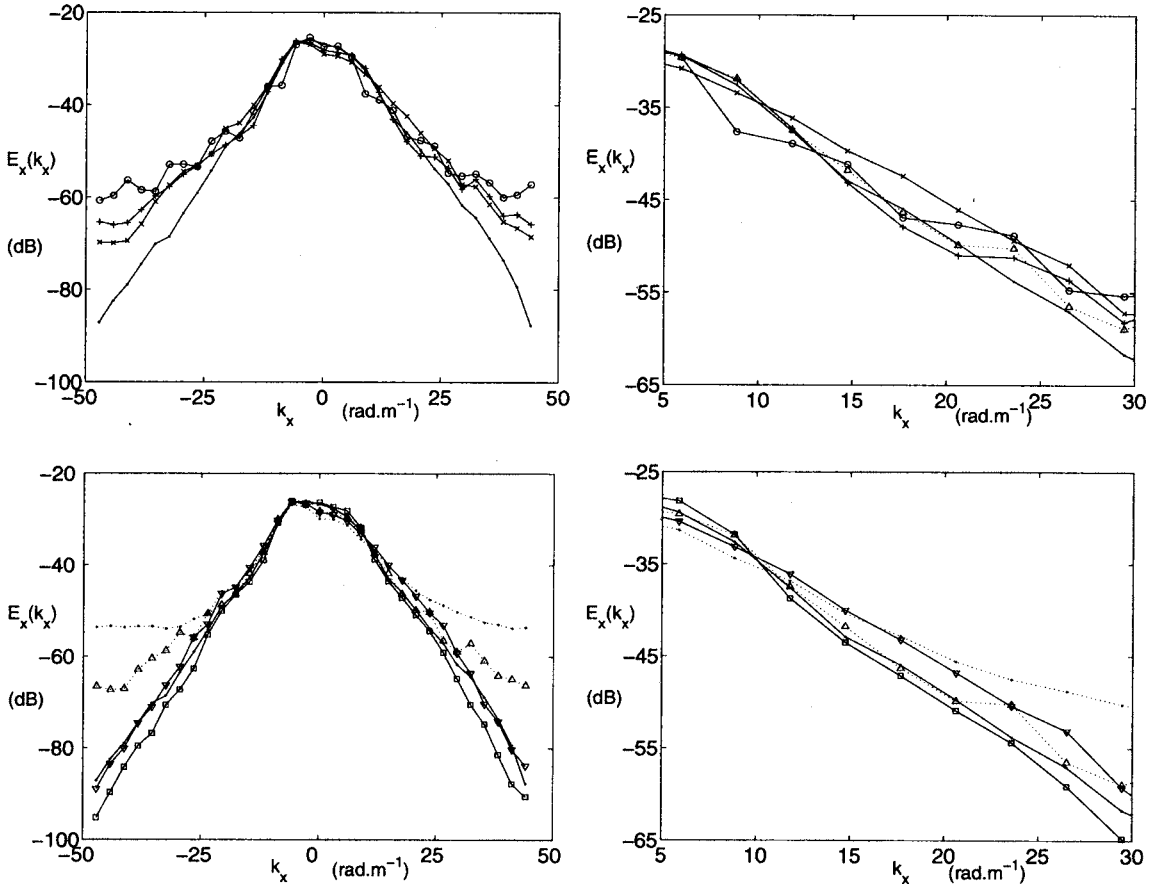


FIG. 10. Marginal wave number distribution $E_x(k_x)$ in decibels for k-spectra K_R (solid .), K_{W_1} ($\times$), K_{W_2} ($+$), K_{W_3} ($\bigcirc$), $K_{W_{2d}}$ (dotted $\triangle$), K_B (dotted .), K_{V_1} ($\triangledown$), and K_{VL_1} ($\square$) with zoom in on [5 30] on the right.

The wavelet method is to be considered as a preprocessing routine which is appropriate to lessen the effects due to the truncation of the hologram. It does not enter into competition with other methods working in the wave number domain, but rather in complement to these methods. From this point of view, wavelet preprocessing could be used in conjunction with regularization techniques; patch formulation, in particular, because the wavelet technique does not take into account the backward propagation distance and its influence on evanescent waves. However, wavelet processing is particularly effective for reducing distortion due to the truncation of acoustic fields because it works selectively in space and wave number domains. The methods of regularization are, on the other hand, built from a noisy k spectrum model, thus particularly effective in the case of measurement noise.

Furthermore, the pattern recognition approach we used to compare the wavelet method to standard NAH demonstrates the relevance of wave number marginal distribution which provides an objective criterion to assess the viability of wavelet preprocessing. Experimentation using this approach may well prove the relevance of any processing method involving acoustic field images.

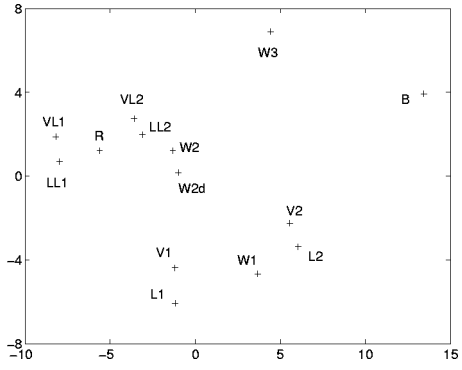


FIG. 11. Representative space for comparing patterns composed of wave number marginal distribution, resulting from principal component analysis (78% information kept).

[1]H. S. Kwon and Y. H. Kim, "Minimization of bias error due to windows in planar acoustic holography using a minimum error window," J. Acoust. Soc. Am. **98**, 2104–2111 (1995).
[2]P. A. Nelson and S. H. Yoon, "Estimation of acoustic source strength by inverse methods: Part i, conditioning of the inverse problem," J. Sound Vib. **233**, 643–668 (2000).
[3]S. H. Yoon and P. A. Nelson, "Estimation of acoustic source strength by inverse methods: Part ii, experimental investigation of methods for choosing regularization parameters," J. Sound Vib. **233**, 669–705 (2000).
[4]J. Hald, "Reduction of spatial windowing effects in acoustical holography," Proceedings of Inter-Noise 94, Yokohama, Japan, 29–31 August 1994, pp. 1887–1890.
[5]E. G. Williams, "Regularization methods for near-field acoustical holography," J. Acoust. Soc. Am. **110**, 1976–1988 (2001).
[6]W. A. Veronesi and J. D. Maynard, "Nearfield acoustic holography (NAH): II. Holographic reconstruction algorithms and computer implementation," J. Acoust. Soc. Am. **81**, 1307–1322 (1987).
[7]A. Schuhmacher, J. Hald, K. B. Rasmussen, and P. C. Hansen, "Sound source reconstruction using inverse boundary element calculations," J. Acoust. Soc. Am. **113**, 114–127 (2003).
[8]X. Zhao and S. F. Wu, "Reconstruction of vibro-acoustic fields using

hybrid nearfield acoustic holography," J. Sound Vib. **282**, 1183–1199 (2005).
[9]E. G. Williams, "Continuation of acoustic near-fields," J. Acoust. Soc. Am. **113**, 1273–1281 (2003).
[10]E. G. Williams, B. H. Houston, and P. C. Herdic, "Fast Fourier transform and singular value decomposition formulations for patch nearfield acoustical holography," J. Acoust. Soc. Am. **114**, 1322–1333 (2003).
[11]J. R. F. Arruda, "Surface smoothing and partial spatial derivatives computation using a regressive discrete Fourier series," Mech. Syst. Signal Process. **6**, 41–50 (1992).
[12]J.-C. Pascal, J.-F. Li, and X. Carniel, "Wavenumber processing techniques to determine structural intensity and its divergence from optical measurements without leakage effects," Shock Vib. **9**(1-2), 57–66 (2002).
[13]Z. El-Khoury and C. Nouals, "Utilisation de l'analyse multirésolution en holographie acoustique champ proche," (Acoustic holography using wavelet transform), Trait. Signal **11**, 257–270 (1994).
[14]J.-H. Thomas and J.-C. Pascal, "Using wavelets to reduce distortion problems in near field acoustic holography," Proceedings of Inter-Noise 01, The Hague, The Netherlands, 27–30 August 2001, pp. 2175–2178.
[15]J.-H. Thomas and J.-C. Pascal, "Acoustic holography experiments using a wavelet preprocessing method," Proceedings of ISMA 2002, Leuven, Belgium, 16–18 September 2002, pp. 1825–1833.
[16]S. Mallat, *A Wavelet Tour of Signal Processing* (Academic, New York, 1998).
[17]M.-H. Masson, B. Dubuisson, and C. Frélicot, "Conception d'un module de reconnaissance des formes floue pour le diagnostic," (Design of a fuzzy pattern recognition system for the diagnosis), J. Eur. Syst. Automat. (RAIRO-APII-JESA), 319–341 (1996).
[18]I. Daubechies, "Ten Lectures on Wavelets," Regional Conference Series in Applied Mathematics, SIAM (1992).
[19]J. F. Li, J.-C. Pascal, and C. Carles, "A new K-space optimal filter for acoustic holography," Proceedings of the 3rd International Congress on Air and Structure Borne Sound and Vibration, Montreal, Canada, 13–15 June 1994, pp. 1059–1066.
[20]E. G. Williams, *Fourier Acoustics* (Academic, New York, 1999).
[21]M. Davy, C. Doncarli, and G. F. Boudreaux-Bartels, "Improved optimization of Time-frequency based signal classifiers," IEEE Signal Process. Lett. **8**(2), 52–57 (2001).
[22]J.-H. Thomas and J.-C. Pascal, "A study of the relevance of an acoustic holography method using wavelets based on features extracted from wavenumber spectra," Proceedings of Inter-Noise 03, Jeju, Republic of Korea, 25–28 August 2003, pp. 2667–2674.

Acoustics of the human middle-ear air space

Cara E. Stepp
Picker Engineering Program, Smith College, 51 College Lane, Northampton, Massachusetts 01063 and Speech and Hearing Biosciences and Technology Program, Harvard-M.I.T. Division of Health Sciences and Technology, Cambridge, Massachusetts 02139

Susan E. Voss[a)]
Picker Engineering Program, Smith College, 51 College Lane, Northampton, Massachusetts 01063

(Received 18 January 2005; revised 19 May 2005; accepted 26 May 2005)

The impedance of the middle-ear air space was measured on three human cadaver ears with complete mastoid air-cell systems. Below 500 Hz, the impedance is approximately compliance-like, and at higher frequencies (500–6000 Hz) the impedance magnitude has several (five to nine) extrema. Mechanisms for these extrema are identified and described through circuit models of the middle-ear air space. The measurements demonstrate that the middle-ear air space impedance can affect the middle-ear impedance at the tympanic membrane by as much as 10 dB at frequencies greater than 1000 Hz. Thus, variations in the middle-ear air space impedance that result from variations in anatomy of the middle-ear air space can contribute to inter-ear variations in both impedance measurements and otoacoustic emissions, when measured at the tympanic membrane. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1974730]

PACS number(s): 43.64.Bt, 43.64.Ha [WPS] Pages: 861–871

I. INTRODUCTION

The human middle-ear air space, schematized in Fig. 1 (left), consists of the tympanic cavity, aditus ad antrum, the antrum of the mastoid, and the mastoid air cells (e.g., Donaldson *et al.*, 1992, p. 151). The tympanic cavity houses the ossicular system and lies between the tympanic membrane and the inner ear. The posterior-superior portion of the tympanic cavity narrows into the passage called the aditus ad antrum which extends to the antrum. Attached to the antrum is a system of mastoid air cells that communicate with one another and vary in size (Donaldson *et al.*, 1992). The total volume of the middle-ear air space is highly variable among normal-hearing ears. The tympanic cavity has a volume ranging from 0.5 to 1 cm^3 (i.e., Gyo *et al.*, 1986; Whittemore *et al.*, 1998); the mastoid air cell system has a much wider volume range, reported[1] by Molvaer *et al.* (1978) to be about 1 to 21 cm^3 and by Koç *et al.* (2003) to be 4 to 14 cm^3.

Even though the large range in middle-ear air space volume is well documented, the acoustical effects of this air space on sound transmission through the human middle ear have not been characterized. Mathematical models of the human middle ear (e.g., Zwislocki, 1962; Kringlebotn, 1988) include the effects of the middle-ear air space (i.e., middle-ear cavity), but because the influence of the air space within these models has a minimal effect on the impedance at the tympanic membrane, the acoustical effects of the middle-ear air space have been largely ignored. For example, as stated by Zwislocki (1962):

> "Since the impedance of the middle-ear cavities is low by comparison to other parts of the middle ear, its effect on the impedance at the eardrum and on the sound transmission to the inner ear is not critical. As a consequence, the analog [model] may be regarded as a sufficient approximation."

The analog circuit model descriptions of the middle-ear cavity of both Zwislocki (1962) and Kringlebotn (1988) are based on measurements of the middle-ear air space impedance that were made on one cadaver ear by Onchi (1961). The only subsequent measurements of middle-ear air space impedance were made on cadaver ears that lacked a portion of their mastoid air-cell networks (Voss *et al.*, 2000b). The goal of this work is to describe the acoustics of the normal (unaltered) middle-ear air space.

Although the role of the human middle-ear air space has been largely ignored in normal ears, its acoustical effects are known to be important in determining middle-ear sound transmission in several pathological middle-ear conditions. For example, Merchant and colleagues have demonstrated that in type IV and type V tympanoplasties, the air space between the graft shield and the round window (termed the cavum minor) should be maximized in order to maximize hearing (Rosowski and Merchant, 1995; Rosowski *et al.*, 1995; Merchant *et al.*, 1995, 1997, 1998). Voss and colleagues have shown that (1) the volume of the middle-ear air space plays a major role in middle-ear function in ears with tympanic membrane perforations (or tubes) at frequencies below about 1000 Hz (Voss, 1998; Voss *et al.*, 2001a, b, c) and (2) in some pathological ears, the air volume and anatomy of the middle-ear air space can influence the sound produced by audiological earphones (Voss *et al.*, 2000a, c).

The work reported here is a step toward a description of the acoustics of the human middle-ear air space. Specifically, we report measurements, made on human cadaver ears, of the impedance of the human middle-ear air space, and we use these measurements in combination with previous measurements and models to predict how the middle-ear air

[a)]Electronic mail: svoss@email.smith.edu

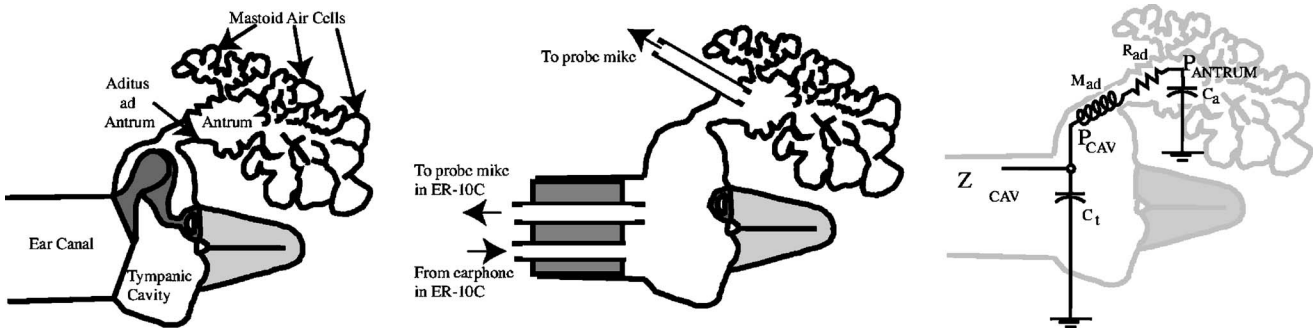


FIG. 1. **Left Panel:** Schematic of the air spaces of the human middle-ear. The tympanic cavity houses the ossicles and is connected to the mastoid air cells and antrum by the tube-like aditus ad antrum. Adapted from Onchi (1961). **Middle Panel:** Schematic of the measurement setup. The tympanic membrane, malleus, and incus are removed. The sound source and microphone assembly sit in the ear canal with the probe tube of the microphone flush with the entrance to the tympanic cavity and extended about 3 mm from the sound tube. To measure the antrum pressure, a probe tube connected to a microphone sits in the antrum. **Right Panel:** An analog circuit model schematic of the middle-ear air space (e.g., Zwislocki, 1962; Kringlebotn, 1988; Voss *et al.*, 2000b). The circuit model relates to the ear structure through the location of circuit elements on the structural outline. The voltages are analogous to sound pressures and currents are analogous to volume velocities. The capacitor in the tympanic cavity represents the volume of air in the tympanic cavity, the resistor and inductor within the additus ad antrum represent the tube-like passage, and the capacitor in the antrum represents the volume of air in the antrum and mastoid air cells.

space does affect the impedance at the tympanic membrane. Our results suggest that the middle-ear air space in normal ears may play a role in some of the variability observed in ear-canal based acoustical measurements on normal populations; such variations among normal ears are essential to understand as otoacoustic emissions and middle-ear power flow measurements are developed as clinical tools (e.g., Stinson, 1990; Keefe *et al.*, 1993; Voss and Allen, 1994; Keefe and Levi, 1996; Hunter and Margolis, 1997; Margolis *et al.*, 1999; Piskorski *et al.*, 1999; Feeney and Keefe, 1999, 2001; Farmer-Fedor and Rabbitt, 2002; Feeney *et al.*, 2003; Magliulo *et al.*, 2004). A preliminary account of this work has been presented elsewhere (Stepp and Voss, 2004).

II. METHODS

A. Cadaver-ear preparation

Comparisons of measurements of acoustic impedance (Rosowski *et al.*, 1990) and of umbo velocity (Goode *et al.*, 1993) between populations of cadaver ears and living ears show no significant differences between the two populations. This apparent normality of acousto-mechanical properties for cadaver ears supports our assumption that the middle-ear air space impedance in the cadaver preparation is equivalent to that in live human ears.

Four adult human hemi-heads that included the entire temporal bone, mastoid air space, and external ear were obtained through the nonprofit organization Life Legacy. Specimens were frozen after removal from their donor, shipped on dry ice, and put in the freezer upon arrival. The first ear was used to develop the methods, and measurements are reported on the final three ears, labeled ears 1, 2, and 3 here. We note that ears 2 and 3 are the right and left ears, respectively, from the same subject. To prepare the ears for measurement, each specimen was thawed and reduced to a manageable size by sawing away bony areas not relevant for the measurements. The entire mastoid space, including all mastoid air cells, was left intact. Next, the bony ear canal was drilled away so that the tympanic ring was fully exposed. Using an otologic-operating microscope, the tympanic membrane, malleus, and incus were removed with forceps, and the stapes was left in the oval window of the cochlea to prevent leakage of cochlear fluid (Fig. 1 middle). In all cases, the ears appeared normal. At this point in the preparation, the ears were refrozen. Several weeks later they were thawed by placing them in a saline-filled container for several hours and measurements were made after they thawed and the saline was suctioned away.

A cylindrical coupling ring designed to couple the sound source to the tympanic ring was attached to the tympanic ring with carboxylate dental cement (ESPE, Durelon, Norristown, PA). This coupling ring had an inner diameter of 11 mm, which approximated the size of the tympanic rings in our specimens. Any irregular spaces between the ring and the bone were filled with dental cement. When the sound source was coupled to the tympanic ring via the coupling ring, the tip of the probe-tube microphone extension (Sec. II C) was approximately at the same depth as the tympanic ring (Fig. 1 middle). To ensure that the middle-ear was sealed acoustically, much of the specimen was coated with additional dental cement, and the specimen was also placed in the finger of a latex glove that provided a tight fit around it. Prior to making acoustic measurements, suction was applied to both the tympanic cavity and the antrum (via the aditus ad antrum) to remove any fluid (i.e., saline) that had collected in the middle-ear air space.

In one ear (ear 3), a probe-tube for a microphone (Etymōtic Research ER-7C) was secured in the antrum in order to measure the sound pressure at this location. In this case, a narrow (less than 1 mm in diameter) tube-like hole was drilled in the roof of the antrum. A flexible probe tube (Etymōtic Research) was secured with dental cement through this hole in the antrum. The location was confirmed visually by looking through the tympanic ring and aditus ad antrum and observing that the tip of the probe-tube sat within the antrum.

B. Stimulus generation and response

The software package SYSid was used to obtain complex frequency response measurements. The stimulus was a broad-band chirp (25 Hz to 25 kHz), and the response mag-

nitudes and angles were calculated as the 2048-point FFT of the time-domain average of N responses. Measurements were sampled at 50 kHz, with $N=1000$ and an input voltage of 0.4 V, which was found to be within the linear operation range for all measurements made.

The acoustic assembly coupled to the tympanic ring was an ER-10C (Etymōtic Research), which contains both a sound source to transduce the computer-generated voltage signal to an acoustic signal and a microphone to transduce the sensed sound to a voltage. The ER-10C was driven by SYSid, and the microphone's response was recorded via this same system. The ER-10C was coupled to either the ear or the calibration cavities (Sec. II C).

C. Impedance measurements

The acoustic assembly was characterized by its Thévenin equivalents, P_{TH} and Z_{TH}, using a method similar to that of Allen (1986), Voss and Allen (1994), and Neely and Gorga (1998). The acoustic assembly was coupled to loads with known acoustic impedance, and microphone responses to the chirp stimulus were measured for each load. The loads were cylindrical closed-ended tubes of different lengths, each with a known theoretical impedance determined by the equations developed by Keefe (1984). The loads all had an inner diameter of 11 mm and lengths ranged from 1.14 to 4.41 cm. The acoustic assembly was coupled to the loads via a cylindrical plastic adapter that housed both the sound-delivery tube and the microphone probe tube of the ER-10C and was machined to fit snugly into each load while maintaining a constant inner diameter of 11 mm. The same plastic adapter fit snugly to the acoustic coupler that was cemented to the tympanic ring. This coupling arrangement allowed for measurements on ears and cylindrical loads without changing the relative position of the microphone with respect to the sound-delivery tube. To minimize the contribution of nonuniform waves generated at the earphone tube tip (e.g., Huang *et al.*, 2000), a small piece of tygon tubing was used to extend the microphone probe tube beyond the sound-delivery tube by 3 mm for all measurements (in loads and in ears).

Measurements were made in nine loads (cylindrical tubes). Four of these responses were used to determine the Thévenin equivalents, P_{TH} and Z_{TH}. These four responses, in combination with the theoretical load impedances, provided an over-specified set of four equations that were solved in MATLAB by optimizing the lengths of the cavities in order to minimize the least-squares error between the four equations and to solve for P_{TH} and Z_{TH} (e.g., Allen, 1986; Voss and Allen, 1994; Neely and Gorga, 1998). Measurements made on the additional five loads were used as controls; measurements in these five loads were used in combination with the determined P_{TH} and Z_{TH} to calculate a measured impedance for each load, which was compared with the theoretical impedance. The ratio between the measurements and the theory generally had a magnitude of less than 1 dB and an angle difference of less than 0.025 cycles; larger differences occurred only at the frequencies where the impedance magnitude was a maximum or minimum. These comparisons suggest that the impedance measurements are generally accurate to within about 10% in magnitude (about 1 dB) and 10° in angle (0.025 cycles).

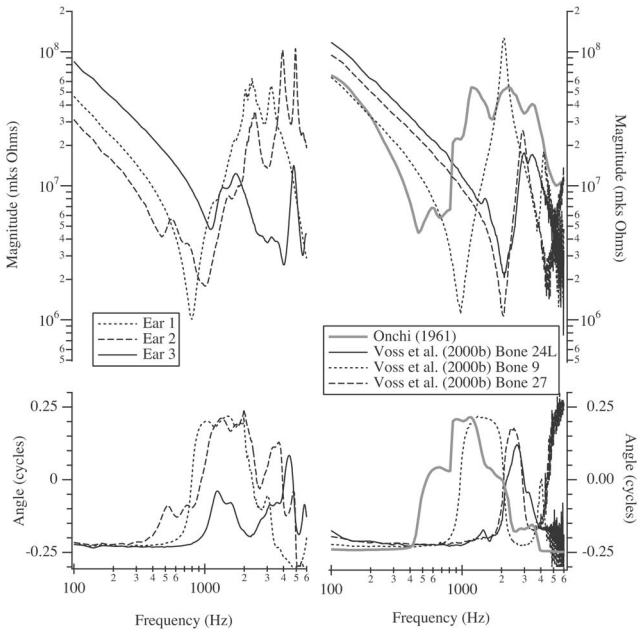


FIG. 2. Acoustic impedances of the middle-ear air space Z_{CAV}. **Upper Panel:** Impedance magnitudes. **Lower Panel:** Impedance angles. **Left Panel:** Impedances measured on the three ears reported here. **Right Panel:** Measurements of middle-ear air space impedance from Onchi (1961) ($N=1$) and Voss *et al.* (2000b). Three representative measurements of 11 total measurements from Voss *et al.* (2000b) are shown; all other measurements exhibit similar features to these three.

D. Measurement of antrum pressure

Responses to a chirp stimulus delivered to ear 3 were recorded simultaneously from a probe-tube microphone (ER-7C) inserted into the antrum and the ER-10C microphone at the tympanic ring. The antrum microphone (ER-7C) was calibrated with respect to the ER-10C microphone (microphone at the tympanic ring) via a measurement made with the two microphones adjacent to each other in a single airtight cavity. The voltages from both microphones were measured simultaneously in response to a chirp stimulus, and the ratio between the two microphone responses was used to calibrate the antrum microphone (ER-7C) relative to the tympanic-ring microphone (ER-10C).

III. RESULTS

A. Impedance of the middle-ear air space: Z_{CAV}

The acoustic impedances (Z_{CAV}) of the middle-ear air spaces have similar features for all three ears (Fig. 2 left). Below 500 Hz, all ears are compliance dominated with a decrease in impedance magnitude of 20 dB per decade and a near constant angle of about −0.25 cycles. At frequencies between 500 and 6000 Hz, the magnitudes have five to nine local extrema associated with steep changes in phase angle. Two of the three impedances have a mass-dominated region centered at about 1000 Hz, with a magnitude that increases at 20 dB per decade and an angle that approaches 0.25 cycles.

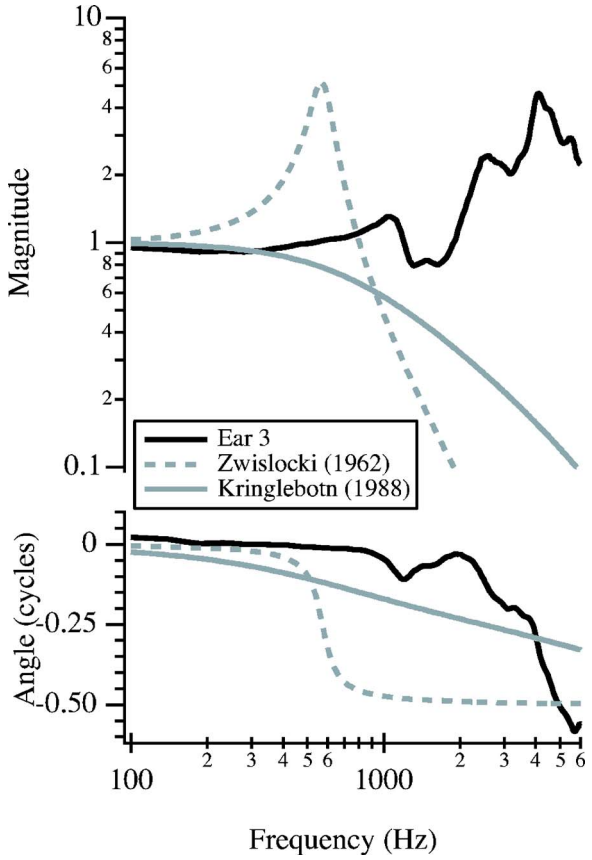


FIG. 3. A measurement of H_P, the ratio between the pressure in the antrum, P_{ANTRUM}, and the pressure at the tympanic ring, P_{CAV}, where $H_P = P_{\mathrm{ANTRUM}}/P_{\mathrm{CAV}}$, for ear 3. Also plotted are the model results of H_P from both Zwislocki (1962) and Kringlebotn (1988).

As frequency increases above 1000 Hz, none of the impedances can be described in terms of simple behavior (i.e., compliance, resistance, or mass dominated).

B. The antrum transfer function

In one ear (ear 3), the pressure in the antrum of the mastoid, P_{ANTRUM}, and the pressure at the tympanic ring, P_{CAV}, were measured simultaneously. The ratio between these pressures,

$$H_P = \frac{P_{\mathrm{ANTRUM}}}{P_{\mathrm{CAV}}}, \tag{1}$$

is plotted in Fig. 3. At low frequencies, the magnitude of H_P is nearly one and the angle is nearly zero. Thus, for these low frequencies, $P_{\mathrm{ANTRUM}} \approx P_{\mathrm{CAV}}$. As frequency increases, P_{ANTRUM} and P_{CAV} differ, and $|H_P|$ exhibits several local extrema.

IV. DISCUSSION

A. Measurements of Z_{CAV}

Previously published measurements of the middle-ear air space impedance (Z_{CAV}) include one measurement from Onchi (1961) and eleven measurements from Voss *et al.* (2000b). The measurement from Onchi (1961) appears to have included the entire mastoid space.[2] The eleven ears used by Voss *et al.* (2000b) were removed with a circular saw (Schuknecht, 1968) such that much of the mastoid air-cell network was left with the cadaver and was not part of the ear specimen used for measurement; thus, these eleven ears had smaller-than-normal mastoid volumes and air-cell networks—we will refer to this condition as "altered mastoid."

The middle-ear air space impedances measured here (Fig. 2 left) show both similarities and differences to the other measurements (Fig. 2 right). Many features are consistent among all measurements: (1) low-frequency behavior is compliance dominated; (2) low-frequency magnitudes are similar; and (3) above 500 Hz there are extrema in magnitude that are associated with steep changes in phase angle. There are also some differences between the measurements made with the unaltered mastoid system [i.e., measurements made here and by Onchi (1961)] as compared to the measurements from Voss *et al.* (2000b) with altered mastoids. Above 500 Hz, the measurements with unaltered mastoids have a mixed impedance, characterized by several (i.e., five to nine) extrema in magnitude with associated steep changes in phase angle. In contrast, the measurements with the altered mastoids (Voss *et al.*, 2000b) show fewer extrema in magnitude; specifically, there are two extrema between 100 and 4000 Hz—one sharp magnitude minimum and one sharp magnitude maximum, both with associated steep changes in phase angle.[3] Thus, the unaltered and altered mastoid cavity configurations are not equivalent to one another; in the following, we explore how their differences affect both middle-ear models and measurements of impedance.

B. Modeling Z_{CAV}

Analog circuit models of the middle-ear air space were reviewed by Voss *et al.* (2000b) and are briefly summarized here. The models of Zwislocki (1962), Kringlebotn (1988), and Voss *et al.* (2000b) are designed to represent the structure/function relationship of the middle-ear air space, and a summary of these models is shown in the right panel of Fig. 1.[4] The models employ a capacitor to represent the compliance of air in the tympanic cavity. This capacitor is in parallel with a series resistor-inductor-capacitor connection. The resistor-inductor path represents volume velocity flow through the tube-like additus ad antrum and the capacitor represents the compliance of the air in the antrum and mastoid air cavities. The models do not include components to represent the acoustics of the air-cell tracts throughout the mastoid. With three energy storage elements, this four-element topology includes two distinct resonances in the impedance of the middle-ear air space: (1) a minimum in impedance magnitude is determined by a series resonance between the inductor and the capacitor that represents the antrum air volume, and (2) a maximum in magnitude impedance is determined by a parallel resonance between the inductor and the combination of the two capacitors.[5] This general four-element model topology has been successfully applied to represent impedance measurements on both human cadaver ears with the mastoid air-cell tract mostly re-

TABLE I. Middle-ear cavity model parameters. Volumes are estimated from the model-compliance values using the relationship $V=\rho c^2 C$ where C is the compliance of air within the volume V, ρ is the density of air, c is the speed of sound in air, and f is frequency in Hz. Bone 24L is used as a representative bone (of 11 total) because it was the one analyzed in detail in Voss *et al.* (2000b).

Ear(mks units)	R_{ad}	M_{ad}	C_a	C_t
Ear 1 Stepp and Voss	$0.03\times10^6\sqrt{f}$	1276	2.9×10^{-11} $V=4.0$ cm^3	5.2×10^{-12} $V=0.73$ cm^3
Ear 2 Stepp and Voss	$0.06\times10^6\sqrt{f}$	550	4.2×10^{-11} $V=5.9$ cm^3	8.8×10^{-12} $V=1.24$ cm^3
Ear 3 Stepp and Voss	$0.15\times10^6\sqrt{f}$	1857	1.0×10^{-11} $V=1.5$ cm^3	6.7×10^{-12} $V=0.94$ cm^3
Bone 24L Voss *et al.* (2000b)	$0.05\times10^6\sqrt{f}$	890	6.3×10^{-12} $V=0.9$ cm^3	6.3×10^{-12} $V=0.9$ cm^3
Range ($N=11$) Voss *et al.* (2000b)	$0.02\times10^6\sqrt{f}-0.08\times10^6\sqrt{f}$	520–1420	$4.2\times10^{-12}-18.8\times10^{-12}$ $V=0.6$ to 2.6 cm^3	$2.5\times10^{-12}-8.0\times10^{-12}$ $V=0.4$ to 1.1 cm^3

moved (Voss *et al.*, 2000b) and on cat ears (Huang *et al.*, 1997), in which two cavities are connected by an additus-like tube and no mastoid air cells are present.

We fit the four-element model (Fig. 1 right) to the measurements made here using the same procedure outlined by Voss *et al.* (2000b), which is summarized in the Appendix. Voss *et al.* (2000b) applied this procedure to each of the 11 bones for which Z_{CAV} with an altered mastoid space was measured. Here, we apply the same procedure to the 3 ears for which Z_{CAV} was measured with an unaltered mastoid space; the specific model elements are listed in Table I along with the predicted volumes for the tympanic and mastoid air spaces.[6] Figure 4 shows comparisons between the four-element model and the measured Z_{CAV} from the three ears here with unaltered mastoids and two ears (B24L and B13) with altered mastoids from Voss *et al.* (2000b). Voss *et al.* (2000b) describe the model fit to the data for ears with altered mastoids:

> In general, below 3000 Hz, the model captures most salient features of the data. At the lowest frequencies, the model angle is nearly −0.25 while the data angle is about −0.20 cycles. Across our population of 11 ears, some of the low-frequency Z_{CAV} angles are at −0.25 cycles while others approach −0.20 cycles; reasons for the difference are not clear. Above 3000 Hz, our model and the data diverge. The local maximum in the data, near 3000 Hz, is overestimated by the model. Thus, up to 3000 Hz, the model is a reasonable description of the measured Z_{CAV}.

In contrast to the measurements with the altered mastoid, the four-element model is unable to fit the Z_{CAV} measured on ears with unaltered mastoids as well as it fits ears with altered mastoids. The three measurements with Z_{CAV} from the unaltered mastoid state (Fig. 4, ears 1, 2, and 3) show that the four-element model is able to represent the Z_{CAV} behavior at the lower frequencies (e.g., less than 500 Hz) and is able to capture the general behavior of the first minimum in magnitude. However, at frequencies near and above the first minimum in magnitude, the four-element model is unable to capture any of the fine structure of the multiple extrema in the measured Z_{CAV}.

In summary, the model topology of Fig. 1 (right) describes the general features of Z_{CAV} for ears with both intact and altered mastoid cavities, but it does not describe all features of the measurements with unaltered mastoids [made both here and by Onchi (1961)[7]], with the mastoid air-cell tract intact. The three energy-storage elements in the model limit the model to two distinct resonances. However, all measurements made with an intact mastoid air-cell system suggest multiple resonances above about 1000 Hz. It appears

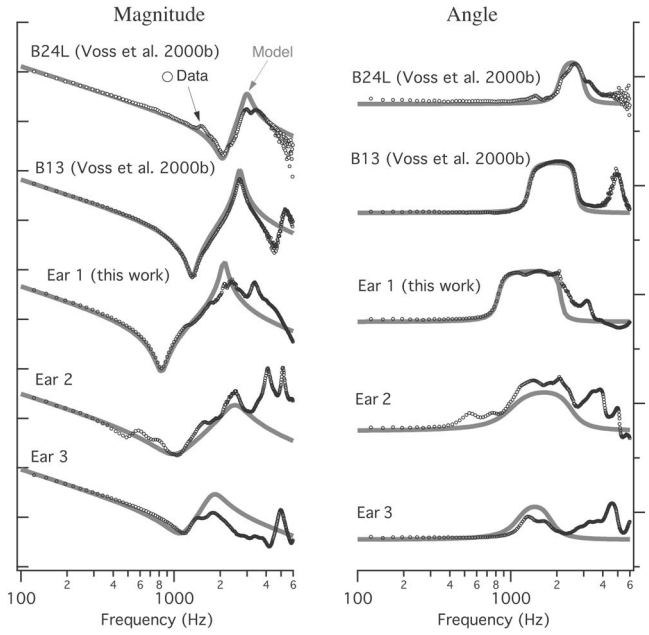


FIG. 4. Comparisons of measurements of Z_{CAV} to the four-element model (Fig. 1 right). Model element values were calculated for each ear via the process outlined in the Appendix. The measured Z_{CAV} (data) are plotted with open black circles and the corresponding model fits are plotted with a gray solid line. The upper two examples are from ears with altered mastoid cavities (Voss *et al.*, 2000b), and the lower three examples are from the ears presented here with normal mastoid cavities. To improve visibility, the plots were shifted on the same graphs and thus the exact values are not labeled. **Left Panel:** Magnitudes. **Right Panel:** Angles.

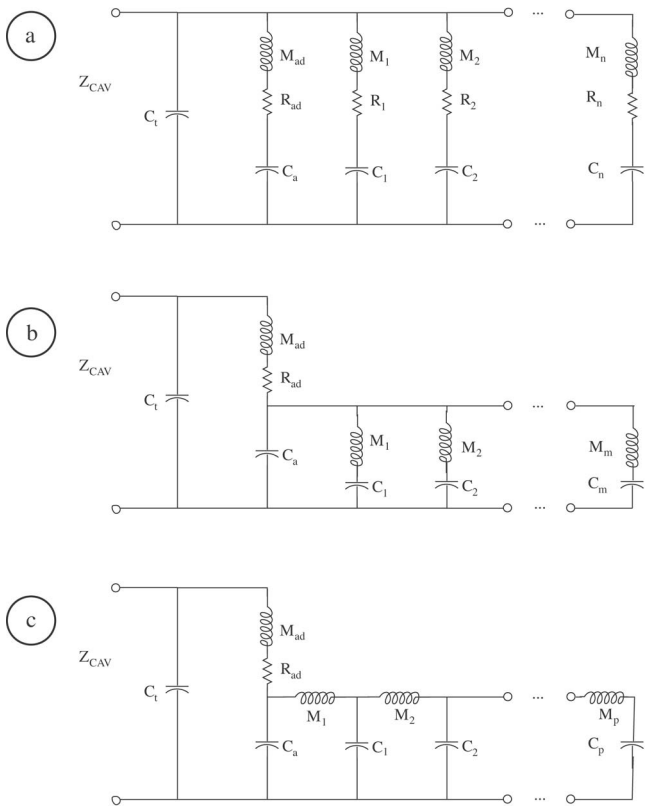


FIG. 5. Analog circuit models that represent features of the middle-ear air space. All three models (a, b, and c) include the four elements (C_t, M_{ad}, R_{ad}, and C_a) of the model described in Fig. 1 (right). A model for the middle-ear air space might include all three mechanisms represented here through three individual topologies. **(a)** Connections between the tympanic cavity and the mastoid air-cell system in addition to the aditus ad antrum are represented by series connections between a mass, a resistor, and a compliance. The model represents a total of n such connections. **(b)** The model represents m tracts within the mastoid air-cell system that originate at the antrum and terminate with a volume of air. **(c)** The model represents a single air-cell tract that originates at the antrum and progressively narrows. This model represents the narrowing of a single air-cell tract through connections of p masses and compliances.

that a more complicated model—possibly such as the "transmission line model" suggested by Onchi (1961)—is needed if the goal is to represent all extrema in the impedance magnitude of the intact middle-ear air space of the human ear.

A model of the middle-ear air space that accounts for more features than the simple four-element model accounts for would be highly individual, as the structure of each individual's mastoid air-cell system is unique (Molvaer *et al.*, 1978). The measurements here show that such a model would have several extrema in the impedance magnitude, and the specific number would depend on the individual ear (e.g., Fig. 2). Figure 5 schematizes three possible model topologies that relate structure to function of the middle-ear air space; these models are described to represent possible structure/function relationships within the middle-ear air space, but they are not detailed precise models of any particular middle-ear air space. In fact, a model of a specific middle-ear air space would most likely be a combination of all three of these model topologies. All three models of Fig. 5 include the four elements (C_t, M_{ad}, R_{ad}, and C_a) of the model described in Fig. 1 (right), which has been successfully used to represent the impedance of ears with altered mastoids. Figure 5(a) represents connections between the tympanic cavity and the mastoid air-cell system that are in addition to the aditus ad antrum. Specifically, each series connection between a mass, resistor, and compliance represents a "tube-like" connection from the tympanic cavity to a volume of air within the mastoid air-cell system. The model represents a total of n such connections.[8] Figure 5(b) is the "transmission line model" suggested by Onchi (1961). The model represents m tracts within the mastoid air-cell system that originate at the antrum and terminate with a volume of air; the model does not represent that each air-cell tract generally gets smaller as it moves away from the antrum. Figure 5(c) represents a single air-cell tract that originates at the antrum; each mass attached to a compliance represents a short tube-like structure that terminates with a volume in which air can compress and expand. Attached to that volume is an additional short tube-like structure that in turn is terminated with a volume. This model can represent the narrowing of a single air-cell tract through connections of p masses and compliances, and the specific values for the masses and compliances depend upon the dimensions of the air-cell tract being modeled. In summary, a model of the mastoid air-cell system that represents all extrema in the measured impedance would most likely include a combination of the mechanisms represented by the three models of Fig. 5. We do not propose that such a model is particularly useful, as we expect the model would be substantially different for each individual ear. However, the topologies in Fig. 5 provide an outline for how one might think about the acoustics of the middle-ear air space and the multiple and variable number of extrema in the measured impedances.

C. Effect of the middle-ear air space on the impedance at the tympanic membrane

To date, our knowledge of how the middle-ear air space impedance affects the impedance of the middle ear at the tympanic membrane has been based on middle-ear models. As discussed earlier (Sec. IV B), these models do not represent all features of the impedance of the middle-ear air space. Here, we use the measurements of the middle-ear air space impedance Z_{CAV} to estimate how variations in middle-ear air space impedance affect the impedance of the ear at the tympanic membrane. We assume that

$$Z_{TM} = Z_{TOC} + Z_{CAV}, \tag{2}$$

where Z_{TM} is the impedance at the tympanic membrane, Z_{TOC} is the impedance of the tympanic membrane, ossicles, and cochlea, and Z_{CAV} is the impedance of the middle-ear air space. Equation (2) is equivalent to middle-ear models in which the middle-ear air space (e.g., cavity) is represented by Z_{CAV} and is in series with the rest of the ear, represented as Z_{TOC} (e.g., Zwislocki, 1962; Kringlebotn, 1988; Voss *et al.*, 2001c). The use of Eq. (2) allows us to determine how Z_{TM} is affected by Z_{CAV}; in particular, estimates of Z_{TM} from the same ear or model can be made using a measured or modeled Z_{TOC} with a variety of representations of Z_{CAV} so that the effect of Z_{CAV} can be understood.

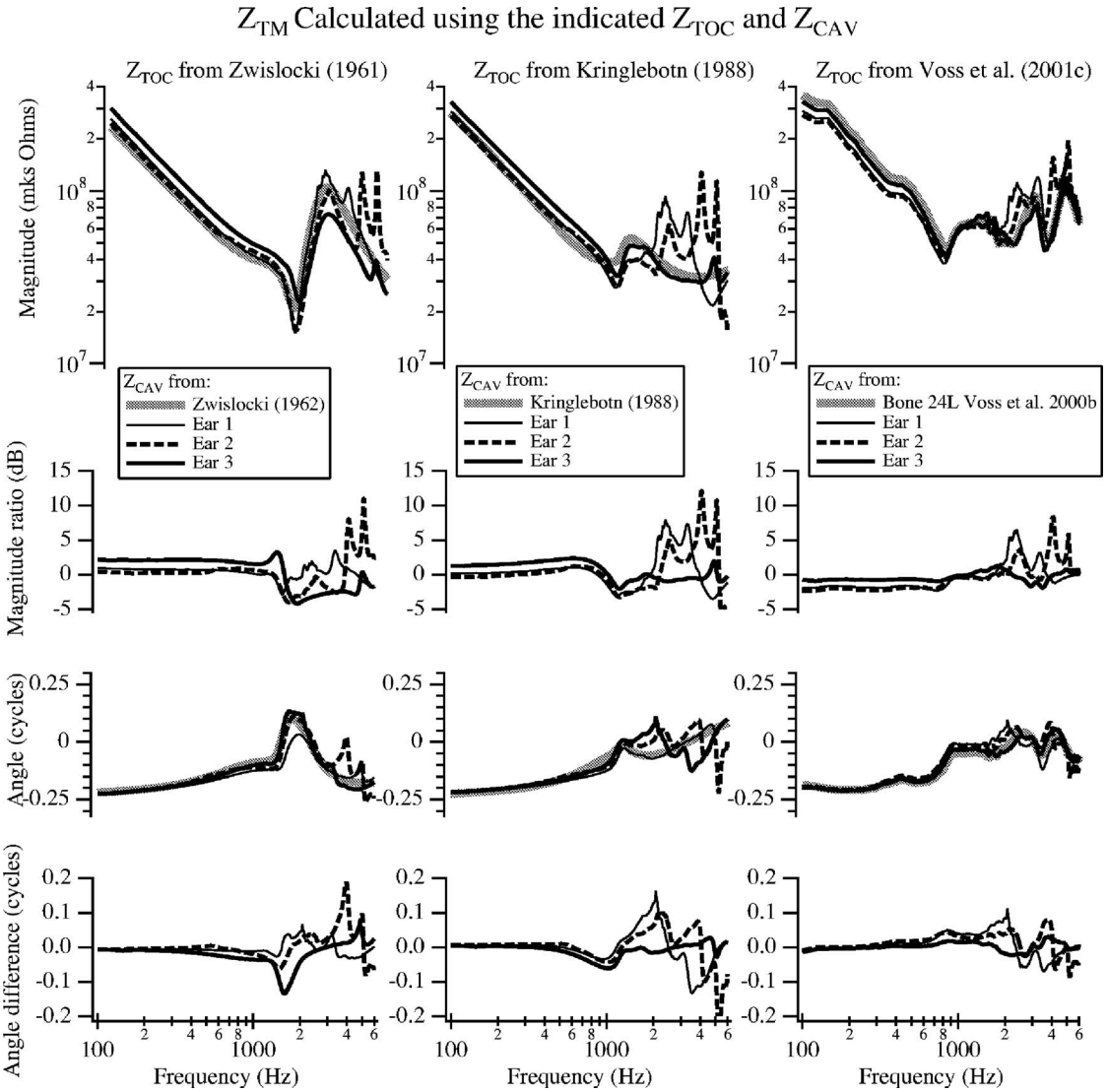


FIG. 6. Model results of the impedance at the tympanic membrane, Z_{TM}, calculated from Eq. (2), using different values for the impedance of the tympanic membrane, ossicles, and cochlea, Z_{TOC}, and the impedance of the middle-ear air space, Z_{CAV}. The values of Z_{TOC} were calculated from: **Left Panel:** the model of Zwislocki (1962); **Middle Panel:** the model of Kringlebotn (1988); and **Right Panel:** the measurements of Voss *et al.* (2001c). The values of Z_{CAV} used for each Z_{TM} calculation are indicated in the legends and were either calculated using the relevant model formulation or are measurements from this work (ears 1, 2, and 3) or from Bone 24L of Voss *et al.* (2001b). **Upper:** Magnitude of the impedance at the tympanic membrane with the indicated Z_{TOC} and Z_{CAV}. **Middle-upper:** Magnitude ratio between the calculation with the Z_{CAV} from the ears with unaltered mastoids (ears 1, 2, and 3) and the baseline load (Zwislocki model, Kringlebotn model, and Bone 24L), where the ratio is plotted in dB where the ratio is computed as 20 $\log_{10}(Z_{CAV}/Z_{CAV}^{baseline})$. **Middle-lower:** Angle of the impedance at the tympanic membrane with the indicated Z_{TOC} and Z_{CAV}. **Lower:** Difference in angle between the calculation with the Z_{CAV} from the ears with unaltered mastoids (ears 1, 2, and 3) and the baseline load (Zwislocki model, Kringlebotn model, and Bone 24L).

We calculate the impedance at the tympanic membrane (Z_{TM}) from Eq. (2) using a combination of models and measurements of Z_{TOC} and Z_{CAV}. Specifically, we use three different estimates of Z_{TOC}: (1) Z_{TOC} from the Zwislocki (1962) model, (2) Z_{TOC} from the Kringlebotn (1988) model, and (3) the measurements of Z_{TOC} from Voss *et al.* (2001c). In combination with these estimates of Z_{TOC}, we use the middle-ear air space impedance Z_{CAV} as determined from the Zwislocki (1962) model, the Kringlebotn (1988) model, the measurements of Voss *et al.* (2000b), and the measurements of Z_{CAV} presented here. Figure 6 plots the impedance at the tympanic membrane, Z_{TM}, using these combinations of Z_{TOC} and Z_{CAV}. The middle-ear air space does affect the impedance at the tympanic membrane. Low-frequency differences between the results with the model air-space impedance and the measured air-space impedance occur because the total volume of the air space differs between the three measurements and the target volume of each model. Above 1000 Hz the major results include: (1) the introduction of multiple maxima and minima in the impedance at the tympanic membrane when Z_{CAV} comes from the impedance measured on ears with unaltered mastoid cavities (i.e., ears 1, 2, 3 versus the models or the altered mastoid cavity of bone 24L), and (2) variations of more than 10 dB in magnitude and 0.1 cycles in angle from the impedances predicted by either model. The introduction of multiple maxima and minima is consistent with many of the impedance measurements of Voss and Allen (1994, Figs. 6 to 10) in which the advanced impedance (estimate of the impedance at the tympanic membrane) demonstrates multiple local maxima and minima, similar to those seen here in

the model results of Fig. 6. Thus, it appears that variations in human middle-ear air spaces might influence the impedance at the tympanic membrane: (1) below 1000 Hz the effect depends on the total volume of the middle-ear air space and systematically increases or decreases the total magnitude by a few dB at all low frequencies, and (2) above 1000 Hz, the effect is complicated, depends on the specific anatomy of a particular ear, and can introduce multiple maxima and minima as a fine structure in the impedance.

D. What does the antrum transfer function H_P tell us?

1. H_P: Models and measurements

Figure 3 compares our measurement of H_P [Eq.(1)] to the middle-ear circuit models of both Zwislocki (1962) and Kringlebotn (1988). Both models predict one maximum in the magnitude of $|H_P|$ with a decrease in magnitude of 40 dB per decade at frequencies above the maximum. The measurement, on the other hand, has multiple extrema that are not represented by these models. The measurement is consistent with the model topologies of Fig. 5, where multiple extrema result from additional energy-storage elements in the model. We are not aware of any other measurements that compare antrum pressure to other pressures within the middle-ear air space; thus, our single measurement of H_P provides a test of the structure-based middle-ear cavity model that suggests that that mastoid air-cell network is important. This measurement-based result is new and invites more work.

2. Impedance of the antrum and mastoid air-cell system

We use our measurement of H_P, the transfer function between the pressures P_{ANTRUM} and P_{CAV}, to characterize the impedance of the antrum and mastoid air-cell system. Specifically, we propose the model topology of Fig. 7 (upper), which differs from the previous models (i.e., Fig. 1 right) in that the antrum and mastoid air-cell system are not represented by a single capacitor but instead by the impedance Z_{MASTOID}. Using the model topology of Fig. 7 (upper), a measurement-based estimate for Z_{MASTOID} can be calculated as

$$Z_{\mathrm{MASTOID}} = \frac{H_P}{\frac{1}{Z_{\mathrm{CAV}}} - \frac{1}{Z_T}}, \tag{3}$$

where H_P is the measured transfer function between the pressures P_{ANTRUM} and P_{CAV}, Z_{CAV} is the measured impedance of the middle-ear air space impedance, and $Z_T = 1/j\omega C_t$ is the impedance of the tympanic cavity with the compliance of the tympanic cavity C_t determined from the volume of the tympanic cavity[9] and calculated for volumes of 0.5, 1.0, and 1.5 cm^3, which spans the range of tympanic-cavity volumes reported in the literature (Gyo *et al.*, 1986; Whittemore *et al.*, 1998) and estimated from model fits (Table I).

The measurement-based estimate of Z_{MASTOID} (Fig. 7 lower) is not consistent with the previous modeling approach of representing the antrum and mastoid air-cell system by a single compliance (Fig. 1 right). At low frequencies Z_{MASTOID} is compliance dominated, but above about 800 Hz

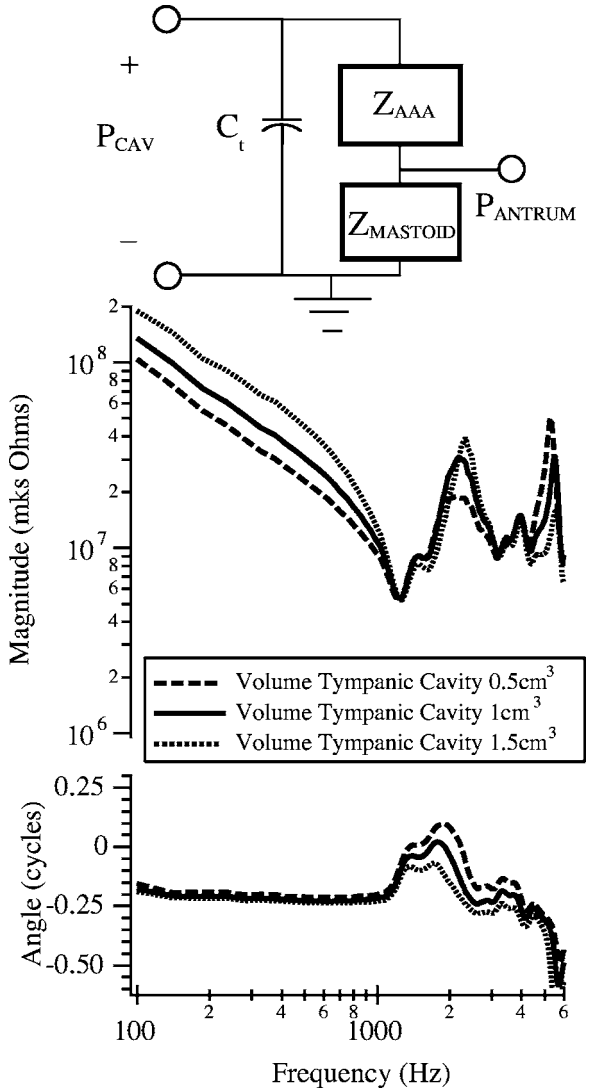


FIG. 7. Estimates of the impedance of the antrum and mastoid air-cell system, Z_{MASTOID}. Z_{MASTOID} is calculated using the model topology shown in the upper panel combined with measurements from ear 3 of the pressure at the tympanic ring, P_{CAV}, and within the antrum, P_{ANTRUM}, and the representation of the tympanic cavity by a compliance C_t. The magnitude (middle panel) and angle (lower panel) of Z_{MASTOID} is plotted for tympanic cavity volumes of 0.5, 1.0, and 1.5 cm^3.

Z_{MASTOID} would be better represented by a model with multiple resonances, such as a combination of the possibilities proposed here in Fig. 5. Above 4000 Hz, our measurement-based estimate of Z_{MASTOID} clearly includes errors, as the angle is less than −0.25 cycles. This error could be due in part to the fact that as frequency increases, the tympanic cavity cannot be represented as a pure compliance. In other words, as the wavelength of sound approaches 10%–20% of the dimensions of the tympanic cavity, it is inappropriate to represent the components as lumped-circuit elements.

E. Clinical implications of results

Many researchers are working to develop noninvasive ear-canal based acoustical measurements as diagnostic tools for middle- and inner-ear dysfunction (e.g., impedance and reflectance measurements, ossicular motion measurements, otoacoustic emissions). It has been noted that substantial

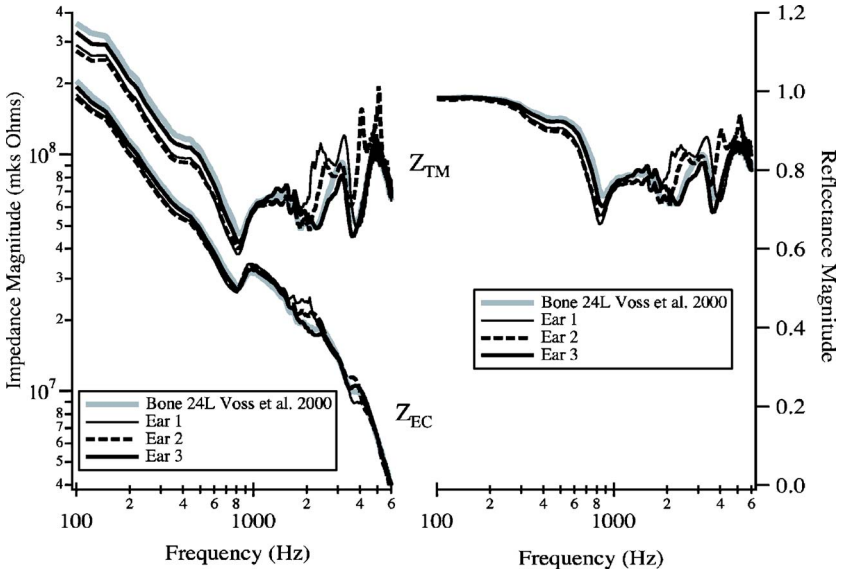


FIG. 8. **Left Panel:** Impedance magnitudes at the tympanic membrane ($|Z_{TM}|$) and 10 mm from the tympanic membrane in the ear canal ($|Z_{EC}|$). The $|Z_{TM}|$ are reproduced from the right column of Fig. 6 and were calculated from Eq. (2) from the Z_{TOC} of bone 24L of Voss *et al.* (2001c) and the Z_{CAV} from the same bone, and ears 1, 2, and 3 here. The $|Z_{EC}|$ was calculated for each $|Z_{TM}|$ by modeling the ear canal as a 10-mm-long air-filled cylindrical tube with a diameter of 8 mm [e.g., Voss and Shera, 2004, Eq. (3)]. **Right Panel:** Reflectance magnitudes calculated from the Z_{EC} at the left, using Eq. (2) of Voss and Allen (1994) (i.e., $R=(Z_{EC}^n-1)/(Z_{EC}^n+1)$, where Z_{EC}^n is the ear-canal impedance normalized by the characteristic impedance of the ear canal).

variability occurs in these measurements in normal-hearing ears (e.g., Zwislocki and Feldman, 1970; Voss and Allen, 1994; Møller, 2000). The work presented here demonstrates that variations in middle-ear air space anatomy contribute to some of the observed intersubject variability in ear-canal based acoustic measurements. Figure 6 demonstrates the effects of variations in middle-ear air space on the impedance at the tympanic membrane: Above 1000 Hz, there are variations of up to 10 dB in magnitude and 0.1 cycles in angle.

Ear-canal based impedance measurements are typically made within the ear canal several millimeters away from the tympanic membrane and not at the tympanic membrane; in this case, the air space between the measurement location and the tympanic membrane contributes substantially to the impedance measurement. Figure 8 (left) compares the impedance magnitude at the tympanic membrane Z_{TM} to the impedance magnitude within the ear canal 10 mm from the tympanic membrane Z_{EC}; variations in middle-ear air space impedance have nearly indistinguishable effects on the impedance at the location 10 mm from the tympanic membrane. Indeed, impedance measured within the ear canal is heavily influenced by the impedance location measurement, and as such, ear-canal impedances are limited in the information they provide. As detailed by others, including Stinson *et al.* (1982), Stinson (1990), and Voss and Allen (1994), the transformation of impedance to reflectance provides a measure of middle-ear response that does not depend on measurement location, assuming acoustic losses in the ear canal are negligible. Figure 8 (right) demonstrates that the reflectance magnitude, measured at any location within the ear canal, is sensitive to variations in the impedance of the middle-ear air space. Thus, as ear-canal middle-ear diagnostic tests are developed and evaluated, intersubject variations in middle-ear air space anatomy should be recognized as a source of variability in some ear-canal based measurements.

V. SUMMARY

Measurements of the impedance of the middle-ear air space from three cadaver ears with intact mastoid air spaces show a compliance-dominated impedance for frequencies lower than about 500 Hz and magnitudes with multiple extrema at higher frequencies. These measurements differ from previous measurements made on cadaver ears with altered mastoid spaces; ears with intact mastoid spaces have impedance magnitudes with multiple extrema, while ears with an enlarged antrum but no mastoid air-cell networks have impedances with fewer and sharper extrema. Previous models of the impedance of the middle-ear air space are consistent with the measurements on ears with no mastoid air-cell networks; the models represent the antrum and mastoid air cell system as a single volume of air. These models are able to capture the gross features of the impedance measurements presented here (e.g., the first magnitude miminum), but they fail to represent the multiple extrema in magnitude and corresponding transitions in angle. An alternative topology for a circuit model of the middle-ear air space is suggested as a guide for describing the acoustics of the middle-ear air space. However, the model's complexity and the substantial inter-ear variability in the anatomy of the middle-ear air space make a specific model for the middle-ear air space impractical. In general, the previous models (Zwislocki, 1962; Kringlebotn, 1988; Voss *et al.*, 2000b) lose some of the fine structure in the impedance of the middle-ear air space but they provide a simple and reasonable approximation of this impedance.

Analysis of the measurements shows that variations in middle-ear air space impedance do affect the impedance at the tympanic membrane for frequencies above 1000 Hz. It is suggested that intersubject variability in ear-canal based measurements may result partially from variability in middle-ear air space anatomy.

ACKNOWLEDGMENTS

We gratefully thank Diane Jones of the Otopathology Laboratory of the Massachusetts Eye and Ear Infirmary for her assistance with preparing the temporal bones for measurements, John Kosakowski of the Smith College Machine Shop for his help in the design and building of the acoustic

cavities used to calibrate our system, and Rindy Northrop of The Temporal Bone Foundation, Inc. for sharing and viewing with us her sections prepared for microscopic observation of a human temporal bone that included the mastoid. We also thank John J. Rosowski, William T. Peake, and Christopher A. Shera for helpful discussions and comments on the manuscript. An anonymous reviewer also provided helpful comments on the manuscript. This work was the undergraduate Honor's Thesis of C.E.S. under the direction of Susan E. Voss. The research was supported by Smith College and a grant from the Ford Motor Company.

APPENDIX: CALCULATION OF MODEL ELEMENTS FOR THE FOUR-ELEMENT MODEL

We outline the process for determining the values for the four circuit elements (C_t, M_{ad}, R_{ad}, and C_a) of the model described in Fig. 1 (right). As presented by Voss *et al.* (2000b), we assume that at low frequencies, the cavity impedance Z_{CAV} can be approximated as a pure compliance C_{cav} which can be calculated from our measurements and used to constrain the compliances C_t and C_a such that

$$C_{cav} = C_t + C_a. \tag{A1}$$

We solve for C_{cav} from our measurements of Z_{CAV} at 244 Hz,

$$C_{cav}(f = 244\ \text{Hz}) = \left| \frac{1}{2\pi f Z_{CAV}} \right|. \tag{A2}$$

The parallel resonance between the mass M_{ad} and the compliances C_t and C_a leads to a maximum in $|Z_{CAV}|$ at the frequency f_{max} where[10]

$$f_{max} = \frac{1}{2\pi} \sqrt{\frac{C_t + C_a}{M_{ad} C_t C_a}}. \tag{A3}$$

The series resonance between M_{ad} and C_a leads to a minimum in $|Z_{CAV}|$ at the frequency f_{min} where

$$f_{min} = \frac{1}{2\pi} \sqrt{\frac{1}{M_{ad} C_a}}. \tag{A4}$$

Division of Eq. (A3) by Eq. (A4) gives

$$\frac{f_{max}}{f_{min}} = \sqrt{1 + (C_a/C_t)}. \tag{A5}$$

We can then compute values for C_a and C_t from Eqs. (A1) and (A5), with C_{CAV}, f_{max}, and f_{min} determined from the experimental data. With values for C_a and C_t, we use the frequency of the minimum in $|Z_{CAV}|$ and Eq. (A4) to solve for M_{ad}. Finally, we use a frequency-dependent resistance (Beranek 1986, pp. 137–138) that results in the magnitude of the measured $|Z_{CAV}|$ and the model $|Z_{CAV}|$ matching at the magnitude minimum, i.e., the resonant frequency between the mass M_{ad} and the compliances C_t and C_a.

[1]The measurements of middle-ear volume by Molvaer *et al.* (1978) include the tympanic cavity and the mastoid air cell system. We approximate the measurements of the mastoid air cell system to be approximately one cm^3 smaller than the reported measured volume of the entire space.

[2]The description provided by Onchi (1961) of the cadaver ear reads: "Pure tones were conducted by a metal tube into the external auditory canal of the temporal bones removed from fresh cadavers." Figure 3 of Onchi (1961) includes a schematic drawing of the antrum and mastoid-cell network which leads us to assume that the mastoid air cells were part of the ear on which Onchi (1961) measured Z_{CAV}.

[3]The measurements are noisy above 4000 Hz because the source used to measure the impedance did not generate sound pressure levels that were high enough to be sensed adequately by the microphone (Voss *et al.*, 2000b).

[4]The Zwislocki (1962) middle-ear air space model includes an additional resistor in parallel with the capacitor that represents the tympanic cavity within Fig. 1 (right). The additional resistor has been ignored for the discussion here, as it controls the sharpness of the resonances and does not add additional ones. This resistor is further discussed by Voss *et al.* (2000b).

[5]The two resonances are not obvious in the Kringlebotn (1988) model results (Fig. 2) because the model inductor value appears to be artificially low (Voss *et al.*, 2000b).

[6]We note that the two ears from the same cadaver resulted in the mastoid volumes of 5.9 cm^3 (ear 2) and 1.5 cm^3 (ear 3). A possible explanation for such asymmetry could be the observation that ears affected by secretory otitis media for substantial periods during the first years of life are smaller in volume than ventilated ears (Robinson *et al.*, 1993); perhaps the left ear of our cadaver donor was afflicted by otitis media during its developmental years.

[7]The impedance measurement made by Onchi (1961) on a single cadaver middle-ear air space was reproduced here as a magnitude and angle (Fig. 2) from the author's representation with a real and imaginary part. The original plot (Onchi, 1961, Fig. 10) suggests a noncausal response: Near 3000 Hz the real part of the impedance appears to be forced to maintain a positive value while there is no corresponding abrupt transition in the derivative of the imaginary part of the impedance at this frequency.

[8]We are unaware of systematic anatomical examinations to determine if connections between the tympanic cavity and the mastoid air-cell system in addition to the aditus ad antrum exist. However, in one set of serial temporal bone sections that we viewed, there appears to be a connection between the superior-anterior portion of the tympanic cavity and several air cells.

[9]Volume V is related to compliance C via the equation $V = \rho c^2 C$ where ρ is the density of air and c is the speed of sound in air.

[10]To simplify our expressions here, we initially ignore the effect of R_{ad}, which is small as shown by the sharp resonances in the measurements of Z_{CAV}.

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.

Beranek, L. L. (**1986**). *Acoustics* (American Institute of Physics, New York).

Donaldson, J. A., Duckert, L. G., Lambert, P. M., and Rube, E. W. (**1992**). *Anson Donaldson Surgical Anatomy of the Temporal Bone*, 4th ed. (Raven, New York).

Farmer-Fedor, B. L. and Rabbitt, R. D. (**2002**). "Acoustic intensity, impedance and reflection coefficient in the human ear canal," J. Acoust. Soc. Am. **112**, 600–620.

Feeney, M. P., Grant, I. L., and Marryott, L. P. (**2003**). "Wideband energy reflectance measurements in adults with middle-ear disorders," J. Speech Lang. Hear. Res. **46**, 901–911.

Feeney, M. P. and Keefe, D. H. (**1999**). "Acoustic reflex detection using wideband acoustic reflectance, admittance, and power measurements," J. Speech Lang. Hear. Res. **42**, 1029–1041.

Feeney, M. P. and Keefe, D. H. (**2001**). "Estimating the acoustic reflex threshold from wideband measures of reflectance, admittance, and power," Ear Hear.. **22**, 316–332.

Goode, R. L., Ball, G., and Nishihara, S. (**1993**). "Measurement of umbo vibration in human subjects—method and possible clinical applications," Am. J. Otol. **14**, 247–251.

Gyo, K., Goode, R. L., and Miller, C. (**1986**). "Effect of middle-ear modification on umbo vibration—human temporal bone experiments with a new vibration measuring system," Arch. Otolaryngol. Head Neck Surg. **112**, 1262–1268.

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.

Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**2000**). "A non-

invasive method for estimating acoustic admittance at the tympanic membrane," J. Acoust. Soc. Am. **108**, 1128–1146.
Hunter, L. L. and Margolis, R. H. (**1997**). "Effects of tympanic membrane abnormalities on auditory function," J. Am. Acad. Audiol **8**, 431–446.
Keefe, D. H. (**1984**). "Acoustical wave propagation in cylindrical ducts: Transmission line parameter approximations for isothermal and nonisothermal boundary conditions," J. Acoust. Soc. Am. **75**, 58–62.
Keefe, D. H., 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. C. (**1996**). "Maturation of the middle and external ears: Acoustic power-based responses and reflectance tympanometry," Ear Hear. **17**, 361–373.
Koç, A., Ekinci, G., Bilgili, A. M., Akpinar, I. N., Yakut, H., and Han, T. (**2003**). "Evaluation of the mastoid air cell system by high resolution computed tomography: Three-dimensional multiplanar volume rendering technique," J. Laryngol. Otol. **117**, 595–598.
Kringlebotn, M. (**1988**). "Network model for the human middle ear," Scand. Audiol. **17**, 75–85.
Magliulo, G., Cianfrone, G., Gagliardi, M., Cuiuli, G., and D'Amico, R. (**2004**). "Vestibular evoked myogenic potentials and distortion-product otoacoustic emissions combined with glycerol testing in endolymphatic hydrops: Their value in early diagnosis," Ann. Otol. Rhinol. Laryngol. **15**, 1000–1005.
Margolis, R. H., Saly, G. L., and Keefe, D. H. (**1999**). "Wideband reflectance tympanometry," J. Acoust. Soc. Am. **106**, 265–280.
Merchant, S. N., Ravicz, M. E., Puria, S., Voss, S. E., Whittemore, Jr., K. R., Peake, W. T., and Rosowski, J. J. (**1997**). "Analysis of middle-ear mechanics and application to diseased and reconstructed ears," Am. J. Otol. **18**, 139–154.
Merchant, S. N., Ravicz, M. E., Voss, S. E., , Peake, W. T., and Rosowski, J. J. (**1998**). "Middle ear mechanics in normal, diseased, and reconstructed ears," J. Laryngol. Otol. **112**, 715–731.
Merchant, S. N., Rosowski, J. J., and Ravicz, M. E. (**1995**). "Middle ear mechanics of type IV and type V tympanoplasty. II. Clinical analysis and surgical implications," Am. J. Otol. **16**, 565–575.
Møller, A. R. (**2000**). *Hearing: Its Physiology and Pathophysiology* (Academic, San Diego).
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.
Neely, S. T. and Gorga, M. P. (**1998**). "Comparison between intensity and pressure as measures of sound level in the ear canal," J. Acoust. Soc. Am. **104**, 2925–2934.
Onchi, Y. (**1961**). "Mechanism of the middle ear," J. Acoust. Soc. Am. **33**, 794–805.
Piskorski, P., Keefe, D. H., Simmons, J. L., and Gorga, M. P. (**1999**). "Prediction of conductive hearing loss based on acoustic ear-canal response using a multivariate clinical decision theory," J. Acoust. Soc. Am. **105**, 1749–1764.
Robinson, P. J., Lodge, S., Goligher, J., Bowley, N., and Grant, H. R. (**1993**). "Secretory otitis media and mastoid air cell development," Int. J. Pediatr. Otorhinolaryngol. **25**, 13–18.
Rosowski, J. J., Davis, P. J., Merchant, S. N., Donahue, K. M., and Coltrera, M. D. (**1990**). "Cadaver middle ears as models for living ears: Comparisons of middle ear input impedance," Ann. Otol. Rhinol. Laryngol. **99**, 403–412.
Rosowski, J. J. and Merchant, S. N. (**1995**). "Mechanical and acoustic analysis of middle ear reconstruction," Am. J. Otol. **16**, 486–497.
Rosowski, J. J., Merchant, S. N., and Ravicz, M. E. (**1995**). "Middle ear mechanics of type IV and type V tympanoplasty. I. Model analysis and predictions," Am. J. Otol. **16**, 555–564.
Schuknecht, H. (**1968**). "Temporal bone removal at autopsy," Arch. Otolaryngol. **87**, 129–137.
Stepp, C. E. and Voss, S. E. (**2004**). "Acoustics of the middle-ear air space in human ears," Abstracts of the American Auditory Society Annual Meeting, Vol. **29**.
Stinson, M. R. (**1990**). "Revision of estimates of acoustic energy reflectance at the human eardrum," J. Acoust. Soc. Am. **88**, 1773–1778.
Stinson, M. R., Shaw, E., and Lawton, B. W. (**1982**). "Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal," J. Acoust. Soc. Am. **72**, 766–773.
Voss, S. E. (**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. and Allen, J. B. (**1994**). "Measurement of acoustic impedance and reflectance in the human ear canal," J. Acoust. Soc. Am. **95**, 372–384.
Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2001a**). "How do tympanic-membrane perforations affect human middle-ear sound transmission?," Acta Oto-Laryngol. **121**, 169–173.
Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2001b**). "Middle-ear function with tympanic-membrane perforations. I. Measurements and mechanisms," J. Acoust. Soc. Am. **110**, 1432–1444.
Voss, S. E., Rosowski, J. J., Merchant, S. N., and Peake, W. T. (**2001c**). "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., 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., Thornton, A. R., Shera, C. A., and Peake, W. T. (**2000c**). "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.
Voss, S. E. and Shera, C. A. (**2004**). "Simultaneous measurement of middle-ear input impedance and forward/reverse transmission in cat," J. Acoust. Soc. Am. **110**, 2187–2198.
Whittemore, 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.
Zwislocki, J. and Feldman, A. (**1970**). "Acoustic impedance in normal and pathological ears," Am. Speech Hear. Assoc. Monograph **15**, 1–42.

Acoustical cues for sound localization by the Mongolian gerbil, *Meriones unguiculatus*

Katuhiro Maki[a] and Shigeto Furukawa
Human and Information Science Laboratory, NTT Communication Science Laboratories, NTT Corporation, 3-1, Morinosato-Wakamiya, Atsugi, Kanagawa, 243-0198 Japan

(Received 11 November 2004; revised 9 May 2005; accepted 10 May 2005)

The present study measured the head-related transfer functions (HRTFs) of the Mongolian gerbil for various sound-source directions, and explored acoustical cues for sound localization that could be available to the animals. The HRTF exhibited spectral notches for frequencies above 25 kHz. The notch frequency varied systematically with source direction, and thereby characterized the source directions well. The frequency dependence of the acoustical axis, the direction for which the HRTF amplitude was maximal, was relatively irregular and inconsistent between ears and animals. The frequency-by-frequency plot of the interaural level difference (ILD) exhibited positive and negative peaks, with maximum values of 30 dB at around 30 kHz. The ILD peak frequency had a relatively irregular spatial distribution, implying a poor sound localization cue. The binaural acoustical axis (the direction with the maximum ILD magnitude) showed relatively orderly clustering around certain frequencies, the pattern being fairly consistent among animals. The interaural time differences (ITDs) were also measured and fell in a ±120 μs range. When two different animal postures were compared (i.e., the animal was standing on its hind legs and prone), small but consistent differences were found for the lower rear directions on the HRTF amplitudes, the ILDs, and the ITDs. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1944647]

PACS number(s): 43.64.Ha, 43.66.Pn, 43.66.Qp [WPS] Pages: 872–886

I. INTRODUCTION

For most animals, the ability to localize a sound source plays a critical role in their survival by assisting them to capture a prey or to escape from predators. Spectral information (the spectral envelope at the eardrum), interaural level and time differences (ILD and ITD, respectively) are known to be major cues for sound localization (e.g., Knudsen and Konishi, 1979; Middlebrooks and Green, 1991). Recent modeling and physiological studies on sound localization have suggested that the neural mechanisms for sound localization differ among animal species, and are optimized by adapting to the characteristics of the acoustical information available to the animal (Brand *et al.*, 2002; Grothe, 2003; McAlpine and Grothe, 2003; Hârper and McAlpine, 2004). Thus, when conducting studies on the sound localization of a given animal species, it is important to base the experimental design or the interpretation of the study on the properties of the acoustical information available to the target research species. The properties of these acoustical cues have been investigated for many animal species including humans (Searle *et al.*, 1975; Middlebrooks *et al.*, 1989; Middlebrooks and Green, 1990), rhesus monkeys (Spezio *et al.*, 2000), cats (Roth *et al.*, 1980; Phillips *et al.*, 1982; Irvine, 1987; Musicant *et al.*, 1990; Rice *et al.*, 1992), ferrets (Carlile, 1990), Tammar wallabies (Coles and Guppy, 1986), various kinds of bat (Jen and Chen, 1988; Obrist *et al.*, 1993; Fuzessery, 1996; Firzlaff and Schuller, 2003), guinea pigs (Carlile and Pettigrew, 1987), and barn owls (Moiseff, 1989; Keller *et al.*, 1998). Unlike the earlier animal species, however, there have been no detailed studies on the acoustical cues available to the Mongolian gerbil (*Meriones unguiculatus*), which is another widely used experimental animal for behavioral and physiological studies on sound localization (e.g., Kelly and Potas, 1986; Heffner and Heffner, 1988; Sanes, 1990; Brückner and Rübsamen, 1995; Spitzer and Semple, 1995; Behrend *et al.*, 2002).

The goals of the present study were to measure gerbils' head-related transfer functions (HRTFs acoustical transfer function from a sound source to a point in an ear canal), to identify acoustical features in the HRTFs as potential cues for sound localization, and to examine the properties of the cues. We hoped that the present results would not only serve as a database of gerbils' HRTFs, but also provide a guideline for designing future experiments on the gerbil's sound localization ability.

In the present study, the monaural cues were analyzed in terms of the directional transfer functions (DTFs) (Middlebrooks and Green, 1990), which are the sound-source-direction sensitive components of HRTF amplitude spectra. Specifically, we examined the following features in the DTF that varied with the sound source direction: the frequency of the sharp spectral notch (e.g., Rice *et al.*, 1992; Wotton *et al.*, 1995); and the DTF gains at individual frequencies. The direction for which the DTF gain at a given frequency was maximal across directions is referred to as the acoustical axis (Middlebrooks and Pettigrew, 1981; Phillips *et al.*, 1982). Equivalent analyses were performed on a binaural cue, namely, the ILD spectrum. The ILD spectrum was the frequency-by-frequency difference between the DTFs of the left and right ears, and exhibited direction sensitive features

[a]Author to whom correspondence should be addressed; electronic mail: maki@avg.brl.ntt.co.jp

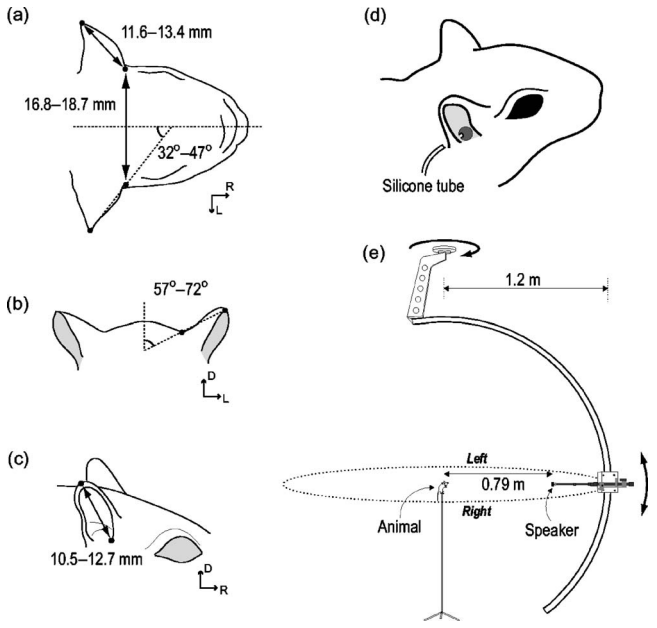


FIG. 1. The dimensions of the animal subjects (a–c), and an illustration of the recording setup (d, e). (d) A silicone tube was inserted into the ear canal from an incision near the ventro-posterior fringe of the pinna. (e) A speaker system for presenting a sound to the animal that allowed the speaker to move between −40° and 90° elevation and in all azimuthal directions. The distance between the sound source and the interaural center of the animal was fixed at 79 cm.

such as negative or positive peaks in the ILD spectrum (comparable to the DTF notches), and the ILD "gains" at given frequencies (comparable to the DTF gains). We also examined the direction and frequency dependence of the ITD, another binaural cue. At the end of our analyses, we compared the acoustical features of two animal states, i.e., the standing and prone states, simulating the animal's natural change of posture (Ågren *et al.*, 1989). In the present paper, the terms standing and prone indicate that the animal was upright on its hind legs and on all four feet, respectively.

II. MATERIALS AND METHODS

A. Animal preparation

The experiments were performed on eight adult Mongolian gerbils (*Meriones unguiculatus*) of both sexes, weighing 62–102 g. Figures 1(a)–1(c) show the range of the size and the angle of the animal's head and pinna. Initially, the animal was anesthetized with pentobarbital sodium (25 mg/kg, i.p.). The tympanic membranes of both ears were solidified by coating them with cyanoacrylate adhesive. This was done to remove the effects of the variation of acoustic impedance in the ear canal over the measurement period, which was probably due to the changes in the mechanical characteristics of the middle-ear system. A small incision was made in the skin near the ventro-posterior fringe of the pinna, and a small hole was made in the wall of the bony meatus, through which a silicone tube (Etymotic Res., ER7-14C, 2.5 cm long, 0.95 mm outer diameter, 0.5 mm inner diameter) was inserted [Fig. 1(d)]. The tube was fixed to the bone with cyanoacrylate adhesive so that the tip of the tube in the ear canal was usually located 0.3–0.7 mm from the entrance of the ear canal. For two animals, the distance between the tube tip and the ear-canal entrance was varied between 0.0 and 3.0 mm, in order to examine the effects on the DTF of the tube-tip position. After fixing the tube in the ear canal, the head and pinna of the animal were held in a natural position, and then the animal was euthanized with a lethal dose of pentobarbital sodium. The head and body of the animal were secured by fixing its lower incisors, waist and chest on a custom-made wire frame so that it was close to its natural standing or prone posture.

B. Stimulus presentation

Experiments were performed in a double-walled, soundproof anechoic room ($4.8 \times 5.4 \times 4.7$ m). The room temperature was maintained at 15 °C by an air conditioner. A movable speaker system was used for sound presentation [Fig. 1(e)]. The system consisted of a semicircular arm (3/4 circular arc, 1.2 m radius), along which a speaker mount could move when driven by a stepping motor controlled by a personal computer (PC). The speaker mount had an extendable radial arm to which a loudspeaker was attached. The system allowed us to vary the sound source location on an imaginary sphere with various radii. The center of interaural axis of the animal's head was positioned at the center of the imaginary sphere. The source-location coordinate system was formed so that the median sagittal plane corresponded to 0° azimuth and the plane formed by the ear canal entrances and the eyes corresponded to 0° elevation. Thus, a source direction of 0° elevation and 0° azimuth indicates a source directly in front of the animal. The azimuth is the angle subtended at the center of the animal's head to the right and left of the midline plane. Positive and negative azimuths indicate positions to the right and left of the midline plane, respectively. The elevation is the angle above and below the horizontal plane. Positive and negative elevations represent positions above and below the horizontal plane, respectively. The distance from the sound source to the center of the interaural axis was fixed at 79 cm.

The stimulus used for estimating the transfer functions was a time-stretched pulse (TSP), which was a kind of frequency sweep with a flat and broadband power spectrum (Suzuki *et al.*, 1995) that covered the gerbil's audible frequency range (up to 45–50 kHz) (Ryan, 1976). The TSP was synthesized digitally by the PC at a sampling rate of 97 656.25 Hz with a resolution of 24 bits. The TSP signal consisted of a total of 65 536 (2^{16}) sample points. The TSP signal was generated through a digital-to-analog converter (Tucker-Davis Technology, RP 2.1), amplified (Nittobo Acoust. Eng., HA-94C), and emitted from a 2.5 cm diameter tweeter (Sony, SS-TW100ED) on the movable speaker system.

C. Measurement system

The end of the silicone tube (opposite to that inserted in the ear canal) was connected to a probe-tube microphone (Brüel and Kjær, type 4182). The output of the probe microphone was amplified (Brüel and Kjær, type 2636) and stored on the hard disk of the PC through an analog-to-digital con-

verter (Tucker-Davis Tech., RP 2.1) at a sampling rate of 97 656.25 Hz with a resolution of 24 bits. Recorded signals of 15 repetitions were averaged in the time domain to increase the signal-to-noise ratio. We measured microphone signals from the left and right ears for a total of 937 (or 865) sound directions ranging from −40° (or −30°) to 90° in 5° steps in elevation and from −180° to 180° in 10° azimuth steps. Before measuring the animal, we measured the frequency response of the speaker-microphone system using the microphone with an extension silicone tube (2.5 cm in length) located at the center of the animal's interaural axis, in the absence of the animal.

D. Data analysis

We calculated the impulse response from a recorded TSP signal based on the time-stretched-pulse theory (Suzuki *et al.*, 1995). Each impulse response was then truncated to 512 points by a Hamming window. The center of the window was always set at the 230th point of the impulse response, so that the maximum point of the impulse responses lay around the window's center. We calculated the frequency response from the impulse response using a 512-point fast Fourier transform (FFT) for analyzing amplitude spectra. For *phase* spectra analysis, zeros were padded after the 512-point impulse response, and a 32 768-point FFT was computed.

We derived the HRTF amplitude spectrum by dividing the amplitude spectrum of the frequency response derived from the animal's ear by that of the speaker-microphone system. We also calculated the DTFs, which we obtained by dividing the HRTF amplitude spectrum for each direction by the geometric average of the HRTF amplitude spectrum across all measurement locations (Middlebrooks and Green, 1990). The DTFs can be considered the source-direction dependent components of the HRTFs.

Monaural DTFs exhibited certain spectral features that depended on the source direction. One such feature consisted of one or more sharp spectral notches on the DTF, and the notch frequency often changed systematically with the sound direction. We used the following semiautomatic algorithm to define a single notch frequency for each DTF. First, for each animal and each azimuth, we defined the notch frequency of the DTF for −40° elevation as the frequency with the lowest DTF amplitude, for frequencies above 20 kHz. Next, for the DTF for the elevation adjacent to −40° (i.e., higher in elevation), the notch frequency was defined as the frequency with the lowest DTF amplitude within a certain frequency range around the notch frequency for the −40° elevation. To determine the notch frequencies for higher elevations, the frequency range was progressively adjusted so as to span by a fixed width above and below the notch frequency for the lower adjacent elevation. We adopted different frequency widths for different animals (range 0.5–2 octaves), so that the obtained notch-frequency-versus-direction plot appeared smooth and close to our subjective judgments by a visual inspection of the raw DTF data. Another type of feature in the DTF was the DTF amplitude gain at each frequency. We found that the gain at a given frequency varied with the source direction, and defined the direction of the maximum DTF gain as the acoustical axis (Middlebrooks and Pettigrew, 1981; Phillips *et al.*, 1982). In determining the acoustical axis for a given frequency, we first interpolated the DTFs with a spline approximation in 1° step for azimuths and elevations, and then searched for the maximum gain.

To characterize the binaural property of the transfer functions, we derived the *ILD spectrum* by computing the difference in the dB scale, frequency by frequency, between the DTFs of the left and right ears. A positive ILD indicates that the level in the right ear was higher than that in the left ear. The ILD spectrum generally exhibited negative and positive peaks whose frequencies depended on the sound-source direction. Thus, we defined the peak frequency of an ILD spectrum as the frequency with the maximum ILD magnitude. When the absolute value of the ILD peak was less than 5 dB, the peak frequency was excluded from the analyses. Also, similarly to the acoustical axis defined for the monaural DTF, we defined the *binaural acoustical axis* at each frequency as the direction for which the ILD was maximal (or minimal) for that frequency.

To calculate ITDs, which are another binaural property of the ear, we calculated the cumulative phase spectrum of each ear from the original phase spectrum by adding and subtracting increments of 2π radian as needed so that the phase at adjacent frequency components never differed by more than $\pm\pi$ radian. Next, the phase component common to all directions was subtracted from each obtained cumulative phase spectrum. The common phase component was calculated from the minimum phase of an average HRTF amplitude spectrum (Middlebrooks and Green, 1990). Finally, the difference phase spectra were calculated by subtracting the cumulative phase spectra of the left ear from those of the right ear for each matching source direction. The ITD at each frequency was derived by dividing the phase difference at each frequency, f, by $2\pi f$.

III. RESULTS

A. Effects of microphone position

Unless otherwise stated, our analyses were performed on the HRTFs of standing animals. It is known that the HRTF amplitude spectra vary with the position of the microphone in the ear canal (Mehrgardt and Mellert, 1977; Middlebrooks *et al.*, 1989; Chan and Geisler, 1990; Hammershøi and Møller, 1996; Keller *et al.*, 1998). Figures 2(a)–2(d) show the HRTF amplitude spectra recorded at four different microphone depths, i.e., the distance from the ear canal entrance (indicated by the numbers in the top left corners of the panels), for five different speaker elevations. A close inspection of the figures shows that the HRTFs consisted of three components. One is the spectral peak at around 8 kHz, which was probably due to the ear canal resonance, and was invariant with the microphone depth and source direction. The second is the spectral notch appearing for frequencies below 30 kHz (large arrows). The frequency of that notch increased with increasing microphone depth. The third type consisted of sharp spectral notches around 20–40 kHz (small arrows). It can be seen that the frequency of the notch moved towards higher frequencies as the source direction varied from −30°

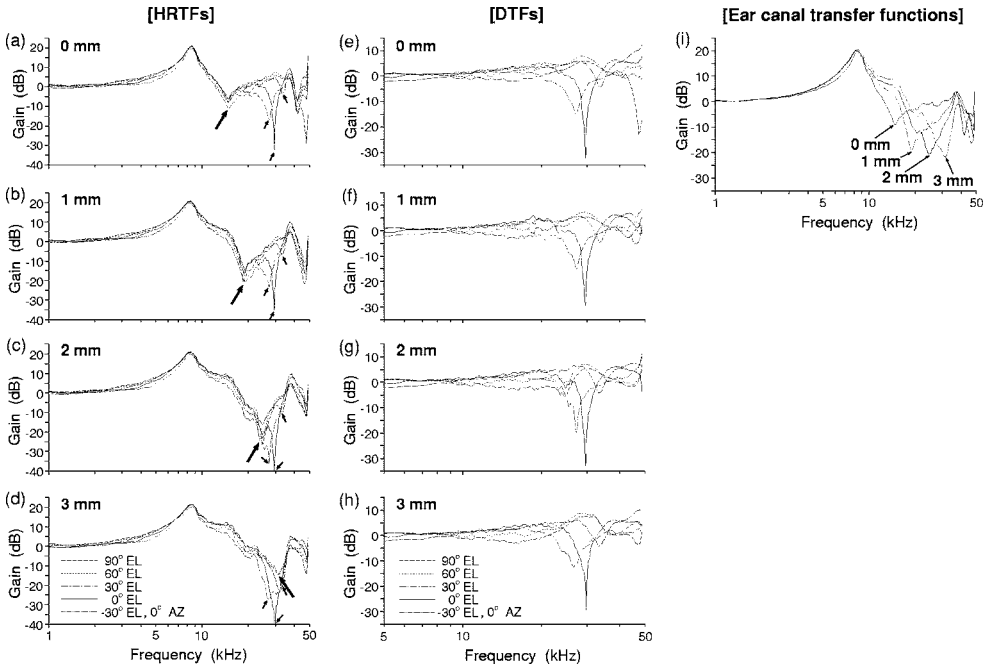


FIG. 2. The effects of the microphone depth (i.e., distance from the ear canal entrance) and sound source elevation of the HRTF amplitude spectra. (a–d) HRTF amplitude spectra. Each panel represents one microphone depth as indicated in the top left corner of each panel in millimeters. In each panel, the sound source directions are indicated by different line styles (see the key). (e–h) Directional transfer functions (DTFs). The conventions are the same as in panels (a–h). (i) The HRTFs averaged across the source directions, representing the direction-insensitive component of the HRTF. See the text for details.

to 30° in elevation. Our primary interest in this study was this source-direction dependent component, which is independent of microphone position. We extracted this component by computing the DTFs (Middlebrooks and Green, 1990), and plotted them in Figs. 2(e)–2(h). It can be seen that the DTFs were largely insensitive to the microphone depth, and that the pattern of the elevation dependence was very similar across all microphone depths. The results indicate that the effect of the difference in microphone position among the animals was negligible when analyzing the DTFs. Figure 2(i) shows the average HRTF amplitude spectra used for deriving the DTFs (i.e., the source-direction insensitive component). In the figure, there is a spectral peak at ~8 kHz, attributable to the ear-canal resonance (Ravicz *et al.*, 1996), and deep spectral notches (indicated by the arrows) whose frequency varied monotonically with the microphone depth [cf. large arrows in Figs. 2(a)–2(d)]. We consider that these spectral notches can be attributed to longitudinal standing waves formed in the ear canal.

B. Monaural DTF

Figure 3 shows the DTFs of the left and right ears of a representative animal (No. 521M). As expected, the DTFs of these ears were essentially bilaterally symmetric; compare, for example, the panels for ±30° azimuths. The most evident spectral features were deep spectral notches at frequencies above 25 kHz. In reference to the source directly in front of the animal (0° azimuth and 0° elevation), the center frequency of the spectral notch increased as the sound-source elevation increased, and for the right ear, as the sound-source azimuth moved from the front to the right side (and the reverse for the left ear). By contrast, the spectral features were less evident for the rear and upper directions and exhibited a relatively flat spectrum. Another feature in the DTF was direction dependent gain, and this was particularly apparent for frequencies above 15 kHz. For example, the gain for the right ear was maximal when the direction was about +90° azimuth (ipsilateral) and 30° elevation. In the following sections, we quantify those DTF features that bear source-direction-related information.

1. Notch frequency

The top panels in Fig. 4 show DTFs for right ear of animal No. 521M for elevations ranging from −40° to 90° in 5 deg steps at 0° azimuth [Fig. 4(a)] and for azimuths ranging from −90° to 90° in 10 deg steps and at 30° elevation [Fig. 4(b)]. Figures 4(a) and 4(b) show the deep spectral notches for frequencies above 25 kHz. Figure 4(a) indicates that the notch frequency generally increased monotonically with increasing source elevation (dotted lines are provided as a guide to the eye), although another notch appeared at a higher frequency for elevations below ~−20°. The notch frequency varied fairly systematically also with the azimuth [Fig. 4(b)]. With the examples shown in Fig. 4(b), the notch frequency tended to increase as the sound source moved from a contralateral to an ipsilateral direction relative to the ear. Figures 4(c) and 4(d) show examples of DTFs that exhibited less distinct spectral notches (animal No. 1020F): The notches for animal No. 1020F were generally shallower than those for animal No. 521M, and the notch frequency change was harder to track. Nonetheless, it can be seen that the source-direction dependence of the notch frequency had a similar pattern to that observed for animal No. 521M.

Figures 5(a) and 5(b) plot the notch center frequencies, observed in the DTFs for both ears of the animals Nos. 521M and 1020F (Fig. 4), as a function of both sound-source elevations and azimuths. The Materials and Methods section provides details of our semiautomatic algorithm for defining notch frequencies. The direction dependence of the notch frequency showed a similar tendency for animals Nos. 521M and 1020F, despite the difference in the notch depth. As the sound elevation changed from −40° to 90°, the notch fre-

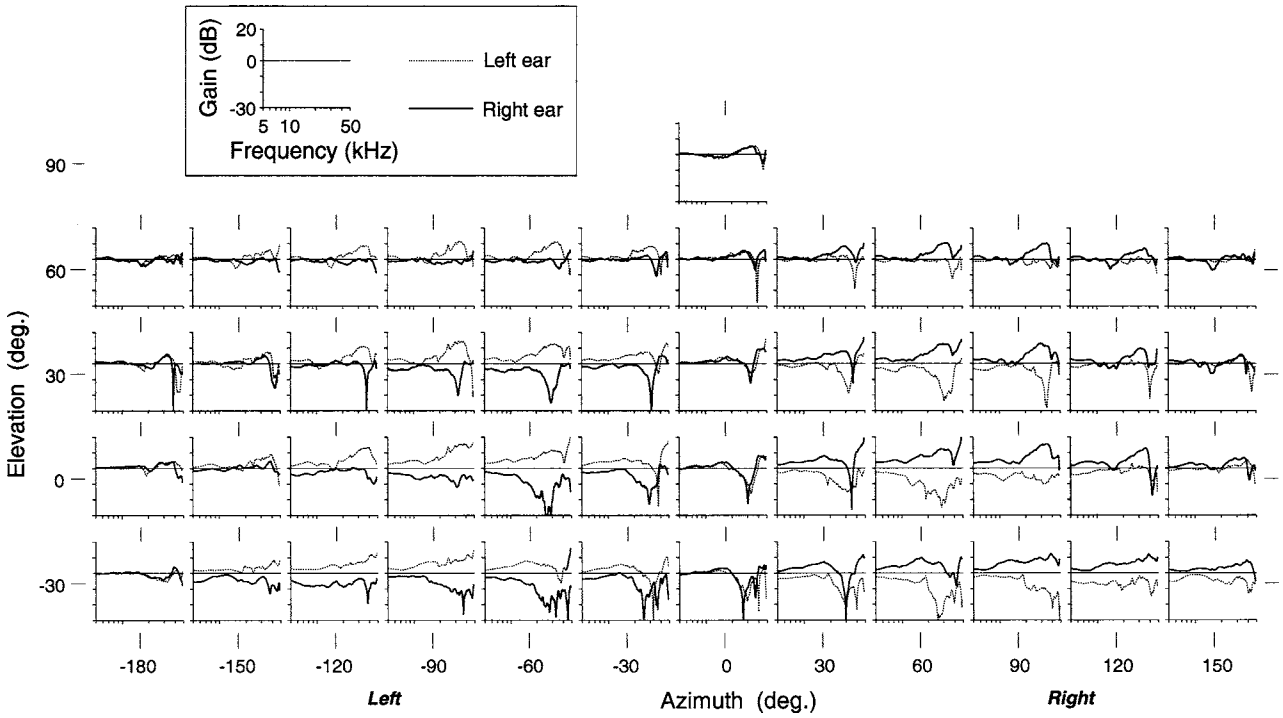


FIG. 3. The DTFs of a representative animal No. 521M. Each panel shows the DTFs of the left and the right ears (as indicated by the key) for one source direction. The panels are arranged vertically by source elevation, and horizontally by source azimuth, as indicated at the left and bottom of the panels, respectively.

quencies changed almost linearly from 25–30 kHz to about 45 kHz. The discontinuity observed for the left ear of animal No. 521M at elevations of −35° and −30° was due to the fact that our notch detection algorithm (see Materials and Methods section) chose the notch with the higher frequency of the two notches in the DTF [e.g., elevations below ~−20° in Fig. 4(a)]. If lower frequency notches were allowed [indicated by asterisks in Fig. 5(a)], the function would look more monotonic. When the sound azimuth changed from −90° to 90° [Fig. 5(b)], the notch frequencies for the two animals moved continuously and systematically from 30–35 to 45–50 kHz for the right ear, and the reverse was true for the left ear.

The simultaneous notch frequency dependence on azimuth and elevation is represented with iso-notch-frequency contours on a Mollweide projection (Fig. 6). The distributions of the notch frequencies were roughly symmetrical between the left and right ears. The lower frequency notches tended to distribute in contralateral azimuths and lower elevations, and vice versa. For both animals, the contours were most regular and continuous in the ipsilateral (e.g., positive azimuth for the right ear) and the upper (above 0° elevation) quadrant. By contrast, the notch frequencies in the lower hemisphere or in the contralateral hemisphere were less orderly, indicating that they did not serve as reliable cues. Behind the animal, (not shown in Fig. 6), there was no notch or it was not clearly defined (see Fig. 3). The earlier properties of the spectral notches were commonly observed in the DTFs for other animals.

2. Spatial distribution of DTF amplitude gain

The DTF gain at a given frequency tended to vary with source direction. Figure 7 shows the spatial distributions of the DTF gain at various frequencies for the right ears of two animals (Nos. 521M and 1020F). The DTF pattern for the left ear was almost mirror-symmetrical with that for the right ear about the midline, and thus the data are not shown. For frequencies between 5 and 30 kHz the DTF gains were symmetrical around 15° azimuth: positive and negative DTF gains for animal No. 521M are distributed on the left and right sides of the boundary, respectively. The spatial distributions of the DTF gain were relatively complex for frequencies above 35 kHz. For example, there were multiple positive peaks for 40 kHz [Fig. 7(f), positive azimuth]. The negative DTF gains for 40 kHz appeared in both hemispheres, in contrast to those for lower frequencies (≤30 kHz), which were restricted within the negative azimuth. This was also observed for animal No. 1020F [Figs. 7(h)–7(n)] and others. The maximum DTF gains across the directions depended on the frequencies. Those for animal No. 521M were about 2 dB at 5 kHz, 6–8 dB at 10–20 kHz, and 15 dB above 30 kHz. The maximum DTF gains for animal No. 1020F were generally 2–4 dB lower than those for animal No. 521M.

3. Acoustical axis

As indicated in Fig. 7, the direction of the maximum DTF gain varied across the frequencies. For example, the gain at 30 kHz for animal No. 521M was maximal at about 80° azimuth and 10° elevation, and that at 40 kHz was maximal at about 105° azimuth and −25° elevation. The direction of the maximum gain at a given frequency in the DTF or HRTF amplitude spectra is known as the acoustical axis (Middlebrooks and Pettigrew, 1981; Phillips *et al.*, 1982). The acoustical axes were calculated from DTFs for both ears of two animals (Nos. 521M and 1020F), and the results on

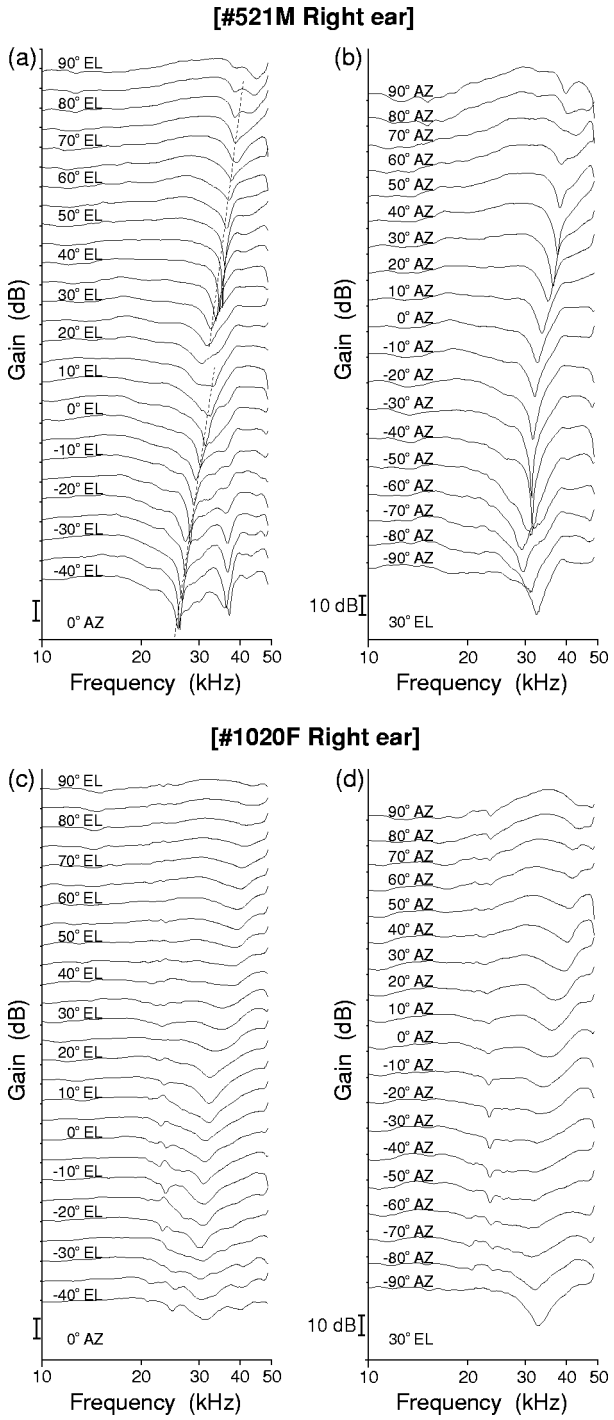


FIG. 4. DTFs of two animals (Nos. 512M and 1020F). Each panel represents DTFs for various elevations (left column) or azimuths (right column), for the right ear. The DTF gain is shown on an arbitrary scale; the vertical bar beside the ordinate indicates a 10 dB gain. The top and bottom rows of panels represent animals Nos. 521M and 1020F, respectively.

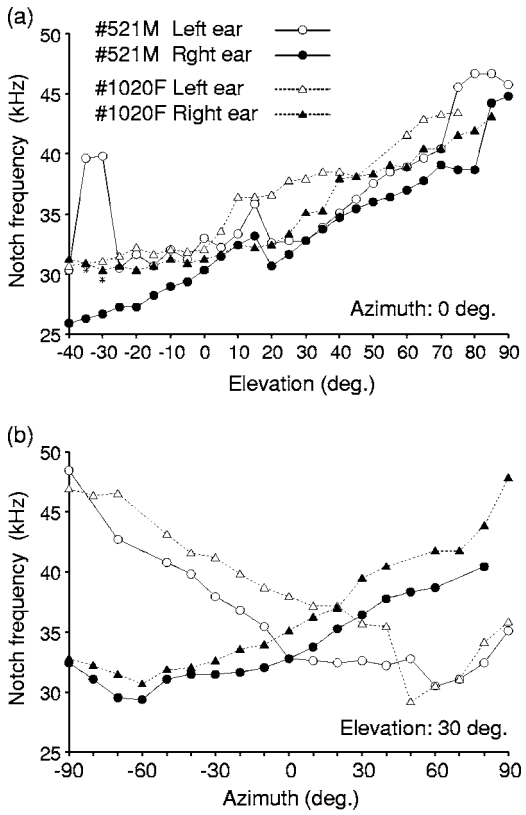


FIG. 5. The elevation and azimuth dependence of notch frequency in DTFs. The notch frequencies were calculated for the left and right ears of two animals (Nos. 512M and 1020F) by a semiautomatic notch-detection algorithm (see text). Each panel shows the notch frequency as a function of elevation [panel (a); 0° azimuth] or azimuth [panel (b); 30° elevation].

the median and horizontal planes are shown in Fig. 8. Figures 8(a)–8(d) indicate that the patterns of the frequency dependence of the acoustical axes as regards the azimuth were rather complex. The azimuth of the acoustical axis increased or decreased with an abrupt shift from lateral to medial or from medial to lateral as the frequency increased. There was no apparent similarity in the azimuthal changes of the acoustical axis among the animals. On the other hand, the elevation of the acoustical axis [Figs. 8(e)–8(h)] changed progressively from above to below the horizontal plane with an abrupt upward shift as the frequency increased. These characteristics have commonly been observed in other animals. For frequencies above 20 kHz, the profile of the elevation of the acoustical axis differed somewhat between ears or between animals. For example, the elevation of the acoustical axis for animal No. 1020F shifted upward for the right ear, but downward for the left ear with increasing frequency. Figure 9 is a two-dimensional representation of the acoustical axis. To indicate an overall pattern of association between the acoustical axis and frequency, the acoustical axes are divided into several frequency bands and are represented with different symbols. As expected from the results shown in Fig. 8, the plots reveal a mixture of smooth and abrupt transitions of the acoustical axis as a function of frequency. However, neither the animals nor the ears had any clearly common features.

C. Interaural level difference

The difference between the monaural DTFs of the left and right ears would result in ILDs that vary with frequency. In this section, we computed frequency-by-frequency ILDs for each source direction, which we refer to as the *ILD spectrum* (see Materials and Methods Section). Examples of ILD spectra for animals Nos. 521M and 1020F are shown in Fig. 10. Positive and negative ILDs indicate higher levels in the right ear and the left ear, respectively. Figure 10(a) shows

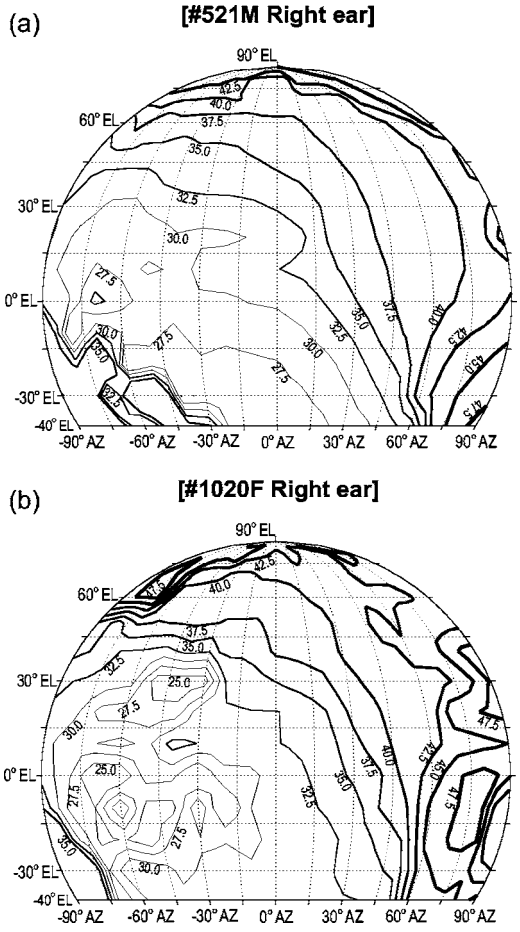


FIG. 6. Isofrequency contour of spectral notches in DTFs. Each panel represents the data for the right ear of one animal. The data are shown only for the frontal hemisphere. The notch frequency is indicated by the number beside each line (in kilohertz), and by the thickness of the line. The top and bottom panels represent animals Nos. 512M and 1020F, respectively.

that the ILDs for animal No. 521M had distinct positive and negative spectral peaks for a range of frequencies between 20 and 45 kHz. These positive and negative peaks in the ILDs systematically moved to higher frequencies as the sound elevation changed from −30° to 90° at the same azimuth of −30°. The ILDs shown in Fig. 10(b) did not have distinct positive spectral peaks (>15 dB) below 45 kHz, but the negative peak ILD value for each azimuth below −30 dB, except for 0° azimuth. By contrast, the ILDs for animal No. 1020F in Figs. 10(c) and 10(d) had no distinct positive or negative spectral peaks. The difference between the animals was mainly due to the difference of the notch depth in monaural DTFs (see Fig. 4).

1. Spatial distribution of positive and negative peak frequencies in the ILD

The positive and negative peaks observed in the ILDs are possible cues for sound localization as well as the spectral notches in the monaural DTFs. In this section, we examine the ILD spectrum, applying a similar analysis to that employed for monaural DTFs. Figure 11 shows isofrequency contours for ILD positive and negative peaks of animals Nos. 521M and 1020F (cf. Fig. 6). The center frequencies of the positive and negative peaks changed in a fairly orderly fashion with respect to both azimuth and elevation, as observed for the monaural DTFs. However, the distribution patterns were less regular than those of the monaural DTFs. The distribution of the ILD peak frequencies of animals Nos. 521M and 1020F were very different, unlike the monaural DTFs (Fig. 6). For example, the negative peak frequencies for animal No. 521M [Fig. 11(b)] were distributed in an orderly way for azimuths of 0°–90°, which was not the case for animal No. 1020F [Fig. 11(d)].

2. Spatial distribution of ILD at various frequencies

In this section, we examine the spatial distribution of the ILDs at various frequencies. This analysis is equivalent to that for the monaural DTF gain shown in Fig. 7. Figure 12 shows the spatial distribution of the ILDs calculated at seven frequencies for animals Nos. 521M and 1020F. For both animals, the spatial distribution of the ILD was symmetrical between the left and the right hemispheres at any given frequencies with inverted signs. As expected from the results shown in Fig. 10, the maximum ILDs for animal No. 1020F were generally 5–10 dB smaller than those for animal No. 521M at frequencies above 20 kHz. The maximum ILDs across all directions varied with frequency. For example, the maximum ILDs for animal No. 521M across all directions were about 5 dB at 5 kHz, 10 dB at 10 kHz, 20–25 dB at 20 kHz, 30–40 dB at 30 kHz and 25–30 dB above 35 kHz. The results were similar for the absolute values of the minimum ILDs. A similar pattern was commonly seen for all the animals.

3. Spatial distribution of directions with maximum ILD

Figure 12 shows that the directions of the spatial maxima or minima in an ILD, defined as the *binaural acoustical axis*, varied with frequency. For example, the binaural acoustical axis for 30 kHz had an azimuth of about −85° and an elevation of 10° for animal No. 1020F, and for 45 kHz, it had an azimuth of about −100° and about an elevation of −10°.

The binaural acoustical axes for animals Nos. 521M and 1020F are shown in Fig. 13. The results for the maximum ILD are not shown because they were generally mirror images of those for the minimum ILD. As seen in Fig. 13, the binaural acoustical axis was widely distributed in the frontal space with an azimuth ranging between −150° and −30°, as opposed to the monaural acoustical axis shown in Fig. 9 (the left ear) where the azimuth ranged between −135° and −45°. The distribution of the binaural acoustical axis showed some similarities between animals, unlike that of the monaural acoustical axis, in that the distribution could be divided into three regions in space: the binaural acoustical axis for low frequencies (5–15 kHz, indicated with x and *) distributed in a spatial region lateral to a −105° azimuth and below −30° elevation. In middle frequencies (15–35 kHz, indicated with triangles), the binaural acoustical axis was distributed above a 90° azimuth across all elevations. The binaural acoustical

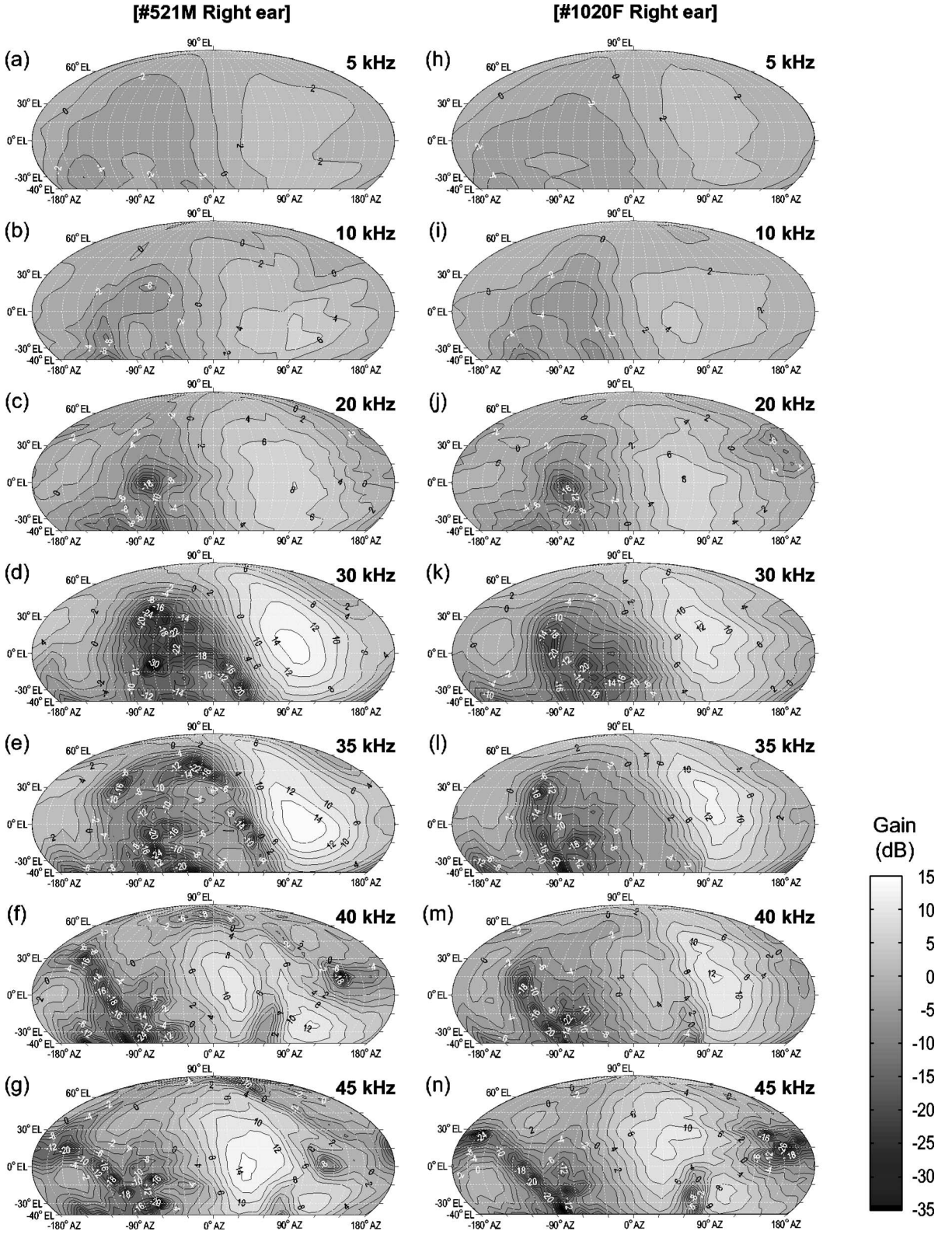


FIG. 7. Spatial distributions of DTF amplitude gain at various frequencies. The gain is indicated by the gray scale and by the isogain contours with numbers on the Mollweide projection. The isogain contour lines are drawn at 2 dB intervals. The panels (a–g) and (h–n) represent the right ear of animals Nos. 512M and 1020F, respectively. The frequencies are indicated in the upper right corners of individual panels.

axis for high frequencies (35–50 kHz, indicated with solid gray symbols) was distributed across low and middle elevations. A similar pattern was also commonly observed for the binaural acoustical axis in other animals.

D. Interaural time difference

In this section, we examine the direction- and frequency-dependent properties of ITDs. Figure 14 shows the spatial

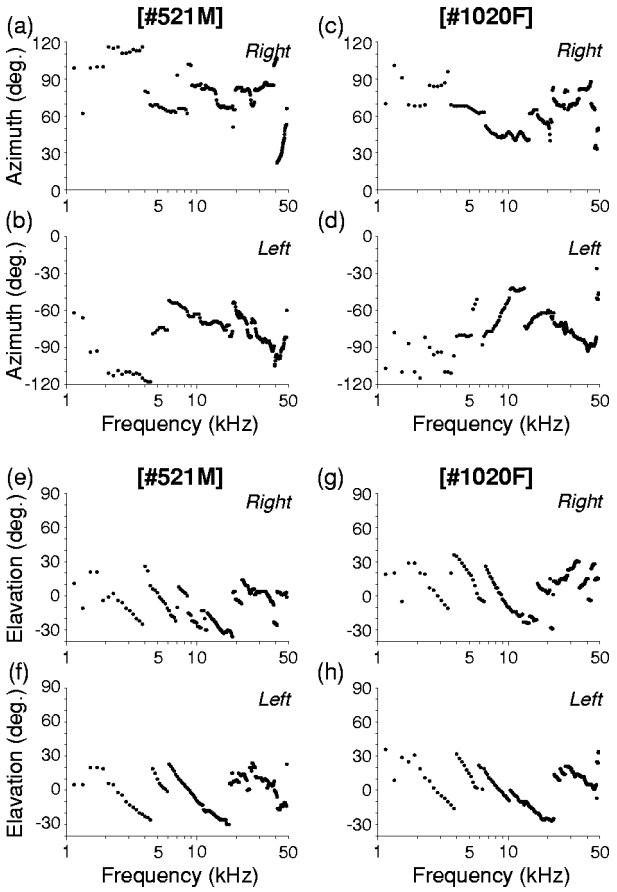


FIG. 8. Acoustical axis as a function of frequency. The acoustical axis is defined as the direction for which the DTF gain is maximal. The top (a–d) and the bottom (e–h) four panels represent the acoustical axis in azimuth and in elevation, respectively. The data from animals Nos. 512M and 1020F are shown in the left and right columns, respectively. The recorded ear (left or right) is indicated in the upper right corner of each panel.

distribution of the ITDs calculated at four frequencies for animals Nos. 521M and 1020F. Positive and negative values indicate that the sound is leading in the right ear and left ear, respectively. For both animals, the ITDs were symmetric between the left and the right hemispheres at any given frequencies with opposite signs. The ITDs were roughly symmetric also between the front and the rear hemispheres around the 90° azimuth. The maximum ITDs for animal No. 1020F were generally 10–20 μs smaller than those for animal No. 521M. This is because animal No. 1020F had a smaller head width than animal No. 521M (the head widths were 16.8 and 18.3 mm, respectively). The maximum ITD was located at around a 90° azimuth and −15° elevation for any frequencies. The maximum ITD magnitudes varied with frequency. For example, the maximum ITD magnitudes for animal No. 521M across all directions were about 135 μs at 2 kHz, 115 μs at 5 kHz, and 110 μs above 10 kHz. These ITD properties were consistent for all the animals.

Figure 15 represents another three animals with similar head widths (about 17 mm), showing ITDs for four azimuths as a function of frequency. For all the animals, the ITDs decreased monotonically with increasing frequency below 15 kHz at any given azimuth. The arrows on the far right of the figures indicate ITDs derived from Woodworth's spherical-head model (Woodworth and Schlosberg, 1962) assuming a head diameter of 17 mm. The measured ITDs of the gerbils were generally larger than the theoretical ITDs obtained from the spherical-head model. This is because the gerbil's head deviated markedly from a globular shape, especially in the frontal hemisphere, and thus, sound waves must travel over the top of the protruding nose to reach the ear contralateral to the sound source [see Fig. 1(a)].

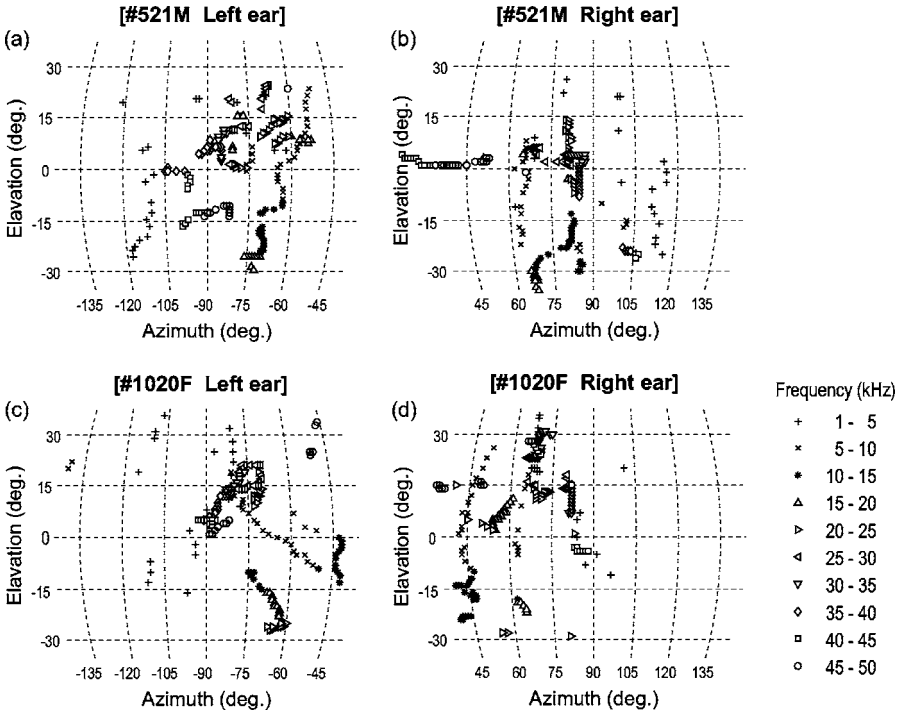


FIG. 9. Spatial distribution of the acoustical axis. The acoustical axes for a certain frequency band are indicated by the same symbols as shown in the key. The left and the right panels represent the left and right ears, respectively. The top and the bottom panels represent animals Nos. 512M and 1020F, respectively.

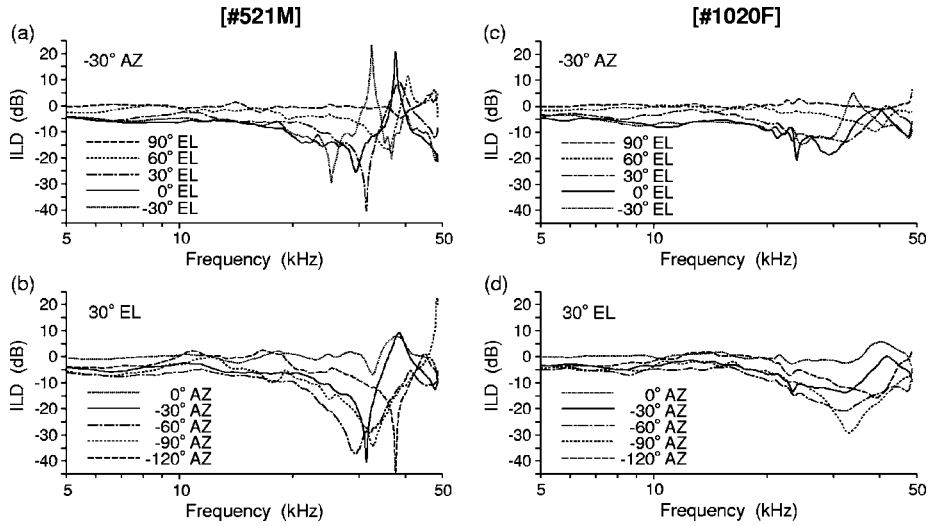


FIG. 10. ILD spectra for two animals (Nos. 512M and 1020F). The ILD spectrum represents the difference on a decibel scale, frequency by frequency, between the DTFs of the left and right ears. A positive ILD indicates that the level in the right ear was higher than that in the left ear. Each panel represents ILD spectra of one animal (indicated at the top of the plots) for various elevations (top panels; at −30° azimuth) and azimuths (bottom panels; at 30° elevation).

E. Effects of posture

The analyses described in the previous sections were performed for HRTFs recorded from standing gerbils (i.e., standing on their hind legs), as is often seen in nature when a gerbil is in an active exploring state. Since the HRTFs of many previously studied species were recorded from animals in a prone state, it is interesting to determine the extent to which the results of the present analyses depend on the choice of animal posture (i.e., standing or prone). Thus, in the present section, we compared the DTFs, ILDs, and ITDs for different postures.

Figures 16(a) and 16(b) show monaural DTFs for the left ear of animal No. 1121M measured when the animal was standing (solid lines) and prone (dotted lines). The overall features of these DTFs are generally similar to those described earlier (Figs. 3 and 4 in Sec. III B). In the frontal and ipsilateral hemispheres [Fig. 16(a)], the DTFs of the prone animal were almost identical to those of the standing animal. Appreciable differences in the DTFs were observed around 5–15 kHz for lower rear directions [indicated with arrows in Fig. 16(b)]. There, the difference in DTF gain was as much as about 10 dB.

Figure 16(c) shows the ILD difference distribution for the two postures. The ILD difference was averaged across frequencies of 5–15 kHz, where differences resulting from posture were most apparent in the monaural DTFs [see Fig.

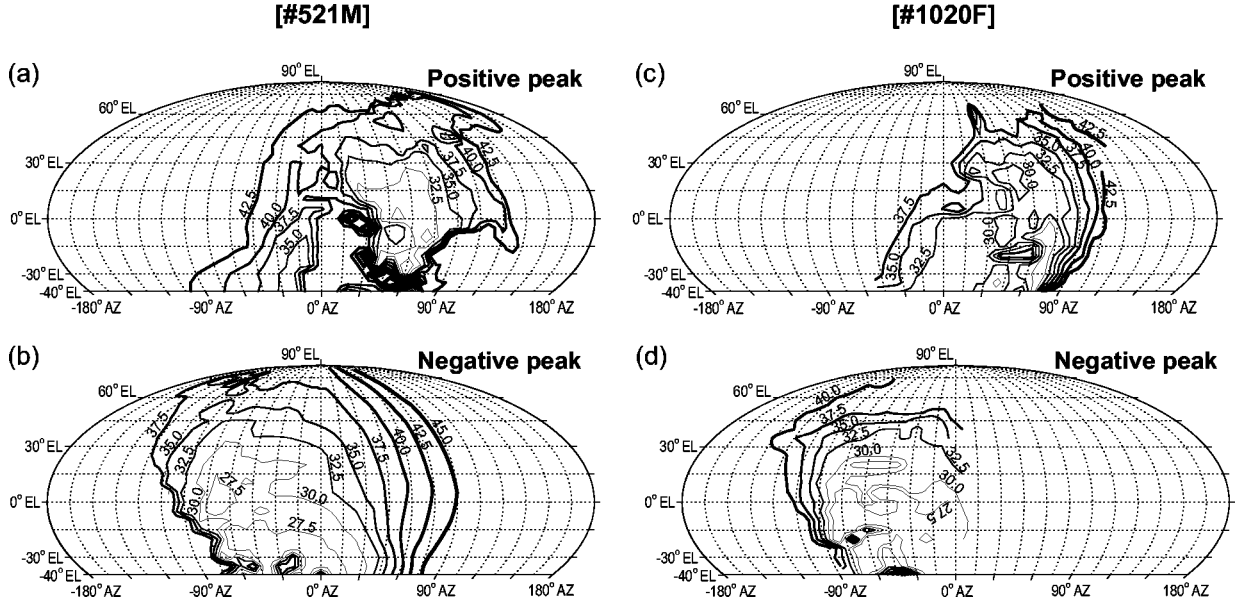


FIG. 11. Isofrequency contour plots of positive and negative peaks in the ILD spectra. In each plot, the peak frequency is indicated by the number beside each line (in kilohertz), and by the thickness of the line. High and low frequencies are shown by thick and thin lines, respectively. The contours are derived from positive or negative peaks whose absolute values exceed 5 dB. The data for animals Nos. 512M and 1020F are shown in the left and right columns, respectively. The top and the bottom panels represent positive and negative peaks, respectively.

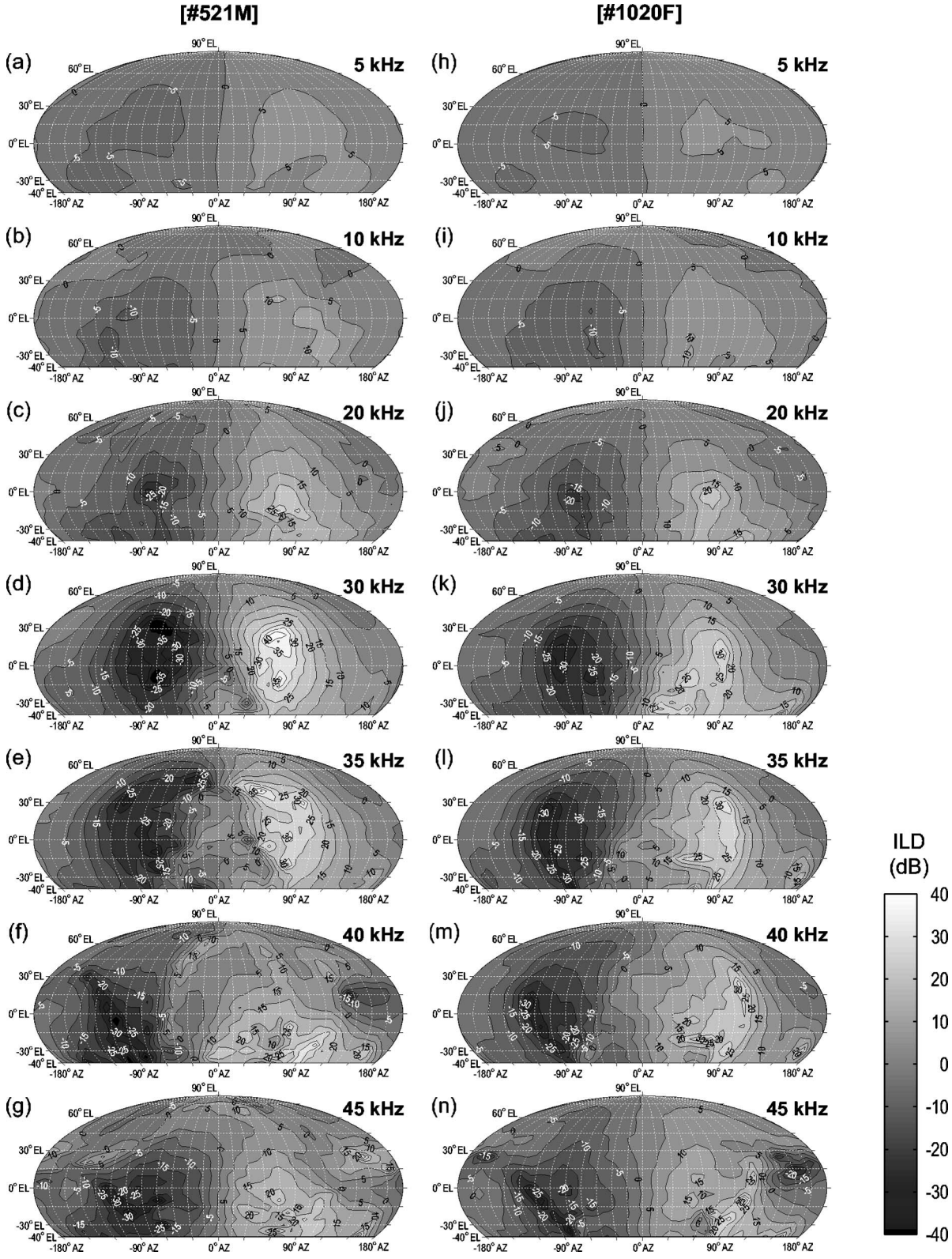


FIG. 12. Spatial distribution of ILDs for seven different frequencies. The ILD is indicated by the gray scale and by the iso-ILD contours with numbers on the Mollweide projection. The iso-ILD contour lines are drawn at 5 dB intervals. Positive ILD values indicate that level in the right ear is greater than in the left ear. Panels (a–g) and (h–n) represent animals Nos. 512M and 1020F, respectively. The frequencies are indicated in the upper right corners of each panel.

16(b)]. Appreciable ILD differences were restricted to the region with azimuths of ±90° to ±165°, and elevations below 0°. This difference reflected the difference in the monaural DTF on the contralateral side [see Fig. 16(b)]. In this region, the absolute ILD of the prone animal was generally larger than that of the standing animal.

Figure 16(d) shows the ITD difference distribution for the two postures. The ITD difference was averaged across frequencies of 2–10 kHz, where the posture-related differences were most apparent. The ITD difference spread concentrically around a ±140° azimuth and −20° elevation. In this region, the absolute ITD of the prone animal was larger

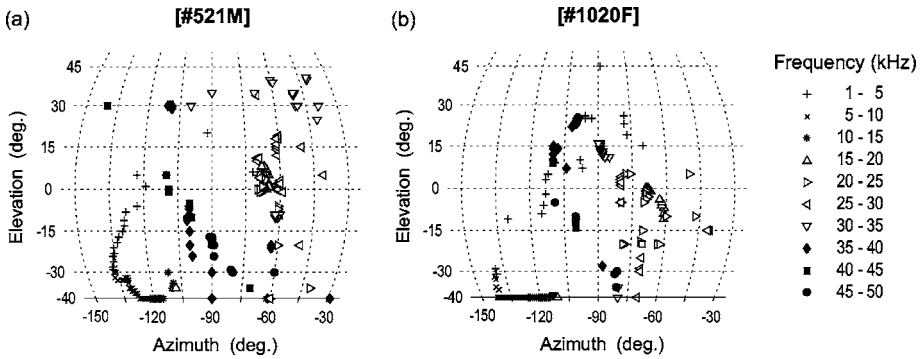


FIG. 13. Spatial distribution of the binaural acoustical axis. The binaural acoustical axis at a given frequency is defined as the direction for which the ILD was minimal (or maximal) for that frequency. The binaural acoustical axes for a certain frequency band are indicated by the same symbols as shown in the key. Panels (a) and (b) represent animals Nos. 512M and 1020F, respectively.

than that of the standing animal. This ITD region overlapped to a large extent the region in which the ILD differed for the standing and prone animal. The maximum ITD difference between the two postures was about 30 μs.

Overall, the most apparent differences were seen for directions in rear and low-elevation regions. This was as expected because for those directions, because the sound shadowing effect caused by the animal body in the prone state would be maximal. These effects of the animal posture were also consistently observed in two other animals that we examined.

IV. DISCUSSION

A. Comparison with previous studies

The present results showed that in gerbils, the DTF notches appeared for a frequency range above 25 kHz. For individual directions, the notches with the lowest frequencies generally fell within the audible frequency range of the animal (up to 45–50 kHz) (Ryan, 1976), and thus could serve as a cue for sound localization. The frequency range for the DTF notches in gerbils largely overlapped with those in other species, such as the big brown bat and the pallid bat

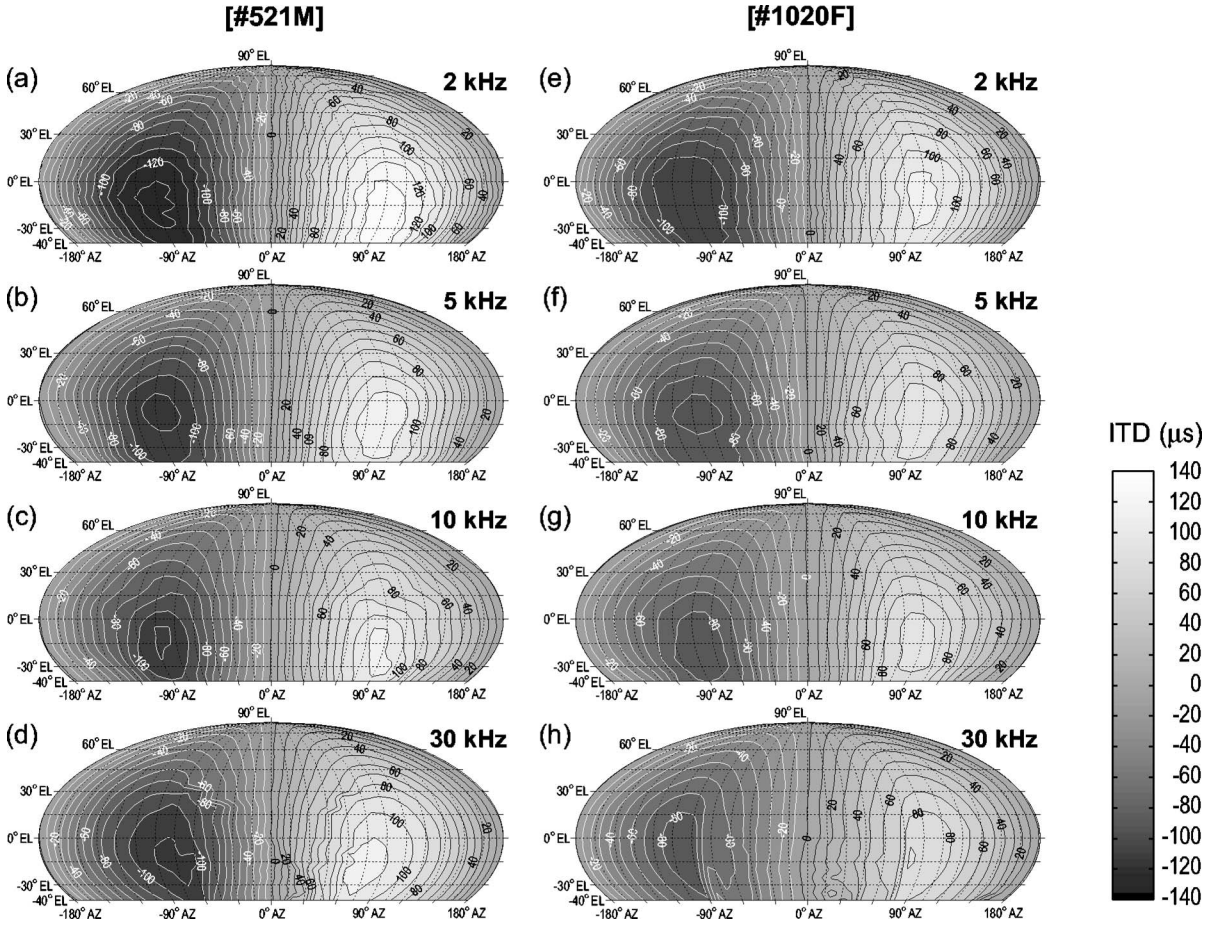


FIG. 14. Spatial distribution of ITDs for four different frequencies. The ITD is indicated by the gray scale and by the iso-ITD contour lines at 10 μs intervals on the Mollweide projection. Positive ITD values indicate that the right signal is leading. Panels (a–d) and (e–h) represent animals Nos. 512M and 1020F, respectively.

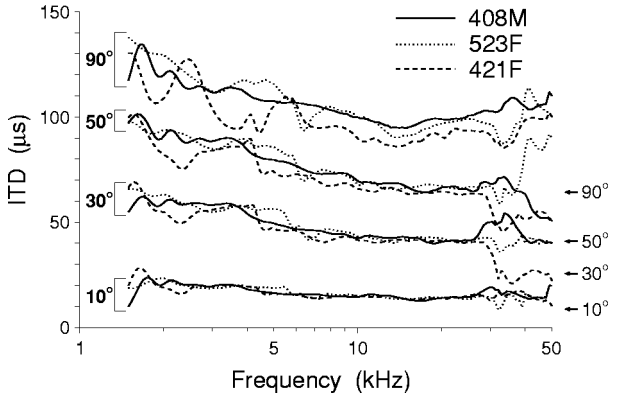


FIG. 15. Frequency dependence of ITDs. ITDs at four azimuths (10°, 30° 50°, and 90°) are drawn as a function of frequency. Source elevation was always 0 degree. Three animals with similar head widths (about 17 mm) are represented (see the key). The arrows on the far right in the figures indicate ITDs derived from a spherical-head model (Woodworth and Schlosberg, 1962). See the text for details.

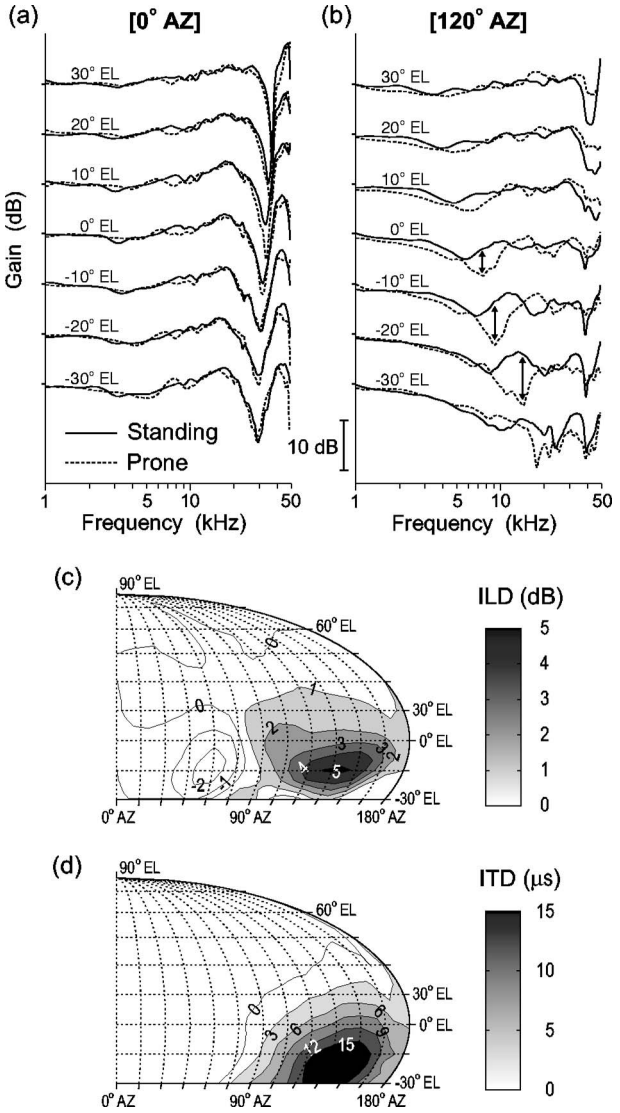


FIG. 16. Effects of different postures on monaural DTF, ILD, and ITD. Panels (a) and (b) show monaural DTFs obtained from the left ear of a standing (solid line) or prone (dotted line) animal. Panels (a) and (b) represent DTFs for 0° and 120° azimuths, respectively, and each panel represents various elevations. Panels (c) and (d) represent the posture-induced differences in ILD and ITD, respectively. The differences are indicated by the gray scale and by the isodifference contour lines at 1-dB intervals for ILD and at 3 μs intervals for ITD on the Mollweide projection. The results were mirror-symmetric around the midline, and thus are shown only for one hemisphere. The ILD and ITD differences were averaged across frequencies from 5 to 15 kHz and from 2 to 10 kHz, respectively, where the posture effects were most apparent. All data were derived from animal No. 1121M.

(Wotton *et al.*, 1995; Fuzessery, 1996), for which the pinna size was comparable to that of the gerbil (15–30 mm) (Jen and Chen, 1988; Fuzessery, 1996). For species with larger pinna sizes, the notch frequencies were generally lower than those for the gerbils (human—Mehrgardt and Mellert, 1977; Shaw, 1982; rhesus monkey—Spezio *et al.*, 2000; cat—Musicant *et al.*, 1990; Rice *et al.*, 1992; ferret—Carlile, 1990; barn owl—Keller *et al.*, 1998). We found that the pattern of the correspondence between the notch frequency and the source direction (Fig. 6) was similar for gerbils and cats (Fig. 8, Rice *et al.*, 1992; Figs 5 and 6, Young *et al.*, 1996), in that the iso-frequency contours of the notch center frequencies run diagonally at a angle of approximately 45° in front of the animals and that the notch center frequency increases as either the azimuth or elevation increases. A similar analysis of the pallid bat (Fig. 12, Fuzessery, 1996), however, failed to indicate such a systematic pattern. It may be that the anatomical features of the gerbil's pinna that determine the DTF notch frequencies are similar to those of the cat's pinna, and unlike those of the pallid bat's pinna.

Previous studies indicated that the acoustical axis changed rather smoothly on the horizontal or median planes as a function of frequency, with occasional abrupt jumps, for most species that have been studied (cat—Musicant *et al.*, 1990; Tammar wallaby—Coles and Guppy, 1986; guinea pig—Carlile and Pettigrew, 1987; bat—Jen and Chen, 1988; Obrist *et al.*, 1993; Firzlaff and Schuller 2003), except for the rhesus monkey (Spezio *et al.*, 2000). For the gerbil, we found a smooth trend along the median plane, although the direction of the trend (decreasing in elevation with increasing frequency) was opposite to that for other species. On the horizontal plane, the smooth trend was much less compelling, and the pattern of the frequency dependence was generally inconsistent between the ears within an animal and among animals. Thus, we could reasonably conclude that the association between acoustic axis and frequency is not robustly represented in the monaural DTFs of the gerbils.

We found that the ILD spectrum [Figs. 10(a) and 10(b)] often had positive and negative peaks, whose frequencies depended on the source direction. Such elevation-dependent ILD peaks were also observed in the cat (Musicant *et al.*, 1990). The maximum ILD of the gerbil was 30–40 dB for a frequency of 30 kHz, which fell within the maximum ILD range (20–50 dB) of other previously studied species (human—Searle *et al.*, 1975; Middlebrooks *et al.*, 1989; rhesus monkey—Spezio *et al.*, 2000; Tammar wallaby—Coles and Guppy, 1986; cat—Irvine, 1987; Musicant *et al.*, 1990; ferret—Carlile, 1990; bat—Obrist *et al.*, 1993; Fuzessery, 1996; Firzlaff and Schuller 2003; barn owl—Moiseff, 1989; Keller *et al.*, 1998).

The maximum ITD of the gerbil was 110–135 μs, which is as predicted from the linear relationship between the maximum ITD and the head size (human—Middlebrooks

and Green, 1990; rhesus monkey—Spezio *et al.*, 2000; cat—Roth *et al.*, 1980; barn owl—Moiseff, 1989; Keller *et al.*, 1998). For a given azimuth, the ITD generally decreased with increasing frequency up to a certain frequency (about 15 kHz for the animals represented in the Fig. 15). This pattern of frequency dependence was similar to those for humans (Kuhn, 1977) and cats (Roth, 1980).

B. Contribution of acoustical features to sound localization

The notches appearing in the monaural DTF changed systematically in frequency with both the azimuth and elevation of sound in the frontal hemisphere (Fig. 6), indicating that these spectral notches could be used by the animal for frontal sound field localization. A possible neural mechanism for coding spectral notches in the gerbil is the type III neuron in the gerbil dorsal cochlear nucleus in the monaural auditory pathway. The frequency response map of the type III unit consists of a narrow excitatory frequency band fringed with narrow inhibitory bands, and thus the unit has been suggested as a spectral notch detector (Parsons *et al.*, 2001). On the other hand, the acoustical axis, another index derived from the monaural DTFs, had a narrow and less orderly distribution in space. Thus, we suspect that the monaural acoustical axis provides relatively poor information for sound localization.

The distribution of the ILD peak frequencies of the ILD spectrum was less regular (Fig. 11) than that of the notch frequencies of monaural DTFs (Fig. 6). This implies that the ILD peak frequency is poor at providing sound direction related information. Compared with the monaural acoustical axis, the binaural acoustical axis exhibited a relatively wide distribution of source directions, and showed a somewhat more orderly dependence on frequency (Fig. 13). Thus, the binaural acoustic axis appears to be a viable sound localization cue in the gerbil. The "EI" neurons, which receive excitatory inputs from one ear and inhibitory inputs from the other ear, are likely to be responsible for representing ILD in the brain. In the gerbil, EI neurons have been found in the lateral superior olive (Sanes, 1990) and inferior colliculus (IC) (Brückner and Rübsamen, 1995). The EI neurons are generally tuned to certain best frequencies (BFs). Thus, given an association between frequency and the binaural acoustical axis, source directions may be mapped on the BF of the most active group of EI neurons in the auditory system. It should be noted that the maximum ILD was largest (30–40 dB) for a frequency range around 30 kHz (Fig. 12), and that the range was close to the BF range in which the EI units are most often found in the gerbil IC (Brückner and Rübsamen, 1995).

We found that the changes in DTF, ILD, and ITD due to the change in the animal's posture were relatively minor: appreciable differences were restricted to rear and low-elevation directions. It should be noted, however, that the present data were obtained in a free field for relatively distal sound sources. Thus, the posture effect shown in the present report underestimates the impact of posture-related changes on acoustical cues when the animal is on a substrate (e.g., the ground or floor) as in naturalistic conditions and/or when the sound source is relatively proximal <25 cm) to the animal (Maki and Furukawa, 2003). In such naturalistic conditions, the posture-induced changes in the acoustical properties may play a positive role in sound localization. It is indicative that the gerbil appears to base its visual distance judgment on a posture-related change in visual information, i.e., the motion parallax cues provided by vertical head bobbing (Ellard *et al.*, 1984).

ACKNOWLEDGMENTS

The authors thank Dr. Tatsuhiko Harada at Tokai University for providing them with experimental facilities, and the Interdepartmental Laboratories of Tokai University for assistance in conducting the experiments.

Ågren, G., Zhou, Q., and Zhong, W. (**1989**). "Ecology and social behavior of Mongolian gerbils, *Meriones unguiculatus*, at Xilinhot, Inner Mongolia, China," Anim. Behav. **37**, 11–27.

Behrend, O., Brand, A., Kapfer, C., and Grothe, B. (**2002**). "Auditory response properties in the superior paraolivary nucleus of the gerbil," J. Neurophysiol. **87**, 2915–2928.

Brand, A., Behrend, O., Marquardt. T., McAlpine, D., and Grothe, B. (**2002**). "Precise inhibition is essential for microsecond interaural time difference coding," Nature (London) **417**, 543–547.

Brückner, S., and Rübsamen, R. (**1995**). "Binaural response characteristics in isofrequency sheets of the gerbil inferior colliculus," Hear. Res. **86**, 1–14.

Carlile, S., and Pettigrew, A. G. (**1987**). "Directional properties of the auditory periphery in the guinea pig," Hear. Res. **31**, 111–122.

Carlile, S. (**1990**). "The auditory periphery of the ferret. I: Directional response properties and the pattern of interaural level differences," J. Acoust. Soc. Am. **88**, 2180–2195.

Chan, J. C., and Geisler, C. D. (**1990**). "Estimation of eardrum acoustical pressure and of ear canal length from remote points in the canal," J. Acoust. Soc. Am. **87**, 1237–1247.

Coles, R. B., and Guppy, A. (**1986**). "Biophysical aspects of directional hearing in the Tammar wallaby, *Macropus Eugenii*," J. Exp. Biol. **121**, 371–394.

Ellard, C. G., Goodale, M. A., and Timney, B. (**1984**). "Distance estimation in the Mongolian gerbil: The role of dynamic depth cues," Behav. Brain Res. **14**, 29–39.

Firzlaff, U., and Schuller. G. (**2003**). "Spectral directionality of the external ear of the lesser spear-nosed bat, *Phyllostomus discolor*," Hear. Res. **181**, 27–39.

Fuzessery, Z. M. (**1996**), "Monaural and binaural spectral cues created by the external ears of the pallid bat," Hear. Res. **95**, 1–17.

Grothe, B. (**2003**). "New roles for synaptic inhibition in sound localization," Nat. Rev. Neurosci. 4, 540–550.

Hammershøi, D., and Møller, H. (**1996**). "Sound transmission to and within the human ear canal," J. Acoust. Soc. Am. **100**, 408–427.

Hârper N. S., and McAlpine, D. (**2004**). "Optimal neural population coding of an auditory spatial cue," Nature (London) **430**, 682–686.

Heffner, R. S., and Heffner, H. E. (**1988**). "Sound localization and use of binaural cues by the gerbil (*Meriones unguiculatus*)," Behav. Neurosci. **102**, 422–428.

Irvine, D. R. (**1987**). "Interaural intensity differences in the cat: Changes in sound pressure level at the two ears associated with azimuthal displacements in the frontal horizontal plane," Hear. Res. **26**, 267–286.

Jen, P. H., and Chen, D. M. (**1988**). "Directionality of sound pressure transformation at the pinna of echolocating bats," Hear. Res. **34**, 101–117.

Keller, C. H., Hartung, K., and Takahashi, T. T. (**1998**). "Head-related transfer functions of the barn owl: Measurement and neural responses," Hear. Res. **118**, 13–34.

Kelly, J. B., and Potas, M. (**1986**). "Directional responses to sounds in young gerbils (*Meriones unguiculatus*)," J. Comp. Psychol. **100**, 37–45.

Knudsen, E. I., and Konishi, M (**1979**). "Mechanisms of sound localization in the barn owl (*Tyto alba*)," J. Comp. Physiol. **133**, 13–21.

Kuhn, G. F. (**1977**). "Model for the interaural time differences in the azimuthal plane," J. Acoust. Soc. Am. **62**, 157-167.

Maki, K., and Furukawa, S. (**2003**). "Acoustical cues for sound localization

by gerbils in an ecologically realistic environment," Assoc. Res. Otolaryngol. Abs. **26**, 89.
McAlpine, D., and Grothe, B. (**2003**). "Sound localization and delay lines—do mammals fit the model?," Trends Neurosci. **26**, 347–350.
Mehrgardt, S., and Mellert, V. (**1977**). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. **61**, 1567–1576.
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.
Middlebrooks, J. C., Makous, J. C., and Green, D. M. (**1989**). "Directional sensitivity of sound-pressure levels in the human ear canal," J. Acoust. Soc. Am. **86**, 89–108.
Middlebrooks, J. C., and Green, D. M. (**1990**). "Directional dependence of interaural envelope delays," J. Acoust. Soc. Am. **87**, 2149–2162.
Middlebrooks, J. C., and Green, D. M. (**1991**). "Sound localization by human listeners," Annu. Rev. Psychol. **42**, 135–159.
Moiseff, A. (**1989**). "Binaural disparity cues available to the barn owl for sound localization," J. Acoust. Soc. Am. **59**, 1222–1226.
Musicant, A. D., Chan, J. C., and Hind, J. E. (**1990**). "Direction-dependent spectral properties of cat external ear: New data and cross-species comparisons," J. Acoust. Soc. Am. **87**, 757–781.
Obrist, M. K., Fenton, M. B., Eger, J. L., and Schlegel, P. A. (**1993**). "What ears do for bats: A comparative study of pinna sound pressure transformation in chiroptera," J. Exp. Biol. **180**, 119-152.
Parsons, J. E., and Lim, E., and Voigt, H. F. (**2001**). "Type III units in the gerbil dorsal cochlear nucleus may be spectral notch detectors," Ann. Biomed. Eng. **29**, 887–896.
Phillips, D. P., Calford, M. B., Pettigrew, J. D., Aitkin, L. M., and Semple, M. N. (**1982**). "Directionality of sound pressure transformation at the cat's pinna," Hear. Res. **8**, 13–28.
Ravicz, M. E., Rosowski, J. J., and Voigt, H. F. (**1996**). "Sound-power collection by the auditory periphery of the Mongolian gerbil *Meriones unguiculatus*. II. External-ear radiation impedance and power collection," J. Acoust. Soc. Am. **99**, 3044–3063.
Rice, J. J., May, B. J., Spirou, G. A., and Young, E. D. (**1992**). "Pinna-based spectral cues for sound localization in cat," Hear. Res. **58**, 132–152.
Roth, G. L., Kochhar, R. K., and Hind, J. E. (**1980**). "Interaural time differences: Implications regarding the neurophysiology of sound localization," J. Acoust. Soc. Am. **68**, 1643–1651.
Ryan, A. (**1976**). "Hearing sensitivity of the mongolian gerbil, *Meriones unguiculatus*," J. Acoust. Soc. Am. **59**, 1222–1226.
Sanes, D. H. (**1990**). "An in vitro analysis of sound localization mechanisms in the gerbil lateral superior olive," J. Neurosci. **110**, 3494–3506.
Searle, C. L., Braida, L. D., Cuddy, D. R., and Davis, M. F. (**1975**). "Binaural pinna disparity: Another auditory localization cue," J. Acoust. Soc. Am. **57**, 448–455.
Shaw, E. A. G. (**1982**). "External ear response and sound localization," in *Localization of Sound: Theory and Applications* (Amphora, Groton, CT), pp. 30–41.
Spezio, M. L., Keller, C. H., Marrocco, R. T., and Takahashi, T. T. (**2000**). "Head-related transfer functions of the rhesus monkey," Hear. Res. **144**, 73–88.
Spitzer, M. W., and Semple, M. N. (**1995**). "Neurons sensitive to interaural phase disparity in gerbil superior olive: Diverse monaural and temporal response properties," J. Neurophysiol. **73**, 1668–1690.
Suzuki, Y., Asano, F., Kim, H.-Y., and Sone, T. (**1995**). "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," J. Acoust. Soc. Am. **97**, 1119–1123.
Woodworth, R. S., and Schlosberg, H. (**1962**). *Experimental Psychology* (Holt, Rinehard and Winston, New York), pp. 349–361.
Wotton, J. M., Haresign, T., and Simmons, J. A. (**1995**). "Spatially dependent acoustical cues generated by the external ear of the big brown bat, *Eptesicus fuscus*," J. Acoust. Soc. Am. **98**, 1423—1445.
Young, E. D., Rice, J. J., and Tong, S. C. (**1996**). "Effects of pinna position on head-related transfer functions in the cat," J. Acoust. Soc. Am. **99**, 3064–3076.

Multiresolution spectrotemporal analysis of complex sounds

Taishih Chi,[a] Powen Ru,[b] and Shihab A. Shamma[c]
Center for Auditory and Acoustics Research, Institute for Systems Research Electrical and Computer Engineering Department, University of Maryland, College Park, Maryland 20742

(Received 22 June 2004; revised 2 May 2005; accepted 12 May 2005)

A computational model of auditory analysis is described that is inspired by psychoacoustical and neurophysiological findings in early and central stages of the auditory system. The model provides a unified multiresolution representation of the spectral and temporal features likely critical in the perception of sound. Simplified, more specifically tailored versions of this model have already been validated by successful application in the assessment of speech intelligibility [Elhilali *et al.*, Speech Commun. **41**(2-3), 331–348 (2003); Chi *et al.*, J. Acoust. Soc. Am. **106**, 2719–2732 (1999)] and in explaining the perception of monaural phase sensitivity [R. Carlyon and S. Shamma, J. Acoust. Soc. Am. **114**, 333–348 (2003)]. Here we provide a more complete mathematical formulation of the model, illustrating how complex signals are transformed through various stages of the model, and relating it to comparable existing models of auditory processing. Furthermore, we outline several reconstruction algorithms to resynthesize the sound from the model output so as to evaluate the fidelity of the representation and contribution of different features and cues to the sound percept. [DOI: 10.1121/1.1945807]

PACS number(s): 43.66.Ba, 43.71.An, 43.71.Gv [KWG] Pages: 887–906

I. INTRODUCTION

Cochlear frequency analysis has for decades influenced the development of algorithms and perceptual measures for the analysis and recognition of speech and audio. Examples include the formulation of the articulation index (Kryter, 1962) to estimate the effect of noise on speech intelligibility, and the exploitation of models of psychoacoustical masking for the efficient coding of speech and music (Pan, 1995). However, cochlear analysis of sound and the extraction of the acoustic spectrum in the cochlear nucleus are only the earliest stages in a sequence of substantial transformations of the neural representation of sound as it journeys up to the auditory cortex via the midbrain and thalamus. And, while much is known about the neural correlates of sound pitch, location, loudness, and the representation of the spectral profile in these early stages, the response properties and functional organization in the more central structures of the inferior colliculus, medial geniculate body, and the cortex have only begun to be uncovered relatively recently (deRibaupierre and Rouiller, 1981; Kowalski *et al.*, 1996; Schreiner and Urbas, 1988b; Miller *et al.*, 2002; Lu *et al.*, 2001; Eggermont, 2002; Ulanovsky *et al.* 2003). Consequently, it is less common that one finds ideas from central auditory processing being applied in psychoacoustics (Houtgast, 1989; Dau *et al.*, 1997a, b; Ewert and Dau, 2000; Grimault *et al.*, 2002) and in design of speech and audio processing systems (Arai *et al.*, 1996; Pitton *et al.*, 1996; Greenberg and Kingsbury, 1997; Tchorz and Kollmeier, 1999; Hansen and Kollmeier, 1999; Kleinschmidt *et al.*, 2001; Atlas and Shamma, 2003). Interestingly, the opposite has occurred, that is, numerous useful algorithms and representations that were developed decades ago, based only on engineering intuition, have turned out to be in hindsight grounded on solid auditory neural processing strategies (Hermansky and Morgan, 1994; Atal, 1974).

To exploit the accumulating physiological findings from the central auditory system and from psychoacoustic experiments, it is essential that they be reformulated as mathematical models and signal processing algorithms. To achieve this objective, this paper provides two specific contributions:

(1) It describes a detailed computational model of central auditory processing. Simplified, specifically tailored, versions of this model have already appeared in previous publications from our group where we demonstrated its successful applications in the objective evaluation of speech intelligibility (Elhilali *et al.*, 2003; Chi *et al.*, 1999) and the perception of phase of complex sounds (Carlyon and Shamma, 2003). Here we provide a more complete mathematical formulation of the model, illustrating how complex signals are transformed through various stages of the model, and relating it to comparable existing models of auditory processing. This expanded version of the model is completely consistent with the earlier versions and has been validated to account for the types of signals and distortions considered in earlier publications.
(2) It outlines algorithms for reconstructing the input acoustic signal from its final model outputs. These algorithms are important in that they demonstrate the sufficiency of this model representation by reconstructing faithful replica of the original inputs. They also enable the model to be used in assessing the perceptual significance of various output features, as well as in applications where a modified final acoustic waveform is necessary such as noise suppression for general audio and hearing aids.

[a]Present address: Department of Communication Engineering, National Chiao Tung University, Hsinchu, Taiwan, Republic of China
[b]Present address: Cybernetics InfoTech Inc.
[c]Electronic mail: sas@eng.umd.edu

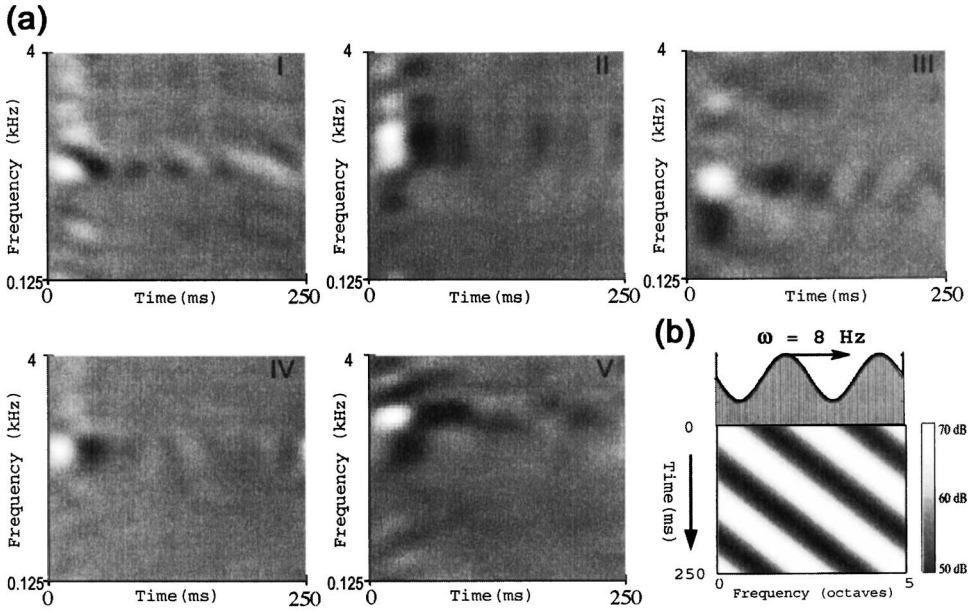


FIG. 1. Details of the dynamic ripple stimulus and examples of spectrotemporal response fields (STRFs) in primary auditory cortex (A1). (a) Example STRFs recorded from A1 of the ferret. White (black) color indicates regions of strongly excitatory (suppressed) responses. The STRFs display a wide range of properties from temporally fast (iv) to slow (iii, v), spectrally sharp (iv) to broad (i, ii), with symmetric (iv) or asymmetric (iii) inhibition. (b) The moving ripple spectral profile $[S(t,x)]$ is defined by the expression: $S(t,x)=1+A\cdot\sin(2\pi\cdot(\omega\cdot t+\Omega\cdot x)+\Phi)$, where A is the modulation depth, Φ is the phase of the profile, ω is called ripple velocity (in Hz), and Ω controls the spectral variation (or modulation)—also called ripple density (in cycles/octave). It usually consists of many simultaneously presented tones, depicted schematically by the vertical lines along the frequency axis. The tones are usually equally spaced along the logarithmic frequency axis and spanning 5 oct (e.g., 0.25–8 kHz or 0.5–16 kHz). The sinusoidal spectral profile $S(t,x)$ is depicted by the dashed curve. The spectrogram of one ripple profile is shown in the bottom panel (Ω=0.4 cycles/octave, ω=8 Hz).

The model we describe is not biophysical in spirit, but rather it abstracts from the physiological data an interpretation that we believe is likely to be relevant in the design of sound engineering systems. Two particularly important physiological observations are incorporated. The first is the apparent progressive loss of temporal dynamics from the periphery to the cortex. Thus, on the auditory nerve, rapid phase locking to individual spectral components of the stimulus survives up to 4–9 kHz. It diminishes to moderate rates of synchrony in the midbrain (under 1 kHz), and to the much lower rates of modulations in the cortex (less than 30 Hz)[1] (Kowalski *et al.*, 1996; Miller *et al.*, 2002; Schreiner and Urbas, 1988a; Langner, 1992). Another important change in the nature of the neural responses is the emergence of elaborate selectivity to combined spectral and temporal features, selectivity that is typically much more complex than the relatively simple tuning curves and dynamics of auditory-nerve fiber responses (Nelken and Versnel, 2000; Shamma *et al.*, 1993; Edamatsu *et al.*, 1989).

The computational model consists of two major auditory transformations. An *early* stage captures monaural processing from the cochlea to the midbrain. It transforms the acoustic stimulus to an auditory time-frequency spectrogramlike representation that combines relatively simple bandpass spectral selectivity with moderate temporal dynamics (<1000 Hz). The second is called the *cortical* stage because it reflects the more complex spectrotemporal analysis presumed to take place in mammalian AI. In the following section, we review the cortical physiological data and psychoacoustical results that motivated and justified this model's development. The mathematical formulation of the early and cortical stages are summarized in Secs. III and IV, together with an illustration of the way in which a variety of complex sounds are represented at each stage. In Sec. V, algorithms to *reconstruct* audible approximate versions of the original sounds from the model's representations are described. We also provide an example of how the reconstructed signals can be used to assess the contribution of different ranges of spectro-temporal modulations to the intelligibility of speech. Finally, we end in Sec. VI with a summary and a brief assessment of the utility of the model in a variety of potential applications.

II. AUDITORY CORTICAL PHYSIOLOGY

The *cortical* stage of the model is strongly inspired by extensive data and ideas gained from physiological and psychoacoustical experiments over the last decade. Specifically, much insight has been gained from measurements of the so-called spectro-temporal response fields (STRF) of AI cells. Examples of a variety of measured STRFs are shown in Fig. 1(a). A STRF summarizes the way a cell responds to the stimulus. Along its ordinate—"frequency axis"—the color white depicts acoustic frequencies that excite responses, black denotes frequencies that suppress (or inhibit) responses, while gray denotes frequency regions of no response. Thus, some STRFs are responsive (excited or suppressed) over a broad range of frequencies, exceeding an octave (**ii**), while others are quite narrowly tuned (**iv**). Along its abscissa—"Time axis"—the STRF depicts the response dynamics to an "impulse" of energy delivered at each frequency. In most STRFs in Fig. 1, the impulse response con-

sists of a damped wave of alternating excitatory (white) and inhibitory (black) responses. The response fades rapidly in some STRFs (**iv**), while it lasts twice as long in others (**v**). Finally, this combined time-frequency sensitivity can take more complex forms that are "inseparable" as in the *oriented* STRFs of **i** and **iii**.

Another way to understand the STRF of a cell is through the implied response selectivity to special test stimuli. STRFs have been measured in many ways (Calhoun and Schreiner, 1995; deCharms *et al.*, 1998), one of which is the "ripple analysis method" (Shamma *et al.*, 1995; Kowalski *et al.*, 1996; Klein *et al.*, 2000). Ripples are broadband noise with sinusoidally modulated spectrotemporal envelopes with different parameters [Fig. 1(b)]. They serve the same function as regular sinusoids in measuring the transfer function of linear filters, except that they are two dimensional (spectral and temporal). AI cells respond well to ripples and are usually selective to a narrow range of ripple parameters that reflect details of their *spectrotemporal transfer functions*. By compiling a complete description of the responses of a cell to all ripple densities and velocities it is possible by an inverse Fourier transform to compute the corresponding STRF.

Therefore, a cell's STRF and its ripple spectrotemporal transfer functions are uniquely related through the Fourier transform. For instance, broadly tuned cells are most responsive to low ripple densities, whereas the opposite is true for narrowly tuned cells. Similarly, STRFs with relatively sluggish dynamics respond poorly to fast ripple rates. Finally, oriented STRFs imply strong selectivity to correspondingly oriented ripples (i.e., of an appropriate rate-density combination). From a functional perspective, the rich variety of STRFs found in AI implies that each STRF acts as a *modulation selective filter* of its input spectrogram, specifically tuned to a particular range of spectral resolutions (also called *scales*) and a limited range of temporal modulations (or *rates*). The collection of all such STRFs then would constitute a filterbank spanning the broad range of psychoacoustically observed scale and rate sensitivity in humans and animals (Viemeister, 1979; Green, 1986; Dau *et al.*, 1997a; Amagai *et al.*, 1999; Chi *et al.*, 1999).

Evidence of the importance of spectrotemporal modulations in the perception of complex sounds has come from experiments in which systematic degradations of the speech signal were correlated with the gradual loss of intelligibility (Drullman *et al.*, 1994; Shannon *et al.*, 1995). All such experiments have consistently pointed to the importance of the slow temporal (<30 Hz) and broad spectral modulations in conveying a robust level of intelligibility (Drullman, 1995; Fu and Shannon, 2000). In fact, the relationship between the temporal modulations and speech intelligibility has long been codified in the formulation of the widely used speech transmission index (STI) (Houtgast *et al.*, 1980). In an extension of such ideas, and inspired by the neurophysiological data briefly reviewed here, we formulated and tested a spectro-temporal modulation index (STMI) (Chi *et al.*, 1999; Elhilali *et al.*, 2003), which assesses the integrity of *both* the spectral and temporal modulations in a signal as a measure of intelligibility. The STMI proved reliable in capturing the deleterious effects of noise and reverberations, as well as of previously difficult to characterize distortions such as nonlinear compression, phase jitter, and phase shifts (Elhilali *et al.*, 2003).

In summary, there is physiological and psychoacoustical evidence that the auditory system, particularly at the level of AI, analyzes the dynamic acoustic spectrum of the stimulus extracted at its earlier stages. It does so by explicitly representing its spectrotemporal modulations by employing arrays of spectrally and temporally selective STRFs. In the remainder of this paper, we elaborate on the mathematical formulation of these computations, and detail a method to invert the representations back to the acoustic stimulus so as to hear the effects of arbitrary manipulations.

III. THE EARLY STAGE: THE AUDITORY SPECTROGRAM

Sound signals undergo a series of transformations in the early auditory system and are converted from a one-dimensional pressure time waveform to a two-dimensional pattern of neural activity distributed along the tonotopic (roughly a logarithmic frequency) axis. This two-dimensional pattern, which we shall call the *auditory spectrogram*, represents an enhanced and noise-robust estimate of the Fourier-based spectrogram (Wang and Shamma, 1994). Details of the biophysical basis and anatomical structures involved are available (Shamma, 1985b; Shamma *et al.*, 1986; Yang *et al.*, 1992).

A. Mathematical formulation

The stages of the early auditory model are illustrated in Fig. 2. In brief, the first operation is an affine wavelet transform of the acoustic signal $s(t)$. It represents the spectral analysis performed by the cochlear filter bank. This analysis stage is implemented by a bank of 128 overlapping constant-Q ($Q_{10dB} \simeq 3$) bandpass filters with center frequencies (CFs) that are uniformly distributed along a logarithmic frequency axis (x), over 5.3 oct (24 filters/octave). The impulse response of each filter[2] is denoted by $h(t;x)$. These cochlear filter outputs $y_{coch}(t,x)$ are transduced into auditory-nerve patterns $y_{AN}(t,x)$ by a hair cell stage consisting of a high-pass filter, a nonlinear compression $g(\cdot)$, and a membrane leakage low-pass filter $w(t)$ accounting for decrease of phase-locking on the auditory nerve beyond 2 kHz. The final transformation simulates the action of a lateral inhibitory network (LIN) postulated to exist in the cochlear nucleus (Shamma, 1989), which effectively enhances the frequency selectivity of the cochlear filter bank (Lyon and Shamma, 1996; Shamma, 1985b). The LIN is simply approximated by a first-order derivative with respect to the tonotopic axis and followed by a half-wave rectifier to produce $y_{LIN}(t,x)$. The final output of this stage is obtained by integrating $y_{LIN}(t,x)$ over a short window, $\mu(t;\tau)=e^{-t/\tau}u(t)$, with time constant $\tau = 8$ ms mimicking the further loss of phase locking observed in the midbrain. The mathematical formulation for this model can be summarized as follows:

$$y_{coch}(t,x) = s(t) \otimes_t h(t;x), \tag{1}$$

$$y_{AN}(t,x) = g(\partial_t y_{coch}(t,x)) \otimes_t w(t), \tag{2}$$

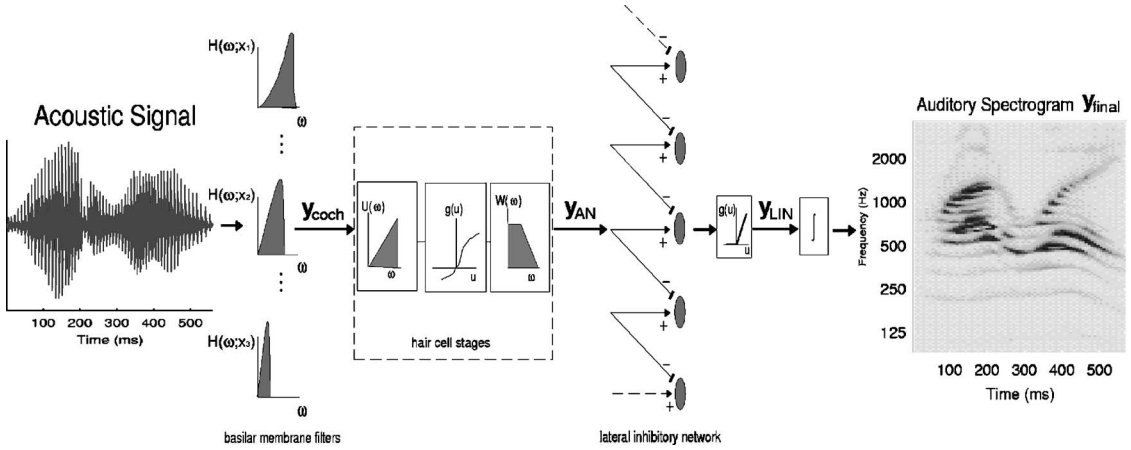


FIG. 2. Schematic of early auditory stages. The acoustic signal is analyzed by a bank of constant-Q cochlear-like filters. The output of each filter (y_{coch}) is processed by a hair cell model (y_{AN}) followed by a lateral inhibitory network, and is finally rectified (y_{LIN}) and integrated to produce the auditory spectrogram (y_{final}).

$$y_{LIN}(t,x) = \max(\partial_x y_{AN}(t,x), 0), \tag{3}$$

$$y_{final}(t,x) = y_{LIN}(t,x) \otimes_t \mu(t;\tau), \tag{4}$$

where $\otimes_t$ denotes convolution operation in the time domain.

The model described above attempts to capture many of the important properties of auditory processing that are critical for our objectives and further detailed in the following sections. In creating such a computational model, one has to balance many conflicting requirements and hence make compromises on what simplifications to apply and what details to include. For instance, our cochlear filtering is essentially linear, lacking such phenomena as two-tone suppression and level-dependent tuning, which are critical in some applications (Carney, 1993). The lateral inhibition model is very schematic and lacks details of single neurons (Shamma, 1989). We also have no explicit adaptive properties in our current model (Westerman and Smith, 1984; Meddis *et al.*, 1990; Dau *et al.*, 1996). All of these details are likely to be important in certain circumstances and should be added when needed (Cohen, 1989).

B. Examples of the auditory spectrogram

Examples of the information preserved at the LIN output $[y_{LIN}(t,x)]$ and midbrain levels $[y_{final}(t,x)]$ of the model are described for five types of progressively more complex stimuli; a three-tone combination, noise, a harmonic complex, ripples, and speech and music segments. Understanding details of the auditory spectrogram $y_{final}(t,x)$ is important since it serves as the input to the cortical analysis stage as we discuss in the next section.

1. Three tones: 250, 1000, and 4000 Hz

Figure 3(a) illustrate the response patterns due to a low-, medium-, and high-frequency tones. The low-frequency tones (250 and 1000 Hz) evoke the typical traveling-wave phase-locked patterns observed experimentally in the auditory nerve (Pfeiffer and Kim, 1975; Shamma, 1985a). For the high-frequency tone, phase locking is lost and only the envelope is preserved. These patterns remain the same at the midbrain stage except that the upper limit of phase locking decreases to below 1000 Hz. Thus, in the right panel of Fig. 3(a), substantial phase locking is only seen for the 250-Hz tone.

2. Noise

Figure 3(b) (left panel) depicts the $y_{final}(t,x)$ generated by a broadband noise constructed with 59 random-phase tones that are equally spaced (0.1 oct) on a logarithmic frequency axis (135–7465 Hz). At this intertone spacing, two to four tones interact within each constant-Q cochlear filter, producing a modulated carrier at the CF of each filter. The envelope modulations at each filter reflect its bandwidth and the intertone spacing in the stimulus. In the low frequency regions (<1000 Hz), the output $[y_{final}(t,x)]$ captures both the carrier and envelope. At higher CF regions, the predominant representation is that of the envelopes as carrier phase-locking diminishes. Note that the modulation rates of the envelope increase (in Hz) with CF as filter bandwidths and stimulus intertone spacing become wider. Maximum rates are limited by maximum filter bandwidths, and hence do not exceed a few hundred Hertz in most mammals (Joris and Yin, 1992).

3. Harmonic complexes

Unlike broadband noise, harmonic complexes have uniform intertone spacing equal to the fundamental frequency of the harmonic series. Consequently, the fundamental component and low-order harmonics remain well resolved by the filters, whereas many high-order harmonics fall within the bandwidth of a cochlear filter at high CFs. Figure 3(b) (middle panel) illustrates the responses to *in-phase* harmonic series stimulus with the fundamental at 80 Hz. Low-order harmonics (<8th) are well resolved (as indicated by the arrows), each dominating the response within one filter, and hence there are little envelope modulations. At high CFs, the unresolved higher harmonics interact, producing the strong 80-Hz periodic envelope modulations. When the harmonics are random phase [Fig. 3(b), right panel], the envelope modulations become irregular and less peaked, but still pre-

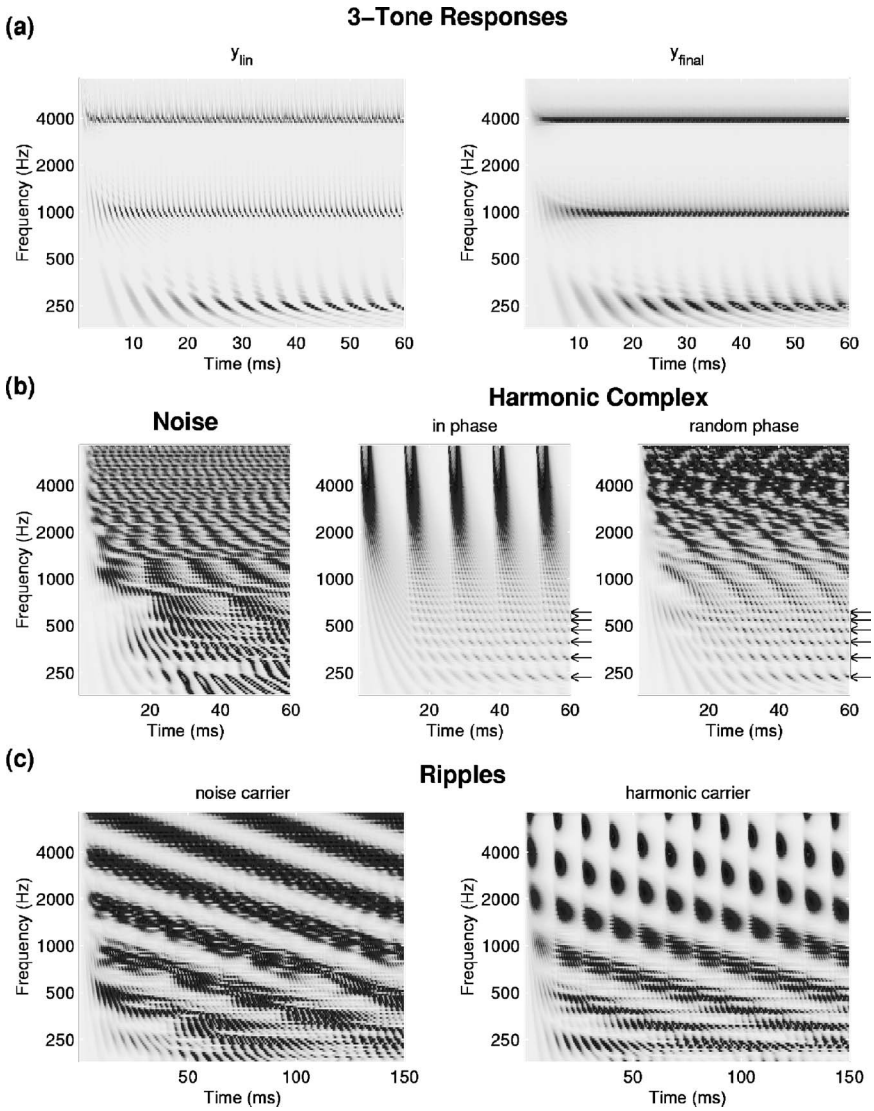


FIG. 3. Examples of early auditory responses for progressively more complex stimuli. (a) A three-tone (250, 1000, 4000 Hz) combination; left panel shows the response at the LIN output $[y_{LIN}(t,x)]$ and right panel shows the response at midbrain level of the model $[y_{final}(t,x)]$. (b) The midbrain output $y_{final}(t,x)$ to a broadband noise (left), broadband in-phase harmonic complex (middle), and a broadband random-phase harmonic complex (right). (c) The $y_{final}(t,x)$ output to a spectro-temporally modulated noise (left) and spectro-temporally modulated in-phase harmonic series (right). All stimuli are sampled at 16 kHz.

serve their periodicity of 80 Hz. The key general observation to make about these envelope modulations is that they relate to intercomponent interactions, and hence are affected by the spacing, phase, and relative amplitudes of the components—factors reflecting the perceptual timber of the sound. In the next two example stimuli, we distinguish these intermediate rate modulations from *slow modulations* created by production mechanisms which, in speech and music, strongly determine the intelligibility of speech and identity of an instrument.

4. Ripples: Spectrotemporally modulated noise

The model's outputs for a spectro-temporal modulated broadband noise—also called a *ripple*—are shown in Fig. 3(c) (left panel). The stimulus is generated by amplitude modulating each of the 59 components of the noise described earlier in Fig. 3(b) (left panel) so as to produce a spectrotemporal profile as depicted in Fig. 1(b). Detailed definition and description of these stimuli can be found in Chi *et al.* (1999) and Kowalski *et al.* (1996).

The left panel of Fig. 3(c) displays the y_{final} output for a downward sweeping ripple (ω=16 Hz, Ω=1 cycle/octave). At low CFs ($\ll$1000 Hz), the responses exhibit temporal modulations at three different time scales simultaneously. The *slow* modulations (16 Hz) reflect the spectrotemporal sinusoidal envelope of the ripple. They ride on top of the *intermediate* modulations due to component interactions (30–400 Hz). These in turn ride on one top of the *fast* responses phase locked to the tones of the stimulus. At high CFs, only the slow and intermediate modulations survive. At very low CFs ($<$250 Hz), slow and intermediate modulation rates may become comparable due to the narrower bandwidths of the filters, and hence the distinct view of the ripple modulations deteriorates.

Figure 3(c) (right panel) illustrates the responses to the *same* ripple spectrotemporal envelope, but this time carried by the harmonic series of Fig. 3(b) (middle panel). The slow modulations are again well represented in the responses, but this time riding on a totally different pattern of intermediate modulations that reflect the 80-Hz periodicity of the fundamental. It is in this sense that we distinguish between these two types of envelope modulations: the intermediate are strictly due to component interactions whereas the slow modulations are superimposed on top and are related to the evolution of the spectrum, e.g., from one syllable to another

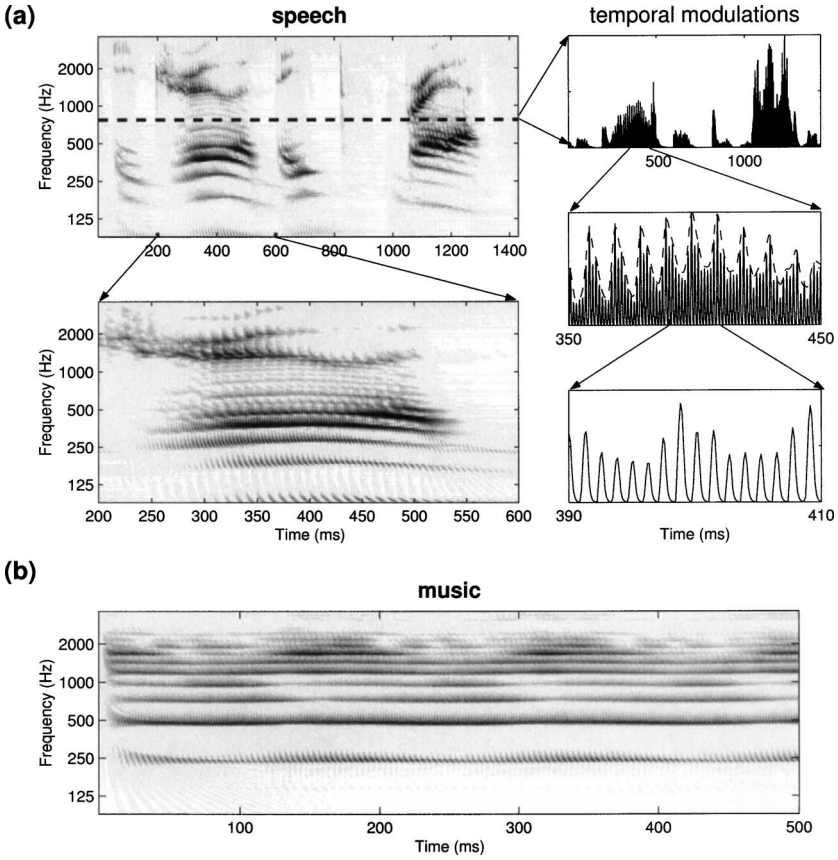


FIG. 4. Examples of auditory spectrograms $[y_{final}(t,x)]$ for speech and music stimuli. (a) The auditory spectrogram of the utterance /He drew a deep breath/ spoken by a male with a pitch of approximately 100 Hz. The dashed line marks the auditory channel at 750 Hz whose temporal modulations are depicted to the right at different time scales. At the coarsest scale (top panel), the slow modulations (few Hz) roughly correlate with the different syllabic segments of the utterance. At an intermediate scale (middle panel), modulations due to interharmonic interactions occur at a rate that reflects the fundamental (100 Hz) of the signal. This is clearly shown by the dashed line envelope of the response. At the finest scale (bottom panel), the fast temporal modulations are due to the frequency component driving this channel best (around 750 Hz). (b) The auditory spectrogram of the note (B3) played on a violin. Again, note the modulations of the energy in time, especially at the higher CF channels (>1500 Hz).

in speech, or from one note or instrument to another in music (see next example).

5. Speech and music

Speech and music are an elaboration of harmonic or noise ripples in that they are conceptually constructed of a spectrotemporal envelope superimposed on a broadband noise or harmonic complex. Figure 4(a) shows the $y_{final}(t,x)$ responses in detail to the utterance /He drew a deep breath/ spoken by a male speaker. Figure 4(b) displays the responses to the B3 note played on a bowed violin. Both responses exhibit similar features to those of the ripple. For example, it is possible to see in Fig. 4(a) the three kinds of *temporal* modulations, as highlighted for one channel (at 750 Hz) in the three right panels. Here the slow modulations that reflect the syllabic rates of speech (top panel) are superimposed upon the intermediate rate modulations due to unresolved harmonics ($\approx$100 Hz) of the fundamental pitch (middle panel), which in turn are riding upon the phase-locked responses to the acoustic energy near 750 Hz (bottom panel). Also evident in the spectrograms are the *spectral* modulations created by the resolved harmonics (<500 Hz), and the second and third formants (>750 Hz). The same types of modulations are seen in the violin sound in Fig. 4(b). Note especially the slow modulations encoding the gradual onset of the note, and the *periodic* modulations at $\approx$6 Hz seen in most channels responses. As in speech, these slow features reflect primarily motor production mechanisms due to the fingering (vibrato) and bowing characteristics.

The distinction between these three types of temporal scales (fast, intermediate, and slow) is essentially identical to one already proposed by Stuart Rosen (Rosen, 1992). In an incisive article, he dissected the acoustic speech waveform into these three time scales and related them to the various auditory and production aspects just as described above. The one point to emphasize here is that the temporal scales defined here are made with respect to the channel responses *after* the early auditory analysis rather than the original acoustic waveform [or as Rosen calls it, the normal hearing case (Rosen, 1992)].

IV. THE CORTICAL STAGE: SPECTROTEMPORAL ANALYSIS

The second analysis stage mimics aspects of the responses of higher central auditory stages (especially the primary auditory cortex). Functionally, this stage estimates the spectral and temporal modulation content of the auditory spectrogram. It does so computationally via a bank of filters that are selective to different spectrotemporal modulation parameters that range from slow to fast *rates* temporally, and from narrow to broad *scales* spectrally. The spectrotemporal receptive fields (STRFs) of these filters are also centered at different frequencies along the tonotopic axis (Chi *et al.*, 1999).

An example of the STRF of such a filter in the bank is shown in Fig. 5(a). Three features are of particular interest: (*i*) it is centered on a particular center frequency (CF). The location of the excitatory (white) and inhibitory (black) stripes on the vertical axis indicates that it is sensitive to

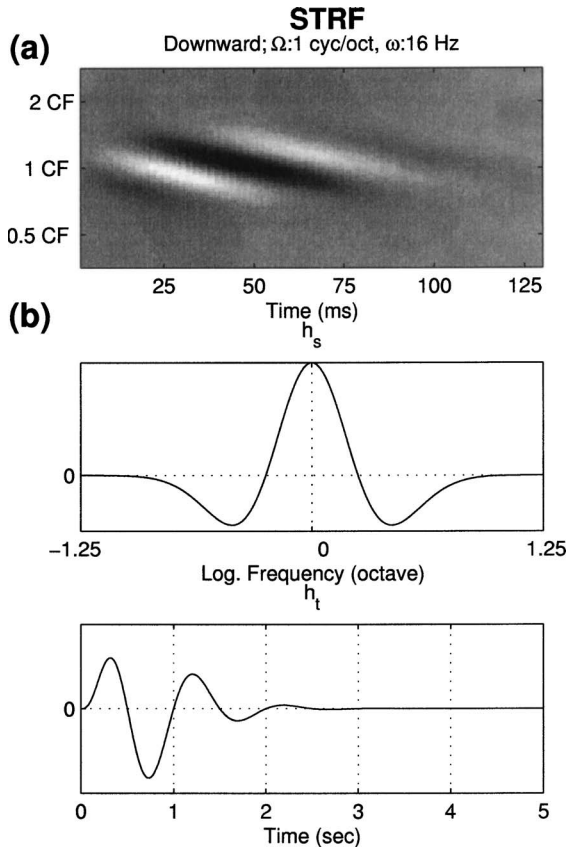


FIG. 5. A representative STRF and the seed functions of the spectrotemporal multiresolution cortical processing model. (a) An example of a STRF in the model. It is upward selective and tuned to (1 cycle/octave, 16 Hz). (b) Seed functions (noncausal h_s and causal h_t) used to generate all STRFs of the model. The abscissa of each figure is normalized to correspond to the tuning scale of 1 cycle/octave or rate of 1 Hz.

frequencies of about a 2-oct range around the CF (between about 0.5 CF and 2 CF); (*ii*) the modulation rate along the time axis is about 16 Hz; and (*iii*) the excitatory portions are separated on the vertical axis by about 1 oct, giving rise to a spectral "scale" sensitivity to peaks separated by 1 oct, or a scale of 1 cycle/octave. Finally, the bars sweep downwards diagonally from the top left, which is denoted in the model by assigning a positive sign to the rate parameter; bars sweeping up from bottom left to top right are designated by negative rate values. This distinction reflects the differential sensitivity of neurons in the auditory cortex to the direction in which spectral peaks move (Depireux *et al.*, 2001).

The filter output is computed by a convolution of its STRF with the input auditory spectrogram $[y_{final}(t,x)]$, i.e., it is a modified spectrogram. Note that the spectral and temporal cross sections of a filter's STRF are typical of a bandpass impulse response in having alternating excitatory (positive) and inhibitory (negative) fields. Consequently, the filter output is large only if the spectrotemporal modulations are commensurate with the rate, scale, and direction of the STRF. That is, each filter will respond best to a narrow range of these modulations. The output of the model consists of a map of the responses across the filter bank, with different stimuli being differentiated by which filters they activate best. The response map provides a unique characterization of the spectrogram, one that is sensitive to the spectral shape and dynamics of the entire stimulus. We now provide a mathematical formulation of the STRFs and procedures to compute, display, and interpret the model outputs.

A. Mathematical formulation

We assume a bank of "idealized" STRFs as depicted in Fig. 5(a). Each STRF is selective to a narrow range of temporal and spectral modulations and is also directionally sensitive to either upward or downward drifting modulations. A complete set of such STRFs with a range of temporal and spectral selectivity (e.g., 1–300 Hz, and 0.25–8 peaks or cycles/octave) would be sufficient to decompose and characterize the modulations in the auditory spectrogram. More realistic complex STRFs can be readily formed by superposition of these basic STRFs.

We define the STRF as a real function that is formed by combining two *complex* functions in a manner consistent with extensive physiological data. Specifically, experimental STRFs are not necessarily time-frequency separable. Instead, we have found that they are almost always so-called "quadrant separable."[3] This requires that the STRF be represented as the real of the product of a complex temporal and a complex spectral "impulse response" function, $h_{IRT}(t)$ and $h_{IRS}(x)$, as follows: $\mathrm{STRF} \equiv \mathcal{R}\{h_{IRT}(t) \cdot h_{IRS}(x)\}$, where

$$h_{IRS}(x;\Omega,\phi) = h_{irs}(x;\Omega,\phi) + j\hat{h}_{irs}(x;\Omega,\phi), \tag{5}$$

$$h_{IRT}(t;\omega,\theta) = h_{irt}(t;\omega,\theta) + j\hat{h}_{irt}(t;\omega,\theta). \tag{6}$$

$\mathcal{R}\{\cdot\}$ denotes the real part, and $h(\cdot)$ and $\hat{h}(\cdot)$ denote Hilbert transform pairs. The real functions $h_{irs}(\cdot)$ and $h_{irt}(\cdot)$ are defined by sinusoidally interpolating seed functions $h_s(\cdot)$, $h_t(\cdot)$ and their Hilbert transforms (Wang and Shamma, 1995),

$$h_{irs}(x;\Omega,\phi) = h_s(x;\Omega)\cos\phi + \hat{h}_s(x;\Omega)\sin\phi, \tag{7}$$

$$h_{irt}(t;\omega,\theta) = h_t(t;\omega)\cos\theta + \hat{h}_t(t;\omega)\sin\theta, \tag{8}$$

where Ω and ω are the spectral density and velocity parameters of the filters; ϕ and θ are characteristic phases; $h_s(\cdot)$ and $h_t(\cdot)$ are the spectral and temporal functions that determine the modulation selectivity of the STRF, and $\hat{h}_s(\cdot)$ and $\hat{h}_t(\cdot)$ are their Hilbert transforms. In addition, the directional sensitivity of the STRF is modeled as

$$\mathrm{STRF}_{\Downarrow} = \mathcal{R}\{h_{IRT}(t) \cdot h_{IRS}(x)\},$$

$$\mathrm{STRF}_{\Uparrow} = \mathcal{R}\{h^{*}_{IRT}(t) \cdot h_{IRS}(x)\},$$

where $*$ denotes the complex conjugate; $\Downarrow$ and $\Uparrow$ denote downward and upward moving direction respectively. Note, the downward STRF shown in Fig. 5(a) is a special case of $\theta = \phi = 0$.

We choose $h_s(\cdot)$ to be a Gabor-like function [commonly used in the vision literature to describe the analogous spatial aspect of a receptive field (Jones and Palmer, 1987)]. It is defined as the second derivative of a Gaussian function; $h_t(\cdot)$ is assumed to be a gamma function [e.g., as in Slaney (1998)]. Both are depicted in Fig. 5(b),

$$h_s(x) = (1 - x^2)e^{-x^2/2},$$

$$h_t(t) = t^2 e^{-3.5t}\sin(2\pi t),$$

and for different scales and rates,

$$h_s(x;\Omega) = \Omega h_s(\Omega x),$$

$$h_t(t;\omega) = \omega h_t(\omega t).$$

Therefore, the STRF in general is an inseparable spectrotemporal function of $h_s(\cdot)$ and $h_t(\cdot)$, with a specific highly constrained spectrotemporal structure known as "quadrant separable."

The spectrotemporal response of a downward (upward) cell c for an input spectrogram $y(t,s)$ is then given by

$$r_{c\Downarrow(\Uparrow)}(t,x;\omega_c,\Omega_c,\theta_c,\phi_c) = y(t,x)\otimes_{tx}\mathcal{R}\{[h_{IRT}^{(*)}(t;\omega_c,\theta_c)\cdot h_{IRS}(x;\Omega_c,\phi_c)]\}, \quad (9)$$

where $\otimes_{tx}$ denotes convolution with respect to both t and x. This multiscale multirate (or *multiresolution spectrotemporal*) response is called "cortical representation." Substituting Eqs. (5)–(8) into Eq. (9), the cortical representation at downward or upward cell c can be rewritten as

$$r_{c\Downarrow}(t,x;\omega_c,\Omega_c,\theta_c,\phi_c) = y(t,x)\otimes_{tx}[(h_t h_s - \hat{h}_t\hat{h}_s)\cos(\theta_c+\phi_c) + (\hat{h}_t h_s + h_t\hat{h}_s)\sin(\theta_c+\phi_c)] \quad (10)$$

and

$$r_{c\Uparrow}(t,x;\omega_c,\Omega_c,\theta_c,\phi_c) = y(t,x)\otimes_{tx}[(h_t h_s + \hat{h}_t\hat{h}_s)\cos(\theta_c-\phi_c) + (\hat{h}_t h_s - h_t\hat{h}_s)\sin(\theta_c-\phi_c)], \quad (11)$$

where $h_t \equiv h_t(t;\omega_c)$ and $h_s \equiv h_s(x;\Omega_c)$ to simplify notation.

A useful reformulation of the response r_c is in terms of the output *magnitude* and *phase* of a two-dimensional complex wavelet transform as follows. Let

$$z_{\Downarrow}(t,x;\omega_c,\Omega_c) = y(t,x)\otimes_{tx}[h_{TW}(t;\omega_c)h_{SW}(x;\Omega_c)] = |z_{\Downarrow}(t,x;\omega_c,\Omega_c)|e^{j\psi_{\Downarrow}(t,x;\omega_c,\Omega_c)}, \quad (12)$$

$$z_{\Uparrow}(t,x;\omega_c,\Omega_c) = y(t,x)\otimes_{tx}[h^*_{TW}(t;\omega_c)h_{SW}(x;\Omega_c)] = |z_{\Uparrow}(t,x;\omega_c,\Omega_c)|e^{j\psi_{\Uparrow}(t,x;\omega_c,\Omega_c)}, \quad (13)$$

with $h_{SW}(\cdot)$ and $h_{TW}(\cdot)$ defined as

$$h_{SW}(x;\Omega_c) = h_s(x;\Omega_c) + j\hat{h}_s(x;\Omega_c), \quad (14)$$

$$h_{TW}(t;\omega_c) = h_t(t;\omega_c) + j\hat{h}_t(t;\omega_c). \quad (15)$$

Substituting Eqs. (14) and (15) into Eqs. (12) and (13) and comparing with Eqs. (10) and (11), the cortical response at cell c can be simplified to

$$r_{c\Downarrow}(t,x;\omega_c,\Omega_c,\theta_c,\phi_c) = \mathcal{R}\{z_{\Downarrow}\}\cos(\theta_c+\phi_c) + \mathcal{I}\{z_{\Downarrow}\}\sin(\theta_c+\phi_c) = |z_{\Downarrow}|\cos(\psi_{\Downarrow}-\theta_c-\phi_c) \quad (16)$$

and

$$r_{c\Uparrow}(t,x;\omega_c,\Omega_c,\theta_c,\phi_c) = \mathcal{R}\{z_{\Uparrow}\}\cos(\theta_c-\phi_c) - \mathcal{I}\{z_{\Uparrow}\}\sin(\theta_c-\phi_c) = |z_{\Uparrow}|\cos(\psi_{\Uparrow}+\theta_c-\phi_c) \quad (17)$$

where $z_{\Downarrow} \equiv z_{\Downarrow}(t,x;\omega_c,\Omega_c)$, $z_{\Uparrow} \equiv z_{\Uparrow}(t,x;\omega_c,\Omega_c)$, $\psi_{\Downarrow} \equiv \psi_{\Downarrow}(t,x;\omega_c,\Omega_c)$, and $\psi_{\Uparrow} \equiv \psi_{\Uparrow}(t,x;\omega_c,\Omega_c)$ for short notation; $\mathcal{R}\{\cdot\}$ and $\mathcal{I}\{\cdot\}$ denote the real part and imaginary part, respectively.

The expressions above show that the cortical model response r_c can be reexpressed in terms of magnitude responses $|z_{\Downarrow}|, |z_{\Uparrow}|$ and phase responses $\psi_{\Downarrow}, \psi_{\Uparrow}$, which are obtained by complex wavelet transform [Eqs. (12) and (13)]. Clearly, the magnitude responses $|z_{\Downarrow}(t,x;\omega_c,\Omega_c)|$ and $|z_{\Uparrow}(t,x;\omega_c,\Omega_c)|$ represent the maximal downward ($\psi_{\Downarrow} = \theta_c + \phi_c$) and upward ($\psi_{\Uparrow} = -\theta_c + \phi_c$) cortical responses at location $(t,x;\omega_c,\Omega_c)$.

B. Examples of cortical representations

Because of the multidimensionality of the cortical response r_c, displaying it in an intuitive manner is not trivial, requiring user judgment as to which dimensional views provide the best insights. We illustrate next a variety of such views for the stimuli discussed earlier in Sec. III.

1. Three tones

Figure 6(a) shows three particularly useful summary views of the cortical responses to the three-tone auditory spectrogram in Fig. 3(a). These three displays are generated by first integrating $|z_{\Downarrow}|, |z_{\Uparrow}|$ over their duration, i.e., removing their dependence on t and becoming three dimensional. Next, to generate each of the 2-D panels in Fig. 6(a), the remaining third variable is integrated out over its domain. For example, in the left panel, the dependence on scale (Ω_c) is removed by integrating all STRF outputs along this dimension, hence emphasizing the representation of temporal modulations (rate) at each CF. Since this stimulus is stationary (sustained tones), it evokes only very low rate outputs ($\omega_c \leqslant 4$ Hz) at each of the three tone frequencies. There is, however, a strong output at x and ω_c of 250 Hz due to the phase-locked responses of this tone [seen in the auditory spectrogram of the stimulus in Fig. 3(a)]; a weaker output due to phase locking is also seen at 1 kHz. Note also that both phase-locked responses are much stronger in the "Downward" panel of the display because of the traveling wave delay evident in the spectrograms of Fig. 3(a).

The center panel of Fig. 6(a) displays the output in the scale-frequency plane, integrating all filter outputs along the rate axis. STRFs with fine resolution relative to the intertone 2-oct spacing (i.e., tuned to $\Omega_c > 0.5$ cycle/octave) respond to each tone separately. STRFs with broad bandwidths ($\Omega_c < 0.5$ cycle/octave) smear the representation of the tones

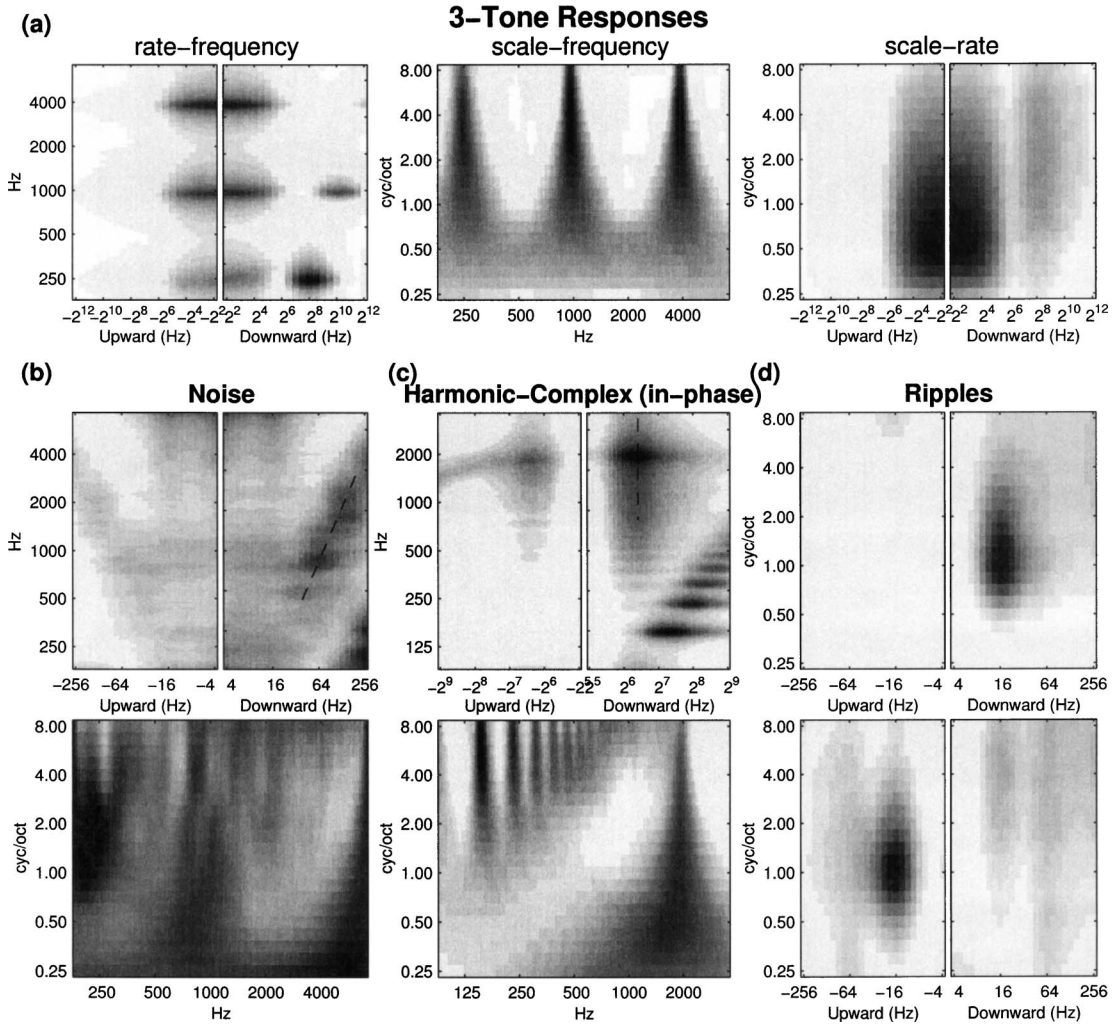


FIG. 6. Examples of cortical representations for stimuli as in Fig. 3: (a) a three-tone (250, 1000, 4000 Hz) combination, (b) a broadband noise, (c) broadband in-phase harmonic complex, and (d) ripples. For each of these stationary stimuli, the four-dimensional representation $|z_{\Downarrow}|$, $|z_{\Uparrow}|$ is first integrated over time to generate a three-dimensional representation. For the three-tone combinations, each of the remaining three variables (scale, rate, frequency) is integrated out over its domain to display these 2-D representations at left, center, and right panels of (a), respectively. For the broadband noise [in (b)] and in-phase harmonic complexes [in (c)], the top and bottom panels demonstrate the rate-frequency and scale-frequency cortical representations. The top and bottom panels of (d) show the scale-rate representations of a downward noise ripple (top) and an upward harmonic ripple (bottom), both modulated at 16 Hz, 1 cycle/octave. In each plot, the negative (positive) rate denotes upward (downward) moving direction.

into one broad peak. A "bifurcation" point emerges around the scale at which the peaks become resolved ($\Omega_c \approx 0.5$).

The right panel is particularly useful in summarizing the conjunction between the temporal and spectral modulations in a spectrogram. As expected, strong response can be seen at very low rate $\leqslant 4$ Hz and at 0.5 cycle/octave (since the tones are separated by 2 oct). A strong 250-Hz phase-locked response is also seen here but has been smeared out along the scale axis. Note, the frequency axis is integrated out, and hence the display is insensitive to pure translations of the spectrum along the x axis.

2. Noise and harmonic complexes

Like the tones, both stimuli here are stationary. However, the drastically different nature of their envelope modulations and underlying spectra creates distinctive cortical outputs as shown in Figs. 6(b) and 6(c).

The noise evokes a rate-frequency response [Fig. 6(b), top panel] which captures the increase in intermediate-rate temporal modulations with increasing CF (marked by the dashed line) due to the increasing bandwidth of the cochlear filters as discussed in Fig. 3(b) earlier. By contrast, the response to the harmonic stimulus [top panel of Fig. 6(c)] is dominated by the phase-locked responses to the resolved low-order harmonics, and all intermediate-rate modulations at high CF ($\geqslant$1000 Hz) occur at a rate=80 Hz (marked by the dashed line). Finally, both panels exhibit larger energy in the "downward"-half of the plot due the accumulating phase lag of the cochlear filters (the well-known "traveling waves").

The scale-frequency panels (bottom panels) of Figs. 6(b) and 6(c) illustrate the contrast between the irregular versus regular nature of the two stimulus spectra. Note especially the distinctive and typical pattern associated with harmonic spectra in which "bifurcation" points shift systematically upwards, indicating the increasing crowding of the higher harmonics along the x axis.

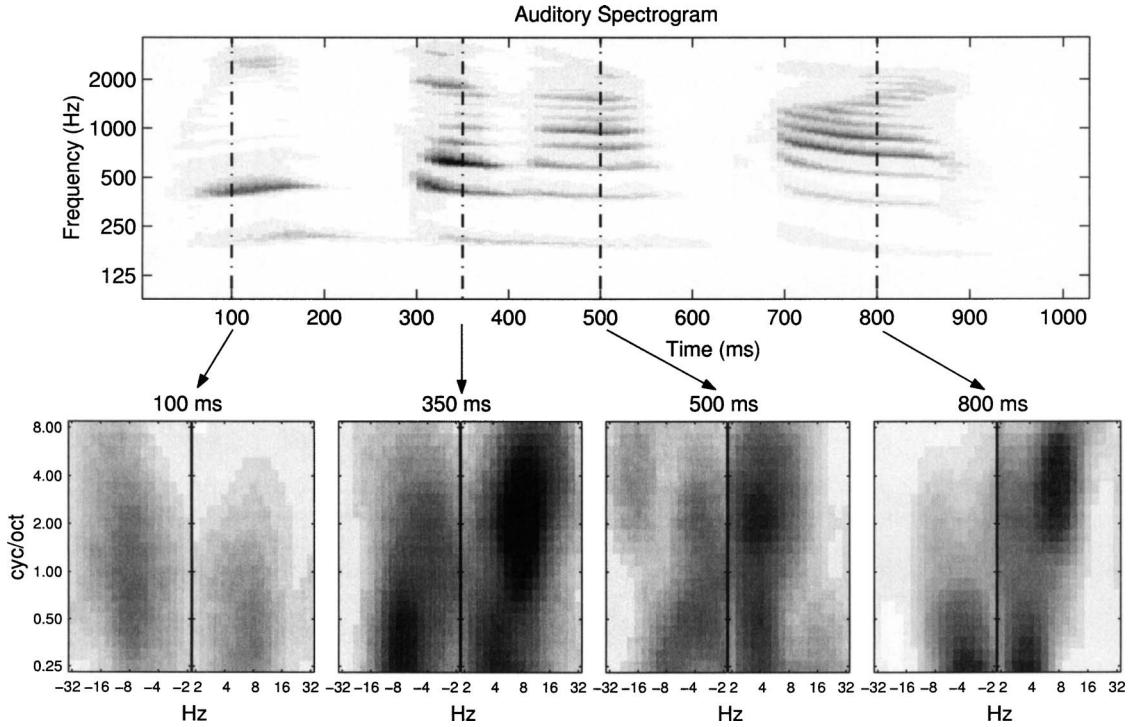


FIG. 7. The cortical multiresolution spectrotemporal representation of speech. The auditory spectrogram of the speech utterance /We've done our part/ spoken by a female speaker. The four bottom panels display scale-rate representation of the model output at the time instants marked by the vertical dashed lines in the auditory spectrogram. Each panel displays the spectrotemporal distribution of responses over the recent past (several 100 ms). For instance, the asymmetric responses at 350 ms reflect the downward shift in the pitch or frequency of all harmonics near the onset of the syllable (300 ms). They peak near 6–10 Hz because of the intersyllable time interval of about 120–180 ms (between the first and second syllables—/we've/ and /done/). They also peak at 2 cycle/octave because most of the spectral energy occurs near the second and third harmonics (which are separated by about 0.5 oct).

3. Ripples

Ripples with a single sinusoidal spectrotemporal modulation activate mostly STRFs with the corresponding selectivity. This is best illustrated by the localized response pattern in the scale-rate views of Fig. 6(d) due to a downward noise ripple (top panel) and an upward harmonic ripple (bottom panel). Regardless of the carrier, both ripples activate a localized response that captures the rate and scale of the slow modulations in the stimulus. Details of other views, however, would distinguish the two ripples from each other.

4. Speech and music

Speech and music are typically nonstationary, with spectrotemporal modulations that change their parameters. Consequently, it is often important to view the time evolution of the response patterns. Figure 7 illustrates one possible representation of the model outputs as a distribution of activity in the scale-rate plane as different phonemes and syllables are analyzed by the model. As before, these panels are computed by first integrating $|z_{\Downarrow}|, |z_{\Uparrow}|$ over frequency x, and then plotting the scale rate as a function of the third axis t.

These plots can uniquely summarize the salient features of the underlying spectrogram, and hence may potentially serve as efficient descriptors of the underlying speech segments. For instance, the downward-sweeping harmonic peaks near 350 and 800 ms generate strongly asymmetric patterns in the second and forth panels. The opposite symmetry is seen near 100 ms where the formant is sweeping upwards. Along the spectral dimensions, the main concentration of energy in the spectrogram shifts upwards from near 500 Hz at 100 ms (second harmonic) to 700 Hz at 350 ms (third harmonic), to near 1000 Hz at 800 ms (fourth and fifth harmonics). Consequently, the concentration of energy along the scale axis (illustrated in the lower series of panels) shifts upwards from near 1 cycle/octave (at 100 ms), to 2 cycles/octave (at 350 ms), to about 4 cycles/octave (at 800 ms). For further examples of such an analysis, please refer to Shamma (2003).

V. RECONSTRUCTION

We derive in this section computational procedures to resynthesize the original input stimulus from the output of early auditory and cortical stages. While the nonlinear operations in the early stage make it impossible to have perfect reconstruction, perceptually acceptable renditions are still feasible as we shall demonstrate. The ability to reconstruct the audio signal from the final representation is extremely useful in building the intuition of the role of different spectrotemporal cues in shaping the timbre percept as we shall elaborate in this section. Furthermore, it provides indirect measure of the fidelity and completeness of the representation as well as a potential means for manipulating timbre of musical instruments, morphing speech, and changing voice quality.

A. Reconstruction from auditory spectrogram

The most important component of the forward analysis stage—the *linear* filter bank operation [Eq. (1)]—is invert-

ible and the inverse operation can be derived as follows (Akansu and Haddad, 1992). From Eq. (1),

$$\begin{aligned} Y_{coch}(\omega,x) &= S(\omega)H(\omega;x) \\ &\Rightarrow \sum_x Y_{coch}(\omega,x)H^*(\omega;x) \\ &= S(\omega)\sum_x H(\omega;x)H^*(\omega;x) \Rightarrow S(\omega) \\ &= \sum_x Y_{coch}(\omega,x)H^*(\omega;x) \Big/ \sum_x |H(\omega;x)|^2, \end{aligned} \tag{18}$$

where $Y_{coch}(\omega,x)$, $S(\omega)$, and $H(\omega;x)$ are the Fourier transforms of $y_{coch}(t,x)$, $s(t)$, and $h(t;x)$ respectively. The overall response of the filter bank, $\Sigma_x|H(\omega;x)|^2$, is flat except at the lowest and highest frequency skirts where it drops precipitously, causing large noise and numerical errors in the inversion procedures. To avoid this problem, we shall simply ignore the response at these extreme frequencies and make the overall response unitary within the remaining band by introducing a real-valued weighting function $W(x)$:

$$H_1(\omega;x) = W(x)H(\omega;x)$$

such that

$$\sum_x |H(\omega;x)|^2 W(x) \simeq 1$$

within the effective band. Therefore, the time waveform $\tilde{s}(t)$ can be computed from the projected filter bank response $\tilde{y}_{coch}(t,x)$ [Eq. (18)]:

$$\begin{aligned} \tilde{S}(\omega) &= \sum_x \tilde{Y}_{coch}(\omega,x)H_1^*(\omega;x), \\ \tilde{s}(t) &= \sum_x \tilde{y}_{coch}(t,x) \otimes_t h_1^*(-t;x) = \sum_x \tilde{y}_{coch}(t,x) \otimes_t h_1(-t;x). \end{aligned} \tag{19}$$

The reconstruction from the envelope $y_{final}(t,x)$ back to $y_{coch}(t,x)$ is difficult to derive directly through the two nonlinear functions $g(\cdot)$ and $\max(\cdot,0)$. Instead, an iterative method based on the *convex projection* algorithm proposed in Yang *et al.* (1992) is used to reconstruct $s(t)$. The method is summarized in the following steps:

(1) Initialize a Gaussian distributed white noise with zero-mean and unit variance, i.e., $\tilde{s}^{(k)}(t) \sim \mathcal{N}(0,1)$, and set the iteration counter $k=1$.
(2) Compute $\tilde{y}_{coch}^{(k)}(t,x)$ and all the way to $\tilde{y}_{final}^{(k)}(t,x)$ with respect to $\tilde{s}^{(k)}(t)$.
(3) Find the ratio $r^{(k)}(t,x)$ between the target $y_{final}(t,x)$ and $\tilde{y}_{final}^{(k)}(t,x)$.
(4) Scale the filter-bank response, i.e., $\tilde{y}_{coch}^{(k)}(t,x) \leftarrow r^{(k)}(t,x) \times \tilde{y}_{coch}^{(k)}(t,x)$.
(5) Reconstruct time waveform $\tilde{s}^{(k+1)}(t)$ by inverse filtering [Eq. (19)], and update counter $k=k+1$.
(6) Go to step 2 unless certain criteria are met [e.g., the distortion rate of $\tilde{y}_{final}^{(k)}(t,x)$ or the number of iteration].

Note, the auditory spectrogram $y_{final}(t,x)$ is assumed roughly representing a local time-frequency (TF) energy distribution, and hence the estimated $\tilde{y}_{coch}(t,x)$ can be adjusted by the ratio of the target $y_{final}(t,x)$ divided by the computed spectrogram $\tilde{y}_{final}(t,x)$ pertaining to $\tilde{y}_{coch}(t,x)$. Figure 8 illustrates the similarity between original and reconstructed auditory spectrograms of two speech utterances after 100 iterations. Note that although this iterative algorithm does not give a unique reconstructed waveform because of the loss of the phase of the original components, the quality of reconstructed sounds using different initial conditions is very close and is reasonably similar to the original signal as can be heard at http://www.isr.umd.edu/CAAR/pubs.html. We shall discuss later in this section an objective assessment of the quality of this reconstructed speech using the mean opinion score (MOS) as quantified by the "perceptual evaluation of speech quality" (PESQ) index available from http://www.itu.int/ under "ITU Publications" (ITU-T, 2001).

B. Reconstruction from the cortical representation

The cortical stage is modeled by a bank of spectrotemporal filters which produce multiscale, multirate (or multiresolution) time-frequency cortical representations from an auditory spectrogram. This linear spectro-temporal filtering process is implemented by a two-dimensional complex wavelet transform [Eqs. (12), (13), (16), and (17)]. This stage is formally identical to the cochlear analysis stage [Eq. (1) versus Eq. (9)], and hence the one-dimensional inverse filtering technique [Eq. (18)] can be extended to solve the inverse problem of two-dimensional cortical filtering process.

The Fourier representations of Eqs. (12) and (13) can be written as

$$Z_{\Downarrow}(\omega,\Omega;\omega_c,\Omega_c) = Y(\omega,\Omega)H_{TW}(\omega;\omega_c)H_{SW}(\Omega;\Omega_c), \tag{20}$$

$$Z_{\Uparrow}(\omega,\Omega;\omega_c,\Omega_c) = Y(\omega,\Omega)H_{TW}^*(-\omega;\omega_c)H_{SW}(\Omega;\Omega_c), \tag{21}$$

and from Eqs. (14) and (15)

$$H_{SW}(\Omega;\Omega_c) = H_s(\Omega;\Omega_c)[1+\mathrm{sgn}(\Omega)], \tag{22}$$

$$H_{TW}(\omega;\omega_c) = H_t(\omega;\omega_c)[1+\mathrm{sgn}(\omega)], \tag{23}$$

where $H_s(\Omega;\Omega_c)$ and $H_t(\omega;\omega_c)$ are the Fourier transform of $h_s(x;\Omega_c)$ and $h_t(t;\omega_c)$, respectively, and

$$\mathrm{sgn}(A) = \begin{cases} 1, & A>0, \\ 0, & A=0, \\ -1, & A<0. \end{cases}$$

Therefore, reconstructing from the cortical representations back to auditory spectrogram is given by

$$\tilde{Y}(\omega,\Omega) = \frac{\Sigma_{\omega_c,\Omega_c} Z_{\Downarrow} H_{TW\Downarrow}^* H_{SW}^* + \Sigma_{\omega_c,\Omega_c} Z_{\Uparrow} H_{TW\Uparrow}^* H_{SW}^*}{\Sigma_{\omega_c,\Omega_c}|H_{TW\Downarrow}H_{SW}|^2 + \Sigma_{\omega_c,\Omega_c}|H_{TW\Uparrow}H_{SW}|^2}, \tag{24}$$

where $Z_{\Downarrow} \equiv Z_{\Downarrow}(\omega,\Omega;\omega_c,\Omega_c)$, $Z_{\Uparrow} \equiv Z_{\Uparrow}(\omega,\Omega;\omega_c,\Omega_c)$, $H_{TW\Downarrow} \equiv H_{TW}(\omega;\omega_c)$, $H_{TW\Uparrow} \equiv H_{TW}^*(-\omega;\omega_c)$, and $H_{SW} \equiv H_{SW}(\Omega;\Omega_c)$ for short notation. With similar considerations given to the lowest and highest frequencies of the overall two-dimensional transfer function, an excellent reconstruction

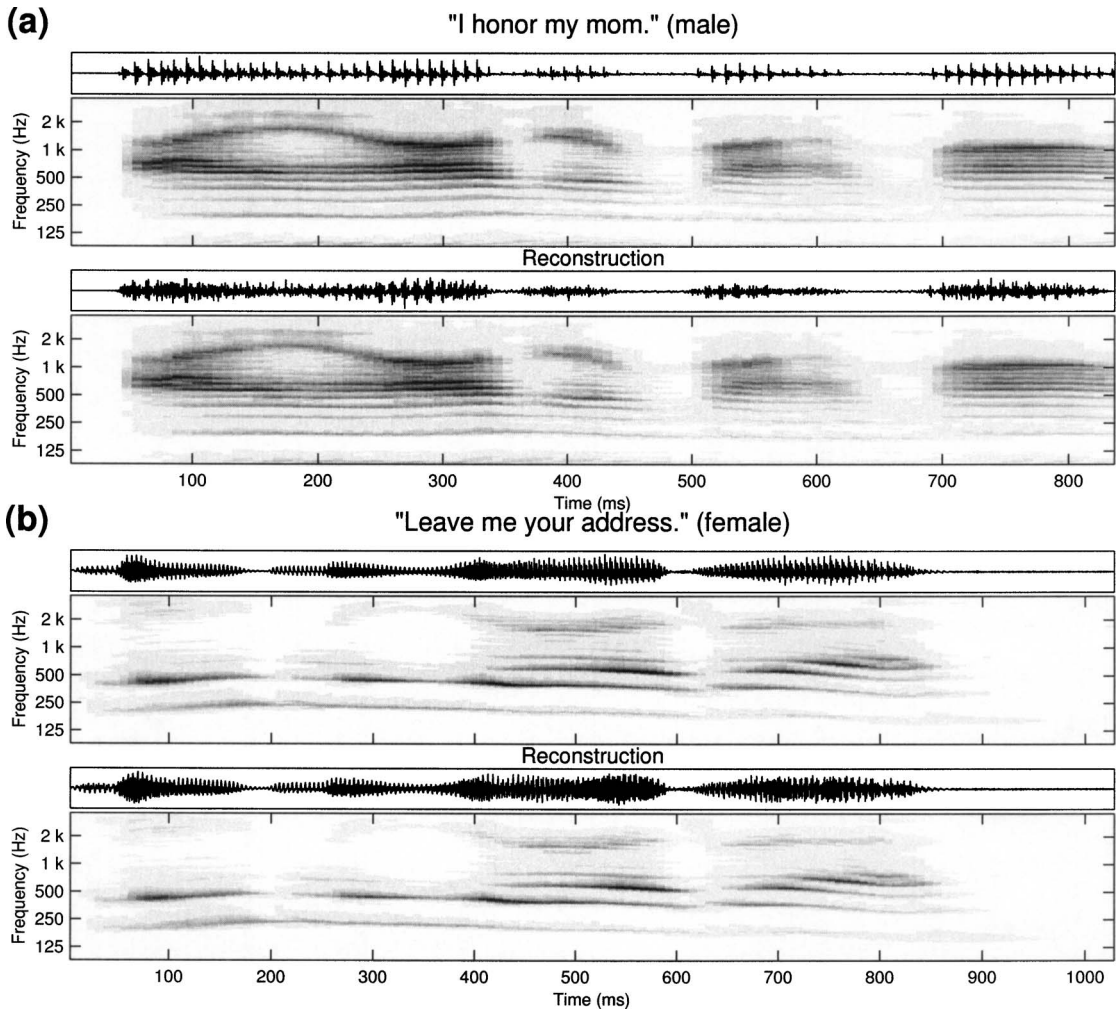


FIG. 8. Two examples of reconstructed acoustic waves from auditory spectrograms: (a) sentence /I honor my mom/ spoken by a male speaker and (b) sentence /Leave me your address/ spoken by a female speaker. The original speech signals are extracted from TIMIT corpus. In each example, the original time waveform $[s(t)]$, the target auditory spectrogram $[y_{final}(t,x)]$, the reconstructed time waveform $[\tilde{s}(t)]$, and the corresponding auditory spectrogram $[\tilde{y}_{final}(t,x)]$ are plotted from top to bottom panels.

within the effective band can be obtained. One example is shown in Fig. 9(b) with the rates up to 32 Hz and scales up to 8 cycles/octave used in the reconstruction. The reconstructed signals can be heard at http://www.isr.umd.edu/CAAR/pubs.html.

It is likely that temporal modulations faster than 20–40 Hz are encoded in the auditory cortex only by their energy distribution or envelope rather than by their actual phase-locked waveforms (Kowalski *et al.*, 1996; Lu *et al.*, 2001). Psychoacoustic experiments and previous models of temporal modulation sensitivity also support this conclusion (Dau *et al.*, 1997b; Sheft and Yost, 1990). Furthermore, in certain applications of the cortical model (Chi *et al.*, 1999), the output magnitude turns out to be an efficient and excellent indicator of the information and percepts of the stimulus. It is therefore useful to demonstrate that the "magnitude" of the response carries sufficient information about the stimulus that generated it. In the Appendix, two algorithms are proposed to reconstruct original speech from the modulation-energy-distributions $[|z_{\Downarrow}|$ and $|z_{\Uparrow}|$ in Eqs. (16) and (17)] only. While the "quality" of the reconstructed signals is worse due to a smaller dynamic range or to propagation of errors in the reconstruction procedures (see the Appendix), they are completely intelligible as can be heard on the website http://www.isr.umd.edu/CAAR/pubs.html.

C. Quality of the reconstructed speech signals

The multiscale auditory model (together with its reconstruction algorithms) can be essentially considered a "coding-decoding" system, and as such we can derive an objective assessment of the "quality" of the reconstructed speech by comparing it to the original clean samples using the standard perceptual evaluation of speech quality (PESQ) metric recommended by ITU (ITU-T, 2001). In this model-based method, we compare samples of clean speech signals to samples reconstructed from the auditory spectrogram and the full cortical representation (magnitude and phase included). The typical PESQ score obtained for the reconstruction from the auditory spectrogram is 4+ (toll quality). For instance, the average of 50 reconstructions of the sentence /Come home right away/ (Fig. 9) starting from different initial conditions and after 200 iterations is 4.04 with $\sigma = 0.075$. The average PESQ score for the reconstruction from

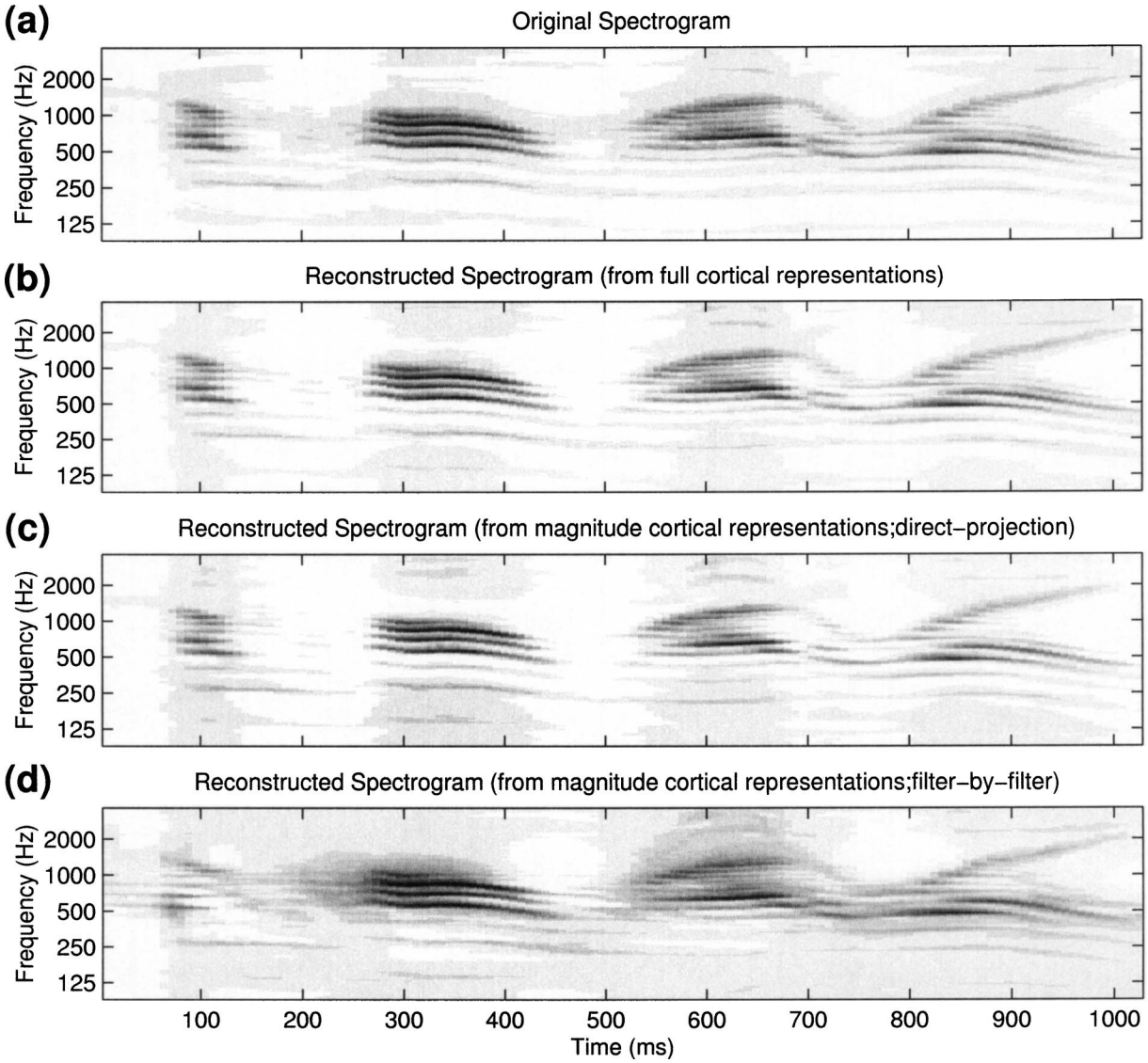


FIG. 9. Examples of reconstructed spectrograms. The top panel shows the original spectrogram of sentence /Come home right away/ spoken by a male speaker. The reconstructed spectrograms from *full* cortical representations and *magnitude* cortical representations (direct-projection and filter-by-filter algorithms) are demonstrated on the second to bottom panel, respectively. All spectrograms are reconstructed from those cortical representations which only include modulation rates up to 32 Hz.

the full cortical representation (with rates up to 128 Hz and scales up to 8 cycles/octave) is 4.02 (toll quality) with $\sigma = 0.069$.

D. Intelligibility of the reconstructed signals

To demonstrate the utility of the reconstructed speech signals from the model, we explore the assertions we made earlier in the Introduction regarding the critical role played by the slow spectrotemporal envelope modulations in preserving intelligibility of the speech signal. Specifically, we use the model to reconstruct a speech sentence after removing from its original version progressively more of its temporal and spectral modulations. We assess in psychoacoustic tests the perceptual effect of such manipulations, and compare the results to the spectrotemporal modulation index (STMI), a measure that was previously demonstrated to be a reliable correlate of human perception of speech intelligibility under a wide variety of interference signals and conditions (Elhilali *et al.*, 2003). We shall specifically employ a particular version of the STMI denoted by STMI^{T} (Elhilali *et al.*, 2003), where the superscript "T" refers to the use of a clean speech signal as the "template" to be compared to each of the "modulation reduced" (or distorted) versions reconstructed from the model.

We first compute the multiscale representation of the clean speech signal through the model [as in Eqs. (16) and (17), $\forall c$]. Temporal modulations are then filtered out by nulling the outputs of the undesired filters (parametrized by their center modulation rates ω_c and Ω_c). This "filtered" representation is then inverted to reconstruct the corresponding "modulation reduced" acoustic signal (as explained in Sec. V B). Figure 10(a) shows the STMI^{T} of the reconstructed speech as a function of the upper limit of *temporal* modulation rates (dashed line). Rates along the abscissa refer to the ω_c's of the cortical filters that are nulled in the STMI^{T} computations. Since the filters are fairly broad, these rates are

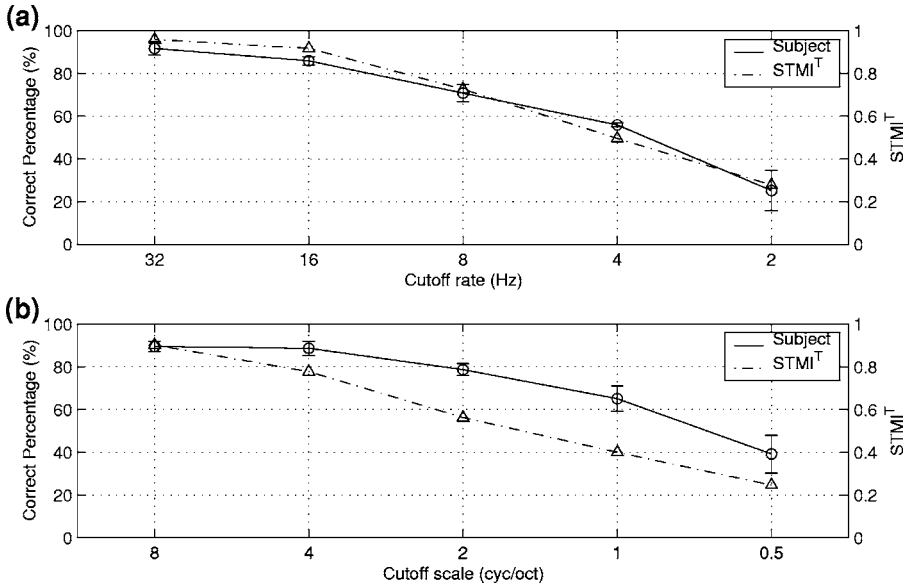


FIG. 10. The spectro-temporal modulation index ($STMI^T$) (Elhilali *et al.*, 2003) of reconstructed speech as a function of the range of spectral and temporal modulations preserved in the signal. (a) The $STMI^T$ (dashed line) and the experimental measurements of the correct phoneme recognition percentage of human subjects (solid line) as a function of the range of temporal modulations preserved. (b) The $STMI^T$ (dashed line) and the human performance (solid line) as function of the scales preserved.

gradual. Each value of the $STMI^T$ shown in the plot is the average of 20 sentences (a mix of males and females) extracted from the TIMIT corpus (the training portion of the New England dialect region). It is evident that intelligibility becomes marginal when temporal modulations around 4 Hz are filtered out, consistent with numerous previous experimental results (Elhilali *et al.*, 2003; Drullman *et al.*, 1994). These results are consistent with the average intelligibility scores measured with four native speakers. In these tests, each subject was to identify 300 reconstructed CVC word samples [see Elhilali *et al.* (2003) for experiment details]. The average percentage "correct phonemes" and the error bars with one standard deviation ranges are plotted in Fig. 10(a).

Figure 10(b) illustrates the $STMI^T$ and intelligibility scores obtained when the spectral profiles of the speech sentence are smoothed by removing progressively higher scales. While the $STMI^T$ and subjects' performance deviate from each other, the overall results confirm that the loss of spectrally sharp features diminishes intelligibility gradually beginning when the filters are effectively wider than about the critical bandwidth (3 cycles/octave). Some intelligibility remains even with filters as broad as 0.5–1 cycles/octave (or about an octave), consistent with previous experimental findings (Shannon *et al.*, 1995).

VI. DISCUSSION

We presented a model of auditory processing that transforms an acoustic signal into a multiresolution spectrotemporal representation inspired by experimental findings from the auditory cortex. The model consists of two major transformations of the acoustic signal:

(1) A frequency analysis stage associated with the cochlea, cochlear nucleus, and response features observed in the midbrain: This stage effectively computes an affine wavelet transform of the acoustic signal with a spectral resolution of about 10% (Lyon and Shamma, 1996).

(2) A spectrotemporal multiresolution analysis stage postulated to conclude in the primary auditory cortex: This stage effectively computes a two-dimensional affine wavelet transform with a Gabor-like spectrotemporal mother-wavelet [see Fig. 5(b)].

The model is intended to be a computational realization of the most basic aspects of auditory processing and not a biophysical description of its stages. Hence, there is only a loose correspondence between any specific structure and model parameters. However, we hypothesize that the model final representation of the acoustic signal captures explicitly and quantitatively the spectral and dynamic aspects that are directly perceived by a listener. Consequently, this representation may be utilized to account for a variety of phenomena, especially those related to the perception of timbre, such as in the assessment of speech quality and intelligibility (Elhilali *et al.*, 2003; Chi *et al.*, 1999), discrimination of musical timbre (Ru and Shamma, 1997), and, more generally, quantifying the perception of complex sounds subjected to arbitrary spectral and temporal changes (Carlyon and Shamma, 2003).

The spirit of this model shares much with others that have been proposed to quantify the perceptual relevance of temporal modulations in acoustic signals (Dau *et al.*, 1997a; Sheft and Yost, 1990; Houtgast, 1989; Bacon and Grantham, 1989; Viemeister, 1979). Dau and colleagues developed the most detailed of these models, consisting of a bank of purely temporal modulation selective filters. They also established its parameters and perceptual relevance in a series of extensive psychoacoustic experiments (Dau *et al.*, 1997a b). Our model is consistent with Dau's model in the details of its analysis of temporal modulations, e.g., possessing similar filter bandwidths in the modulation filterbank ($Q_{3dB}=1.8$ versus $Q_{3dB}=2$ in Dau's model). The two models fundamentally differ in the way temporal modulations from different spectral channels are integrated at the end. Dau's model is *fully separable*, integrating spectral information subsequent to an independent temporal analysis. By contrast the multiscale

cortical model is inseparable (but see footnote 3), postulating a "spectral" modulation filterbank that is fully integrated with the temporal modulation analysis. Under circumstances where *both* temporal and spectral features of the input spectrograms are manipulated [e.g., as in phase jitter or phase shift distortions described in Elhilali *et al.* (2003)], the two models respond differently.

A. Variations on the cortical model

As with the early auditory stage, the multiresolution cortical model is highly schematic and lacks realistic biophysical mechanisms and parameters. Nevertheless, the model aims to capture perceptually significant features in the auditory spectrogram, and hence justify its relevance through its successful application in accounting for a variety of perceptual thresholds and tasks as we have described above.

Many details of the model are somewhat arbitrary and can be probably modified to reflect future physiological and anatomical findings with no significant effect on the computations. For example, real cortical STRFs (Fig. 1) are far more complex than the simple Gabor-like shapes we have employed in the model. They are often tuned to multiple frequencies and are rarely purely selective to upward or downward frequency sweeps but rather are simply more responsive to one direction or the other. In many situations, these differences are not crucial as long as important spectrogram features (e.g., FM sweeps and AM modulations) are still encoded explicitly albeit in a different form.

One potentially interesting variation on our model is to split the spectrotemporal modulation analysis into two stages. The first would be a relatively fast bank of filters mimicking the temporal analysis hypothesized to exist in the inferior colliculus (Langner and Schreiner, 1988) (rates of 30–1000 Hz). The second stage would be slower filters ($\leqslant$30 Hz) operating on *each* output from the earlier stage. This latter stage would then capture all the important slow modulations of the spectrogram explicitly, whereas the earlier stage extracts the intermediate and fast modulations of the auditory spectrogram. The natural split between the dynamic factors involved in intelligibility (the slow rates found in the cortex) from those involved in sound quality (intermediate rates found precortically) becomes particularly advantageous when considering phenomena that contrast these two rate domains such as the streaming of two sounds based purely on their modulation rates (Roberts *et al.*, 2002; Grimault *et al.*, 2002).

B. Relation to previous reconstruction algorithms

The multiresolution representation and associated reconstruction algorithms presented here differ from previous methods for processing spectral and temporal envelopes in two ways. First, its formulation *combines* the spectral and temporal dimensions compared to the purely spectral (e.g., ter Keurs *et al.*, 1992; Baer and Moore, 1993), purely temporal (e.g., Drullman *et al.*, 1994), or a separable cascade of the two (e.g., Dau *et al.*, 1997b). Second, our reconstruction algorithm starts from a random noise signal without any prior information about the original speech. By contrast, previous experiments usually retained the carrier waveform of the speech in each frequency band (Drullman *et al.*, 1994) or the harmonic structure of the speech in each frame (ter Keurs *et al.*, 1992; Baer and Moore, 1993) and used them to resynthesize the filtered speech by superimposing the newly processed envelopes upon them. These carriers improve the quality of the reconstructed speech, but may contain residual intelligible information (Ghitza, 2001; Smith *et al.*, 2002).

Our algorithms are similar in spirit to Slaney's inversion algorithm (Slaney *et al.*, 1994), which also employs the iterative projection method and disposes of the fine structure in reconstructing the stimulus. The algorithm, however, differs fundamentally in all of its details in that it uses for its two-stage representation the cochleagram from a simpler Gammatone filter bank cochlear model (as opposed to the *early stage*) and the correlogram (as opposed to the *cortical multiscale* representation). Consequently, all the constraints imposed during the iterations are completely different.

C. Applications of the multiscale auditory model

The validity of the auditory model stems from its ability to account for psychoacoustic findings and from its successful application in a variety of perceptual tasks. To this end, we have recently adapted and tested the auditory model in several very different contexts. In the first, the auditory model was used to account for the detection of phase of complex sounds such as phase differences between the envelopes of sounds occupying remote frequency regions, and between the fine structures of partials that interact within a single auditory filter (Carlyon and Shamma, 2003). The approach was simply to interpret the discrimination between two stimuli as being proportional to the distance (or difference) measured between their cortical representation in the model (Tchorz and Kollmeier, 1999). Discriminations successfully accounted for phase differences between pairs of bandpass filtered harmonic complexes, and between pairs of sinusoidally amplitude modulated tones, discrimination between amplitude and frequency modulation, and discrimination of transient signals differing only in their phase spectra ("Huffman sequences") (Carlyon and Shamma, 2003).

In a second application, we used the model to analyze the effects of noise, reverberations, and other distortions on the joint spectrotemporal modulations present in speech, and on the ability of a channel to transmit these modulations (Chi *et al.*, 1999; Elhilali *et al.*, 2003). The rationale behind this approach is that the perception of speech is critically dependent on the faithful representation of spectral and temporal modulations in the auditory spectrogram (Hermansky and Morgan, 1994; Drullman *et al.*, 1994; Shannon *et al.*, 1995; Arai *et al.*, 1996; Dau *et al.*, 1996; Greenberg *et al.*, 1998). Therefore, an intelligibility index which reflects the integrity of these modulations can be effective regardless of the source of the degradation. Such a spectrotemporal modulation index (STMI) was derived using the model representation of speech modulations and was validated by comparing its predictions of intelligibility to those of the classical *speech transmission index (STI)* and to error rates reported by human subjects listening to speech contaminated with

combined noise and reverberation. We further demonstrated that the STMI can handle difficult and nonlinear distortions such as phase jitter and shifts, to which the STI is not sensitive (Elhilali *et al.*, 2003).

In another application, the auditory model was used to discriminate speech from nonspeech signals (Mesgarani *et al.*, 2004), a relatively easy task for humans but one that has been very difficult to reliably automate. The proposed algorithm was largely based on learning a template of the unique representation of speech spectrotemporal modulations, a strategy that proved quite effective when compared to state-of-art alternatives. In a further recent extension of this application, it was possible to use the auditory model as a "filter" to remove "noise" modulations that lie outside of the range typical of speech (Mesgarani and Shamma, 2005). Subsequent reconstruction of the filtered signal demonstrated significant enhancement in sound quality.

VII. SUMMARY AND CONCLUSIONS

An auditory model inspired by existing psychophysical and physiological evidence is described. The first module mimics early auditory processing; it consists of a bank of constant-Q bandpass filters, followed by nonlinear compression and derivative across scale (frequency resolution sharpening) mechanisms, and ending with an envelope detector at each frequency band. The resulting output is an estimate of the spectrogram of the input stimulus with noise-robust and feature-enhanced properties (Wang and Shamma, 1994). The second module further analyzes the auditory spectrogram by a bank of linear spectro-temporal modulation filters, which effectively perform a two-dimensional complex wavelet transform. The result is a multiresolution representation which combines information about the temporal and spectral modulations and their distribution in time and frequency.

Several reconstruction algorithms adapted from convex projection methods are proposed to resynthesize the acoustic signals from the full or just the envelope of the auditory spectrogram and the multiresolution representation. The resynthesized sounds imply that these representations carry information critical to the perception of the timbre and the intelligibility of the sound.

To validate our model, the output representations of the model have been adapted for several applications and show promising results when used to measure the perceptual distance between two sounds (Carlyon and Shamma, 2003) or to assess the intelligibility of speech with various types of linear and nonlinear distortions (Elhilali *et al.*, 2003). In addition, we believe this model can be served as a preprocessor to segregate different auditory cues for sound grouping or streaming applications associated with the field of auditory scene analysis.

The proposed model has been implemented in a MATLAB environment, with a variety of computational and graphical modules to allow the user the flexibility of constructing any appropriate sequence of operations. The package also contains demos and help files for users, together with default parameter settings, making it easy learn for the new user. This software is available for download through our website at http://www.isr.umd.edu/CAAR/under "Publications."

APPENDIX: RECONSTRUCTION FROM MAGNITUDE CORTICAL REPRESENTATION

This restoration-from-magnitude problem (also called the phase retrieval problem) is encountered in many fields (Hayes, 1982; Fienup and Wackerman, 1987). Several approaches have been proposed in the past, including a generalized iterative projection algorithm to solve two-dimensional image restoration problems (Levi and Stark, 1984), reconstructing speech from auditory wavelet transform (Irino and Kawahara, 1993), and the error-reduction and extrapolation algorithms (Gerchberg and Saxton, 1972; Fienup, 1982; Papoulis, 1975). All these algorithms essentially perform iterative Fourier and inverse Fourier transforms between the object and Fourier domain, applying specific constraints in each domain. Mathematical convergence of these iterations is not generally guaranteed (Bates, 1984; Hayes, 1987; Seldin and Fienup, 1990). However, combining different algorithms improves the probability of convergence (Fienup, 1982; Mou-yan and Unbehauen, 1997).

In our case, there are no prescribed magnitude constraints in the Fourier domain (ω-Ω domain). Instead, the input and output (envelope) constraints are in the same time-frequency domain [see Eqs. (12) and (13)]. In general, complex signals (such as $z_{\Downarrow}$ and $z_{\Uparrow}$) cannot be uniquely determined from their modulus ($|z_{\Downarrow}|$ and $|z_{\Uparrow}|$) without additional information. Although the *analytical* form of the cortical filters [Eqs. (14) and (15)] narrows down the range of possible phases to be assigned to a given modulus, the lack of additional constraints about the locations of the poles or zeros of the cortical filters precludes a unique solution to our phase retrieval problem (Hayes *et al.*, 1980). The two algorithms proposed below are iterative and are inspired by traditional phase-retrieval and convex projection algorithms. Although the set of magnitude constraints is not convex, the proposed projections are generalized in the sense of Levi-Stark (Levi and Stark, 1983, 1984) and equivalent to the Gerchberg-Saxton algorithm with error-reduction property (Fienup, 1982). Detailed mathematical descriptions of the proposed projection operators can be found in Chi (2003).

1. Algorithm I: Direct projection

The first algorithm considers magnitude constraints of all filters ($|z_{\Downarrow}(t,x;\omega_c,\Omega_c)|$ and $|z_{\Uparrow}(t,x;\omega_c,\Omega_c)|$, $\forall c$) at the same time. It can be summarized as

(1) Initialize a *non-negative* auditory spectrogram $\tilde{y}^{(k)}(t,x)$ randomly and set the iteration counter $k=1$.
(2) Compute magnitude and phase cortical representations $|\tilde{z}_{\Downarrow}^{(k)}|$, $|\tilde{z}_{\Uparrow}^{(k)}|$, $\tilde{\psi}_{\Downarrow}^{(k)}$ and $\tilde{\psi}_{\Uparrow}^{(k)}$ associated with $\tilde{y}^{(k)}(t,x)$ by cortical filtering process [Eqs. (12) and (13)].
(3) Modify cortical representations by keeping phase $\tilde{\psi}_{\Downarrow}^{(k)}$ and $\tilde{\psi}_{\Uparrow}^{(k)}$ intact but replacing magnitude $|\tilde{z}_{\Downarrow}^{(k)}|$ and $|\tilde{z}_{\Uparrow}^{(k)}|$ with the prescribed magnitude responses $|z_{\Downarrow}|$ and $|z_{\Uparrow}|$ (constraints on the cortical output).

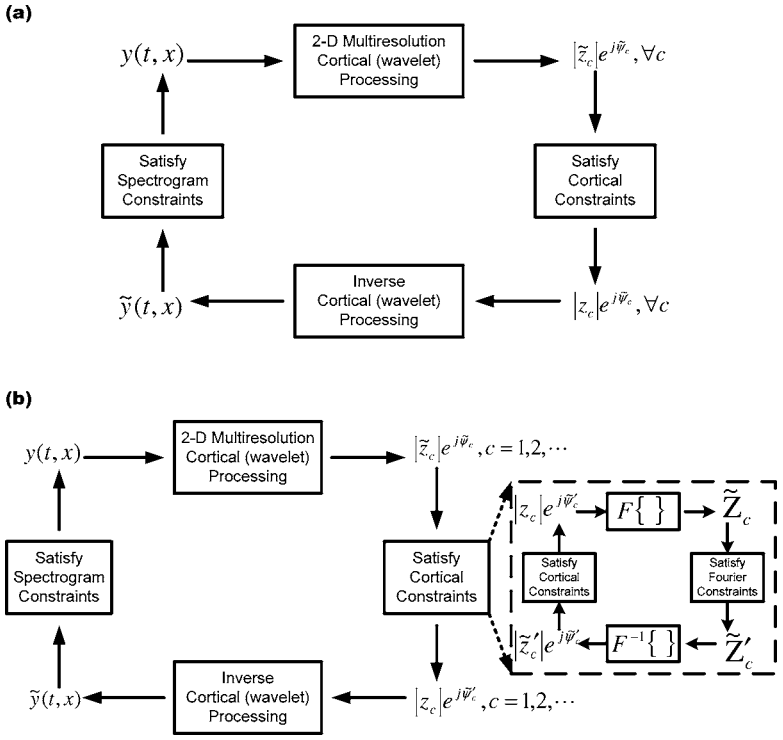


FIG. 11. Block diagrams of two proposed algorithms that reconstruct the spectrograms from magnitude cortical representation. (a) Direct projection algorithm; (b) Filter-by-filter algorithm.

(4) Synthesize $\tilde{y}^{(k+1)}(t,x)$ from modified cortical representations ($|z_{\Downarrow}|$, $|z_{\Uparrow}|$, $\tilde{\psi}_{\Downarrow}^{(k)}$, and $\tilde{\psi}_{\Uparrow}^{(k)}$) by inverse cortical filtering [Eq. (24)].
(5) Half-wave rectify $\tilde{y}^{(k+1)}(t,x)$ (constraints on the cortical input) and update counter $k=k+1$.

Repetitive application of step 2 to step 5 defines the iteration which is depicted in Fig. 11(a). This algorithm has been shown an implementation of the gradient descent search method in solving the nonlinear reconstruction problem (Chi, 2003).

2. Algorithm II: Filter-by-filter

The cortical filters are highly overlapped in both ω and Ω domains, therefore, the magnitude constraints of adjacent filters are redundant. Consequently, Algorithm I yields accurate reconstruction when it converges, but with very high computational cost. Here, taking the analytical form implementation of the cortical filters into account shall reduce the computational load dramatically.

Observed from Eqs. (20)–(23), $Z_{\Downarrow}(\omega,\Omega;\omega_c,\Omega_c)$ and $Z_{\Uparrow}(\omega,\Omega;\omega_c,\Omega_c)$ only have nonzero elements in the first and second quadrants of the (ω,Ω) space, respectively. With these additional implicit constraints and the fact that the frequency responses of the adjacent cortical filters are highly overlapped, the second algorithm is proposed as follows:

(1) Initialize a non-negative auditory spectrogram $\tilde{y}_{(i)}(t,x)$ randomly and set the filter indicator $i=1$.
(2) Compute cortical representations $|\tilde{z}_{\Downarrow}^{(1)}(i)|$, $|\tilde{z}_{\Uparrow}^{(1)}(i)|$, $\tilde{\psi}_{\Downarrow}^{(1)}(i)$ and $\tilde{\psi}_{\Uparrow}^{(1)}(i)$ of filter i, which has the lowest characteristic BF (ω_i,Ω_i) with coverage of DC response $(i=1)$. Here, $|\tilde{z}^{(1)}(i)|$ and $\tilde{\psi}^{(1)}(i)$ are short notations for $|\tilde{z}^{(1)}(t,x;\omega_i,\Omega_i)|$ and $\tilde{\psi}^{(1)}(t,x;\omega_i,\Omega_i)$.
(3) Set iteration counter $k=1$.
   (a) Replace $|\tilde{z}_{\Downarrow}^{(k)}(i)|$, $|\tilde{z}_{\Uparrow}^{(k)}(i)|$ with prescribed $|z_{\Downarrow}(i)|$, $|z_{\Uparrow}(i)|$ and compute $\tilde{Z}_{\Downarrow}^{(k)}(i)$, $\tilde{Z}_{\Uparrow}^{(k)}(i)$ by two-dimensional Fourier transforming $|z_{\Downarrow}(i)|$, $|z_{\Uparrow}(i)|$, $\tilde{\psi}_{\Downarrow}^{(k)}(i)$, and $\tilde{\psi}_{\Uparrow}^{(k)}(i)$.
   (b) Modify $\tilde{Z}_{\Downarrow}^{(k)}(i)$ and $\tilde{Z}_{\Uparrow}^{(k)}(i)$ by keeping the first- and second-quadrant components intact, respectively, and resetting all components in the other quadrants to zero.
   (c) Compute $|\tilde{z}_{\Downarrow}^{(k+1)}(i)|$, $|\tilde{z}_{\Uparrow}^{(k+1)}(i)|$, $\tilde{\psi}_{\Downarrow}^{(k+1)}(i)$, and $\tilde{\psi}_{\Uparrow}^{(k+1)}(i)$ by two-dimensional inverse Fourier transforming modified $\tilde{Z}_{\Downarrow}^{(k)}(i)$ and $\tilde{Z}_{\Uparrow}^{(k)}(i)$.
   (d) Update counter $k=k+1$; go to step 3 (a) when $k<N_i$ (predetermined number of iterations).
(4) Compute $\tilde{y}_{(i+1)}(t,x)$ by Eq. (24) from cortical responses up to filter $i(\tilde{Z}^{(N_i)}(1),\ldots,\tilde{Z}^{(N_i)}(i))$ and half-rectify it (constraint on the cortical input).
(5) Estimate cortical representations ($|\tilde{z}_{\Downarrow}^{(1)}(i+1)|$, $|\tilde{z}_{\Uparrow}^{(1)}(i+1)|$, $\tilde{\psi}_{\Downarrow}^{(1)}(i+1)$ and $\tilde{\psi}_{\Uparrow}^{(1)}(i+1)$) for adjacent filter $i+1$ by cortical forward filtering process [Eqs. (12) and (13)] when $i<N_f$ (number of filters).
(6) Go to step 3 and update filter indicator $i=i+1$.

Note, for each filter i, the starting pattern [initial estimate $\tilde{z}^{(1)}(i)$] shall strongly affect the fidelity of the reconstruction since the generalized projection algorithms do not guarantee a unique solution for nonconvex sets. The block diagram of this filter-by-filter algorithm is depicted in Fig. 11(b).

3. Comparing Algorithms I and II

Algorithm II resolves constraints of one filter at a time (step 3) and thus consumes much less computational time

than Algorithm I. The initial phase for filter i [$\tilde{\psi}_{\Downarrow}^{(1)}(i)$ and $\tilde{\psi}_{\Uparrow}^{(1)}(i)$ in step 3] is estimated recursively from the reconstruction result of previous $i-1$ filters (step 5). This is justified by the assumption that cortical filters have highly overlapped frequency responses, and hence the output phases of one filter and adjacent filters do not change rapidly. However, the overall performance primarily depends on the reconstruction result from the first filter because the errors propagate and are magnified through the iterations. The reconstructed spectrograms from both algorithms are plotted in the bottom two panels of Fig. 9. The processing time of Algorithm I (the third panel from top; 100 iterations) is 150 times longer than the processing time of Algorithm II (bottom panel; $N_i=10$ for each filter). Note, the reconstructed spectrogram at bottom panel shows a smaller dynamic range with apparent distortions near onsets, offsets, lower harmonics, and other weak features in the original spectrogram.

A hybrid algorithm can be used to balance the disadvantages of proposed algorithms, i.e., high computational load and propagation of errors. For example, the output after several iterations of the first direct-projection algorithm is a much better starting pattern than the random pattern to initialize the second algorithm for all filters.

The STMI$^{\mathrm{T}}$s of the reconstructed speech (second to bottom panel) in Fig. 9 are 0.97, 0.97, and 0.91, respectively. These scores indicate that all reconstruction algorithms preserve the slow temporal modulations very well, as can be seen in the figure. However, the quality of the reconstructed sounds is not very good due to distortions as discussed above.

[1]Cortical cells may respond to transient stimuli with high precision (<1 ms), and at times phase lock to high rates exceeding 200 Hz for short intervals. These response patterns reflect the influence of complex mechanisms such as synaptic depression and feedforward inhibition that give rise to the cortical "slow down" in the first place. For details of these phenomena in the auditory cortex, see Elhilali *et al.* (2004).

[2]The cochlear filter is implemented by a minimum-phase signal $h(t)$ with magnitude frequency response

$$|H(x)| = \begin{cases} (x_h - x)^{\alpha} e^{-\beta(x_h - x)}, & 0 \leq x \leq x_h, \\ 0, & x > x_h, \end{cases}$$

where x_h is the cutoff frequency, $\alpha=0.3$, and $\beta=8$. Details of cochlear filter implementations can be found in Ru (2000).

[3]Quadrant-separability implies that in the 2D Fourier transform of the STRF, the temporal and spectral transfer functions are required to be separable only *within each quadrant* (not necessarily across quadrants) (Watson and Ahumada, 1985). This property implies that the STRF temporal cross sections (at different spectral locations) are all composed of the same essential temporal function except for an arbitrary (Hilbert) rotation. In our physiological investigations, we have rarely come across cortical STRFs that violate this property (Depireux *et al.*, 2001). An example of the consequence of such a constraint is that the STRFs cannot be strictly velocity-selective, i.e., respond to any arbitrary spectrum only when it sweeps past at a specific velocity because such STRFs would not be quadrant separable.

Akansu, A. N., and Haddad, R. A. (**1992**). *Multiresolution Signal Decomposition* Academic, Boston.

Amagai, S., Dooling, R., Shamma, S., Kidd, T., and Lohr, B. (**1999**). "Detection of modulation in spectral envelopes and linear-rippled noises by budgerigars," J. Acoust. Soc. Am. **105**, 2029–2035.

Arai, T., Pavel, M., Hermansky, H., and Avendano, C. (**1996**). "Intelligibility of speech with filtered time trajectories of spectral envelopes," *Proc. ICSLP*, pp. 2490–2492.

Atal, B. S. (**1974**). "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identifucation and verification," J. Acoust. Soc. Am. **55**, 1304–1312.

Atlas, L., and Shamma, S. (**2003**). "Joint acoustic and modulation frequency," EURASIP J. Appl. Signal Process. **7**, 668–675.

Bacon, S. P., and Grantham, D. W. (**1989**). "Modulation masking: Effects of modulation frequency, depth, and phase," J. Acoust. Soc. Am. **85**, 2575–2580.

Baer, T., and Moore, B. C. J. (**1993**). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. **94**, 1229–1241.

Bates, R. H. T. (**1984**). "Uniqueness of solutions to two-dimensional fourier phase problems for localized and positive images," Comput. Vis. Graph. Image Process. **25**, 205–217.

Calhoun, B., and Schreiner, C. (**1995**). "Spectral envelope coding in cat primary auditory cortex," J. Aud. Neuroscie. **1**, 39–61.

Carlyon, R., and Shamma, S. (**2003**). "An account of monaural phase sensitivity," J. Acoust. Soc. Am. **114**, 333–348.

Carney, L. H. (**1993**). "A model for the responses of low-frequency auditory-nerve fibers in cat," J. Acoust. Soc. Am. **93**, 401–417.

Chi, T. (**2003**). "Computational Spectro-temporal Auditory Model with Applications to Acoustical Information Processing," Ph.D. thesis, University of Maryland, College Park, MD.

Chi, T., Gao, Y., Guyton, C. G., Ru, P., and Shamma, S. (**1999**). "Spectro-temporal modulation transfer functions and speech intelligibility," J. Acoust. Soc. Am. **106**, 2719–2732.

Cohen, J. R. (**1989**). "Application of an auditory model to speech recognition," J. Acoust. Soc. Am. **85**, 2623–2633.

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., Puschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the effective signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.

deCharms, R. C., Blake, D. T., and Merzenich, M. M. (**1998**). "Optimizing sound features for cortical neurons," Science **280**(5368), 1439–1443.

Depireux, D., Simon, J., Klein, D., and Shamma, S. (**2001**). "Spectro-temporal response field characterization with dynamic ripples in ferret primary auditory cortex," J. Neurophysiol. **85**(3), 1220–1234.

deRibaupierre, F., and Rouiller, E. (**1981**). "Temporal coding of repetitive clicks: presence of rate selective units in the cat's medial geniculate body (mgb)," J. Physiol. (London) **318**, 23–24.

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

Drullman, R., Festen, J., and Plomp, R. (**1994**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

Edamatsu, H., Kawasaki, M., and Suga, N. (**1989**). "Distribution of combination-sensitive neurons in the ventral fringe area of the auditory cortex of the mustached bat," J. Neurophysiol. **61**(1), 202–207.

Eggermont, J. J. (**2002**). "Temporal modulation transfer functions in cat primary auditory cortex: Separating stimulus effects from neural mechanisms," J. Neurophysiol. **87**, 305–321.

Elhilali, M., Chi, T., and Shamma, S. A. (**2003**). "A spectro-temporal modulation index (stmi) for assessment of speech intelligibility," Speech Commun. **41**(2–3), 331–348.

Elhilali, M., Fritz, J. B., Klein, D. J., Simon, J. Z., and Shamma, S. A. (**2004**). "Dynamics of precise spike timing in primary auditory cortex," J. Neurosci. **24**(5), 1159–1172.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.

Fienup, J. R. (**1982**). "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769.

Fienup, J. R., and Wackerman, C. C. (**1987**). "Phase-retrieval stagnation problems and solutions," J. Opt. Soc. Am. A **3**(11), 1897–1907.

Fu, Q.-J., and Shannon, R. V. (**2000**). "Effect of stimulation rate on phoneme recognition by nucleus-22 cochlear implant listeners," J. Acoust. Soc. Am. **107**, 589–597.

Gerchberg, R. W., and Saxton, W. O. (**1972**). "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik (Jena) **35**, 237–246.

Ghitza, O. (**2001**). "On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception," J. Acoust.

Soc. Am. **110**, 1628–1640.
Green, D. M. (**1986**). "Frequency and the detection of spectral shape change," in *Auditory Frequency Selectivity* (Plenum, New York), pp. 351–359.
Greenberg, S., and Kingsbury, B. (**1997**). "The modulation spectrogram: In pursuit of an invariant representation of speech," in *Proc. ICASSP*, pp. 1647–1650.
Greenberg, S., Arai, T., and Silipo, R. (**1998**). "Speech intelligibility derived from exceedingly sparse spectral information," in *Proc. of the Intl. Conf. on Spoken Language Processing*, Sydney, pp. 2803–2806.
Grimault, N., Bacon, S. P., and Micheyl, C. (**2002**). "Auditory stream segregation on the basis of amplitude-modulation rate," J. Acoust. Soc. Am. **111**, 1340–1348.
Hansen, M., and Kollmeier, B. (**1999**). "Continuous assessment of time-varying speech quality," J. Acoust. Soc. Am. **106**, 2888–2899.
Hayes, M. H. (**1982**). "The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-30**(2), 140–154.
Hayes, M. H. (**1987**). "The unique reconstruction of multidimensional sequences from fourier transform magnitude or phase," in *Image Recovery: Theory and Application*, edited by H. Stark (Academic, San Diego), pp. 195–230.
Hayes, M. H., Lim, J. S., and Oppenheim, A. V. (**1980**). "Signal reconstruction from phase or magnitude," IEEE Trans. Acoust., Speech, Signal Process. **ASSP-28**(6), 672–680.
Hermansky, H., and Morgan, N. (**1994**). "Rasta processing of speech," IEEE Trans. Speech Audio Process. **2**(4), 578–589.
Houtgast, T. (**1989**). "Frequency selectivity in amplitude-modulation detection," J. Acoust. Soc. Am. **85**(4), 1676–1680.
Houtgast, T., Steeneken, H. J. M., and Plomp, R. (**1980**). "Predicting speech intelligibility in rooms from the modulation transfer function. i. general room acoustics," Acustica **46**, 60–72.
Irino, T., and Kawahara, H. (**1993**). "Signal reconstruction from modified auditory wavelet transform," IEEE Trans. Signal Process. **41**(12), 3549–3554.
ITU-T (**2001**). "Perceptual evaluation of speech quality (pesq): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," ITU-T Recommendation P.862, February.
Jones, J. P., and Palmer, L. A. (**1987**). "An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex," J. Neurophysiol. **58**(6), 1233–1258.
Joris, P., and Yin, T. C. (**1992**). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J. Acoust. Soc. Am. **91**, 215–232.
Klein, D. J., Depireux, D. A., Simon, J. Z., and Shamma, S. A. (**2000**). "Robust spectro temporal reverse correlation for the auditory system: Optimizing stimulus design," J. Comput. Neurosci. **9**, 85–111.
Kleinschmidt, M., Tchorz, J., and Kollmeier, B. (**2001**). "Combining speech enhancement and auditory feature extraction for robust speech recognition," Speech Commun. **34**(1–2), 75–91.
Kowalski, N., Depireux, D., and Shamma, S. A. (**1996**). "Analysis of dynamic spectra in ferret primary auditory cortex: I. Characteristics of single unit responses to moving ripple spectra," J. Neurophysiol. **76**(5), 3503–3523.
Kryter, K. (**1962**). "Methods for the calculation and use of the articulation index," J. Acoust. Soc. Am. **34**, 1689–2147.
Langner, G. (**1992**). "Periodicity coding in the auditory system," Hear. Res. **60**, 115–142.
Langner, G., and Schreiner, C. E. (**1988**). "Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms," J. Neurophysiol. **60**(6), 1799–1822.
Levi, A., and Stark, H. (**1983**). "Signal restoration from phase by projections onto convex sets," J. Opt. Soc. Am. **73**(6), 810–822.
Levi, A., and Stark, H. (**1984**). "Image restoration by the method of generalized projections with application to restoration from magnitude," J. Opt. Soc. Am. A **1**(9), 932–943.
Lu, T., Liang, L., and Wang, X. (**2001**). "Temporal and rate representations of time-varying signals in the auditory cortex of awake primates," Nat. Neurosci. **11**, 1131–1138.
Lyon, R., and Shamma, S. (**1996**). "Auditory representations of timbre and pitch," in *Auditory Computation*, edited by H. Hawkins, E. T. McMullen, A. Popper, and R. Fay (Springer Verlag, New York), pp. 221–270.
Meddis, R., Hewitt, M. J., and Shackleton, T. M. (**1990**). "Implementation details of a computation model of the inner hair-cell/auditory-nerve synapse," J. Acoust. Soc. Am. **87**, 1813–1816.
Mesgarani, N., and Shamma, S. (**2005**). "Speech enhancement based on filtering the spectrotemporal modulations," in *Proc. ICASSP*. Vol. 1, pp. 1105–1108.
Mesgarani, N., Slaney, M., and Shamma, S. (**2004**). "Discrimination of speech from non-speech based on multiscale spectro-temporal modulations," IEEE Trans. Speech Audio Process. (accepted for publication).
Miller, L. M., Escabi, M. A., Read, H. L., and Schreiner, C. E. (**2002**). "Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex," J. Neurophysiol. **87**(1), 516–527.
Mou-yan, Z., and Unbehauen, R. (**1997**). "Methods for reconstruction of 2-d sequences from fourier transform magnitude," IEEE Trans. Image Process. **6**(2), 222–233.
Nelken, I., and Versnel, H. (**2000**). "Responses to linear and logarithmic frequency-modulated sweeps in ferret primary auditory cortex," Eur. J. Neurosci. **12**(2), 549–562.
Pan, D. (**1995**). "A tutorial on mpeg audio compression," IEEE Multimedia **2**(2), 60–74.
Papoulis, A. (**1975**). "A new algorithm in spectral analysis and band-limited extrapolation," IEEE Trans. Circuits Syst. **CAS-22**(9), 735–742.
Pfeiffer, R. R., and Kim, D. O. (**1975**). "Cochlear nerve fiber responses: distributing along the cochlear partition," J. Acoust. Soc. Am. **58**, 867–869.
Pitton, J. W., Wang, K., and Juang, B.-H. (**1996**). "Time-frequency analysis and auditory modeling for automatic recognition of speech," Proc. IEEE **84**(9), 1199–1215.
Roberts, B., Glasberg, B. R., and Moore, B. C. J. (**2002**). "Primitive stream segregation of tone sequences without differences in fundamental frequency or passband," J. Acoust. Soc. Am. **112**(5), 2074–2085.
Rosen, S. (**1992**). "Temporal information in speech: acoustic, auditory, and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**(10), 367–373.
Ru, P. (**2000**). "Perception-Based Multi-resolution Auditory Processing of Acoustic Signal," Ph.D. thesis, University of Maryland, College Park, MD.
Ru, P., and Shamma, S. A. (**1997**). "Presentation of musical timbre in the auditory cortex," J. New Music Res. **26**(2), 154–169.
Schreiner, C. E., and Urbas, J. V. (**1988a**). "Representation of amplitude modulation in the auditory cortex of the cat. i: The anterior field," Hear. Res. **21**, 227–241.
Schreiner, C. E., and Urbas, J. V. (**1988b**). "Representation of amplitude modulation in the auditory cortex of the cat. ii: Comparison between cortical fields," Hear. Res. **32**, 49–63.
Seldin, J. H., and Fienup, J. R. (**1990**). "Numerical investigation of the uniqueness of phase retrieval," J. Opt. Soc. Am. A **7**(3), 412–427.
Shamma, S. (**2003**). "Physiological foundations of temporal integration in the perception of speech," J. Phonetics **31**, 495–501.
Shamma, S., Chadwick, R., Wilbur, J., Morrish, K., and Rinzel, J. (**1986**). "A biophysical model of cochlear processing: Intensity dependence of pure tone responses," J. Acoust. Soc. Am. **80**, 133–145.
Shamma, S. A. (**1985a**). "Speech processing in the auditory system I: The representation of speech in the response of the auditory nerve," J. Acoust. Soc. Am. **78**, 1612–1621.
Shamma, S. A. (**1985b**). "Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve," J. Acoust. Soc. Am. **78**, 1622–1632.
Shamma, S. A. (**1989**). "Spatial and temporal processing in central auditory networks," in *Methods in Neuronal Modeling*, edited by C. Koch and I. Segev (MIT, Cambridge, MA), pp. 247–289.
Shamma, S. A., Versnel, H., and Kowalski, N. (**1995**). "Ripple analysis in the ferret auditory cortex: I. Response characteristics of single units to sinusoidally rippled spectra," J. Aud. Neurosci. **1**(2), 233–254.
Shamma, S. A., Fleshman, J. W., Wiser, P. R., and Versnal, H. (**1993**). "Organization of the response areas in ferret primary auditory cortex," J. Neurophysiol. **69**(2), 367–383.
Shannon, R. V., Zeng, F.-G., Wygonski, J., Kamath, V., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Sheft, S., and Yost, W. (**1990**). "Temporal integration in amplitude modulation detection," J. Acoust. Soc. Am. **88**, 796–805.
Slaney, M. (**1998**). "Auditory toolbox: Version 2," Technical Report 1998-010, Interval Research Corporation.
Slaney, M., Naar, D., and Lyon, R. F. (**1994**). "Auditory model inversion for sound separation," in *Proc. ICASSP*, Vol. **II**, pp. 77–80.

Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**(6876), 87–90.
Tchorz, J., and Kollmeier, B. (**1999**). "A model of auditory perception as front end for automatic speech recognition," J. Acoust. Soc. Am. **106**, 2040–2050.
ter Keurs, M., Festen, J. M., and Plomp, R. (**1992**). "Effect of spectral envelope smearing on speech reception. I," J. Acoust. Soc. Am. **91**, 2872–2880.
Ulanovsky, N., Las, L., and Nelken, I. (**2003**). "Processing of low-probability sounds by cortical neurons," Nat. Neurosci. **6**, 391–398.
Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.
Wang, K., and Shamma, S. A. (**1994**). "Self-normalization and noise-robustness in early auditory representations," IEEE Trans. Speech Audio Process. **2**(3), 421–435.
Wang, K., and Shamma, S. A. (**1995**). "Representation of spectral profiles in primary auditory cortex," IEEE Trans. Speech Audio Process. **3**(5), 382–395.
Watson, A. B., and Ahumada, A. J. (**1985**). "Model of human visual-motion sensing," J. Opt. Soc. Am. A **2**(2), 322–342.
Westerman, L. A., and Smith, R. L. (**1984**). "Rapid and short term adaptation in auditory nerve responses," Hear. Res. **15**, 249–260.
Yang, X., Wang, K., and Shamma, S. A. (**1992**). "Auditory representations of acoustic signals," IEEE Trans. Inf. Theory **38**(2), 824–839.

A test of the Equal-Loudness-Ratio hypothesis using cross-modality matching functions[a)]

Michael Epstein[b)]
Institute for Hearing, Speech, and Language and Communications and Digital Signal Processing Center, ECE Department (440 DA), Northeastern University, 360 Huntington Avenue, Boston, Massachusetts 02115

Mary Florentine[c)]
Institute for Hearing, Speech, and Language and Department of Speech-Language Pathology and Audiology (133 FR), Northeastern University, 360 Huntington Avenue, Boston, Massachussets 02115

(Received 9 July 2004; revised 13 May 2005; accepted 25 May 2005)

This study tests the Equal-Loudness-Ratio hypothesis [Florentine *et al.*, J. Acoust. Soc. Am. **99**, 1633–1644 (1996)], which states that the loudness ratio between equal-SPL long and short tones is independent of SPL. The amount of temporal integration (i.e., the level difference between equally loud short and long sounds) is maximal at moderate levels. Therefore, the Equal-Loudness-Ratio hypothesis predicts that the loudness function is shallower at moderate levels than at low and high levels. Equal-loudness matches and cross-modality string-length matches were used to assess the form of the loudness function for 5 and 200 ms tones at 1 kHz and the loudness ratio between them. Results from nine normal listeners show that (1) the amount of temporal integration is largest at moderate levels, in agreement with previous studies, and (2) the loudness functions are shallowest at moderate levels. For eight of the nine listeners, the loudness ratio between the 200 and 5 ms tones is approximately constant, except at low levels where it tends to increase. The average data show good agreement between the two methods, but discrepancies are apparent for some individuals. These findings support the Equal-Loudness-Ratio hypothesis, except at low levels. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1954547]

PACS number(s): 43.66.Cb, 43.66.Mk, 43.66.Ba [AJO] Pages: 907–913

I. INTRODUCTION

The Equal-Loudness-Ratio hypothesis states that the loudness ratio between equal-SPL long and short tones is independent of level (Florentine *et al.*, 1996). If the Equal-Loudness-Ratio hypothesis is valid, then loudness functions must be shallower at moderate levels than at low and high levels. This follows because studies have shown that the amount of temporal integration—defined as the level difference between equally loud short and long tones—is greater at moderate levels than at low and high levels (e.g., Florentine *et al.*, 1996, 1998). Support for this notion comes from loudness functions derived from assuming the Equal-Loudness-Ratio hypothesis is true; the loudness functions agree with loudness functions derived from measurements of loudness summation across frequency (Buus, 1999).

To further clarify, if loudness functions for short and long tones are plotted on logarithmic scales with respect to level and are assumed to be parallel, then the horizontal distance between the two functions would indicate the amount of temporal integration at a given level. If a simple power-function model of the growth of loudness is used, temporal integration would have to be constant at all levels because the horizontal distance between the parallel functions would be constant with respect to level. Because measurements of the amount of temporal integration indicate that it varies with level, the slope of the loudness function must also vary in order to create changes in the horizontal distance between the two functions. Specifically, because maximum temporal integration occurs at moderate levels, the slope of the parallel loudness functions must be shallower at those levels.

Some theories, such as the theory of auditory intensity resolution (Durlach and Braida, 1969; Braida and Durlach, 1972) can be applied to attribute the flattening of the loudness function at moderate levels to decisional complexity increases when the test range is large. Ward *et al.* (1996) showed that the slopes of the loudness function decreased when the range was widened using several methods including brightness cross-modality matching. Still, several studies have indicated that loudness functions closely resemble basilar-membrane input/output functions, which are known to be flatter at moderate levels (Buus *et al.*, 1997; Schlauch *et al.*, 1998; Epstein and Florentine, 2005). This makes it unlikely that the mid-level flattening of the loudness functions is only a result of decisional ambiguity resulting from the test range.

The Equal-Loudness-Ratio hypothesis has never been tested directly. Direct measurements of loudness functions for short and long tones are necessary to validate the Equal-Loudness-Ratio hypothesis and its prediction of a shallow

[a)]A portion of this work was presented at the meeting of the International Society for Psychophysics in 2001 [Florentine, M., Epstein, M., and Buus, S. (2001). "Loudness functions for long and short tones," in Fechner Day 2001, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (Pabst, Berlin).]

[b)]Electronic mail: mepstein@ece.neu.edu
[c)]Electronic mail: florentin@neu.edu

mid-level segment of the loudness function. Although direct measurements of loudness functions for short tones have been made using magnitude estimation in four listeners (McFadden, 1975), the data vary widely and agree poorly with other investigators' loudness-balance measurements of temporal integration (see Florentine *et al.*, 1996). Given the dearth and uncertainty of magnitude-estimation data for the loudness of short tones, the present study employs a cross-modality matching procedure to measure loudness functions. Such cross-modality matching procedures have been shown to yield reliable data (Teghtsoonian and Teghtsoonian, 1983; Hellman and Meiselman, 1988, 1990, 1993; Hellman, 1999). Additionally, the present study also employs loudness matches between long and short tones to evaluate the transitivity and reliability of the present cross-modality-matching data with another commonly used procedure.

II. METHOD

A. Stimuli

The stimuli were 1 kHz tones with equivalent rectangular durations of 5 and 200 ms. The levels ranged from 5 dB SL to 100 dB SPL for the 200 ms tones and 110 dB SPL for the 5 ms tones. The 1 kHz test frequency and the durations of 5 and 200 ms were chosen to make the measurements directly comparable with several previous studies (Florentine *et al.*, 1996, 1998; Buus *et al.*, 1999). The tones had a 6.67 ms raised-cosine rise and fall. Durations measured between the half-amplitude points were 1.67 ms longer than the nominal durations. Accordingly, the 5 ms stimuli consisted only of the rise and fall, whereas the 200 ms stimuli had a 195 ms steady-state portion. These envelope shapes ensured that almost all the energy of the tone bursts was contained within the 160-Hz-wide critical band centered at 1 kHz (Scharf, 1970; Zwicker and Fastl, 1990). Even for the 5 ms tone burst, the energy within the critical band was only 0.3 dB less than the overall energy.

B. Apparatus

A PC-compatible computer with a signal processor (TDT AP2) generated the stimuli via a 16 bit D/A converter (TDT DD1) with a 50 kHz sample rate. It also recorded the listeners' responses and executed the adaptive procedure. The output of the D/A converter was attenuated (TDT PA4), lowpass filtered (TDT FT5, fc=20 kHz, 135 dB/octave), attenuated again (TDT PA4), and led to a headphone amplifier (TDT HB6), which fed one earphone of the Sony MDR-V6 headset. This setup ensured that the stimulus level could be controlled linearly by the attenuators over at least a 130 dB range. For routine calibration, the output of the headphone amplifier was led to an A/D converter (TDT DD1), such that the computer could sample the waveform, calculate its spectrum and rms voltage, and display the results before each block of trials.

C. Procedure

The experiment consisted of three parts. In the first part, absolute thresholds were measured for the test stimuli. In the second part, cross-modality matches were made for the 5 and 200 ms tones to obtain loudness functions. In the third part, equal-loudness matches were made to assess the reliability of the cross-modality matches.

1. Absolute thresholds

Absolute thresholds were measured monaurally for 5 and 200 ms tones at 1 kHz using a two-interval, two-alternative forced-choice paradigm with feedback. On each trial, two observation intervals, marked visually, were presented with an interstimulus interval of 500 ms. The stimulus was presented in the first or second observation interval with equal *a priori* probability for each interval. The listener's task was to press one of two buttons corresponding to the interval containing the stimulus. One hundred milliseconds after the listener's response, the correct answer was indicated by a 200 ms light. Following the feedback, the next trial began after a 500 ms delay.

For each listener and duration, three measurements were made using an adaptive method. A single threshold measurement consisted of three interleaved tracks, each of which ended after five reversals. Reversals occurred when the signal level changed from increasing to decreasing or *vice versa*. On each trial, the track was selected at random among the tracks that had not yet ended. For each track, the level of the signal was initially set approximately 15 dB above the listener's threshold. It decreased following three consecutive correct responses and increased following one incorrect response, such that the signal converged on the level yielding 79.4% correct responses (Levitt, 1971). The step size was 5 dB until the second reversal, after which it decreased to 2 dB.

The threshold for each track was calculated as the average signal level of the fourth and fifth reversals and the average of the three tracks was used as one estimate of the absolute threshold. This method has been shown to provide highly reliable measurements of threshold (Hicks and Buus, 2000). Three such estimates (for a total of nine tracks) were obtained for each listener and duration. The average of the three estimates was used as the reference to set the sensation level, SL, for each listener in the remaining portions of the experiment.

2. Cross-modality matching

Each listener was asked to match the length of a string to the loudness of a sound. The listener was given a virtually unbounded ball of very thin, but strong string (i.e., embroidery floss) and was instructed to cut a piece that was as long as the sound was loud following each stimulus presentation. The 5 and 200 ms tones at the various test levels were presented in mixed order. One block of trials contained five trials at each level and duration. To accustom the listener to the task, a single cross-modality match for each stimulus and level was given as training. Then, two blocks of trials were completed such that ten cross-modality matches were made for each level and duration.

The trials were randomized by selecting each new tone level and duration randomly from a set of possibilities that

met the following criteria: the SL needed to be within 30 dB of the level in the previous trial for tones of the same duration and within 25 dB for the other duration. In addition, the stimulus (i.e., a level and duration) must have been presented fewer than five times within the current block of trials. If no stimuli fulfilled these criteria, but some other stimuli still had been presented fewer than five times, a dummy trial was inserted. The dummy trial had the same duration and a level 30 dB above or below the preceding level, depending on the levels of the stimuli that remained to be presented. Such dummy trials were excluded from the data analysis. Large level differences between trials were avoided in order to prevent surprise from a sudden level increase or a missed stimulus from a sudden level decrease.

Each trial was presented in the middle of a 250 ms visually marked interval. After each presentation, the listener cut the string to match the loudness of the presented tone, taped the string segment into a book, turned the page, and pressed a button to indicate completion of the response. The next trial began 700 ms after the listener completed the response. The final cross-modality matches were obtained as the geometric mean of the ten string lengths that were cut to match a given duration and level.

3. Loudness matching

In the final part of the experiment, loudness matches were obtained between 5 and 200 ms tones using a roving-level two-alternative, forced-choice adaptive procedure. This procedure obtained ten concurrent loudness matches by randomly interleaving ten adaptive tracks. Five of these tracks varied the 5 ms tone and five varied the 200 ms tone. The fixed stimulus for each of the five tracks was set to different SLs between 5 and 85 dB in 20 dB steps. (Tracks were eliminated when they exceeded 100 dB SPL for the 200 ms tone or 110 dB SPL for the 5 ms tone.) This procedure ensured that listeners could not identify the stimulus being varied and forced them to use only the two stimuli presented in the current trial to make the loudness judgment (for further discussion, see Buus *et al.*, 1997, 1998).

On each trial, the listener heard two tones separated by a 600 ms interstimulus interval. The fixed-level tone followed the variable tones or the reverse with equal *a priori* probability. The listener's task was to indicate which sound was louder by pressing a key on a response terminal. The next trial began after a 1 s delay. The level of the variable tone was adjusted according to a simple up-down procedure. If the listener indicated that the variable tone was the louder one, its level was reduced, otherwise it was increased. The step size was 5 dB until the second reversal, after which it was 2 dB. This procedure made the variable tone converge toward a level at which it was judged louder than the fixed tone in 50% of the trials (Levitt, 1971).

For each track, the variable stimulus was initially set approximately 15 dB below the expected equal-loudness level. (If this was below threshold, the variable stimulus was set to threshold.) This starting level ensured that the listener would initially hear some trials in which the short tone was definitely louder and some trials in which the long tone was definitely louder. On each trial, the track was chosen at random among those that had not yet ended, which they did after nine reversals. The average level of the last four reversals of each track was used as an estimate of the level at which the loudness of the variable tone was equal to that of the fixed-level tone. Three such loudness matches were obtained for each fixed stimulus and the average was used to estimate the level difference needed for equal loudness. The level difference between short and long tones that are judged to be equally loud will be compared to the level difference between short and long tones that yield equal string lengths.

D. Listeners

A total of nine listeners were tested on all conditions. All listeners had bilaterally normal thresholds and medical histories consistent with normal hearing. They ranged in age from 20 to 46 years. Test ears had audiometric thresholds within 10 dB of ANSI (1989) standard at octave frequencies from 250 to 8000 Hz. Most of the listeners had previous experience making loudness judgments, except for L5 and L7. Listener L4 is the first author.

E. Data analysis

For each listener, one data point was the geometric mean of ten string lengths. The standard error of the mean was determined from the logarithms of the string lengths. The group mean and standard deviations were calculated across the nine individual listeners' geometric means for each duration and SL. The resulting data were transformed back into the string-length domain to show the average and probable range of individual listeners' responses.

To examine the effects of stimulus variables, an analysis of variance (ANOVA) for repeated-measures was performed on the logarithms of the ten string lengths obtained for each listener, SL, and duration. Because the Equal-Loudness-Ratio hypothesis states that the loudness ratio between equal-SPL long and short tones is independent of SPL, the SLs were transformed into approximate "SPLs" for each listener to obtain an indicator of the stimulus level for the ANOVA. However, loudness functions are more similar across listeners when evaluated in terms of SL than in terms of SPL (Hellman and Zwislocki, 1961). In other words, using true SPLs would obscure similarities across listeners. To keep the loudness functions for the different listeners aligned according to SL, while ensuring that the 5 and 200 ms loudness functions were aligned according to approximately equal SPLs, 0 dB SL for the 200 ms tones was equated with 10 dB "SPL" for all listeners. (The true average of all the listeners' thresholds was 9.9 dB SPL.) For the 5 ms tones, each listener's threshold difference between the 5 and 200 ms tones was rounded to the nearest 5 dB and added to 10 dB "SPL" to obtain the "SPL" corresponding to 0 dB SL. For eight of the nine listeners, the rounded threshold difference of 15 dB was between 2 dB higher and 1 dB lower than the true differences. For the ninth listener, the threshold difference was exactly 10 dB. To ensure that the ANOVA encompassed only "SPLs" for which data were available for both 5 and 200 ms

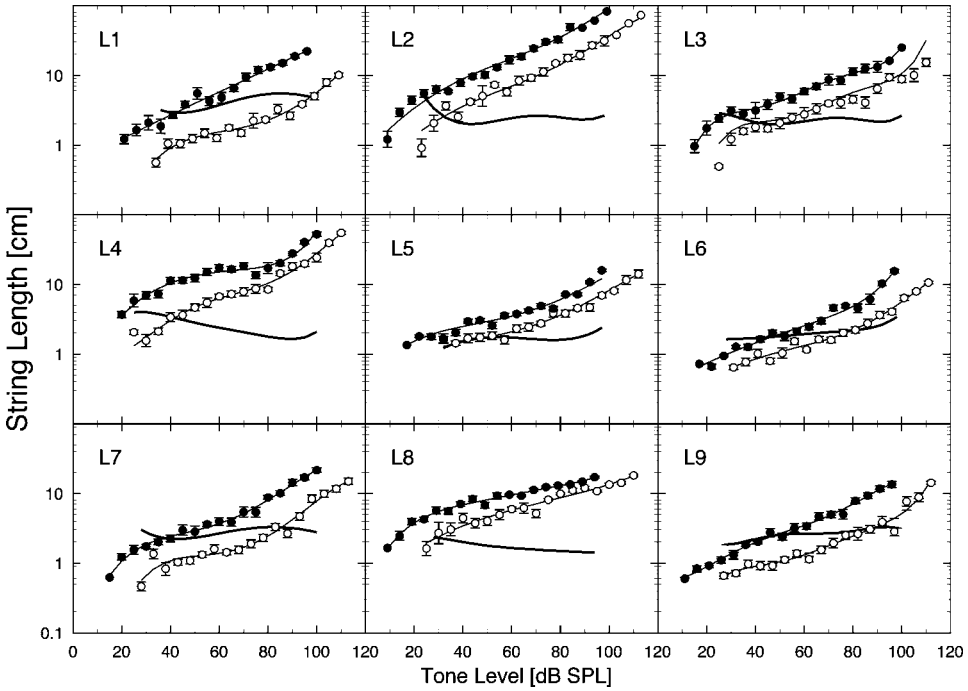


FIG. 1. Individual CMM functions obtained from the nine listeners. The geometric mean of string length is plotted on a log scale as a function of level. The closed circles show data for the 200 ms tones and the open circles show the data for the 5 ms tones. The vertical bars show ± one standard error of the mean. The thin lines show fourth-order polynomials fitted to the data. The thick lines show the ratio of string lengths obtained for equal-SPL 200 and 5 ms tones as estimated from the polynomials.

tones, the analyses encompassed only "SPLs" between 30 and 95 dB. In the following, the sound level corrected in this manner will be referenced simply as SPL.

An initial ANOVA showed that repetition and all interactions with it were not significant factors. Therefore, the main analyses did not include repetition as a factor. The primary ANOVA examined the effects of level and duration. In this analysis, listener was used as a random factor to produce a repeated-measures analysis of variance. For all tests, the outcome was considered significant when $p \leq 0.05$.

III. RESULTS

Figure 1 shows the individual cross-modality-matching (CMM) functions obtained from the nine listeners. The geometric mean of string length is plotted on a log scale as a function of level. The full range of string lengths cut was from 0.1 to 159.1 cm. The data for individual listeners are generally consistent, as indicated by the small standard errors and the general monotonicity. However, there are clear differences among listeners as is typical of measurements of loudness. Despite the individual differences, a few findings are clear. For all levels, the string length matched to the short tone is less than that for the long tone in agreement with classic temporal-integration data (for review, see Florentine *et al.*, 1996). For most of the listeners, the cross-modality-matching functions for both the 5 and 200 ms tones are shallower at moderate levels than at low and high levels. They are also nearly parallel, as indicated by the roughly constant vertical distance between the two functions. The thick, solid lines in Fig. 1, show the ratio of string lengths matched to equal-SPL long and short tones. It is approximately independent of SPL for most of the listeners, although some exceptions are apparent.

The average CMM functions are shown in Fig. 2. They are plotted in the same manner as Fig. 1. The average data show the same general trends as the majority of the individual data. The ratio of string lengths matched to equal-SPL long and short tones is approximately independent of SPL, except for a slight increase below 40 dB SPL. Like the individual data, both loudness functions are shallower at moderate levels than at low and high levels.

These observations are supported by the ANOVA. The effects of SPL and duration are both highly significant ($P < 0.0001$), but the interaction between them is not ($P = 0.68$), as is expected if the effect of duration is independent of SPL. Accordingly, the ANOVA is consistent with the Equal-Loudness-Ratio hypothesis.

Although individual and group data for the CMM functions appear quite orderly, the comparison with the direct loudness matches used to validate them shows considerable variability among the listeners. Figure 3 compares individual listeners' adaptive loudness matches with indirect loudness matches obtained from their CMM data. It shows the amount

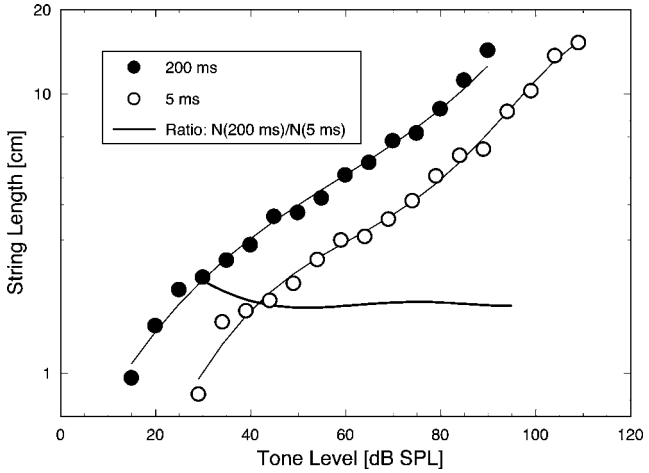


FIG. 2. Mean data and fourth-order polynomials fitted to the data plotted in Fig. 1. The thick line shows the ratio of string lengths obtained for equal-SPL 200 and 5 ms tones as averaged from the polynomials.

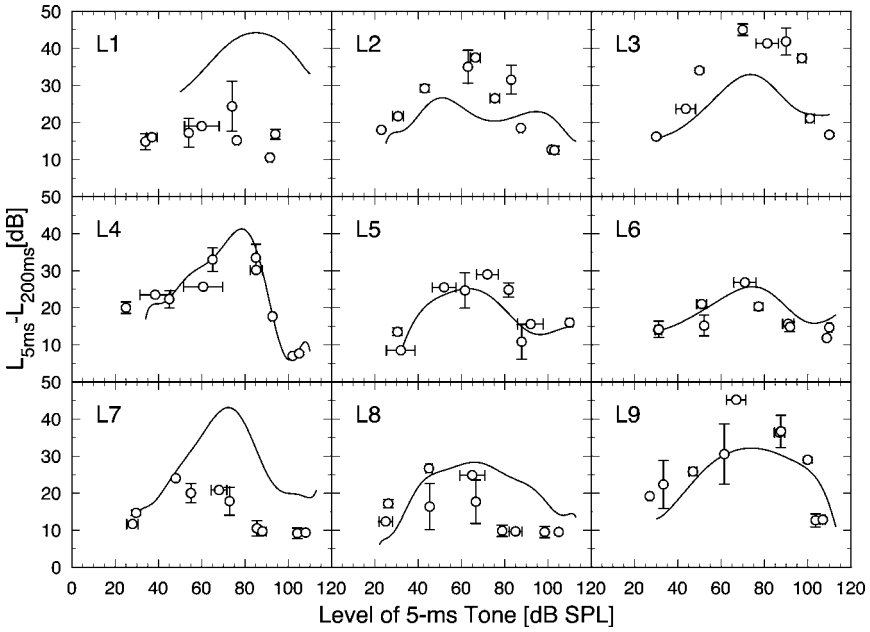


FIG. 3. Temporal integration of loudness for nine individual listeners derived from adaptive loudness (circles) and CMM (lines). The level differences needed to obtain equal loudness between the 5 and 200 ms tones are plotted as a function of the level of the 5 ms tone. The error bars show ± one standard error of the mean.

of temporal integration—defined as the level difference between equally loud 5 and 200 ms tones—plotted as a function of the 5 ms tone's level. Each panel shows data for one listener. The circles show the amount of temporal integration obtained with the adaptive loudness-matching procedure. The error bars indicate the standard error of the mean and are oriented to indicate which tone level was varied. The solid line shows the amount of temporal integration derived from the individual CMM data. The latter function was obtained as the level differences between 5 and 200 ms tones that yielded equal string lengths according to polynomials fitted to the logarithms of the geometric means for each listener and duration. All listeners show a mid-level maximum in the amount of temporal integration in agreement with a number of previous studies (e.g., Florentine *et al.*, 1996, 1998; Buus, 1999). The mid-level maximum is clearly present in the individual data, both for adaptive loudness matches and for the indirect matches obtained from the cross-modality matches. The magnitude of the mid-level maximum varies greatly among individuals. Of the nine listeners, four [L4, L5, L6, and L9 (except for one point)] show excellent agreement between the cross-modality matches and the equal-loudness matches. Four listeners (L2, L3, L7, and L8) show agreement at some levels and one listener (L1) shows quite large disagreement. For the four listeners who show agreement at some levels there are some data points that clearly do not overlap between the adaptive matching procedure and the CMM procedure.

Despite substantial differences between the two sets of data for some listeners, average data were calculated to compare with other average data in the literature. Figure 4 shows the average amount of temporal integration across the nine listeners plotted in the same manner as Fig. 3. The standard deviation across the individual listeners is very large, in contrast to the variability for the individual listeners. Although there are some differences between the two methods at moderate and high levels, the average results show reasonable agreement. This indicates that the CMM and loudness-matching procedures produce a generally consistent assessment of the relation between the loudness of short and long tones for group data, except at the highest levels.

IV. DISCUSSION

A. Agreement between cross-modality matching and loudness matches

Agreement between the measurements of CMM and the adaptive loudness balances varies from reasonable to excellent for seven of the nine listeners. Although it is difficult to estimate the probable error of the amount of temporal integration derived from the CMM functions, the data for two other listeners (L1 and L7) appear to show substantial discrepancies between the two methods. One possible reason for these discrepancies is that estimates of the amount of temporal integration from the CMM data are quite sensitive to perturbations in the polynomial fits used to summarize the CMM data. This is especially true when the slope of the CMM function is shallow, as it is at moderate levels for most

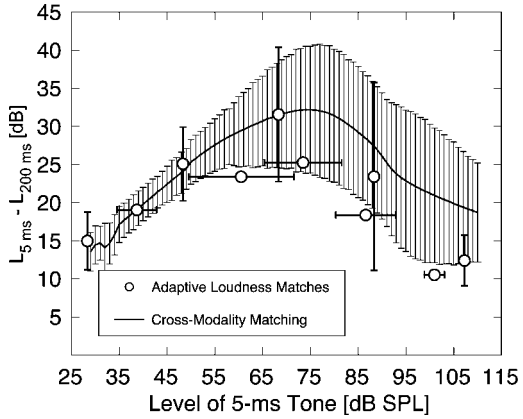


FIG. 4. Average temporal integration of loudness for nine listeners plotted in the same manner as Fig. 3. The error bars show ± one standard deviation calculated across the nine listeners' data for the loudness matching experiment.

listeners, including L7 and L1 (primarily for 5 ms tones). Assuming that variation is uniform across the entire intensity range, a shallower slope increases the sensitivity because a wider range of stimulus levels will result in a statistically identical cross-modality match.

The idea that random error is responsible for most of the discrepancies between CMM and loudness matches noted for individual listeners is supported by the reasonable agreement between the two data sets shown by the average data in Fig. 4 except at high levels. Whereas the two sets of data agree up to about 55 dB SPL, differences are apparent at higher levels. These differences may be due to effects of induced loudness reduction [ILR; Nieder *et al.*, 2003; also called loudness recalibration (Marks, 1994), as described in the following]. Therefore, the difference between the two data sets at high levels should not be taken to indicate that one or the other method yields invalid results, at least when data are averaged across listeners.

Generally, ILR refers to the finding that intense tones reduce the loudness of subsequent weaker tones at or near the same frequency (e.g., Marks, 1994; Nieder *et al.*, 2003; Arieh and Marks, 2003). Recently, it has been shown that ILR is likely to affect measurements of temporal integration of loudness considerably when the stimulus level varies from trial to trial, as it did for both methods used to assess loudness in the present study. Nieder *et al.* (2003) found that 5 ms tones are much less effective in inducing ILR on 200 ms tones than the reverse. In addition, ILR effects are greatest when inducers are at moderate-to-high levels and relatively close in level to the test tone. (The maximum effect occurs with approximately a 10 dB separation.) Thus, when 5 and 200 ms tones at various levels separated by relatively small level differences are presented within a block of trials, the loudness reduction is probably greater for 5 ms tones than for 200 ms tones. In turn, the SPL of the 5 ms tone must be raised more to achieve the same loudness as a 200 ms tone, which leads to an enlarged estimate of the amount of temporal integration.

Because it is highly likely that a high-level presentation of a long tone would have occurred early in the experimental block, the data of Nieder *et al.* (2003) indicate that ILR would have affected both the loudness matches and the cross-modality matches. However, the CMM procedure contained more levels with closer spacing than the loudness-matching procedure. Furthermore, because the present CMM procedure restricted the amount of level change from trial to trial, the listener received moderate and high stimuli closer together in time, offering less opportunity for recovery than in the loudness matches. Therefore, the effect of ILR could be greater for CMM than for adaptive loudness matching. Individual differences in performance on psychophysical tasks make it difficult to see a clear effect of procedure in the individual data. Further understanding of the time course of ILR would be necessary to determine what the precise effect might have been.

B. Comparison with data in the literature

To facilitate the comparison with previous studies, it is useful to approximate the mid-to-high-level portion of the CMM functions with power functions. Above 40 dB SPL, the data for both 5 and 200 ms tones approximate a power function with an exponent of about 0.14 and the ratio between them is approximately 1.8.

The exponent of 0.14 is somewhat smaller than the 0.2 CMM exponent reported by Baird *et al.* (1980) and considerably smaller than the 0.32 reported for matches between line length and loudness (Hellman, 1999). However, some discrepancy is to be expected because Hellman's (1999) listeners adjusted tones to match fixed-length lines, whereas the present listeners had to match a variable string length to a given tone. Given the general tendency of judgments to regress toward the middle of the scale, one would expect the present cross-modality functions to have lower exponents than those obtained by Hellman (1999). Likewise, the ratio 1.8 is considerably smaller than 4.0 estimated in recent loudness-balance experiments (e.g., Florentine *et al.*, 1998; Buus *et al.*, 1999). However, if the present ratio is scaled by the ratio between the present exponent and that of about 0.3 generally used to approximate the growth of loudness at moderate and high levels, the corresponding loudness ratio is about 3.9, which is in excellent agreement with the ratios estimated in loudness-balance experiments.

C. Testing the Equal-Loudness-Ratio hypothesis

The present group data show that the ratio of string lengths matched to equal-SPL long and short tones is approximately constant, except for a slight increase below 40 dB SPL. Accordingly, the data support the Equal-Loudness-Ratio hypothesis, except at low levels.

The finding that the loudness ratio between equal-SPL long and short tones is approximately constant is not unexpected. This relationship is an inherent property of Zwislocki's (1969) theory of temporal integration. Moreover, assuming that loudness bears a simple relation to the overall neural activity evoked by the stimulus, this finding agrees with data on auditory-nerve adaptation. Smith and Zwislocki (1975) showed that the ratios of spike rates measured in the auditory nerve at various times after the onset of a stimulus were approximately independent of the spike rate. This finding indicates that the ratio between the number of spikes evoked by equal-SPL long and short tones is approximately independent of their SPL. Thus, one would expect that the loudness ratio is also independent of SPL, even if loudness may be formed after central transformations of the auditory-nerve activity and may not be directly proportional to nerve-spike count (Relkin and Doucet, 1997). As discussed by Buus and Florentine (2001), the loudness ratio appears to be nearly proportional to the integral of the square of the firing rate in the auditory nerve, rather than being equal to the ratio between the numbers of spikes evoked by the stimuli (Zeng and Shannon, 1994). Nevertheless, it is clear that the present data regarding how loudness changes with duration agree with expectations based on auditory-nerve data and both the present data and auditory-nerve data support the Equal-Loudness-Ratio hypothesis.

V. CONCLUSIONS

The CMM procedure used to measure loudness functions for long and short tones yields reasonably reliable results for most listeners. Group comparisons with loudness-matching data for the same listeners and stimuli indicate that the listeners' average loudness judgments were generally internally consistent, although the individual data are variable. The loudness functions obtained by CMM generally supported the Equal-Loudness-Ratio hypothesis. The loudness functions also show a decrease in slope at moderate levels consistent with previous studies.

ACKNOWLEDGMENTS

Søren Buus contributed substantially to this project and passed away prior to completion of the work. Rhona Hellman, Bert Scharf, and Eva Wagner gave helpful comments on an earlier version of this manuscript. Editor Andrew Oxenham and reviewers Bert Schlauch and Lawrence Ward also provided helpful suggestions. This research was supported by NIH/NIDCD Grant No. R01DC02241.

ANSI (1989). "Specifications for audiometers," ANSI S3.6-1989.

Arieh, Y. and Marks, L. E. (**2003**). "Time course of loudness recalibration: Implications for loudness enhancement," J. Acoust. Soc. Am. **114**, 1550–1556.

Baird, J. C., Green, D. M., and Luce, R. D. (**1980**). "Variability and sequential effects in cross-modality matching of area and loudness," J. Exp. Psychol. Hum. Percept. Perform. **6**, 277–289.

Braida, L. D. and Durlach, N. I. (**1972**). "Intensity perception. II. Resolution in one-interval paradigms," J. Acoust. Soc. Am. **51**, 483–502.

Buus, S. (**1999**). "Loudness functions derived from measurements of temporal and spectral integration of loudness," in *Auditory Models and Nonlinear Hearing Instruments*, edited by A. N. Rasmussen, P. A. Osterhammel, T. Andersen, and T. Poulsen (GN ReSound, Taastrup, Denmark).

Buus, S. and Florentine, M. (**2001**). "Modifications to the power function for loudness," in *Fechner Day 2001*, edited by E. Sommerfeld, R. Kompass, and T. Lachmann (Pabst, Berlin).

Buus, S., Florentine, M., and Poulsen, T. (**1997**). "Temporal integration of loudness, loudness discrimination, and the form of the loudness function," J. Acoust. Soc. Am. **101**, 669–680.

Buus, S., Florentine, M., and Poulsen, T. (**1999**). "Temporal integration of loudness in listeners with hearing losses of primarily cochlear origin," J. Acoust. Soc. Am. **105**, 3464–3480.

Buus, S., Müsch, H., and Florentine, M. (**1998**). "On loudness at threshold," J. Acoust. Soc. Am. **104**, 399–410.

Durlach, N. I. and Braida, L. D. (**1969**). "Intensity perception. I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

Epstein, M. and Florentine, M. (**2005**). "Inferring basilar-membrane motion from tone-burst otoacoustic emissions and psychoacoustic measurements," J. Acoust. Soc. Am. **117**, 263–274.

Florentine, M., Buus, S., and Poulsen, T. (**1996**). "Temporal integration of loudness as a function of level," J. Acoust. Soc. Am. **99**, 1633–1644.

Florentine, M., Buus, S., and Robinson, M. (**1998**). "Temporal integration of loudness under partial masking," J. Acoust. Soc. Am. **104**, 999–1007.

Hellman, R. P. (**1999**). "Cross-modality matching: A tool for measuring loudness in sensorineural impairment," Ear Hear. **20**, 193–213.

Hellman, R. P. and Meiselman, C. H. (**1988**). "Prediction of individual loudness exponents from cross-modality matching," J. Speech Hear. Res. **31**, 605–615.

Hellman, R. P. and Meiselman, C. H. (**1990**). "Loudness relations for individuals and groups in normal and impaired hearing," J. Acoust. Soc. Am. **88**, 2596–2606.

Hellman, R. P. and Meiselman, C. H. (**1993**). "Rate of loudness growth for pure tones in normal and impaired hearing," J. Acoust. Soc. Am. **93**, 966–975.

Hellman, R. P. and Zwislocki, J. J. (**1961**). "Some factors affecting the estimation of loudness," J. Acoust. Soc. Am. **33**, 687–694.

Hicks, M. L. and Buus, S. (**2000**). "Efficient across-frequency integration: Evidence from psychometric functions," J. Acoust. Soc. Am. **107**, 3333–3342.

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

Marks, L. E. (**1994**). " 'Recalibrating' the auditory system: The perception of loudness," J. Exp. Psychol. Hum. Percept. Perform. **20**, 382–396.

McFadden, D. (**1975**). "Duration-intensity reciprocity for equal loudness," J. Acoust. Soc. Am. **57**, 702–704.

Nieder, B., Buus, S., Florentine, M., and Scharf, B. (**2003**). "Interactions between test- and inducer-tone durations in induced loudness reduction," J. Acoust. Soc. Am. **114**, 2846–2855.

Relkin, E. M. and Doucet, J. R. (**1997**). "Is loudness simply proportional to the auditory nerve spike count?," J. Acoust. Soc. Am. **101**, 2735–2740.

Scharf, B. (**1970**). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. I.

Schlauch, R. S., DiGiovanni, J. J., and Ries, D. T. (**1998**). "Basilar membrane nonlinearity and loudness," J. Acoust. Soc. Am. **103**, 2010–2020.

Smith, R. L. and Zwislocki, J. J. (**1975**). "Short-term adaptation and incremental responses in single auditory-nerve fibers," Biol. Cybern. **17**, 169–182.

Teghtsoonian, M. and Teghtsoonian, R. (**1983**). "Consistency of individual exponents in cross-modal matching," Percept. Psychophys. **33**, 203–214.

Ward, L. M., Armstrong, J., and Golestani, N. (**1996**). "Intensity resolution and subjective magnitude in psychophysical scaling," Percept. Psychophys. **58**, 793–801.

Zeng, F.-G. and Shannon, R. V. (**1994**). "Loudness-coding mechanisms inferred from electric stimulation of the human auditory system," Science **264**, 564–566.

Zwicker, E. and Fastl, H. (**1990**). *Psychoacoustics - Facts and Models* (Springer, Berlin).

Zwislocki, J. J. (**1969**). "Temporal summation of loudness: An analysis," J. Acoust. Soc. Am. **46**, 431–441.

Word recognition in noise at higher-than-normal levels: Decreases in scores and increases in masking

Judy R. Dubno,[a] Amy R. Horwitz, and Jayne B. Ahlstrom
Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina, 135 Rutledge Avenue, P.O. Box 250550, Charleston, South Carolina 29425

(Received 17 January 2005; revised 14 April 2005; accepted 14 May 2005)

Under certain conditions, speech recognition in noise decreases above conversational levels when signal-to-noise ratio is held constant. The current study was undertaken to determine if nonlinear growth of masking and the subsequent reduction in "effective" signal-to-noise ratio accounts for this decline. Nine young adults with normal hearing listened to monosyllabic words at three levels in each of three levels of a masker shaped to match the speech spectrum. An additional low-level noise equated audibility by producing equivalent masked thresholds for all subjects. If word recognition was determined entirely by signal-to-noise ratio and was independent of overall speech and masker levels, scores at a given signal-to-noise ratio should remain constant with increasing level. Masked pure-tone thresholds measured in the speech-shaped maskers increased linearly with increasing masker level at lower frequencies but nonlinearly at higher frequencies, consistent with nonlinear growth of upward spread of masking that followed the peaks in the spectrum of the speech-shaped masker. Word recognition declined significantly with increasing level when signal-to-noise ratio was held constant which was attributed to nonlinear growth of masking and reduced "effective" signal-to-noise ratio at high speech-shaped masker levels, as indicated by audibility estimates based on the Articulation Index. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1953107]

PACS number(s): 43.66.Dc, 43.71.Es, 43.71.Pc, 43.66.Sr [JHG] Pages: 914–922

I. INTRODUCTION

The deterioration of speech recognition at high signal and noise levels has been under investigation for many years (e.g., French and Steinberg, 1947; Hawkins and Stevens, 1950; Pollack and Pickett, 1958). This issue was revisited when Studebaker *et al.* (1999) noted that speech recognition in noise by normal-hearing and hearing-impaired listeners decreased at signal levels just exceeding conversational levels when signal-to-noise ratio was held constant. Similarly, when speech was amplified to high levels and processed with high-frequency emphasis, observed speech-recognition scores for normal-hearing subjects (Horwitz *et al.*, 2004) and hearing-impaired subjects (Ching *et al.*, 1998; Hogan and Turner, 1998; Horwitz *et al.*, 2004) were poorer than scores predicted using a simple audibility model. Also using amplification with high-frequency emphasis, Amos and Humes (2001) found that the negative effects of high signal levels were the same for normal-hearing and hearing-impaired subjects for some but not all conditions. Hornsby and Ricketts (2001) reported that consonant recognition decreased as speech level increased for normal-hearing subjects listening to speech processed through a simulated wide-dynamic-range-compression system. Shanks *et al.* (2002) found that, in unaided conditions, recognition of words in sentences declined with increasing speech level for subjects with relatively mild hearing loss but not for subjects with more severe hearing loss; in aided conditions, recognition declined for nearly all subjects.

Somewhat different conclusions regarding the adverse effects of high signal levels, however, were drawn from results of Dubno *et al.* (2000). In this study, recognition of key words in high- and low-context sentences from the Speech Perception in Noise test (SPIN; Kalikow *et al.*, 1977) was measured over a wide range of speech and noise levels by younger adults with normal hearing and older adults with minimally elevated thresholds; effective signal-to-noise ratio for each subject was held constant. Over the range of speech and noise levels used, word recognition generally remained constant or decreased only slightly.

Several issues remain unresolved regarding the detrimental effects on speech recognition of speech and noise at high levels. First, the mechanism responsible for the deterioration of speech recognition at high levels is not yet known. In most studies, pure-tone thresholds at frequencies within the spectrum of the speech either were not measured or were measured only in quiet conditions to assess magnitude of hearing loss. Without thresholds measured in the noises and at the levels used to mask the speech, it is not possible to determine if observed declines in speech recognition at high signal levels are due simply to changes in "effective" signal-to-noise ratio resulting from increases in masked thresholds with increasing signal level, or if more complex, level-dependent changes in suprathreshold processing must be considered.

Second, the magnitude of the reduction in speech recognition at high signal and masker levels varies substantially among studies, perhaps due to differences in speech materials and/or masker spectra. Studebaker *et al.* (1999) noted that effects were largest for nonsense syllables and monosyllabic words and when speech was presented in a masker with a

[a]Electronic mail: dubnojr@musc.edu

 0001-4966/2005/118(2)/914/9/$22.50

spectrum that matched the spectrum of the speech. This is consistent with results of Dubno *et al.* (2000) in which a reduction in speech recognition at high levels was not observed for sentences presented in a masker whose spectrum approximated the shape of the audiogram rather than the shape of the speech spectrum.

Third, there is indirect evidence that the deterioration of speech recognition at high levels may vary with the spectral content of the speech and/or masker. Declines in scores observed in some studies for speech amplified to high levels and processed with high-frequency emphasis suggest that recognition of high-frequency speech may be particularly vulnerable. Indeed, with increasing low- and mid-frequency speech energy, recognition of nonsense syllables that were amplified to high levels improved for hearing-impaired subjects (Turner and Brus, 2001), whereas increasing high-frequency speech energy sometimes resulted in decreasing speech recognition (e.g., Hogan and Turner, 1998; Ching *et al.*, 1998). Molis and Summers (2003) reported that recognition of key words in sentences declined at high levels more for high-pass-filtered sentences than for low-pass-filtered sentences.

Based on these results, the current study was undertaken to explore further speech recognition in noise at higher-than-normal levels. Speech and maskers were selected to maximize the likelihood of observing a decline in speech recognition at high levels. In addition, to maintain consistency among studies, speech tokens and the spectrum of the masker were identical to those used by Studebaker *et al.* (1999).[1] Spectral effects on word recognition in noise at high levels were assessed in another experiment (Dubno *et al.*, 2005). In the current study, nine young adults with normal hearing listened to broadband monosyllabic words from the Northwestern University Test No. 6 (NU#6; Tillman and Carhart, 1966) lists at three levels in each of three levels of a masker shaped to match the spectrum of the NU#6 talker's speech. Speech and masker levels were selected so that, for each of three signal-to-noise ratios, speech varied from moderate to high levels. An additional low-level noise was always present to equate audibility by producing equivalent masked thresholds for all subjects. To quantify audibility, pure-tone thresholds across a wide range of frequencies were measured in each level of the speech-shaped masker. With this design, if word recognition was determined entirely by signal-to-noise ratio and was independent of overall speech and masker levels, scores at a given signal-to-noise ratio should remain constant with increasing level. The degree to which word recognition decreased with increasing signal level provides evidence of the adverse effects of high speech and noise levels.

II. METHODS

A. Subjects

Nine young adults participated, ranging in age from 21 to 28 years (mean age: 24.1 years). All subjects had thresholds $\leq$15 dB HL (ANSI, 1996) at octave frequencies from 0.25 to 8.0 kHz and normal immitance measures. Test ear was selected randomly. Subjects did not have experience with the psychophysical task used in this study and thus received approximately 1 h of practice. Subjects did not have prior experience with the words contained in the NU#6 lists. Data collection was completed in five to six 2-h sessions, including appropriate rest periods. Subjects were paid an hourly rate for their participation.

B. Apparatus and stimuli

1. Tonal signals and noises

Tonal signals were digitally generated (TDT DA3-4) pure tones, sampled at 50 kHz and low-pass filtered at 12 kHz (TDT FT6). Signals were 350 ms in duration, including 10-ms raised-cosine rise/fall ramps.

For masking of pure tones and speech, the masker was a broadband noise that was digitally generated at a sampling rate of 28 kHz. Its spectrum was then adjusted in 0.05-kHz intervals (from 0.05 to 11.5 kHz) using Matlab™ (Version 5.3, The Mathworks, Inc., Natick, MA) and in-house software so that the long-term spectrum of the masker matched the 1% levels of the NU#6 talker's speech ("speech-shaped masker"). As shown in Studebaker *et al.* (1999, Fig. 1), the long-term rms spectra of the speech and speech-shaped masker are similar. For example, in both one-third-octave band spectra, the bands centered at 0.5 and 0.63 kHz contain the most energy. The speech-shaped masker was presented at overall levels of 70, 77, and 84 dB SPL.

Although all subjects had normal hearing, small differences in quiet thresholds were observed. Results of a previous study (Dubno and Ahlstrom, 1997) revealed that differences in audibility among subjects as a result of relatively small threshold differences can produce substantial intersubject variability in word recognition under certain conditions. To minimize the influence of differences in quiet thresholds, a second masker was always present to ensure equal audibility among subjects. This low-level broadband noise was digitally generated at a sampling rate of 28 kHz and then its spectrum adjusted at one-third-octave intervals (Cool Edit Pro™ Version 1.2, Syntrillium Software Corp., Scottsdale, AZ) to achieve masked thresholds of 20 dB HL for all subjects. The overall level of this "threshold-matching noise" (TMN) was 54 dB SPL. For conditions in which the speech-shaped masker was presented at 77 and 84 dB SPL, the TMN was presented at 61 and 68 dB SPL, respectively, to maintain constant audibility across conditions.

The speech-shaped masker and the TMN were output through 16-bit digital-to-analog converters (TDT DA3-4), low-pass filtered at 12 kHz (TDT FT6), and recorded onto separate channels of a digital audio tape (DAT) for later playback. Spectral characteristics of all maskers were verified on an acoustic coupler and a signal analyzer (Stanford Research SR780). Maskers were always present during the measurement of pure-tone thresholds and speech recognition.

2. Speech

Speech tokens were the 200 NU#6 monosyllabic words, originally recorded by a male talker by Auditec of St. Louis and later digitized at 25 kHz and stored (with the carrier phrase) in individual .wav files. These stimuli were identical

to those used by Studebaker *et al.* (1999). Speech and masker levels were selected so that word-recognition scores ranged from approximately 30% to 80%, avoiding floor and ceiling effects. NU#6 words were presented at three levels for each of three signal-to-noise ratios (+8, +3, and −2 dB). For the +8-dB signal-to-noise ratio, speech levels were 78, 85, and 92 dB SPL; for the +3-dB signal-to-noise ratio, speech levels were 73, 80, and 87 dB SPL; for the −2-dB signal-to-noise ratio, speech levels were 68, 75, and 82 dB SPL. Thus, word recognition was measured in nine conditions, corresponding to all combinations of three signal-to-noise ratios and three speech-shaped masker levels (70, 77, and 84 dB SPL).

Digital speech waveforms were output through a 16-bit digital-to-analog converter (TDT DA3-4) and low-pass filtered at 12 kHz (TDT FT6). The amplitudes of all signals and maskers were controlled individually using programmable attenuators (TDT PA4). The signal was added to the speech-shaped masker and the TMN (TDT SM3), band-pass filtered (2 TDT PF1s) from 0.165 to 7.4 kHz, and delivered through one of a pair of TDH-49 earphones mounted in supra-aural cushions.

C. Procedures

For each subject, thresholds for pure tones were measured in the following order: (1) thresholds in quiet; (2) masked thresholds in TMN at levels of 54, 61, and 68 dB SPL; and (3) masked thresholds in the speech-shaped masker at levels of 70, 77, and 84 dB SPL, with TMN at 54, 61, and 68 dB SPL, respectively. In each case, thresholds were measured at 16 frequencies at one-third-octave intervals ranging from 0.2 to 6.3 kHz.

Pure-tone thresholds were obtained using a single-interval (yes-no) maximum-likelihood psychophysical procedure, similar to that described by Green (1993) and discussed in detail in Leek *et al.* (2000). Signal level was varied adaptively. Listen and vote periods were displayed on the screen of a computer monitor. Subjects responded by clicking one of two mouse buttons corresponding to the responses "yes, I heard the tone" and "no, I did not hear the tone."

Following measurement of quiet and masked pure-tone thresholds, recognition of the 200 NU#6 words was measured at three speech levels in each of three levels of the speech-shaped masker (nine conditions). In each condition, the 200 words were presented in four, 50-word blocks; the 50 words in each of the four blocks were the same as those in the four traditional 50-word NU#6 lists. Within each 50-word block, words were selected randomly for presentation by in-house software. Subjects were instructed to respond by repeating aloud the word following the carrier phrase and were encouraged to guess.

Because it was necessary to repeat the same 200 words in nine conditions, it was important to consider the confounding effects of repeated presentations (e.g., Egan, 1948, Fig. 4). In examining recognition of words embedded in sentences using an adaptive psychophysical procedure, Wilson *et al.* (2003) concluded that improved performance with repeated measures was due to subjects becoming more familiar with the test procedure, response task, speaker's voice, and test environment. To minimize effects of procedural learning, prior to data collection, subjects listened to nine 25-word lists from digital recordings of the CID W-22 monosyllabic word test (Hirsh *et al.*, 1952) at speech and masker levels corresponding to each of the nine experimental conditions; the W-22 word lists were recorded by a different speaker than the NU#6 word lists. To minimize effects of familiarization, prior to data collection, subjects had no experience with the NU#6 words and all subjects read an alphabetic list of the NU#6 words. To minimize effects of trial order, a 9×9 Latin-square design determined the order of the nine masker level/signal-to-noise ratio combinations for each of the nine subjects. Finally, to avoid any effects of overall list differences or list differences that may vary with condition difficulty (e.g., Stockley and Green, 2000), the order of the four NU#6 50-word blocks was randomized for each subject and each data point was the percentage of correct responses to all 200 NU#6 words.

D. Data analysis

Outcome measures included pure-tone thresholds in the speech-shaped masker at three levels, observed word-recognition scores as a function of speech-shaped masker level and signal-to-noise ratio, word-recognition scores predicted using the Articulation Index (AI), and differences between observed and predicted word-recognition scores. Differences between predicted and observed scores may reveal a role of nonlinear growth of masking for speech recognition to the extent that these differences are due to changes in "effective" masker levels resulting from nonlinear effects. AI values and predicted scores were computed using procedures similar to ANSI (1997), assuming a 40-dB effective dynamic range of speech (Studebaker *et al.*, 1999), and using the frequency importance function and AI-recognition transfer function for the Auditec of St. Louis recordings of the NU#6 word test (Studebaker *et al.*, 1993b). Using rau-transformed word-recognition scores (Studebaker, 1985), differences due to trial number (1–9) and word list (1–4) were assessed by Latin-square and repeated-measures ANOVAs, respectively. Significant differences in word recognition were found among the nine trials and the four lists. Scores were adjusted to remove the effects of trial and these adjusted scores were submitted to ANOVAs to assess effects of speech level and signal-to-noise ratio on word recognition. Significant differences in word recognition were also observed among the four 50-word NU#6 lists but no adjustments were necessary because subjects' scores were based on all 200 NU#6 words (see the Appendix for additional details on differences due to trial number and word list).

III. RESULTS AND DISCUSSION

A. Masked thresholds

Figure 1 shows mean thresholds for pure tones at one-third-octave frequencies from 0.2 to 6.3 kHz measured in the speech-shaped masker (filled circles) with the masker level at 70 dB SPL (top panel), 77 dB SPL (middle panel), and 84 dB SPL (bottom panel). Also included in the three panels are the

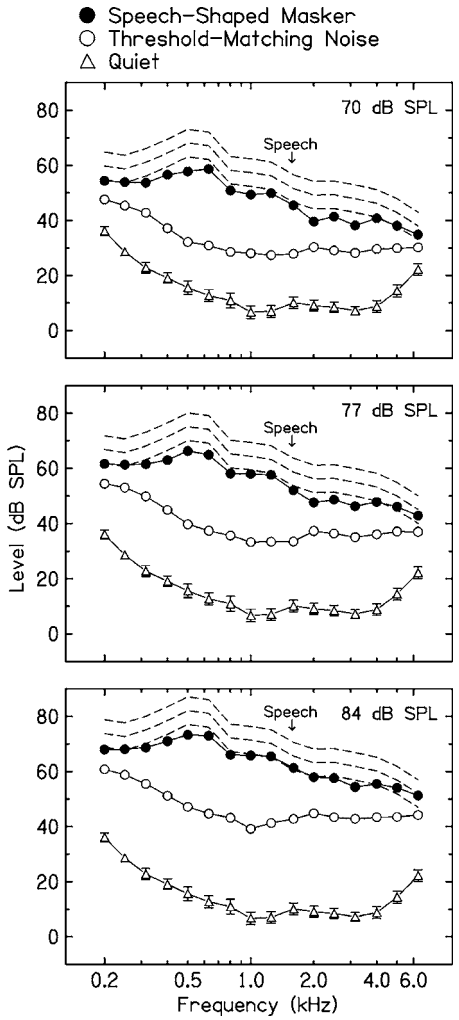


FIG. 1. Mean thresholds (±1 standard error, SE) for pure tones at one-third-octave frequencies from 0.2 to 6.3 kHz measured in the speech-shaped masker (filled circles) with the masker level at 70 dB SPL (top panel), 77 dB SPL (middle panel), and 84 dB SPL (bottom panel). Also included are mean thresholds for the same signals measured in threshold-matching noise (TMN, open circles) at 54, 61, and 68 dB SPL, respectively. Mean thresholds in quiet are also shown in each panel (triangles). Standard error ranges are visible only for quiet thresholds. Each panel also includes the one-third-octave rms spectrum of the NU#6 words plotted at the three levels used with each masker level (dashed lines).

mean thresholds for the same signals measured in the threshold-matching noise (TMN, open circles) with the TMN level at 54, 61, and 68 dB SPL, respectively. Mean thresholds in quiet are also shown in each panel (triangles). Standard error ranges exceed the size of the data points only for quiet thresholds.

In each panel, the dashed line is the one-third-octave rms spectrum of the NU#6 words plotted at the three levels used with each masker level (i.e., 68, 73, and 78 dB for the 70-dB masker; 75, 80, and 85 dB for the 77-dB masker; 82, 87, and 92 dB for the 84-dB masker). Within each panel, as speech level increased, the signal-to-noise ratio and audible speech also increased, as seen by the relative positions of the masked thresholds (solid line with filled symbols) and the speech spectrum (dashed lines). Although these patterns are similar in all three panels, it is apparent that audible speech is decreasing slightly with increasing masker level. This is

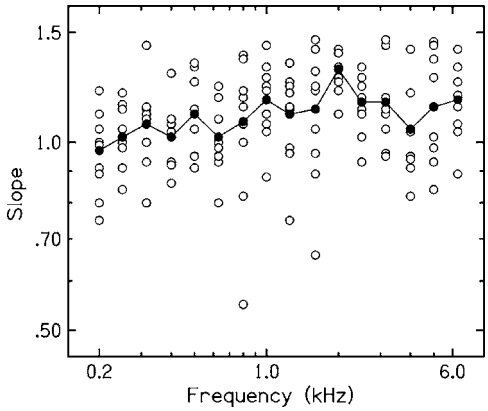


FIG. 2. Slopes of growth-of-masking functions for thresholds at 0.2–6.3 kHz measured in three levels of the speech-shaped maskers (open circles), computed using linear regression. Mean slopes are shown by the filled circles.

because as masker level was increased by 7 dB (from 70 dB to 77 dB to 84 dB), masked thresholds tended to increase by slightly more than 7 dB. The effect of this nonlinear increase in masked thresholds was to reduce "effective" signal-to-noise ratio with increasing masker level.

Another view of these results is displayed in Fig. 2 which shows slopes of growth-of-masking functions for each subject for signal frequencies ranging from 0.2 to 6.3 kHz. Growth-of-masking slopes for thresholds at each frequency measured in three levels of the speech-shaped masker were derived using linear regression. Mean slopes are shown by the filled circles; note the relatively large intersubject variability in growth-of-masking slope, even among these young adults with normal hearing. Nonlinear growth of masking (slopes >1.0) is consistent with upward spread of masking attributed to the lower frequency peaks in the speech-shaped masker. A repeated-measures ANOVA on the growth-of-masking slopes revealed that frequency had a significant effect on slope $[F(15,120)=3.53, p=0.0027]$. A *posthoc* test was significant for slopes peaking at 2.0 kHz, that is, an increase in slopes from 0.2 to 2.0 kHz and a decrease in slopes from 2.0 to 6.3 kHz $[F(1,8)=18.12, p=0.0028]$. A 2.0-kHz peak in growth-of-masking slope coincided with the peak of the frequency importance function for NU#6 words (Studebaker *et al.*, 1993b).

To obtain a single value with which to compare thresholds measured in the three masker levels, weighted average masked thresholds were computed using weights from the frequency importance function for the NU#6 stimuli (Studebaker *et al.*, 1993b). Computed in this way, weighted average thresholds take into account the relative importance of certain frequencies to word recognition. With the speech-shaped masker increasing by 7 dB from 70 to 77 dB and from 77 to 84 dB, weighted average masked thresholds increased by 7.5 and 8.3 dB, respectively, consistent with nonlinear growth of masking. As noted above, these changes in masked thresholds with increasing masker level reduced the "effective" signal-to-noise ratio at the higher masker levels, especially at the middle-to-higher frequencies.

The nonlinear growth of upward spread of masking observed here is consistent with compressive basilar-membrane

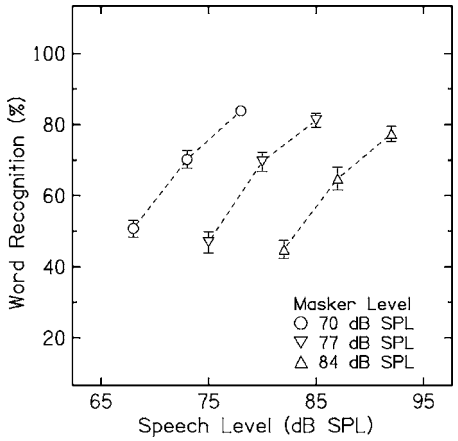


FIG. 3. Psychometric functions relating word recognition measured in speech-shaped maskers to speech level with masker level as the parameter. Mean word-recognition scores (±1 SE) obtained in the 70-dB masker (circles), 77-dB masker (reverse triangles), and 84-dB masker (triangles) are plotted at their respective speech levels.

responses, whereby for moderate-level input signals, input-output functions demonstrate substantial compression (Robles *et al.*, 1986; Yates, 1990; Ruggero *et al.*, 1997). The response to a signal at the signal-frequency place (say, 2.0 kHz) will be compressed for moderate levels but the response at 2.0 kHz to the lower-frequency speech-shaped masker will be linear. Thus, for a given increase in masker level, a larger increase in signal level is required to maintain threshold (e.g., Oxenham and Plack, 1997).

B. Word recognition

The three psychometric functions relating word recognition to speech level measured in three masker levels should be parallel because speech level and masker level were increased by equal amounts. That is, relative to the function for the 70-dB masker, the functions for the 77- and 84-dB maskers should be shifted to the right by 7 and 14 dB, respectively (Studebaker *et al.*, 1993a; Dubno and Ahlstrom, 1997; Dubno *et al.*, 2000). Figure 3 plots these psychometric functions with masker level as the parameter. Mean word-recognition scores obtained in the 70-dB masker (circles), the 77-dB masker (reverse triangles), and the 84-dB masker (triangles) are plotted at their respective speech levels. As expected, for each masker level, scores increased nonlinearly as speech level and signal-to-noise ratio increased.

To illustrate the effect of speech level with signal-to-noise ratio held constant, the data of Fig. 3 are replotted with signal-to-noise ratio as the parameter and are shown in the top panel of Fig. 4 (open symbols). For example, scores for the +8-dB signal-to-noise ratio (open triangles) are identical to the highest scores on the three functions in Fig. 3. The filled symbols in the top panel of Fig. 4 are predicted word-recognition scores. AI values predict scores from signal-to-noise ratios estimated from the levels and spectrum of the speech, each subject's thresholds measured in the TMN, and the levels and spectrum of the speech-shaped masker. Thus, given that these predicted scores do not take into account any changes in "effective" signal-to-noise ratio due to nonlinear growth of masking in the speech-shaped masker, they remain

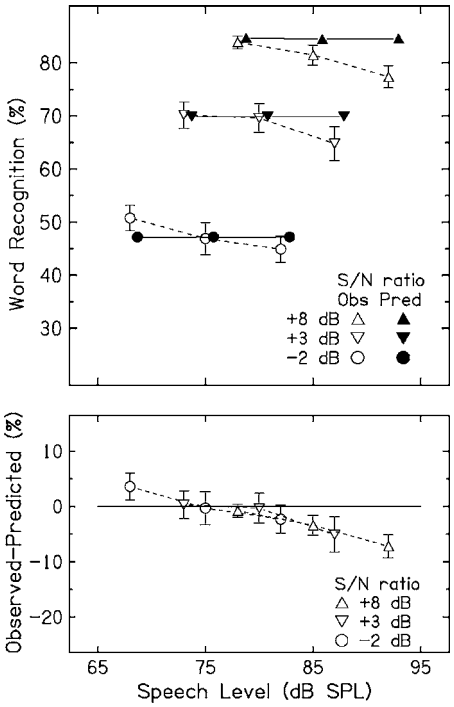


FIG. 4. Top: Observed word-recognition scores from Fig. 3 replotted with signal-to-noise (S/N) ratio as the parameter (Obs, open symbols). Also shown are mean word-recognition scores (±1 SE) predicted from AI values computed using each subject's thresholds in the threshold-matching noise (Pred, filled symbols). Bottom: Mean differences between observed and predicted scores (±1 SE) plotted as a function of speech level with signal-to-noise (S/N) ratio as the parameter. For clarity, a solid line is drawn at an observed-predicted difference of 0% and some data points are offset along the abscissa.

constant because, as speech level increased, masker level also increased, maintaining a constant signal-to-noise ratio. However, observed scores in Fig. 4 (open symbols) declined significantly as a function of speech level $[F(2,16)=10.21, p=0.0014]$. *Posthoc* tests showed a significant linear trend $[F(1,8)=18.22, p=0.0027]$, suggesting that observed scores declined linearly with speech level. The interaction of speech level and signal-to-noise ratio was not statistically significant $[F(4,32)=0.76, p=0.558]$. The significant decline in scores confirmed that word recognition in speech-shaped maskers decreased at high speech levels when signal-to-noise ratio was held constant.

The bottom panel of Fig. 4 shows differences between observed and predicted word-recognition scores plotted as a function of speech level. The parameter is signal-to-noise (S/N) ratio. When word-recognition scores were predicted without taking into account changes in signal-to-noise ratio due to nonlinear growth of masking, observed scores decreased relative to predicted scores as speech level increased. That is, observed minus predicted values decreased significantly as a function of speech level $[F(2,16)=10.01, p=0.0015]$, with *posthoc* tests showing a significant linear trend $[F(1,8)=17.79, p=0.0029]$. The interaction of speech level and signal-to-noise ratio was not statistically significant $[F(4,32)=0.75, p=0.566]$. These results are consistent with word recognition declining at higher signal levels when signal-to-noise ratio is held constant.

To assess whether nonlinear growth of masking in the

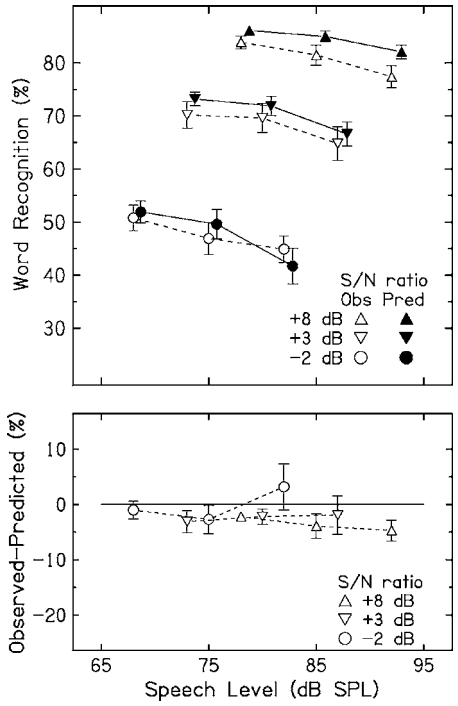


FIG. 5. Same as Fig. 4, but with predicted scores determined from AI values computed using each subject's thresholds measured in the speech-shaped masker.

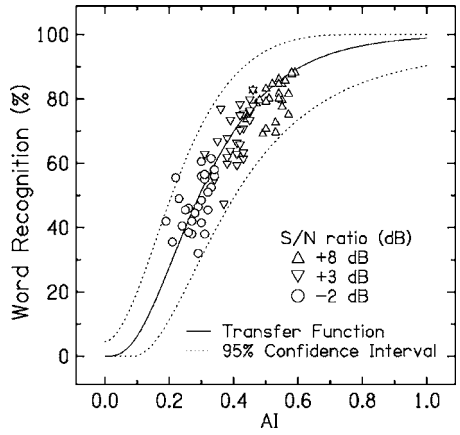


FIG. 6. Word-recognition scores at three signal-to-noise (S/N) ratios plotted against AI, with AI values computed using thresholds measured in the speech-shaped masker. The solid line is the AI-recognition transfer function for the Auditec of St. Louis recordings of the NU#6 word test (Studebaker *et al.*, 1993b); the dotted lines encompass the 95% confidence limits computed for 200-item word-recognition scores.

speech-shaped maskers contributed to the decline in word recognition at high levels, predicted word-recognition scores were also determined using AI values computed using the "effective" noise spectrum measured empirically from subjects' thresholds in the speech-shaped masker, rather than using thresholds measured in the TMN and the levels and spectrum of the speech-shaped masker. If the decline in word recognition with increasing speech level was entirely attributed to reduced "effective" signal-to-noise ratio related to nonlinear growth of masking in the speech-shaped masker, predicted scores determined using speech-shaped masked thresholds should decline as speech level increased because a constant "effective" signal-to-noise ratio was not maintained. Given that this is similar to the pattern seen in the observed scores, observed-predicted differences should remain constant with increasing speech level.

Figure 5 presents these results. The filled symbols in the top panel are predicted word-recognition scores estimated from AI values computed using each subject's thresholds measured in the speech-shaped masker. The open symbols are observed scores, identical to those in the top panel of Fig. 4. The bottom panel of Fig. 5 shows differences between observed and predicted word-recognition scores plotted as a function of speech level. The parameter is signal-to-noise (S/N) ratio. When word-recognition scores were predicted while taking into account changes in "effective" signal-to-noise ratio with increasing speech-shaped masker level, differences between observed and predicted scores remained constant as speech level increased $[F(2,16)=0.32, p=0.73]$ and their slopes were not significantly different from zero $[F(1,8)=0.16, p=0.70]$. Taken together, these results suggest that the decrease in word recognition at higher speech and masker levels may be attributed to nonlinear growth of masked thresholds (and reduced "effective" signal-to-noise ratio) in the speech-shaped masker. Moreover, given evidence that tonal growth-of-masking functions reflect basilar-membrane nonlinearities [see Oxenham and Bacon (2003) for review], these results provide additional support for a role of nonlinear effects on speech recognition at high signal levels.

Figure 6 shows word-recognition scores at three signal-to-noise (S/N) ratios plotted against AI, with AI values computed using thresholds measured in the speech-shaped masker. The solid line is the normal transfer function relating the AI to word recognition established for the NU#6 word test used in this experiment (Studebaker *et al.*, 1993b); the dotted lines encompass the 95% confidence limits computed for 200-item word-recognition scores. Despite significant declines in word recognition with increasing speech level, nearly all of the observed scores fall within the 95% confidence interval of their predicted scores. There were no consistent patterns with regard to the observed scores for the three signal-to-noise ratios, or for specific speech or masker levels (not shown). One exception was a trend for scores obtained in the +8-dB (most advantageous) signal-to-noise ratio to be poorer than predicted, which is also seen in the top panel of Fig. 5 (open and filled triangles). The +8-dB signal-to-noise ratio condition corresponds to the highest speech levels.

Attributing the decline in speech recognition to changes in "effective" signal-to-noise ratio resulting from increased masked thresholds also provides a reasonable explanation to other conclusions of Studebaker *et al.* (1999). These authors found large variability among studies in the magnitude of the reduction in speech recognition at high levels and noted that effects were largest for speech presented in a speech-shaped masker and for nonsense syllables and monosyllabic words. With regard to the masker spectrum, a speech-shaped masker (including a multitalker babble) will likely have a lower frequency peak, unlike a white-noise masker or a masker shaped like the audiogram (as in Dubno *et al.*, 2000). Thus, it

is possible that thresholds measured in a masker with a low-frequency spectral peak may be more susceptible to nonlinear growth of upward spread of masking than thresholds measured in maskers with more uniform spectra. As noted earlier, the decline in speech recognition at high levels may vary with the spectral content of the speech and masker, which is the focus of another experiment (Dubno *et al.*, 2005). With regard to speech materials, because of the absence of contextual information, recognition of nonsense syllables and monosyllabic words may be more sensitive to small changes in signal-to-noise ratio with increasing masker level than recognition of sentences. Finally, the explanation for the absence of a decline in key-word recognition for SPIN sentences in Dubno *et al.* (2000) relates to its experimental design. In that experiment, masker levels within one-third-octave bands were adjusted for each subject to obtain 20- and 40-dB threshold shifts relative to quiet thresholds, thus minimizing individual differences in growth of masking and maintaining precisely equivalent "effective" signal-to-noise ratio as masker level increased. With no nonlinear changes in masked thresholds, no decline in word recognition with increasing level was expected or observed.

IV. SUMMARY AND CONCLUSIONS

Nine young adults with normal hearing listened to 200 NU#6 monosyllabic words at three levels for each of three signal-to-noise ratios. The masker was a broadband noise shaped to match the spectrum of the NU#6 talker's speech. Word recognition was measured in nine conditions, corresponding to all combinations of three signal-to-noise ratios and three speech-shaped masker levels. An additional low-level noise was always present that was shaped to produce equivalent masked thresholds for all subjects. Pure-tone thresholds were measured in quiet and in all maskers. Results may be summarized as follows:

(1) Pure-tone thresholds measured in a speech-shaped masker increased linearly at lower frequencies with increases in masker level. At higher frequencies, increases in masked thresholds were larger than increases in masker level, revealing nonlinear growth of upward spread of masking that followed the low-to-mid-frequency peak in the spectrum of the speech-shaped masker. These changes in masked thresholds with increasing masker level reduced the "effective" signal-to-noise ratio at higher levels, especially at the frequencies known to be important for correct recognition of NU#6 monosyllabic words.
(2) Word recognition in a speech-shaped masker declined significantly with increases in speech level when signal-to-noise ratio was held constant.
(3) Using audibility estimates based on the AI, the decline in word recognition at high speech and masker levels was attributed to nonlinear growth of masking in the speech-shaped masker. These results provide additional support for a role of nonlinear effects on speech recognition at high signal levels.

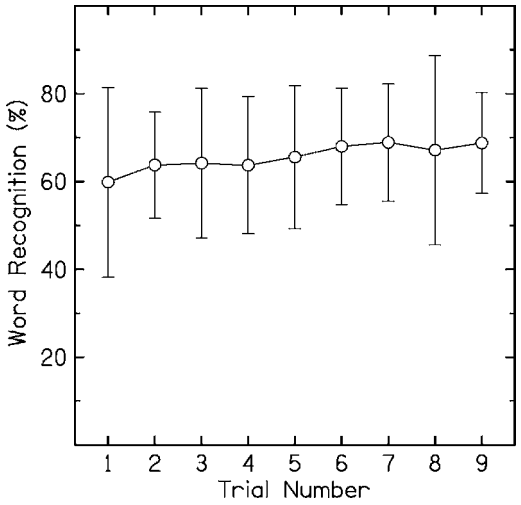


FIG. 7. Mean word-recognition scores (±1 standard deviation, SD) plotted as a function of trial number.

ACKNOWLEDGMENTS

This work was supported (in part) by Grant Nos. R01 DC00184 and P50 DC00422 from NIH/NIDCD, the James E. and Pamela Knowles Foundation, and the MUSC General Clinical Research Center (M01 RR 01070). This investigation was conducted in a facility constructed with support from Research Facilities Improvement Program Grant No. C06 RR14516 from the National Center for Research Resources, National Institutes of Health. The authors thank Chris Ahlstrom for computer and signal-processing support, Fu-Shing Lee for advice on data analysis, and Gerald Studebaker for sharing digitized speech waveforms and the speech spectrum.

APPENDIX: EFFECTS OF TRIAL NUMBER AND WORD LIST

In this experiment, it was necessary to repeat the same 200 words in nine conditions (3 masker levels ×3 signal-to-noise ratios). In addition, in each condition, the 200 words were presented in four 50-word blocks, corresponding to the four NU#6 word lists, which may not be equivalent (e.g., Stockley and Green, 2000). Figure 7 shows mean word-recognition scores (±1 standard deviation, SD) plotted as a function of trial number; for each trial, 200-word scores were averaged over nine subjects with nine different conditions for each subject. Figure 8 shows mean word-recognition scores (+1 SD) for the four NU#6 word lists; for each list, 50-word scores were averaged over trial, condition, and subject.

Before assessing differences in word-recognition score due to signal level or signal-to-noise ratio (i.e., condition effects), it was necessary to determine if condition effects may have been contaminated by trial- or list-related differences in word recognition and, if so, remove their effects. Non-condition-related differences are troublesome because they tend to increase the variance, making it more difficult to detect condition-related differences of interest. Several procedures were put in place to assess differences in word-recognition scores due to trial and word list. With regard to trial, a 9×9 Latin-square design determined the order of the

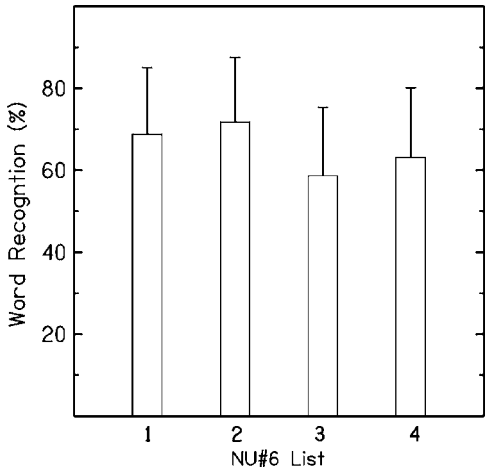


FIG. 8. Mean word-recognition scores (+1 SD) for the four NU#6 word lists.

nine conditions for the nine subjects, providing a completely balanced design. A Latin-square ANOVA revealed significant differences in word recognition due to trial $[F(8,56)=3.62, p=0.0019)]$. *Posthoc* tests showed a significant linear trend $[F(1,56)=25.25, p<0.0001]$, suggesting that observed scores improved linearly across nine trials. To adjust scores to remove the effects of trial, a set of constants was calculated for the nine trials based on the least-square means of all subjects. The word-recognition scores for each subject were then adjusted by adding or subtracting the appropriate constant according to their trial number. Adjusting for trial effects, even if they are small (in this case, ~3%), increases the power of the test because it decreases the mean square error. The adjustment does not affect the mean score for each condition because of the balanced, Latin-square design. The adjusted scores were then submitted to an ANOVA with speech level and signal-to-noise ratio as repeated measures, as reported in Sec. III above.

With regard to list, the order of the four 50-word blocks (corresponding to the four NU#6 word lists) was randomized for each subject. A one-way ANOVA with list as the repeated measure revealed significant differences in word recognition due to list $[F(3,24)=47.31, p<0.0001]$. *Posthoc* analyses revealed that each list was significantly different from all other lists. However, given that each score for each condition was the average of all 200 NU#6 words, any list-related differences would not contaminate condition effects, so it was not necessary to adjust scores to remove any list-related effects.

[1]Digitized speech waveforms for the Auditec of St. Louis recordings of the 200 NU#6 words and values for the talker's speech spectrum in 0.05-kHz intervals were obtained from Dr. Gerald Studebaker.

Amos, N. E. and Humes, L. E. (**2001**). "The contribution of high frequencies to speech recognition in sensorineural hearing loss," in *Physiological and Psychophysical Bases of Auditory Function*, edited by D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs and R. Schoonhoven (Shaker Maastricht, Netherlands), pp. 437–444.

ANSI (**1996**). ANSI S3.6-1996, "American National Standard Specification for Audiometers" (American National Standards Institute, New York).

ANSI (**1997**). ANSI S3.5-1997, "American National Standard Methods for the Calculation of the Speech Intelligibility Index" (American National Standards Institute, New York).

Ching, T., Dillon, H., and Byrne, D. (**1998**). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," J. Acoust. Soc. Am. **103**, 1128–1140.

Dubno, J. R., and Ahlstrom, J. B. (**1997**). "Additivity of multiple maskers of speech," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Mahwah, NJ), pp. 253–272.

Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (**2000**). "Use of context by younger and older adults with normal hearing", J. Acoust. Soc. Am. **107**, 538–546.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2005**). "Recognition of filtered words in noise at higher-than-normal levels: Decreases in scores with and without increases in masking," J. Acoust. Soc. Am. **118**, 923–933.

Egan, J. P. (**1948**). "Articulation testing methods," Laryngoscope **58**, 955–991.

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

Green, D. M. (**1993**). "A maximum-likelihood method for estimating thresholds in a yes-no task," J. Acoust. Soc. Am. **93**, 2096–2105.

Hawkins, J., and Stevens, S. (**1950**). "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.

Hirsh, I., Davis, H., Silverman, S. R., Reynolds, E., Eldert, E., and Benson, R. W. (**1952**). "Development of materials for speech audiometry," J. Speech Hear Disord. **17**, 321–337.

Hogan, C., and Turner, C. (**1998**). "High-frequency audibility: Benefits for hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 432–441.

Hornsby, B., and Ricketts, T. (**2001**). "The effects of compression ratio, signal-to-noise ratio, and level on speech recognition in normal-hearing listeners," J. Acoust. Soc. Am. **109**, 2964–2973.

Horwitz, A. R., Ahlstrom, J. B., and Dubno, J. R. (**2004**). "Factors affecting the benefits of high-frequency speech audibility," poster presented at the International Hearing Aid Conference, Lake Tahoe, CA, August 2004.

Kalikow, D., Stevens, K., and Elliott, L. (**1977**). "Development of a test of speech intelligibility in noise using test material with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.

Leek, M. R., Dubno, J. R., He, N.-j., and Ahlstrom, J. B. (**2000**). "Experience with a yes-no single-interval maximum-likelihood procedure," J. Acoust. Soc. Am. **107**, 2674–2684.

Molis, M. R., and Summers, V. (**2003**). "Effects of high presentation levels on recognition of low- and high-frequency speech," ARLO **4**, 124–128.

Oxenham, A., and Bacon, S. P. (**2003**). "Cochlear compression: Perceptual measures and implications for normal and impaired hearing," Ear Hear. **24**, 352–366.

Oxenham, A., and Plack, C. (**1997**). "A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing," J. Acoust. Soc. Am. **101**, 3666–3675.

Pollack, I., and Pickett, J. (**1958**). "Masking of speech by noise at high sound levels," J. Acoust. Soc. Am. **30**, 127–130.

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 phase responses," J. Acoust. Soc. Am. **80**, 1364–1374.

Ruggero, M., Rich, N., Recio, A., Narayan, S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.

Shanks, J. E., Wilson, R. H., Larson, V., and Williams, D. (**2002**). "Speech recognition performance of patients with sensorineural hearing loss under unaided and aided conditions using linear and compression hearing aids," Ear Hear. **23**, 280–290.

Stockley, K. B., and Green, W. B. (**2000**). "Interlist equivalency of the Northwestern University Auditory Test No. 6 in quiet and in noise with adult hearing-impaired individuals," J. Am. Acad. Audiol **11**, 91–96.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Studebaker, G. A., Gilmore, C., and Sherbecoe, R. L. (**1993a**). "Performance-intensity functions at absolute and masked thresholds," J. Acoust. Soc. Am. **93**, 3418–3421.

Studebaker, G. A., Sherbecoe, R. L., and Gilmore, C. (**1993b**). "Frequency-importance and transfer functions for the Auditec of St. Louis recordings of the NU-6 word test," J. Speech Hear. Res. **36**, 799–807.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431–2444.

Tillman, T. W., and Carhart, R. (**1966**). "An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6," Technical Report No. SAM-TR-66-55, USAF School of Aerospace Medicine, Brooks Air Force Base, TX, pp. 1–12.
Turner, C., and Brus, S. L. (**2001**). "Providing low- and mid-frequency speech information to listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **109**, 2999–3006.
Wilson, R. H., Bell, T. S., and Koslowski, J. A. (**2003**). "Learning effects associated with repeated word-recognition measures using sentence materials," J. Rehabil. Res. Dev. **40**, 329–336.
Yates, G. (**1990**). "Basilar-membrane nonlinearity and its influence on auditory nerve rate-intensity functions," Hear. Res. **50**, 145–162.

Recognition of filtered words in noise at higher-than-normal levels: Decreases in scores with and without increases in masking

Judy R. Dubno,[a] Amy R. Horwitz, and Jayne B. Ahlstrom
Department of Otolaryngology-Head and Neck Surgery, Medical University of South Carolina, 135 Rutledge Avenue, P.O. Box 250550, Charleston, South Carolina 29425

(Received 17 January 2005; revised 14 April 2005; accepted 14 May 2005)

To examine spectral effects on declines in speech recognition in noise at high levels, word recognition for 18 young adults with normal hearing was assessed for low-pass-filtered speech and speech-shaped maskers or high-pass-filtered speech and speech-shaped maskers at three speech levels (70, 77, and 84 dB SPL) for each of three signal-to-noise ratios (+8, +3, and −2 dB). An additional low-level noise produced equivalent masked thresholds for all subjects. Pure-tone thresholds were measured in quiet and in all maskers. If word recognition was determined entirely by signal-to-noise ratio, and was independent of signal levels and the spectral content of speech and maskers, scores should remain constant with increasing level for both low- and high-frequency speech and maskers. Recognition of low-frequency speech in low-frequency maskers and high-frequency speech in high-frequency maskers decreased significantly with increasing speech level when signal-to-noise ratio was held constant. For low-frequency speech and speech-shaped maskers, the decline was attributed to nonlinear growth of masking which reduced the "effective" signal-to-noise ratio at high levels, similar to previous results for broadband speech and speech-shaped maskers. Masking growth and reduced "effective" signal-to-noise ratio accounted for some but not all the decline in recognition of high-frequency speech in high-frequency maskers.
 [DOI: 10.1121/1.1953127]

PACS number(s): 43.66.Dc, 43.71.Es, 43.71.Pc, 43.66.Sr [JHG] Pages: 923–933

I. INTRODUCTION

Under certain conditions, speech recognition decreases as speech and masker levels are increased above conversational levels (e.g., French and Steinberg, 1947; Hawkins and Stevens, 1950; Pollack and Pickett, 1958). Studebaker *et al.* (1999) observed a decline in word recognition at higher-than-normal speech and noise levels for listeners with and without hearing loss, and also noted that the magnitude of the adverse effects of high signal levels varied with speech material and masker type. To explore these effects, Dubno *et al.* (2000) measured recognition of key words in high- and low-context sentences from the Speech Perception in Noise test (SPIN; Kalikow *et al.*, 1977) over a wide range of speech and noise levels with effective signal-to-noise ratio for each subject held constant. The masker was spectrally shaped to match each subject's quiet thresholds and was adjusted as need to maintain equal signal-to-noise ratio with increasing noise level. Subjects were younger adults with normal hearing and older adults with minimally elevated thresholds. The results suggested that over the range of speech and noise levels used, key word recognition in sentences generally remained constant or decreased only slightly. In a second experiment (Dubno *et al.*, 2005), recognition of Northwestern University Test No. 6 (NU#6; Tillman and Carhart, 1966) monosyllabic words in a speech-shaped masker by young adults with normal hearing declined significantly with increasing speech level while a constant signal-to-noise ratio was maintained. Using the Articulation Index (AI), the deterioration in word recognition was attributed to nonlinear growth of masking in the speech-shaped masker which resulted in reduced "effective" signal-to-noise ratio with increasing signal levels.

In the past, to account for declining speech recognition at high signal levels, a correction factor was developed and ultimately incorporated into the Speech Intelligibility Index (ANSI, 1997) as the "level distortion factor." This correction factor is applied uniformly to each analysis band, rather than using different weights that account for the relative importance of various frequency bands to intelligibility. Applying a single correction factor uniformly across frequency assumes that the mechanism responsible for the deterioration of speech recognition at high levels has an equivalent detrimental effect regardless of the spectral content of the speech or masker. However, there is evidence that this assumption may not be correct. Studebaker and Sherbecoe (2002) reported that declines in recognition of narrowly filtered bands of NU#6 words in speech-shaped noise varied as a function of frequency. Consistent with these findings, Molis and Summers (2003) observed that recognition of key words in sentences in quiet by subjects with normal hearing declined at high levels more for high-pass-filtered sentences than for low-pass-filtered sentences. For normal-hearing subjects listening in quiet, Studebaker *et al.* (1999) found only very small declines with level in recognition of relatively broadband NU#6 words filtered from 0.447 to 2.239 kHz.

Additional, but conflicting, evidence concerning the dependence of word recognition at high levels on spectral content comes from studies that assessed benefit of amplification. Some studies suggest that the benefit of increasing

[a]Electronic mail: dubnojr@musc.edu

speech audibility through amplification varies as a function of frequency. For example, when speech in quiet was spectrally shaped and amplified to high levels for hearing-impaired listeners, recognition improved with increasing bandwidth for low and mid frequencies (e.g., Turner and Brus, 2001; Hogan and Turner, 1998; Ching *et al.*, 1998) but did not consistently improve for higher frequencies, and, in some cases, declined (Hogan and Turner, 1998; Ching *et al.*, 1998). Shanks *et al.* (2002) found that, in unaided conditions, recognition of speech in multitalker babble declined with increasing speech level for subjects with relatively mild hearing loss but not for subjects with more severe hearing loss. When speech and babble were amplified with a hearing aid (from ~1.0 to 4.0 kHz), recognition declined with level for nearly all subjects. The presence of background noise may also be a factor affecting the benefit of frequency-specific amplification. For example, contrary to their expectations based on results for speech in quiet, Turner and Henry (2002) found improvements in recognition with increasing bandwidth when speech was presented in background noise.

Several issues remain unresolved with regard to the detrimental effects of high speech and noise levels on speech recognition. First, the decline in speech recognition with increasing signal level varies in a complex manner with the spectral content of the speech and/or masker and may depend on the frequency regions that are audible. Thus, the magnitude of the reduction in speech recognition at high signal levels may differ between broadband and bandlimited speech and maskers, and between low-frequency and high-frequency speech and maskers. Second, although a decline in recognition at high levels for broadband speech in broadband maskers has been attributed to a reduction in speech audibility at high levels, this explanation may not be straightforward, or even suitable, for low-frequency and/or high-frequency speech and maskers. For broadband speech and maskers, the reduction in "effective" signal-to-noise ratio was a result of nonlinear growth of masking in the speech-shaped masker (Dubno *et al.*, 2005); masking patterns were consistent with compressive basilar-membrane responses. However, growth-of-masking slopes vary with signal frequency and masker spectrum, consistent with physiological and psychophysical evidence that compression is greater at higher than lower frequencies (e.g., Cooper and Yates, 1994; Plack and Oxenham, 2000), although there may be substantial compression even at low frequencies (Plack and Drga, 2003). Thus, different mechanisms may be responsible for the deterioration of broadband, low-frequency, and high-frequency speech recognition at high levels.

In this experiment, to examine spectral effects, word recognition was assessed for low-pass-filtered speech and speech-shaped maskers and for high-pass-filtered speech and speech-shaped maskers at three speech levels in each of three masker levels. With the exception of low- and high-pass filtering of speech and maskers, methods were the same as in Dubno *et al.* (2005). Constant signal-to-noise ratios were maintained across subjects and masker levels as speech level was increased. To quantify speech audibility, pure-tone thresholds across a wide range of frequencies were measured in each masker (three levels of the low-frequency masker and three levels of the high-frequency masker).

II. METHODS

A. Subjects

A total of 18 young adults participated, organized into two nine-person groups; one group listened to low-frequency speech and speech-shaped maskers and the other group listened to high-frequency speech and speech-shaped maskers. For the low-frequency group, subjects ranged in age from 19 to 30 years (mean age: 23.9 years); the comparable values for the high-frequency group were 19 to 32 years (mean age: 24.1 years). The 18 subjects were assigned to the two groups on an alternating basis (i.e., subjects 1, 3, 5, etc. were assigned to the low-frequency group and subjects 2, 4, 6, etc. were assigned to the high-frequency group). All subjects had thresholds ⩽15 dB HL (ANSI, 1996) at octave frequencies from 0.25 to 8.0 kHz and normal immitance measures. Test ear was selected randomly. Subjects did not have experience with the psychophysical task used in this study and thus received approximately 1 h of practice. Subjects did not have prior experience with the words contained in the NU#6 lists. Data collection was completed in five to six 2-h sessions, including appropriate rest periods. Subjects were paid an hourly rate for their participation.

B. Apparatus and stimuli

1. Tonal signals and noises

Tonal signals were digitally generated (TDT DA3-4) pure tones, sampled at 50 kHz and low-pass filtered at 12 kHz (TDT FT6). Signals were 350 ms in duration, including 10-ms raised-cosine rise/fall ramps.

For masking of low- or high-frequency speech, the masker was a low-pass-filtered or high-pass-filtered speech-shaped noise, respectively; these maskers were also used to measure masked thresholds for pure tones. To create maskers that were matched to the speech spectrum and were low- or high-pass filtered, first, a broadband noise was digitally generated at a sampling rate of 28 kHz. Its spectrum was then adjusted in 0.05-kHz intervals (from 0.05 to 11.5 kHz) using Matlab™ (Version 5.3, The Mathworks, Inc., Natick, MA) and in-house software so that the long-term spectrum of the masker matched the 1% levels of the NU#6 talker's speech ("speech-shaped masker"). For the low-frequency speech-shaped masker, the masker was then filtered from 0.16 to 1.41 kHz using Matlab™ and in-house software; for the high-frequency speech-shaped masker, the masker was filtered from 1.41 to 7.40 kHz using the same software. These maskers were presented at levels of 70, 77, and 84 dB SPL *re*: broadband level. Due to the shape of the speech spectrum, after filtering, the overall level of the low-frequency masker was 14 dB higher than the overall level of the high-frequency masker; however, within each passband, the one-third-octave band levels were equivalent to those in the broadband, speech-shaped masker.

To minimize the influence of small differences in quiet thresholds among subjects, a second masker was always present to ensure equal audibility. This low-level broadband

noise was digitally generated at a sampling rate of 28 kHz and then its spectrum adjusted at one-third-octave intervals (Cool Edit Pro™ Version 1.2, Syntrillium Software Corp., Scottsdale, AZ) to achieve masked thresholds of 20 dB HL for all subjects. The overall level of this "threshold-matching noise" (TMN) was 54 dB SPL. For conditions in which the speech-shaped masker was presented at 77 and 84 dB SPL, the TMN was presented at 61 and 68 dB SPL, respectively, to maintain constant audibility across conditions.

The low- or high-frequency speech-shaped maskers and the TMN were output through 16-bit digital-to-analog converters (TDT DA3-4), low-pass filtered at 12 kHz (TDT FT6), and recorded onto separate channels of a digital audio tape (DAT) for later playback. Spectral characteristics of all maskers were verified on an acoustic coupler and a signal analyzer (Stanford Research SR780). Maskers were always present during the measurement of pure-tone thresholds and word recognition.

2. Speech

Speech tokens were the 200 NU#6 monosyllabic words recorded by a male talker by Auditec of St. Louis, as used previously in Dubno *et al.* (2005) and in Studebaker *et al.* (1999). Filter cutoffs and software for speech were the same as described above for filtering of speech-shaped maskers. For low-frequency speech, each item was filtered from 0.16 to 1.41 kHz; for high-frequency speech, each item was filtered from 1.41 to 7.40 kHz. Due to the shape of the rms speech spectrum, after filtering, the overall level of the low-frequency speech was 16 dB higher than the overall level of the high-frequency speech; however, within each passband, the one-third-octave band levels were equivalent to those for unfiltered speech. The overall level difference between low- and high-frequency signals was slightly larger for speech than for the masker because the rms levels of the masker were matched to the 1% levels of the speech and resulted in a somewhat more-uniform spectrum than the rms levels of the speech. The 1.41-kHz cutoff frequency was selected to achieve equal word-recognition scores for low- and high-frequency speech and speech-shaped maskers, as determined by the AI. Speech and masker levels were selected so that word-recognition scores ranged from approximately 20% to 70%, avoiding floor and ceiling effects. NU#6 words were presented at three levels for each of three signal-to-noise ratios (+8, +3, and −2 dB). Thus, for low- and high-frequency speech and maskers, word recognition was measured in nine conditions, corresponding to all combinations of three signal-to-noise ratios and three levels of the speech-shaped masker (70, 77, and 84 dB SPL).

Digital speech waveforms were output through a 16-bit digital-to-analog converter (TDT DA3-4) and low-pass filtered at 12 kHz (TDT FT6). The amplitudes of all signals and maskers were controlled individually using programmable attenuators (TDT PA4). The signal was added to the speech-shaped masker and the TMN (TDT SM3), and delivered through one of a pair of TDH-49 earphones mounted in supra-aural cushions.

C. Procedures

Procedures were similar to those in Dubno *et al.* (2005) and are briefly reviewed here. For each subject, thresholds for pure tones from 0.2 to 6.3 kHz were measured in the following order: (1) thresholds in quiet; (2) masked thresholds in TMN at levels of 54, 61, and 68 dB SPL; and, (3) depending on subject group assignment, masked thresholds in the low- *or* high-frequency speech-shaped masker at levels of 70, 77, and 84 dB SPL, with TMN at 54, 61, and 68 dB SPL, respectively. Pure-tone thresholds were obtained using a single-interval (yes-no) maximum-likelihood psychophysical procedure (Green, 1993; Leek *et al.*, 2000). Signal level was varied adaptively.

Following measurement of pure-tone thresholds, recognition of 200 low-frequency or 200 high-frequency NU#6 words was measured at three speech levels in each of three levels of the low- or high-frequency speech-shaped masker (nine conditions for each subject). In each condition, the 200 words were presented randomly within four, 50-word blocks, corresponding to the four traditional 50-word NU#6 lists. Subjects responded by repeating aloud the word following the carrier phrase.

The experimental design required repeated presentations of the same 200 words in various conditions. In the previous study with a similar design and using nine repeated presentations of the 200 NU#6 words, several procedures were put in place to minimize procedural learning, familiarization, and trial number effects [see Dubno *et al.* (2005) for details]. Nevertheless, significant differences in word recognition were found among the nine trials; the largest difference in score for adjacent trials was between the first and second trials. In an attempt to reduce this effect, prior to data collection, subjects listened to 200 low- or high-frequency NU#6 words in the +8-dB signal-to-noise ratio, 70-dB masker condition (either low or high frequency, depending on group) and this score was discarded. This condition was selected because it was predicted to yield the highest word-recognition score. An additional concern was that, with spectrum as a repeated measure, each subject would hear the same words 18 times (nine low-frequency conditions and nine high-frequency conditions). Results of the previous study suggested that word recognition continued to improve over at least nine trials. For this reason, rather than include spectrum as a repeated measure, the 18 subjects were organized into two nine-person groups, with each group providing data for one of the two spectrum conditions. A 9×9 Latin-square design determined the order of the nine masker level/signal-to-noise ratio combinations for the nine subjects in each group to minimize effects of trial number.

D. Data analysis

Outcome measures included: (1) pure-tone thresholds in the low- and high-frequency speech-shaped masker at three levels; (2) observed low- and high-frequency word-recognition scores as a function of low- and high-frequency speech-shaped masker level and signal-to-noise ratio; (3) low- and high-frequency word-recognition scores predicted using the AI; and (4) differences between observed and pre-

dicted word-recognition scores. Differences between predicted and observed scores may reveal a role of nonlinear growth of masking in the decline of speech recognition at high levels to the extent that these differences are due to changes in "effective" masker levels resulting from nonlinear effects. AI values and predicted scores were computed using procedures similar to ANSI (1997), assuming a 40-dB effective dynamic range of speech (Studebaker *et al.*, 1999), and using the frequency importance function and AI-recognition transfer function for the Auditec of St. Louis recordings of the NU#6 word test (Studebaker *et al.*, 1993b). Using low- and high-frequency rau-transformed word-recognition scores (Studebaker, 1985), differences due to trial number were assessed by Latin-square ANOVAs and differences due to list, speech level, and signal-to-noise ratio were assessed by repeated measures ANOVAs (see the Appendix for additional details on effects of trial number and word list).

III. RESULTS AND DISCUSSION

A. Masked thresholds

Figure 1 shows mean thresholds for pure tones at one-third-octave frequencies from 0.2 to 6.3 kHz measured in the low-frequency speech-shaped masker (filled circles) with the masker level at 70 dB SPL (top panel), 77 dB SPL (middle panel), and 84 dB SPL (bottom panel). Also included in the three panels are the mean thresholds for the same signals measured in the TMN (open circles) with the TMN level at 54, 61, and 68 dB SPL, respectively. Mean thresholds in quiet are also shown in each panel (triangles). Standard error ranges exceed the size of the data points only for quiet thresholds.

In each panel, the dashed line is the one-third-octave rms spectrum of the low-frequency NU#6 words plotted at the three levels used with each low-frequency masker level (i.e., 68, 73, and 78 dB for the 70-dB masker; 75, 80, and 85 dB for the 77-dB masker; 82, 87, and 92 dB for the 84-dB masker). Within each panel, as the level of the low-frequency speech increased, the signal-to-noise ratio and audible speech in the low frequencies also increased, as seen by the relative positions of the masked thresholds (solid line with filled symbols) and the speech spectrum (dashed lines); by design, there was no audible speech in the high frequencies. Although these patterns are similar in all three panels, it is apparent that audible speech is decreasing slightly with increasing masker level. This is because as the low-frequency masker was increased by 7 dB (from 70 dB to 77 dB to 84 dB), masked thresholds tended to increase by slightly more than 7 dB at certain frequencies. The effect of this nonlinear increase in masked thresholds was to reduce "effective" signal-to-noise ratio within certain bands as low-frequency masker level was increased.

Figure 2 is organized in the same way as Fig. 1, but includes thresholds in high-frequency speech-shaped maskers (filled circles) and high-frequency speech spectra (dashed lines). Comparing masked thresholds in the high-frequency masker with the spectra of high-frequency speech suggests that audible high-frequency speech remained relatively constant as masker level increased. This is because as

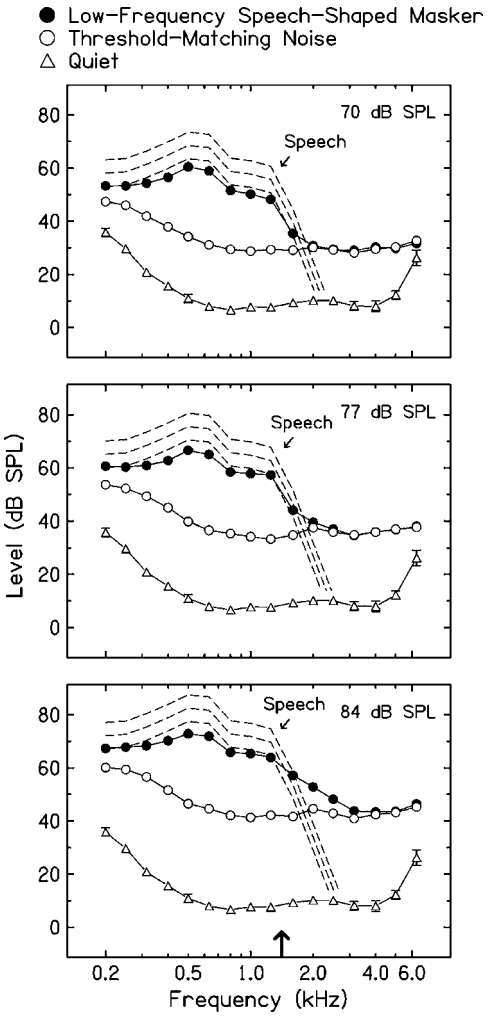


FIG. 1. Mean thresholds (±1 standard error, SE) for pure tones at one-third-octave frequencies from 0.2 to 6.3 kHz measured in the low-frequency speech-shaped masker (filled circles) with the masker level at 70 dB SPL (top panel), 77 dB SPL (middle panel), and 84 dB SPL (bottom panel). Also included are mean thresholds for the same signals measured in threshold-matching noise (open circles) at 54, 61, and 68 dB SPL, respectively. Mean thresholds in quiet are also shown in each panel (triangles). Standard error ranges are visible only for quiet thresholds. Each panel also includes the one-third-octave rms spectrum of the low-frequency NU#6 words plotted at the three levels used with each masker level (dashed lines). In the bottom panel, the arrow along the abscissa is plotted at the low-pass-filter cutoff frequency for the speech and speech-shaped masker.

the high-frequency masker was increased by 7 dB, masked thresholds within the higher frequencies also increased by ~7 dB.

To illustrate differences in growth of masking for the two maskers, Fig. 3 contains slopes of growth-of-masking functions for signal frequencies ranging from 0.2 to 6.3 kHz, measured in low-frequency speech-shaped maskers (top panel) and high-frequency speech-shaped maskers (bottom panel). Growth-of-masking slopes at each frequency measured in three levels of the speech-shaped masker were derived using linear regression. Mean slopes are shown by the filled circles. It is notable that, even for these young adults with normal hearing, there are considerable individual differences in growth-of-masking slope in both maskers.

For low-frequency speech-shaped maskers, linear growth of masking was observed at low frequencies, which

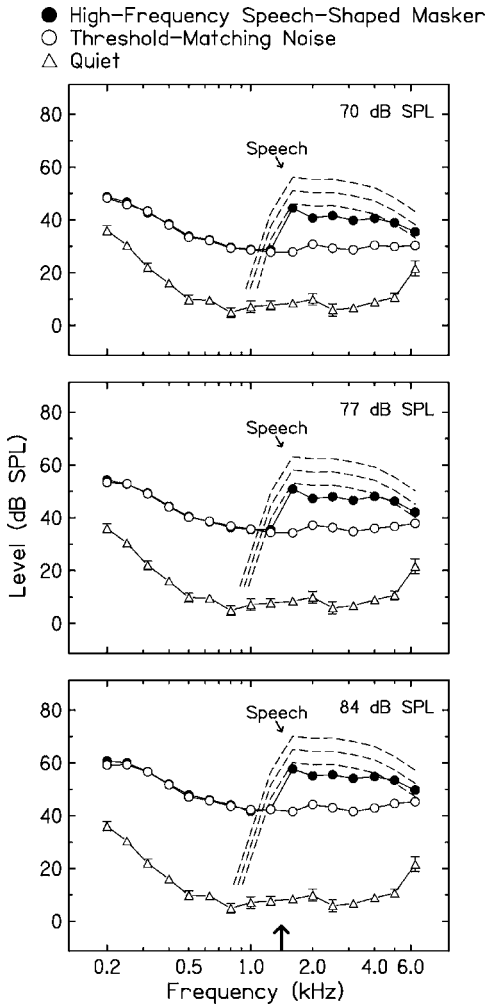


FIG. 2. Same as Fig. 1 but for thresholds in high-frequency speech-shaped maskers (filled circles) and the spectrum of the high-frequency NU#6 words (dashed lines). The arrow along the abscissa in the bottom panel is plotted at the high-pass-filter cutoff frequency for the speech and speech-shaped masker.

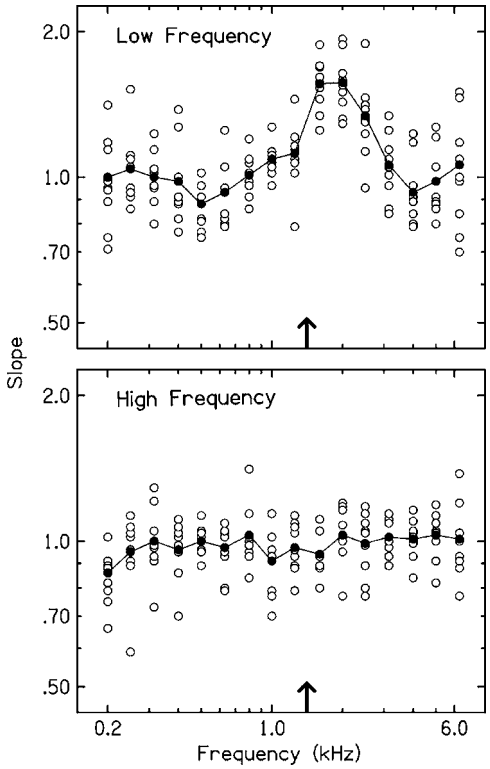


FIG. 3. Slopes of growth-of-masking functions for signal frequencies from 0.2 to 6.3 kHz measured in three levels of the low-frequency speech-shaped masker (top panel) and in three levels of the high-frequency speech-shaped masker (bottom panel), computed using linear regression. In both panels, mean slopes are shown by the filled circles and the arrow along the abscissa is plotted at the filter cutoff for the speech and speech-shaped masker.

is within the passband of the speech-shaped masker, and at very high frequencies, which is remote from the masker and where masked thresholds were likely determined by the TMN. Steep growth-of-masking slopes for the middle frequencies were consistent with nonlinear growth of upward spread of masking attributed to the low-to-mid-frequency peaks in the speech-shaped masker. A repeated-measures ANOVA revealed that frequency had a significant effect on growth-of-masking slope $[F(15,120)=17.09, p<0.0001]$. A *posthoc* test was significant for slopes peaking at 1.6 and 2.0 kHz, that is, slopes increasing from 0.2 to 1.6 kHz and decreasing from 2.0 to 6.3 kHz $[F(1,8)=83.14, p=0.00002]$. Slopes at 1.6 and 2.0 kHz were significantly higher than slopes at the remaining frequencies $[F(1,8)=311.45, p<0.0001]$. These frequencies coincided with the maxima of the frequency importance function for the NU#6 words (Studebaker *et al.*, 1993b), although there was little or no speech energy at 2.0 kHz due to low-pass filtering. In contrast, thresholds measured in the high-frequency speech-shaped masker at three levels generally revealed linear growth of masking across frequency, with slopes varying around 1.0 at frequencies within the passband of the high-frequency masker and at frequencies remote from the masker. Slopes did not vary significantly as a function of frequency $[F(15,120)=1.425, p=0.146]$. There was no evidence of downward spread of masking or remote masking.

Slopes of growth-of-masking functions depend on the relative levels and frequencies of the signal and the masker, with steeper slopes observed when the masker is lower in frequency than the signal and more shallow slopes observed when the masker is higher in frequency than the signal (e.g., Egan and Hake, 1950). Steep growth-of-masking functions have been viewed as reflecting the nonlinear growth of response at the basilar membrane, which may also vary with signal frequency (e.g., Bacon *et al.*, 1999). For the maskers in this experiment, differences in growth of masking for low- and high-frequency maskers may have also been a result of the particular spectral content of the speech-shaped masker, with its lower frequency peak and downslope at the higher frequencies. As a result, after low- and high-pass filtering, the overall level of the low-frequency masker was 14 dB higher than the overall level of the high-frequency masker. In addition, the spectrum of the low-frequency masker had peaks within its passband which may have contributed to nonlinear growth of upward spread of masking, whereas the spectrum of the high-frequency masker was relatively uniform.

In this experiment, growth of masking for low- and high-frequency maskers was important because it may have

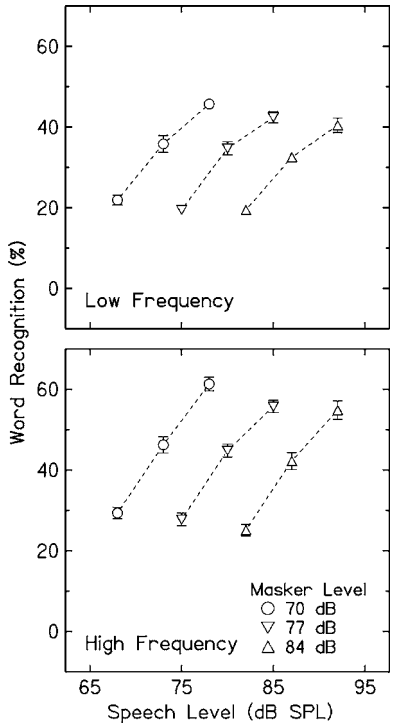


FIG. 4. Psychometric functions relating word recognition measured in speech-shaped maskers to speech level with masker level as the parameter. Scores for low-frequency speech in low-frequency maskers are in the top panel. Scores for high-frequency speech in high-frequency maskers are in the bottom panel. In each panel, mean word-recognition scores (±1 SE) obtained in the 70-dB masker (circles), 77-dB masker (reverse triangles), and 84-dB masker (triangles) are plotted at their respective speech levels.

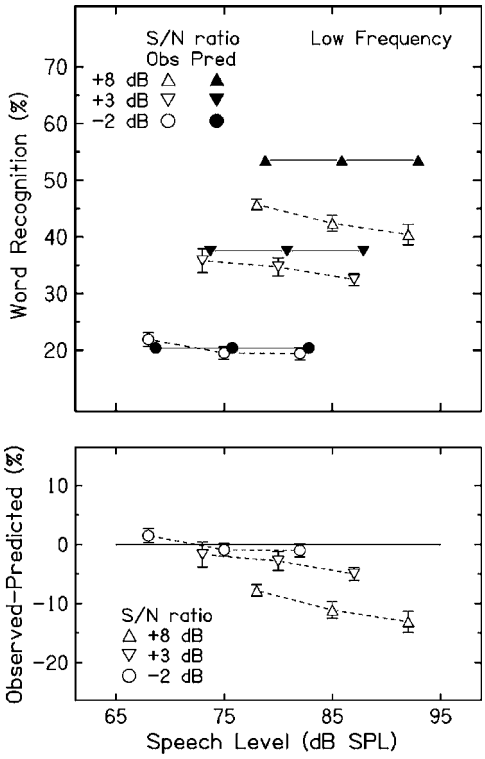


FIG. 5. Top: Observed low-frequency word-recognition scores from Fig. 4 replotted with signal-to-noise (S/N) ratio as the parameter (Obs, open symbols). Also shown are mean low-frequency word-recognition scores (±1 SE) predicted from AI values computed using each subject's thresholds in the *threshold-matching noise* (Pred, filled symbols). Bottom: Mean differences between observed and predicted scores (±1 SE) plotted as a function of low-frequency speech level with signal-to-noise (S/N) ratio as the parameter. For clarity, a solid line is drawn at an observed-predicted difference of 0% and some data points are offset along the abscissa.

differentially affected "effective" signal-to-noise ratios in low- and high-frequency regions and could explain differences between the declines in recognition of low- versus high-frequency speech. To obtain a single value with which to compare thresholds measured in the three low-frequency and three high-frequency masker levels, weighted average masked thresholds were computed using weights from the frequency importance function for the NU#6 stimuli (Studebaker *et al.*, 1993b). Computed in this way, weighted average thresholds take into account the relative importance of certain frequencies to word recognition. With the low-frequency masker increasing by 7 dB from 70 to 77 dB and from 77 to 84 dB, weighted average masked thresholds increased by 7.2 and 9.1 dB, respectively, consistent with nonlinear growth of masking. These changes in masked thresholds with increasing masker level reduced effective speech audibility at higher masker levels in spectral regions with relatively more importance to word recognition. In contrast, with the high-frequency masker increasing by 7 dB from 70 to 77 dB and from 77 to 84 dB, weighted average masked thresholds increased by only 6.5 and 5.8 dB, respectively.

B. Word recognition

1. Speech level and signal-to-noise ratio

Figure 4 shows psychometric functions for low-frequency speech in low-frequency maskers (top panel) and for high-frequency speech in high-frequency maskers (bottom panel), with masker level as the parameter. Mean word-recognition scores obtained in the 70-dB masker (circles), the 77-dB masker (reverse triangles), and the 84-dB masker (triangles) are plotted at their respective speech levels. The functions are generally parallel and, relative to the function for the 70-dB masker, the functions for the 77- and 84-dB maskers are shifted to the right by 7 and 14 dB, respectively, because speech levels were also increased by 7 dB (e.g., Studebaker *et al.*, 1993a). For each masker level, scores increased nonlinearly as speech level and signal-to-noise ratio increased. Although equivalent AI values for low- and high-frequency speech and maskers predicted equivalent scores, recognition of low-frequency speech in low-frequency maskers was always poorer than recognition of high-frequency speech in high-frequency maskers.

To illustrate the effect of speech level with signal-to-noise ratio held constant, the data in the top panel of Fig. 4 (low-frequency speech and maskers) were replotted with signal-to-noise ratio as the parameter and are shown in the top panel of Fig. 5 (open symbols). In the same way, the data in the bottom panel of Fig. 4 (high-frequency speech and maskers) were replotted and are shown in the top panel of Fig. 6 (open symbols). Word-recognition scores in Fig. 5 (top, open symbols) declined significantly as a function of low-frequency speech level $[F(2,16)=9.94, p=0.0016]$. *Posthoc* tests showed a significant linear trend $[F(1,8)=14.61, p=0.0051]$, suggesting that scores declined linearly

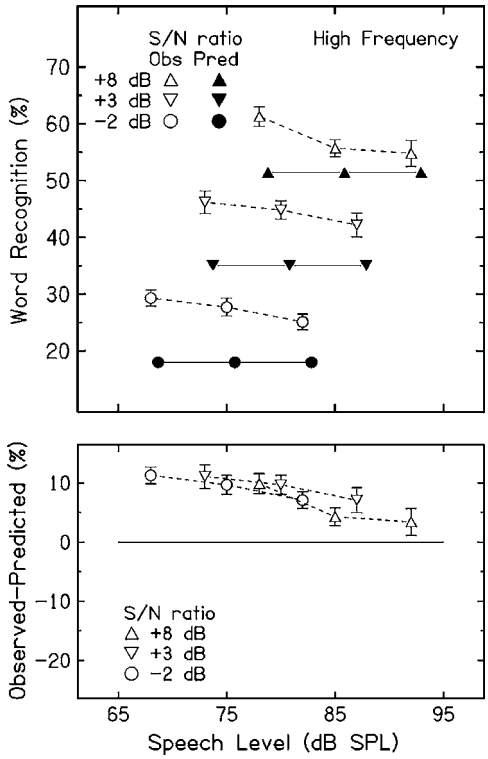


FIG. 6. Same as Fig. 5, but for observed and predicted high-frequency word-recognition scores (top panel) and observed-predicted differences (bottom panel).

with low-frequency speech level. The interaction of low-frequency speech level and signal-to-noise ratio was not statistically significant $[F(4,32)=1.10, p=0.373]$. For recognition of high-frequency speech, observed scores in Fig. 6 (top, open symbols) also declined significantly as a function of high-frequency speech level $[F(2,16)=23.20, p<0.0001]$; *posthoc* tests also showed a significant linear trend $[F(1,8)=30.93, p=0.0005]$. The interaction of high-frequency speech level and signal-to-noise ratio was statistically significant $[F(4,32)=2.81, p=0.042]$; scores declined significantly with high-frequency speech level for each of the three signal-to-noise ratios. The significant decline in scores confirmed that recognition of both low- and high-frequency speech in low- and high-frequency speech-shaped maskers decreased at high levels when signal-to-noise ratio was held constant. Comparing scores at the lowest and highest speech levels averaged across signal-to-noise ratio, the decline in scores was larger for high-frequency speech and maskers than for low-frequency speech and maskers (4.9% vs. 3.7%). If word recognition was determined entirely by signal-to-noise ratio, scores should have remained constant with increasing speech level.

The filled symbols in the top panels of Figs. 5 and 6 are predicted word-recognition scores. AI values predict scores from signal-to-noise ratios estimated from the levels and spectrum of the speech, each subject's thresholds measured in the TMN, and the levels and spectrum of the low- or high-frequency speech-shaped masker. Thus, given that these predicted scores do not take into account any changes in "effective" signal-to-noise ratio due to nonlinear growth of masking, they remain constant because, as low- and high-frequency speech level increased, masker level also increased, maintaining a constant signal-to-noise ratio for low- and high-frequency speech.

The bottom panel of Fig. 5 shows differences between observed and predicted word-recognition scores for low-frequency speech plotted as a function of speech level. The parameter is signal-to-noise (S/N) ratio. As in the top panel, predicted scores were estimated from AI values computed using thresholds measured in the TMN. When low-frequency word-recognition scores were predicted without taking into account changes in low-frequency speech audibility due to nonlinear growth of masking, observed scores decreased relative to predicted scores as speech level increased. That is, observed *minus* predicted values decreased significantly as a function of speech level $[F(2,16)=9.97, p=0.0015]$, with *posthoc* tests showing a significant linear trend $[F(1,8)=14.64, p=0.0050]$. The interaction of speech level and signal-to-noise ratio was not statistically significant $[F(4,32)=1.14, p=0.357]$. These results are consistent with recognition of low-frequency speech in low-frequency maskers declining at higher signal levels. These results are also similar to findings from the earlier study using broadband speech and speech-shaped maskers (Dubno *et al.*, 2005).

The bottom panel of Fig. 6 shows the comparable results for observed-predicted differences for recognition of high-frequency speech in high-frequency maskers. As in the top panel, predicted scores were estimated from AI values computed using thresholds measured in the TMN. The findings were consistent with a decline in word recognition at higher levels. Observed *minus* predicted values decreased significantly as a function of speech level $[F(2,16)=23.20, p<0.0001]$, with *posthoc* tests showing a significant linear trend $[F(1,8)=30.78, p=0.0005]$. The interaction of speech level and signal-to-noise ratio was statistically significant $[F(4,32)=2.78, p=0.043]$, but observed-predicted differences declined significantly with speech level for each of the three signal-to-noise ratios.

To assess how growth of masking in the low- or high-frequency speech-shaped maskers contributed to the decline in low- or high-frequency word recognition at high levels, predicted word-recognition scores were also determined using AI values computed using the "effective" noise spectrum measured empirically from subjects' thresholds in the low- or high-frequency masker. If the decline in word recognition with increasing speech level was entirely attributed to reduced audibility related to growth of masking in the low- or high-frequency maskers, predicted scores determined using low- or high-frequency masked thresholds should decline as speech level increased because constant "effective" signal-to-noise ratios were not maintained. Given that this is similar to the pattern seen in the observed scores, observed-predicted differences should remain constant with increasing level of the low- or high-frequency speech.

Figures 7 and 8 present these results. The filled symbols in the top panels are predicted recognition scores for low-frequency speech (Fig. 7) and high-frequency speech (Fig. 8) estimated from AI values computed using each subject's thresholds measured in the low- or high-frequency speech-

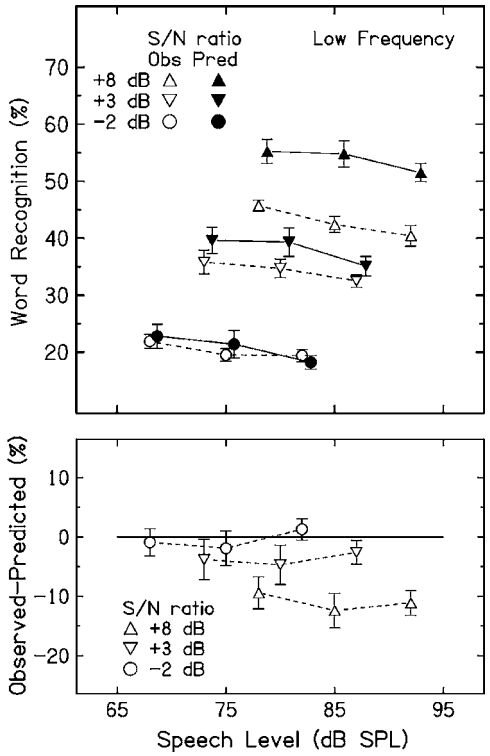


FIG. 7. Same as Fig. 5, but with predicted scores determined from AI values computed using each subject's thresholds measured in the low-frequency *speech-shaped masker.*

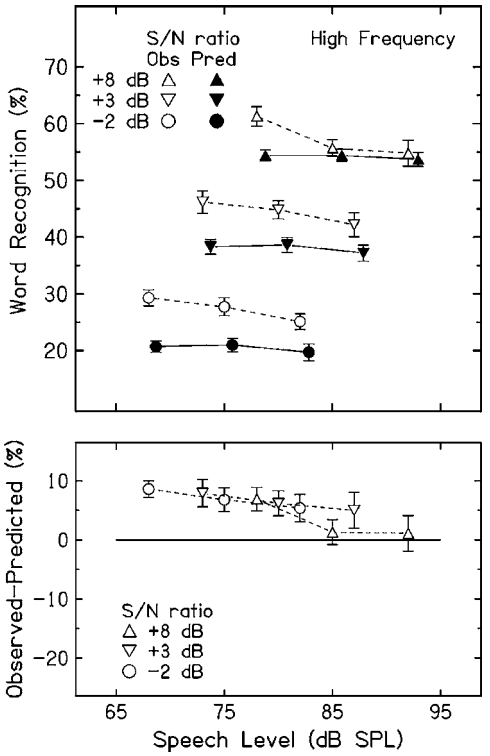


FIG. 8. Same as Fig. 5, but for observed and predicted high-frequency word-recognition scores (top panel) and observed-predicted differences (bottom panel). Predicted scores were determined from AI values computed using each subject's thresholds measured in the high-frequency *speech-shaped masker*.

shaped masker. The open symbols are observed scores, identical to those in the top panels of Figs. 5 and 6. The bottom panels of Figs. 7 and 8 show differences between observed and predicted word-recognition scores plotted as a function of speech level for low-frequency speech and high-frequency speech, respectively. The parameter is signal-to-noise (S/N) ratio.

Reviewing the results for low-frequency speech and maskers first (Fig. 7, bottom panel), when word-recognition scores were predicted while taking into account changes in speech audibility with increasing masker level, observed *minus* predicted values remained constant as speech level increased $[F(2,16)=0.79, p=0.471]$ and the interaction with signal-to-noise ratio was not statistically significant $[F(4,32)=0.149, p=0.229]$. Taken together, these results suggest that the decline in recognition of low-frequency speech at higher levels may be attributed to nonlinear growth of masked thresholds (and reduced "effective" signal-to-noise ratio) in the low-frequency speech-shaped masker, similar to results for recognition of broadband speech in broadband speech-shaped maskers (Dubno *et al.*, 2005). Moreover, given evidence that tonal growth-of-masking functions reflect basilar-membrane nonlinearities [see Oxenham and Bacon (2003) for review], these results also provide additional support for a role of nonlinear effects in the understanding of low-frequency speech in noise at various signal levels.

Turning to results for high-frequency speech and maskers (Fig. 8, bottom panel), when word-recognition scores were predicted using masked thresholds measured in the high-frequency masker rather than TMN, predicted scores declined only slightly at the highest speech level. With observed scores declining significantly in the high-frequency masker, observed *minus* predicted values also declined as speech level increased $[F(2,16)=4.92, p=0.0216]$; however, this significant difference was attributed to a significant decline only for observed-predicted differences for the +8 signal-to-noise ratio $[F(1,8)=13.87, p=0.0058]$. Thus, masking growth and the resulting decrease in "effective" signal-to-noise ratio in the high-frequency region did not entirely account for the decline in recognition at high levels. From another perspective, only very small deviations from linear growth of masking were observed for high-frequency maskers (Fig. 3). Nevertheless, even the relatively small reductions in high-frequency speech audibility accounted for the decline in recognition for two of the three signal-to-noise ratios (−2 and +3 dB, but not +8 dB).

Figure 9 shows word-recognition scores for low-frequency speech and maskers (top panel) and high-frequency speech and maskers (bottom panel) at three signal-to-noise (S/N) ratios plotted against AI, with AI values computed using thresholds measured in the speech-shaped maskers. The solid line is the normal transfer function relating the AI to word recognition established for the NU#6 word test used in this experiment (Studebaker *et al.*, 1993b); the dotted lines encompass the 95% confidence limits computed for 200-item word-recognition scores. As noted earlier, observed scores for low-frequency speech and maskers were poorer than observed scores for high-frequency speech and maskers. Consistent with these results, Fig. 9 shows that ob-

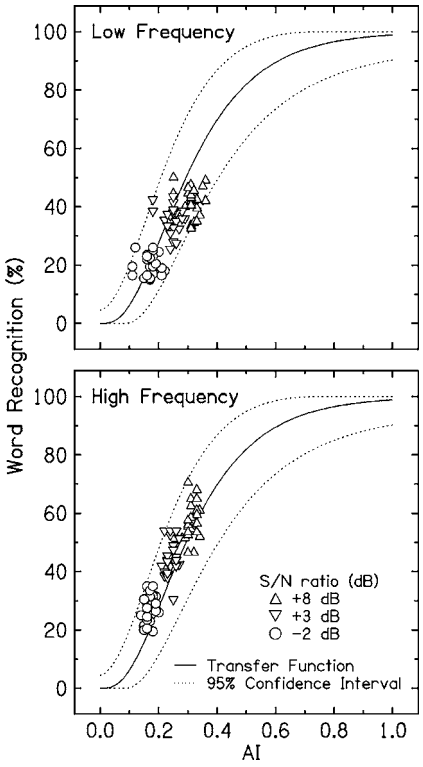


FIG. 9. Top: Low-frequency word-recognition scores at three signal-to-noise (S/N) ratios plotted against AI with AI values computed using thresholds measured in the low-frequency speech-shaped masker. The solid line is the AI-recognition transfer function for the Auditec of St. Louis recordings of the NU#6 word test (Studebaker *et al.*, 1993b); the dotted lines encompass the 95% confidence limits computed for 200-item word-recognition scores. Bottom: Same as the top panel, but for high-frequency word-recognition scores, with AI values computed using thresholds measured in the high-frequency speech-shaped masker.

served scores for low-frequency speech were generally poorer than predicted whereas observed scores for high-frequency speech were generally better than predicted (see also Figs. 7 and 8, bottom panels). Even with these trends and despite significant declines in word recognition with increasing speech level, nearly all of the observed scores fall within the 95% confidence interval of their predicted scores.

2. Differences in recognition of low- and high-frequency speech in noise at high levels

Scores for both low- and high-frequency speech in low- and high-frequency maskers declined with level while signal-to-noise ratio was held constant. Using audibility estimates based on the AI, the decline for low-frequency speech was attributed to nonlinear growth of masking in the low-frequency speech-shaped masker which resulted in reduced "effective" signal-to-noise ratio at high levels, as was previously observed for broadband speech and speech-shaped maskers (Dubno *et al.*, 2005). However, an unexpected finding was that masking growth accounted for some but not all the decline in recognition of high-frequency speech in high-frequency speech-shaped maskers at high levels.

Possible explanations for these results may relate to level-dependent differences in low- and high-frequency speech information and in the function of the auditory system for low- and high-frequency signals. For example, cues available in high-frequency speech may be more susceptible to deterioration at high levels than cues available in low-frequency speech. Higher frequency cues, such as for place information, may require more fine spectral resolution than some lower frequency cues, such as for voicing information or to distinguish among manner classes, which may be conveyed by periodicity or envelope cues. Thus, the normally broadened auditory filters at higher levels and at higher frequencies may have a greater detrimental effect on high-frequency speech cues than on low-frequency speech cues. Results from studies of auditory-nerve responses also suggest frequency-dependent effects for processing of high-level speech. For example, a loss of F2 synchrony, which could relate to auditory-nerve fibers' increasing response to F1, occurs at a lower level for high-frequency fibers than for low-frequency fibers (Wong *et al.*, 1998). Such differential effects of high signal levels on processing of low- and high-frequency speech information may be more easily revealed in recognition of bandlimited speech, rather than in recognition of broadband speech that contains an abundance of redundant cues. Indeed, for broadband speech in speech-shaped noise, declines in word recognition were entirely explained by reduced "effective" signal-to-noise ratio.

3. Differences in observed and predicted recognition scores for low- and high-frequency speech and maskers

Although bandwidths were selected to achieve nearly equal word-recognition scores for low- and high-frequency speech and speech-shaped maskers (i.e., equal AI values predicted equivalent scores), word-recognition scores averaged 10.3% higher for high-frequency speech than for low-frequency speech. Consistent with this result, observed scores for low-frequency speech and maskers were generally worse than predicted. As observed scores declined with increasing low-frequency speech level, they increasingly deviated from the predicted scores (see top panel of Fig. 5). Thus, observed-predicted differences became larger and more negative at higher levels (see bottom panel of Fig. 5), although no significant changes with level were observed when predictions accounted for reduced low-frequency speech audibility (Fig. 7). In contrast, for high-frequency speech and maskers, observed scores were generally better than predicted. As observed scores declined with increasing high-frequency speech level, they approached the predicted scores, making observed-predicted differences smaller and less positive at higher levels (Fig. 6). Changes with level were partially accounted for when predictions included reduced high-frequency speech audibility (Fig. 8).

These results may be related to the method used to compute the AI, from which the predicted scores were obtained. Based on recommendations in Studebaker *et al.* (1999) for NU#6 words and on accurate predictions in the previous study with broadband NU#6 words and speech-shaped maskers (Dubno *et al.*, 2005), speech peaks were fixed at 15 dB across frequency. However, the physical speech peaks (Sherbecoe *et al.*, 1993) and the effective speech peaks (Studebaker and Sherbecoe, 2002) for these materials in-

crease with increasing frequency. As a result, speech audibility may have been larger in the higher frequencies and smaller in the lower frequencies than originally estimated with speech peaks fixed across frequency. When computing AI values for broadband speech, effects of differences in speech peaks between lower and higher frequencies may be offset, which may explain why observed-predicted differences were small for broadband speech and speech-shaped maskers in Dubno *et al.* (2005), despite the use of 15-dB speech peaks. In the current study, however, with audibility limited to low- or high-frequency bands, effects of frequency-related differences in speech peaks were revealed, resulting in less accurate predictions and larger observed-predicted differences for low-frequency and high-frequency speech.

IV. SUMMARY AND CONCLUSIONS

Word recognition was assessed for 18 young adults with normal hearing for 200 NU#6 monosyllabic words in a masker shaped to match the spectrum of the NU#6 talker's speech. The speech and speech-shaped maskers were low- or high-pass filtered at 1.41 kHz. Nine subjects listened to low-frequency speech and maskers and nine subjects listened to high-frequency speech and maskers at three speech levels for each of three signal-to-noise ratios. An additional low-level noise was always present which was shaped to produce equivalent masked thresholds for all subjects. Pure-tone thresholds were measured in quiet and in all maskers. Results may be summarized as follows:

(1) Pure-tone thresholds measured in a low-frequency speech-shaped masker increased linearly at lower frequencies with increases in masker level. However, for mid-frequency signals, growth-of-masking slopes were consistent with nonlinear growth of upward spread of masking attributed to the low-to-mid-frequency peak in the speech-shaped masker. Such changes in masked thresholds with increasing masker level reduced "effective" signal-to-noise ratio at higher levels, especially at frequencies known to be important for recognition of NU#6 monosyllabic words.
(2) Masked thresholds measured in a high-frequency speech-shaped masker generally revealed linear growth of masking across frequency with increases in masker level. In addition to expected differences between upward and downward spread of masking, differences in growth of masking for low- and high-frequency maskers, may be due, in part, to the lower overall level and more uniform spectral shape of the high-frequency masker, as compared to the low-frequency masker.
(3) Recognition of low-frequency speech in low-frequency maskers and high-frequency speech in high-frequency maskers declined significantly with increasing speech level when signal-to-noise ratio was held constant.
(4) Using audibility estimates based on the AI, the decline in recognition of low-frequency speech in low-frequency maskers at high levels may be attributed to nonlinear growth of masking in the speech-shaped masker which resulted in reduced "effective" signal-to-noise ratio at high levels. This result is similar to that observed previously for broadband speech and speech-shaped maskers (Dubno *et al.*, 2005) and provides additional support for a role of nonlinear effects in the understanding of broadband and low-frequency speech in noise at various signal levels.
(5) Masking growth accounted for some but not all the decline in recognition of high-frequency speech in high-frequency maskers at high levels. Spectral effects on word recognition in noise at high levels may relate to level-dependent differences in processing of low- and high-frequency speech information.

ACKNOWLEDGMENTS

This work was supported (in part) by Grant Nos. R01 DC00184 and P50 DC00422 from NIH/NIDCD, the James E. and Pamela Knowles Foundation, and the MUSC General Clinical Research Center (M01 RR 01070). This investigation was conducted in a facility constructed with support from Research Facilities Improvement Program Grant No. C06 RR14516 from the National Center for Research Resources, National Institutes of Health. The authors thank Chris Ahlstrom for computer and signal-processing support, Fu-Shing Lee for advice on data analysis, and Gerald Studebaker for sharing digitized speech waveforms and the speech spectrum.

APPENDIX: EFFECTS OF TRIAL NUMBER AND WORD LIST

Mean word-recognition scores were computed as a function of trial number, for low- and high-frequency conditions. For each trial, 200-word scores were averaged over nine subjects with nine different conditions for each subject. Recall that a 9×9 Latin-square design determined the order of the nine conditions for the nine subjects in each group, providing a completely balanced design. A Latin-square ANOVA revealed no significant differences in word recognition due to trial for low-frequency conditions $[F(8,56)=1.99, p=0.0641]$ and high-frequency conditions $[F(8,56)=0.82, p=0.5902]$. Therefore, no adjustments were made for the trial effect in the repeated measures ANOVAs assessing effects of speech level and signal-to-noise ratio. These results also suggest that having subjects listen to a list of 200 low- or high-frequency NU#6 words prior to data collection was helpful in reducing the trial number effect observed previously.

For each list, 50-word scores were averaged over trial, condition, and subject. The order of the four 50-word blocks (corresponding to the four NU#6 word lists) was randomized for each subject. A one-way ANOVA with list as the repeated measure revealed significant differences in word recognition due to list for low-frequency conditions $[F(3,24)=49.21, p<0.0001]$ and high-frequency conditions $[F(3,24)=4.05, p=0.0183]$. *Posthoc* tests revealed that, for low-frequency conditions, scores for lists 1 and 2 and for lists 3 and 4 were not significantly different from each other, but scores for all other list pairs were significantly different. For high-frequency conditions, lists 2 and 3 were the only pair for which scores were significantly different. Nevertheless,

given that each score for each condition was the average of all 200 NU#6 words, these list-related differences did not contaminate condition effects. Note that when using the same stimuli and experimental design but broadband speech and speech-shaped maskers and higher average scores (Dubno *et al.*, 2005), each of the four lists was found to be significantly different from the other lists. This is consistent with results of other studies that suggest that list equivalence is dependent on condition difficulty (e.g., Stockley and Green, 2000; Stuart, 2004).

ANSI (**1996**). ANSI S3.6-1996, "American National Standard Specification for Audiometers" (American National Standards Institute, New York).

ANSI (**1997**). ANSI S3.5-1997, "American National Standard Methods for the Calculation of the Speech Intelligibility Index" (American National Standards Institute, New York).

Bacon, S. P., Boden, L. N., Lee, J., and Repovsch, J. L. (**1999**). "Growth of simultaneous masking for $f_m < f_s$: Effects of overall frequency and level," J. Acoust. Soc. Am. **106**, 341–350.

Ching, T., Dillon, H., and Byrne, D. (**1998**). "Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification," J. Acoust. Soc. Am. **103**, 1128–1140.

Cooper, N. P., and Yates, G. K. (**1994**). "Nonlinear input-output functions derived from the responses of guinea-pig cochlear nerve fibres: Variations with characteristic frequency," Hear. Res. **78**, 221–234.

Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (**2000**). "Use of context by younger and older adults with normal hearing," J. Acoust. Soc. Am. **107**, 538–546.

Dubno, J. R., Horwitz, A. R., and Ahlstrom, J. B. (**2005**). "Word recognition in noise at higher-than-normal levels: Decreases in scores and increases in masking," J. Acoust. Soc. Am. **118**, 914–922.

Egan, J. P., and Hake, H. W. (**1950**). "On the masking pattern of a simple auditory stimulus," J. Acoust. Soc. Am. **22**, 622–630.

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

Green, D. M. (**1993**). "A maximum-likelihood method for estimating thresholds in a yes-no task," J. Acoust. Soc. Am. **93**, 2096–2105.

Hawkins, J., and Stevens, S. (**1950**). "The masking of pure tones and of speech by white noise," J. Acoust. Soc. Am. **22**, 6–13.

Hogan, C., and Turner, C. (**1998**). "High-frequency audibility: Benefits for hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 432–441.

Kalikow, D., Stevens, K., and Elliott, L. (**1977**). "Development of a test of speech intelligibility in noise using test material with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.

Leek, M. R., Dubno, J. R., He, N.-j., and Ahlstrom, J. B. (**2000**). "Experience with a yes-no single-interval maximum-likelihood procedure," J. Acoust. Soc. Am. **107**, 2674–2684.

Molis, M. R., and Summers, V. (**2003**). "Effects of high presentation levels on recognition of low-and high-frequency speech," ARLO **4**, 124–128.

Oxenham, A., and Bacon, S. P. (**2003**). "Cochlear compression: Perceptual measures and implications for normal and impaired hearing," Ear Hear. **24**, 352–366.

Plack, C., and Drga, V. (**2003**). "Psychophysical evidence for auditory compression at low characteristic frequencies," J. Acoust. Soc. Am. **113**, 1574–1586.

Plack, C. J., and Oxenham, A. (**2000**). "Basilar-membrane nonlinearity estimated by pulsation threshold," J. Acoust. Soc. Am. **107**, 501–507.

Pollack, I., and Pickett, J. (**1958**). "Masking of speech by noise at high sound levels," J. Acoust. Soc. Am. **30**, 127–130.

Shanks, J. E., Wilson, R. H., Larson, V., and Williams, D. (**2002**). "Speech recognition performance of patients with sensorineural hearing loss under unaided and aided conditions using linear and compression hearing aids," Ear Hear. **23**, 280–290.

Sherbecoe, R. L., Studebaker, G. A., and Crawford, M. R. (**1993**). "Speech spectra for six recorded monosyllabic word tests," Ear Hear. **14**, 104–111.

Stockley, K. B., and Green, W. B. (**2000**). "Interlist equivalency of the Northwestern University Auditory Test No. 6 in quiet and in noise with adult hearing-impaired individuals," J. Am. Acad. Audiol **11**, 91–96.

Stuart, A. (**2004**). "An investigation of list equivalency of the Northwestern University Auditory Test No. 6 in interrupted broadband noise," Am. J. Audiol. **13**, 23–28.

Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Studebaker, G. A., and Sherbecoe, R. L. (**2002**). "Intensity-importance functions for bandlimited monosyllabic words," J. Acoust. Soc. Am. **111**, 1422–1436.

Studebaker, G. A., Gilmore, C., and Sherbecoe, R. L. (**1993a**). "Performance-intensity functions at absolute and masked thresholds," J. Acoust. Soc. Am. **93**, 3418–3421.

Studebaker, G. A., Sherbecoe, R. L., and Gilmore, C. (**1993b**). "Frequency-importance and transfer functions for the Auditec of St. Louis recordings of the NU-6 word test," J. Speech Hear. Res. **36**, 799–807.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., and Gwaltney, C. A. (**1999**). "Monosyllabic word recognition at higher-than-normal speech and noise levels," J. Acoust. Soc. Am. **105**, 2431–2444.

Tillman, T. W., and Carhart, R. (**1966**). "An expanded test for speech discrimination utilizing CNC monosyllabic words: Northwestern University Auditory Test No. 6," Technical Report No. SAM-TR-66-55, USAF School of Aerospace Medicine, Brooks Air Force Base, TX, pp. 1–12.

Turner, C., and Brus, S. L. (**2001**). "Providing low- and mid-frequency speech information to listeners with sensorineural hearing loss," J. Acoust. Soc. Am. **109**, 2999–3006.

Turner, C., and Henry, B. A. (**2002**). "Benefits of amplification for speech recognition in background noise," J. Acoust. Soc. Am. **112**, 1675–1680.

Wong, J. C., Miller, R. L., Calhoun, B. M., Sachs, M. B., and Young, E. D. (**1998**). "Effects of high sound levels on responses to the vowel "eh" in cat auditory nerve," Hear. Res. **123**, 61–77.

Pitch shifts for complex tones with unresolved harmonics and the implications for models of pitch perception

Rebecca K. Watkinson,[a] Christopher J. Plack, and Deborah A. Fantini
Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom

(Received 4 May 2004; revised 22 March 2005; accepted 4 May 2005)

Complex tone bursts were bandpass filtered, 22nd–30th harmonic, to produce waveforms with five regularly occurring envelope peaks ("pitch pulses") that evoked pitches associated with their repetition period. Two such tone bursts were presented sequentially and separated by an interpulse interval (IPI). When the IPI was varied, the pitch of the whole sequence was shifted by between +2% and −5%. When the IPI was greater than one period, little effect was seen. This is consistent with a pitch mechanism employing a long integration time for continuous stimuli that resets in response to temporal discontinuities of greater than about one period of the waveform. Similar pitch shifts were observed for fundamental frequencies from 100 to 250 Hz. The pitch shifts depended on the IPI duration *relative* to the period of the complex, not on the *absolute* IPI duration. The pitch shifts are inconsistent with the autocorrelation model of Meddis and O'Mard [J. Acoust. Soc. Am. **102**, 1811–1820 (1997)], although a modified version of the weighted mean-interval model of Carlyon *et al.* [J. Acoust. Soc. Am. **112**, 621–633 (2002)] was successful. The pitch shifts suggest that, when two pulses occur close together, one of the pulses is ignored on a probabilistic basis. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1940467]

PACS number(s): 43.66.Hg, 43.66.Ba [NFV] Pages: 934–945

I. INTRODUCTION

Unresolved harmonics are high-numbered harmonics (greater than about 10) that cannot be separated out by the auditory system (Plomp, 1964; Plomp and Mimpen, 1968; Bernstein and Oxenham, 2003). The auditory system can determine the pitch associated with unresolved harmonics only by using timing information from the interaction of the harmonics in the cochlea (Licklider, 1956). The interaction produces a complex waveform that repeats at the fundamental frequency (*F*0). It is known that auditory-nerve fibers phase lock to the envelope of a complex tone consisting of high harmonics (Joris and Yin, 1992). Phase locking of spikes gives rise to patterns of interspike intervals (ISIs) that reflect stimulus periodicities, including interpulse intervals (IPIs) in the complex tone. IPIs are the time intervals between one peak in the envelope and the next peak, where envelope is defined as a smooth curve passing through the peaks of a waveform.

There is some controversy regarding how the auditory system uses the periodicities in a waveform to determine pitch. Autocorrelation models of pitch (de Cheveigné, 1998; Licklider, 1951; Meddis and Hewitt, 1991; Meddis and O'Mard, 1997; Slaney and Lyon, 1990) propose that the auditory system uses information in the form of distributions of all-order ISIs (that is, intervals between any two spikes in the sequence) to determine the pitch that is heard. The ISIs are dependent on, but not necessarily equal to, the IPIs of the complex tone. Pitch is determined by the ISIs that are most prominent over the entire population of Type 1 auditory nerve fibers at a given time. Mean-interval models (Carlyon, 1997; Carlyon *et al.*, 2002), however, propose that the auditory system effectively *averages* the IPIs (assumed to be coded veridically in the ISIs) in some way to produce pitch. It is reasonable to suggest that the IPIs (the intervals between the peaks of the envelope) may determine pitch. Schouten *et al.* (1962) suggested that pitch was not determined by the period of the envelope. Listeners matched the pitch of a three-component complex tone to another three-component complex tone in which the same three harmonics had all been shifted by a constant amount. They did not match the pitch of three components, for example 1840, 2040, and 2240 Hz, to the pitch associated with the period of the envelope, in this example 200 Hz. However, Moore and Moore (2003) have suggested that this result may be due to listeners matching the excitation patterns of the stimuli, rather than by matching the low pitch. When the spectral region was held constant, shifting the frequencies of the individual unresolved harmonics by the same amount did not produce pitch shifts. Carlyon *et al.* (2002) proposed a mean-interval model that weighted longer intervals more heavily than shorter intervals. They showed that a "4–6" pulse train (that is, a pulse train with alternating IPIs of 4 and 6 ms) bandpass filtered between 3900 and 5300 Hz, is matched to an isochronous pulse train with an IPI of approximately 5.5 to 5.7 ms by normally hearing participants and cochlear implant users. The listeners did not appear to hear a pitch corresponding to the first-order IPIs of 4 or 6 ms or to the second-order IPI of 10 ms. The model of Carlyon *et al.* can account for this finding by determining a weighted mean of the 4- and 6 -ms IPIs. The 6-ms interval is weighted more heavily so that the mean is skewed towards the longer interval.

[a]Electronic mail: rwatki@essex.ac.uk

 0001-4966/2005/118(2)/934/12/$22.50

The Carlyon *et al.* (2002) model is also able to reproduce the data from Plack and White (2000). Plack and White studied the contribution of a single IPI to the pitch of a complex tone with unresolved harmonics by producing stimuli that consisted of two 20-ms tone bursts of a 250-Hz *F*0, that were separated by a 0-, 1-, or 2-period silent interval. The envelope of the second tone burst was phase shifted relative to the first. This caused the IPI between the bursts to be varied. Plack and White (2000) observed that, when the silent interval between the bursts was 0 period, envelope phase shifts of +0.75, +0.25, and −0.75 periods (where positive numbers represent a phase delay) produced a negative pitch shift for the tone-burst pair, and the envelope phase shift of −0.25 periods produced a slightly positive pitch shift. These results indicate that changing a single IPI can cause a shift in pitch. Pitch shifts were not observed when the silent interval between the tone bursts was 1 period or greater.

For a short tone burst, the IPIs are integrated together to produce a single pitch percept. Therefore, not only does the question of how the IPIs are used to determine pitch need to be answered, but also the question of how the information about pitch is integrated over time. Initially, White and Plack (1998) suggested two hypotheses of temporal integration for pitch: that the pitch mechanism uses a fixed, long-duration sampling window (>40 ms) or that the pitch mechanism uses short sampling windows (<20 ms for a 250-Hz *F*0, and <40 ms for a 62.5-Hz *F*0). In the latter hypothesis, the discrete pitch estimates may be combined to improve the pitch estimate, as in the "multiple-looks" model (Viemeister and Wakefield, 1991). White and Plack (1998) found that the improvement in *F*0 discrimination with increasing duration for complex tones with unresolved harmonics is much greater than predicted by the multiple-looks model, suggesting long integration. However, they also found that when a brief silent interval was introduced into a tone burst, then *F*0 discrimination was much worse than when the tone burst was continuous. White and Plack (1998) suggested a third hypothesis: that the pitch mechanism for complex tones with unresolved harmonics uses a long integration window, which is reset in response to temporal discontinuities. Previously, Bregman *et al.* (1994a, 1994b) had provided a theoretical background for this third hypothesis. They suggested that the auditory system may use neural onset and offset responses to reset pitch processing and to aid auditory scene analysis.

Nabelek (1996) showed that a phase change between two short pure tones produced a change in pitch. However, when the silent interval between the two tone bursts reached a "critical pause duration" of between 8 and 16 ms, this effect was no longer seen; the two tone bursts appeared to be processed separately. As described above, Plack and White (2000) employed a similar technique to Nabelek (1996), but used complex tone bursts with unresolved harmonics. They suggested that a silent interval of around 8 ms between two tone bursts of 250-Hz *F*0 may be sufficient for resetting. For silent intervals longer than this, envelope phase shifts had no effect on pitch. The present study investigated whether the critical pause duration has an absolute value or whether it is dependent upon *F*0.

Experiment 1 replicated and extended the findings of Plack and White (2000). A wider range of phase shifts was utilized, including +0.5 and −0.5 periods in addition to those phase shifts used by Plack and White (2000). In addition, complex tone bursts of four *F*0s were used, 100, 125, 166.7, and 250 Hz. Using four *F*0s produces tone bursts with a range of IPI durations, allowing the suggestion of Carlyon *et al.* (2002), that longer intervals are weighted more heavily than shorter ones, to be tested. Experiment 2 was conducted to investigate one of the findings of experiment 1. A method of constant stimuli was used in experiment 2 to ensure that the pattern of pitch shifts produced by the envelope phase changes was not an artifact of the pitch-matching procedure used in experiment 1. Only a subset of the stimuli from experiment 1 was used. Experiment 3 was conducted to investigate whether the negative pitch shifts produced when two pulses occurred close together was due to the IPI being partially filled by basilar-membrane (BM) ringing.

The aims of these experiments were

(i) To determine whether the results of Plack and White (2000) can be extended to a range of *F*0s.
(ii) To test the mean-interval model and the weighted mean-interval model of Carlyon *et al.* (2002).
(iii) To determine whether the critical pause duration has an absolute value or whether it is dependent upon *F*0.

II. EXPERIMENT 1

A. Stimuli

Stimuli consisted of complex-tone bursts with *F*0s of 100, 125, 166.7, and 250 Hz. Each harmonic had a level of 50 dB SPL before filtering. The harmonics were added in sine phase relative to each other (i.e., positive-going zero crossings were aligned) and bandpass filtered (digital FIR) between the 22nd and the 30th harmonic (3-dB downpoints, 90-dB/octave slope). The overall level of the stimuli was therefore about 60 dB SPL. All filtering occurred after gating. Each tone burst had a duration of 5 waveform cycles (i.e., 50, 40, 30, and 20 ms for *F*0s of 100, 125, 166.7, and 250 Hz, respectively), and was gated with no ramps.

The comparison interval consisted of two tone bursts presented consecutively and separated by a silent interval of 0, 1, or 2 waveform periods (where the periods were 10, 8, 6, and 4 ms for *F*0s of 100, 125, 166.7, and 250 Hz, respectively). The standard interval ("experimental" interval) was the same except that the IPI between the bursts (the "central" IPI) was varied by producing a phase change in the envelope of the second tone burst relative to the first. The (nominal) waveform before filtering, for a 250-Hz *F*0 complex tone, was given by

$$a(t) = A\sum_{n=1}^{\infty} \sin[n(2\pi F0t + \theta)], \quad (0 \leqslant t < 5/F0), \qquad (1)$$

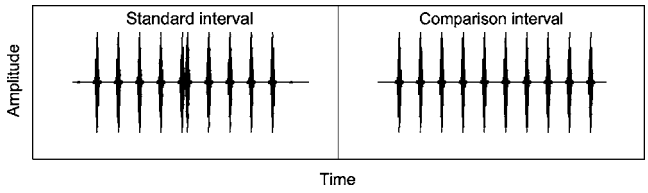


FIG. 1. The waveforms (after filtering) of the comparison and standard intervals used in experiment 1.

$$a(t) = A\sum_{n=1}^{\infty} \sin[n(2\pi F0t + \theta + \varphi)],$$
$$[(G + 5/F0) \leqslant t < (G + 0.04)], \quad (2)$$

where t is time in seconds, $a(t)$ is the waveform of the complex, A is the peak amplitude of each harmonic, n is the harmonic number, $F0$ is $F0$ in Hz, θ is the starting phase of the $F0$ component ($n=1$) of the first tone burst, and φ is the starting phase of the $F0$ component of the second tone burst relative to the first. The relative values of θ and φ were varied to increase or decrease the interval between the last envelope peak of the first burst and the first envelope peak of the second burst. G is the silent interval duration, in seconds. Figure 1 shows the waveforms after filtering of the standard interval (with a −0.75-period phase shift) and the comparison interval. Figure 2 shows the spectra of the comparison interval stimuli for the four $F0$s. Figure 3 shows the waveforms of the 0-period silent interval conditions with an $F0$ of 100 Hz. This figure shows the effect on the waveforms of the various phase shifts.

The tone bursts were presented in noise (generated independently for each interval) with a spectrum level of 15 dB. The noise was low-pass filtered at 1000 Hz for the tones with an $F0$ of 100 Hz, at 1250 Hz for the tones with $F0$s of 125 and 166.7 Hz, and at 2500 Hz for the tones with an $F0$ of 250 Hz. The noise was gated on (no ramps) 50 ms before the onset of the tone bursts, and gated off 50 ms after the offset of the tone bursts. The noise was presented to mask low-frequency combination tones.

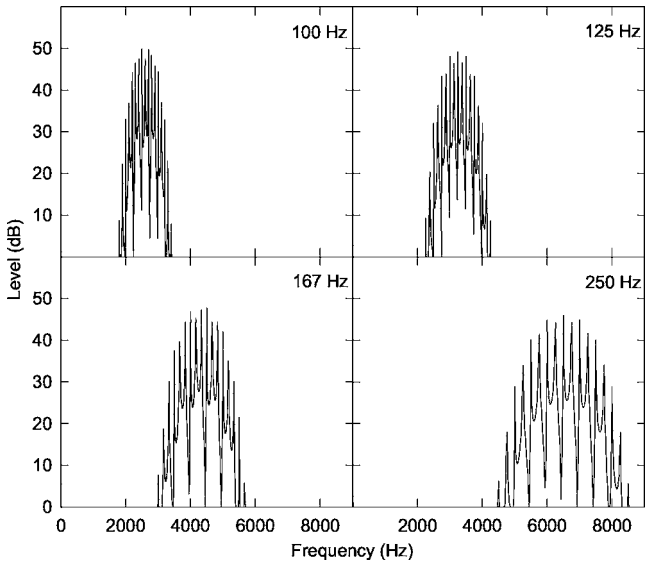


FIG. 2. The spectra of the comparison interval stimuli for the four $F0$s (100, 125, 166.7, and 250 Hz) used in experiment 1.

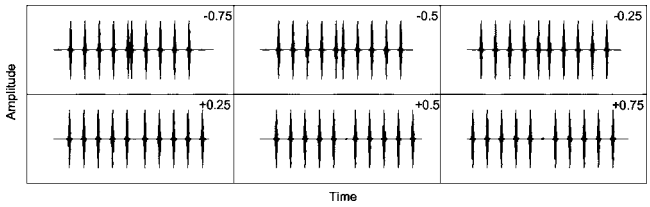


FIG. 3. The waveforms for the 0-period silent interval conditions with an $F0$ of 100 Hz used in experiment 1. Each panel shows the waveform for a single phase-shift condition (indicated, in periods, by the number in the top right of each panel).

Stimuli were generated digitally with 32-bit resolution and were output via an RME Digi96/8 PAD 24-bit soundcard. The sampling rate was 48 kHz. The stimuli were presented to the right ear via Sennheiser HD 580 headphones. The output from the sound card was fed to the headphones via a patch panel in the sound booth without filtering or amplification.

B. Procedure

Pitch matches were obtained using the adaptive procedure described by Jesteadt (1980) and by Plack and White (2000). For each trial, two listening intervals (separated by 500 ms) were presented. Each listening interval contained a tone-burst pair. The standard interval contained the shifted stimulus with an $F0$ of 100, 125, 166.7, or 250 Hz. The comparison interval contained a similar stimulus without the IPI shift (preserved envelope phase), but with the same silent interval as the standard. The $F0$ of the comparison tone burst pair was varied adaptively using two interleaved adaptive tracks. Initially, the $F0$ of the comparison tone burst pair was 20% above (higher track) or 20% below (lower track) the $F0$ of the standard tone burst pair.

Listeners were presented with the standard and a comparison (with either a higher or lower $F0$), in a random order. The listener's task was to decide which interval contained the higher-pitched sound. The higher track used a "two-down, one-up" rule. If the listener chose the comparison interval twice in succession, then the $F0$ of the comparison was decreased for the next trial. If the listener chose the standard interval once, then the $F0$ of the comparison was increased for the next trial. The lower track used a "two-up, one-down" rule. If the listener chose the comparison interval once, then the $F0$ of the comparison was decreased for the next trial. If the listener chose the standard interval twice in succession, then the $F0$ of the comparison was increased for the next trial. The $F0$ of the comparison interval was increased or decreased by 4% of the $F0$ of the standard for the first four reversals, and by 2% thereafter.

Each trial was selected randomly from the higher or lower track. Testing continued until the comparison interval had changed from an increasing to decreasing $F0$, or vice versa, 16 times for each track. The threshold of the track was calculated from the mean of the comparison $F0$s (in percent relative to the $F0$ of the standard tone burst) for the last 12

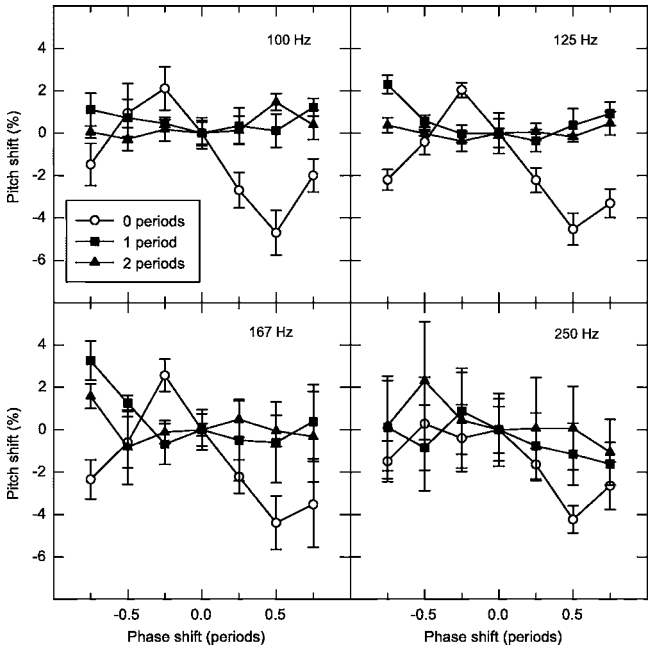


FIG. 4. Results of experiment 1 showing mean pitch shifts (%) for tone burst pairs of *F*0s 100, 125, 166.7, and 250 Hz as a function of the phase shift between the bursts. Error bars show standard errors between listeners.

transitions for that track. The two tracks bracketed the pitch match. Effectively, the upper track converged on the 70.7% point on the psychometric function for detecting an increase in *F*0, and the lower track converged on the 29.3% point for detecting an increase in *F*0 (this is also the 70.7% point for detecting a decrease in *F*0). The pitch match was taken as the mean of the thresholds for the upper and lower tracks. Half the difference between the thresholds from the upper and lower tracks was taken as an estimate of the *F*0 difference limen (*F*0DL, the smallest detectable change in the *F*0).

All the blocks for a single *F*0 were presented together. The order in which the *F*0s were presented was random and, for each *F*0, the conditions were also randomized. Eight pitch matches were collected for each condition. The final estimate was the arithmetic mean of these eight.

Listeners sat in a double-walled sound-attenuated booth and the responses were recorded via a computer keyboard. Listeners could see a computer monitor, through a window in the sound booth, on which "lights" were presented concomitantly with each stimulus interval. No feedback was given. The participants were tested for a maximum of 2 hours consecutively with breaks, although it was more usual for participants to test for 1 hour.

C. Listeners

Four normally hearing listeners were tested. They were aged between 22 and 32 years. Listeners were given at least 2 hours training on the task before data collection began.

D. Results

The pattern of results was consistent across listeners, so the mean pitch matches are shown in Fig. 4. In Fig. 4 the mean percentage difference between the *F*0 of the comparison tone pair and the nominal *F*0 of the shifted tone pair is plotted against the phase shift in periods. The pitch shifts are plotted relative to the control condition, 0 phase shift, to correct for any biases in the pitch matching procedure (see Plack and White, 2000). The mean standard errors for each data series in Fig. 4 are given in Table I.

Only the 0-period silent interval conditions produced pitch matches for the phase-shift conditions that differed greatly from the no-shift control. The −0.75, +0.25, +0.5, and +0.75 phase-shift conditions all produced a negative pitch shift. The −0.25 phase-shift condition produced a positive pitch shift. This pattern is seen for all the *F*0s, except for the 250-Hz *F*0 tone bursts; these data were also more variable. (The *F*0DLs for the 250-Hz *F*0 were much greater than for the other *F*0s, see Fig. 5.) For the three lower *F*0s, the pattern of pitch shifts obtained seems to be dependent on the duration of the central IPI *relative to the waveform period*, not on the absolute IPI in ms. For example, the phase shift that produced the maximum positive pitch shift was −0.25 periods for both 100- and 166.7-Hz *F*0s, even though the respective absolute IPIs between the bursts for this shift were 7.5 and 4.5 ms, respectively. It should be noted that both the duration of the stimulus and the spectral region of the stimuli vary with *F*0, and either of these variables could influence the pitch shifts. It cannot be stated conclusively that the pattern of pitch shifts is determined by the envelope phase alone.

The pattern of results for the 0-period silent interval conditions seems to follow a cubic trend, and this does not appear to be the case for the conditions with a 1- or 2-period silent interval. To test whether the pattern of results is statistically significant, cubic trend analysis was carried out. A significant cubic trend indicates that the values of the pitch shift change with phase shift following a cubic function. Cubic trend analysis showed a significant cubic trend for the 0-period silent interval conditions when the *F*0 of the complex tones was 100, 125, or 166.7 Hz (for the 100-Hz *F*0 condition, $F=220.6$, $p<0.05$; for the 125-Hz *F*0 condition,

TABLE I. The mean standard error (in dB) between listeners for the data in Figs. 4 and 5.

	Pitch shifts (Fig. 4)			*F*0DLs (Fig. 5)		
*F*0	0-period	1-period	2-period	0-period	1-period	2-period
100	0.94	0.66	0.54	1.45	2.27	2.17
125	0.53	0.52	0.42	1.20	1.63	2.05
167	1.04	1.26	1.00	2.94	5.22	5.81
250	1.19	1.71	2.11	1.19	1.70	2.30

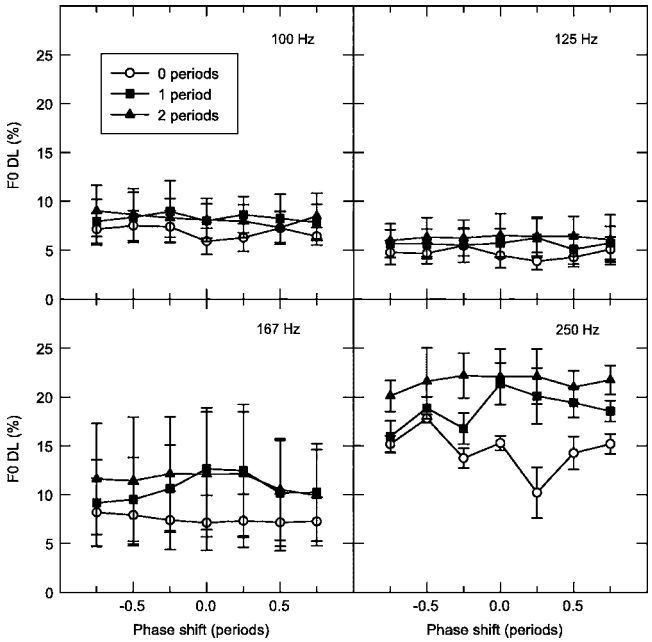


FIG. 5. Results of experiment 1 showing *F*0DLs for tone-burst pairs of *F*0s 100, 125, 166.7, and 250 Hz as a function of the phase shift between the bursts. Error bars show standard errors between listeners.

$F=126.3$, $p<0.05$; for the 166.7-Hz *F*0 condition, $F=114.3$, $p<0.05$). The cubic trends were insignificant for the 1- and 2-period silent interval conditions. When the *F*0 of the complex tone was 250 Hz, none of the cubic trends was statistically significant even when there was no silent interval between the tone bursts.

In Fig. 5 the mean *F*0DLs are plotted against the phase shift in periods. The general pattern suggests the *F*0DLs for the tone bursts separated by a 1- or 2-period silent interval are higher than the *F*0DLs for the tone bursts with no silent interval. The *F*0DLs are considered further in Sec. V.

III. EXPERIMENT 2

A. Rationale

Experiment 2 was conducted to attempt to confirm the finding of experiment 1 that the –0.25 phase shift tends to produce the highest pitch. A different methodology, a method of constant stimuli, was used to ensure that the pattern of pitch shifts produced by the envelope phase changes was not an artifact of the adaptive pitch matching procedure.

B. Stimuli and procedure

The experiment used a subset of the complex-tone bursts used in experiment 1, those that had a 0-period silent interval between the tone bursts. The *F*0s tested were 100, 125, 166.7, and 250 Hz. The basic procedure was the same as in experiment 1 except that the adaptive tracking rule was not employed; instead, the method of constant stimuli was used. The pitch of the −0.25 phase-shift stimulus was compared to four other phase shifts (−0.75, −0.5, 0, and +0.25 periods); the *F*0s of the two tone burst pairs in each trial were the same. A block consisted of 100 trials of the same condition

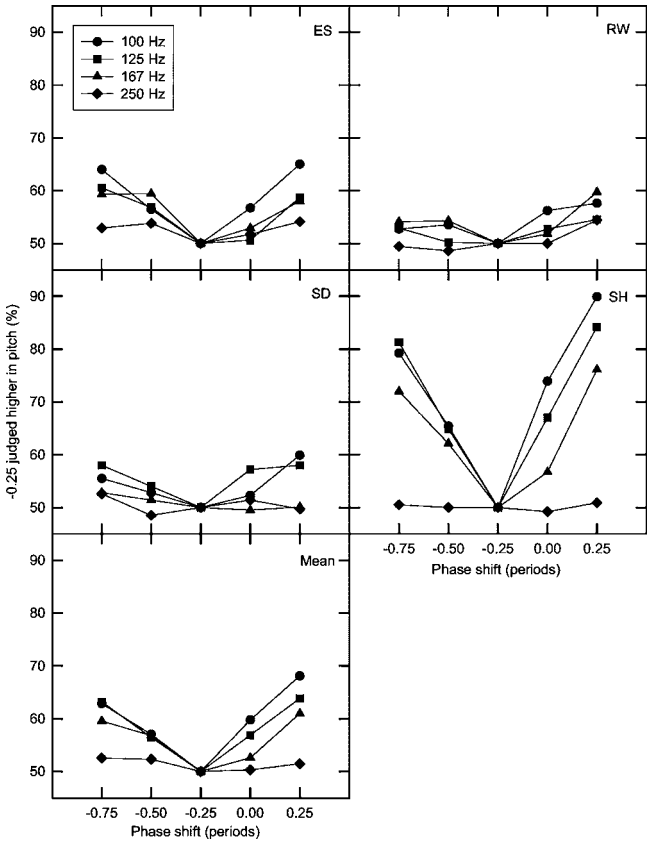


FIG. 6. Mean and individual results of experiment 2 showing the percent of responses where the −0.25 phase-shifted stimulus was judged higher in pitch, as a function of the phase of the comparison stimulus.

and each block was run 10 times. The listeners reported which of the two stimuli presented was higher in pitch.

C. Listeners

Four normally hearing listeners were tested. They were aged between 22 and 32 years. Listeners were given at least 2 hours training on the task before data collection began. Two of the listeners (ES and RW) also participated in experiment 1.

D. Results

The results of experiment 1 suggested that the highest overall pitch was heard for the tone bursts containing a −0.25 phase shift. The results also suggested that it would be easier to differentiate the −0.75 and +0.25 phase shifts from the −0.25 phase shift than it would be to differentiate the −0.5 and 0 phase shifts from the −0.25 phase shift. The results of experiment 2 are consistent with these findings. The individual and mean responses (in percentage of times the −0.25 phase shift was judged higher in pitch) are shown in Fig. 6. Statistical significance was determined by comparing the lower limit of the 95th percentile of the data, determined across individuals, to chance performance (50%). If the lower limit of the 95th percentile was greater than 50%, then this difference was considered to be statistically significant at the $p<0.05$ significance level. All the comparisons were significantly different from chance except for the comparison

between −0.25 and 0 phase shift at the 250-Hz *F*0. Overall, the results suggest that the pitch shifts are caused by the shift in a single IPI and are not an artifact of the pitch-matching procedure.

IV. EXPERIMENT 3

A. Rationale

Experiments 1 and 2 showed that phase shifts of −0.75 or −0.5 periods tended to cause a negative pitch shift. Plack and White (2000) suggested that, when two pulses occur close together, there may be a bias towards processing the first pulse. They suggested that this is a consequence of the refractory period of auditory nerve (Kiang *et al.*, 1965) being similar to the shortest IPIs. Although not all the neurons would fire on the first pulse (so only a proportion of the neuronal population would be in the refractory period when the second pulse arrived), the overall effect may be a decreased response to the second pulse. If one of the pulses is "ignored," then a mean-interval model, modified to take account of this, would predict a negative pitch shift for the stimuli with the smallest IPIs.

On the basis of the present data, it seems unlikely that the pitch shifts caused by the −0.75- and −0.5-period phase shifts were influenced by the refractory periods of neurons. These negative pitch shifts occurred at an IPI that was a constant *proportion of the period*, not a constant absolute time value that might be associated with neural recovery times. However, basilar-membrane (BM) ringing (Wilson and Johnstone, 1972) may provide a more adequate explanation. The impulse response of the BM to the first tone burst may not have "damped down" completely before the second tone burst arrived; therefore, the silent interval between the pulses may have been partially filled and this may have caused the auditory system to regard the two pulses as a single pulse, thereby reducing the pitch. The same pattern of pitch shifts (as a function of phase shift in periods) would be shown at each *F*0 because in experiments 1 and 2 the frequency region increased as *F*0 increased. A higher *F*0 would mean that there were shorter silent intervals between the pulses, but these tone bursts would be filtered into a higher region where there is less BM ringing (Wilson and Johnstone, 1972). This interaction between *F*0 and frequency region may have caused the pattern to be consistent across *F*0. Experiment 3 was conducted to investigate this possibility. If BM ringing is the cause of the negative pitch shifts, then, for a given *F*0, the pattern of pitch shifts should be dependent on frequency region.

B. Stimuli and procedure

The experiment tested a subset of the complex-tone bursts used in experiment 1; those that had an *F*0 of 100 Hz, with a 0-period silent interval. The standard interval contained a phase shift of −0.75, −0.5, −0.25, or 0 periods. The comparison interval did not contain a phase shift. The harmonics were bandpass filtered (digital FIR) between 2200 and 3000 Hz or between 3667 and 5000 Hz (3-dB down-points, 90-dB/octave slope). The adaptive procedure was the same as that used in experiment 1.

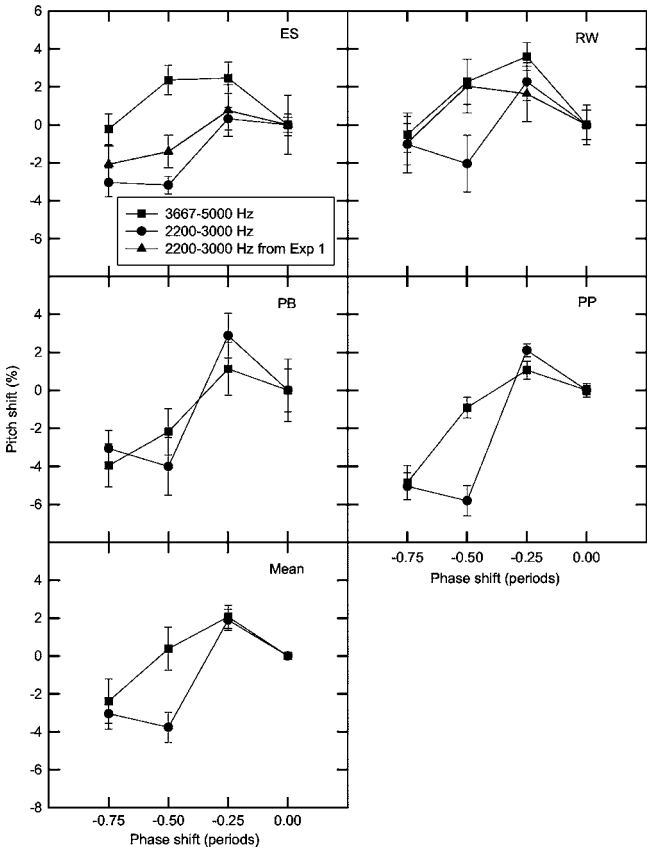


FIG. 7. Mean and individual results of experiment 3 showing mean pitch shifts (%) for 100-Hz *F*0 tone burst pairs, filtered from 2200–3000 Hz or from 3667–5000 Hz, as a function of the phase shift between the bursts. For listeners ES and RW the mean pitch shifts are shown for 100-Hz *F*0 tone burst pairs, filtered from 2200–3000 Hz, from experiment 1. Error bars show standard errors, and, for the mean results, the standard errors between listeners.

C. Listeners

Four normally hearing listeners were tested. They were aged between 22 and 32 years. Listeners were given at least 2 hours training on the task before data collection began. Two of the listeners (ES and RW) also participated in experiments 1 and 2.

D. Results

The mean and individual pitch matches are shown in Fig. 7. The mean percentage difference between the *F*0 of the comparison tone pair and the nominal *F*0 of the shifted tone pair is plotted against the phase shift in periods. As in Fig. 4, the pitch shifts are plotted relative to the control condition with 0 phase shift.

The lower frequency region conditions were identical to the 100-Hz *F*0 conditions in experiment 1. The pitch shifts were measured in the same way and two of the same listeners were used. There is a discrepancy between the pitch shifts produced by the −0.5 phase shift for experiments 1 and 3. In particular, the mean pitch shift for the −0.5-period phase shift was more negative in experiment 3. This is mostly due to interlistener variability caused by using two new listeners. PB and PP both produced large negative shifts for this condition. However, the two listeners that were com-

mon to both experiments also produced more negative shifts in experiment 3. The individual data from experiment 1 are presented in Fig. 7 for the two common listeners. Considering the data for the −0.5-period phase shift stimuli only, listener ES produced a slightly more negative pitch shift for the 2200–3000 Hz frequency region condition when tested in experiment 3 than when tested in experiment 1. Both these pitch shifts were different from the positive pitch shift produced for the 3667–5000 Hz frequency region condition tested in experiment 3. Listener RW produced very similar, positive pitch shifts for the 3667–5000 Hz frequency region condition in experiment 3 and for the 2200–3000 Hz frequency region condition in experiment 1. It is not clear why the same condition (2200–3000 Hz), when tested in experiment 3, produced a negative pitch shift. This suggests that there is not only a degree of intersubject variability, but also that the two listeners common to experiments 1 and 3 may have changed their response. It is important to note that the pitch shifts for the other low-frequency region conditions in experiment 3 are similar to the results for the same conditions in experiment 1.

To test whether there was a significant effect of frequency region, paired-samples t-tests were carried out for each phase shift between the pitch shifts produced by the low- and the high-frequency region conditions. When the phase shift was −0.75 or −0.25 periods, there was no significant difference between the pitch shifts for the low- and high-frequency region conditions [$t(3)=0.847$, $p=0.459$ and $t(3)=0.182$, $p=0.867$]. When the phase shift was −0.5 periods, there was a significant effect of frequency region on the pitch shift [$t(3)=5.074, p<0.05$]. The implications of these results will be considered in Sec. V.

V. DISCUSSION

A. Resetting mechanisms

The results of experiment 1 confirm that changing a single IPI can cause a shift in the pitch of a complex tone burst with unresolved harmonics. This was found consistently when the silent interval between the bursts was 0 period. Generally, pitch shifts were not produced by the pulse trains with 1- or 2-period silent intervals, and this is consistent with the hypothesis of Plack and White (2000) that the auditory system has a pitch mechanism that is reset in response to temporal discontinuities of greater than approximately one waveform period. The *F*0DLs for experiment 1 also suggest this. Generally, the *F*0DLs for the stimuli with a silent interval of 1 or 2 periods were higher than the *F*0DLs for the stimuli without a silent interval. This suggests that the tone bursts with a silent interval of around one period or more between them are processed separately, and the benefits of long integration are not observed (White and Plack, 1998).

Some tone bursts separated by a 1-period silent interval did produce a shift in pitch. The more positive pitch shift for the −0.75 phase shift than for the −0.5 phase shift (1-period silent interval condition) for some of the *F*0s may be taken as evidence that the resetting mechanism is beginning to act (see Fig. 4). The *F*0DLs show this more clearly (see Fig. 5). For the 166.7- and 250-Hz *F*0 stimuli, the *F*0DLs for the 2-period silent interval condition were higher than the *F*0DLs for the 0-period silent interval condition. The tone bursts separated by a 1-period silent interval and containing phase shifts of 0, +0.25, +0.5, and +0.75 periods also produced *F*0DLs similar to the 2-period silent interval condition. The tone bursts in these conditions may have been processed separately. However, the tone bursts separated by a 1-period silent interval and phase shifts of −0.75, −0.5, and −0.25 periods produced *F*0DLs that were more similar to the *F*0DLs produced by the tone bursts without a silent interval than to the *F*0DLs produced by the 2-period silent interval conditions. It is possible that the reduction in the IPI resulting from the phase advance may have been enough to encourage the auditory system to integrate across the silent interval in these conditions, and so produce a more accurate estimate of the pitch. However, one might also predict that this integration would have occurred only for the standard interval that contained the phase shift, and not for the comparison interval with no phase shift.

B. Basilar-membrane ringing

The results of experiment 1 showed that the pitches for the −0.5 and −0.75 phase-shift stimuli were successively lower than the −0.25 phase-shift stimuli, when there was a 0-period silent interval between the tone bursts. Experiment 3 showed that when the −0.5 phase-shift condition was filtered into a higher frequency region, while keeping the *F*0 constant at 100 Hz, the pitch shift was significantly less negative. The auditory filters in a higher frequency region show less ringing than filters in a lower frequency region. Therefore, a pulse is less likely to be partially masked by BM ringing in the higher frequency region. The results of experiment 3 suggest that the negative pitch shifts seen for the smallest IPIs may have been a consequence, in part, of BM ringing. However, the fact that there was no significant effect of frequency region shown in experiment 3 for the −0.75 phase shift indicates that this cannot be a complete explanation (see Sec. V C 2).

C. Models of pitch perception

1. Autocorrelation and other common-interval models

Using the model of Meddis and Hewitt (1991), Plack and White (2000) produced summary autocorrelation functions (SACFs, sums of all the autocorrelation channels across auditory filter center frequency) for stimuli with a 250-Hz *F*0 that contained a single shifted IPI. The SACFs failed to predict the pitch shifts shown experimentally. Similarly, the 100-Hz stimuli used in experiment 1 were processed through a more recent autocorrelation model (Meddis and O'Mard, 1997). The 100-Hz stimuli were chosen because experiment 2 showed these stimuli to be the most easily discriminable. Meddis and O'Mard's (1997) model consists of eight stages of processing: an outer-ear bandpass function, middle-ear low- and high-frequency attenuation, mechanical-to-neural transduction at the hair cell (Meddis, 1986; 1988; Meddis *et al.*, 1990),[1] refractory inhibition of firing of the auditory-nerve fibers, an estimation of the distribution of intervals among all spikes originating from fibers within the same

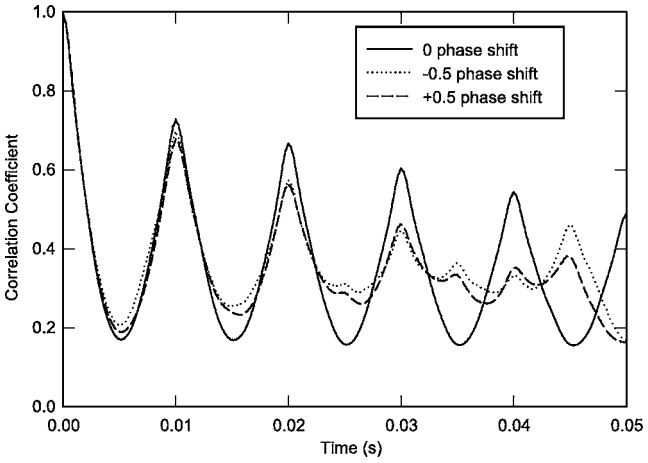


FIG. 8. Summary autocorrelation functions (SACFs) for 0, −0.5, and +0.5 phase-shift stimuli (100-Hz *F*0).

channel, a summation of interval estimates across channels, and finally, pitch extraction by inspection of the summary autocorrelation function (for further details of this model see Meddis and O'Mard, 1997, and Meddis and Hewitt, 1991). The 100-Hz *F*0 stimuli (no silent interval) were processed through the first seven stages of Meddis and O'Mard's (1997) model, up to the summation of interval estimates across channels, and then processed by an autocorrelation algorithm using a rectangular sampling window that included the entire stimulus.

Figure 8 shows the output of the model for a 0 phase shift, a −0.5 phase shift, and a +0.5 phase shift (the SACFs for the other phase shift conditions are illustrated in Plack and White, 2000). The pitch estimate in the original model (Meddis and Hewitt, 1991) was determined by the position of the first major peak along the time axis of the SACF. In the simulation here, however, the first *three* peaks in the SACF were in the same positions for the 0, −0.5, and +0.5 phase-shift conditions. The dominant first-, second-, and third-order IPIs (10, 20, and 30 ms) were affected little by the phase shifts. The SACFs for the −0.5 and +0.5 phase shifts are very similar to each other. This is also not surprising because the only significant difference between these stimuli is that the second half of the −0.5 phase-shift stimulus is advanced by one period relative to the second half of the +0.5 phase-shift stimulus. Consequently, almost all the pitch pulses are coincident for these two stimuli and the distributions of time intervals between pulses are very similar. However, the results of experiment 1 show that these two stimuli are perceived very differently, and that the −0.5-period phase-shifted tone bursts have a pitch about 6% higher than that for the +0.5-period phase-shifted tone bursts at 100 Hz. A pitch shift of this magnitude should be clearly visible in the SACFs if they are to account for the pitches of these stimuli. However, to confirm the result a quantitative analysis was performed.

In the recent version of the model described by Meddis and O'Mard (1997), pitch matching is conducted by calculating the Euclidean distances (D^2, the sum across delay of the squared differences between the correlation values) between the SACFs of the two tones. A small D^2 indicates a close pitch match. It is suggested here that the SACFs for −0.5 and +0.5 phase stimuli are more similar to each other than to the SACFs for the tone bursts to which they were matched in experiment 1. To test this quantitatively, D^2s were calculated between the SACFs for the −0.5 and +0.5 phase-shift stimuli, between the SACFs for the −0.5 phase-shift stimulus and a no-shift 100.95-Hz *F*0 tone burst, and between the SACFs for the +0.5 phase-shift stimulus and a no-shift 95-Hz *F*0 tone burst. (The *F*0s for the no-shift stimuli were equal to the matched *F*0s of the comparison stimulus from experiment 1, *F*0=100 Hz.) D^2s were calculated for SACFs covering autocorrelation delays from 0 up to 15, 25, 35, 45, 55, or 65 ms. If the model is to successfully predict the pitch shifts shown experimentally, D^2 for the latter two comparisons should be smaller than D^2 for the first comparison. In the simulation, however, D^2 for the comparison between the two phase-shifted stimuli was considerably smaller than the D^2 for the comparisons between the phase-shifted stimuli and the experimentally matched tone bursts. This was true for all the delay ranges that were tested. In other words, the SACFs for the phase-shifted tone bursts were more similar to each other than they were to the SACFs for tone bursts with the matched *F*0s of the comparison from experiment 1. However, since there is a discrepancy between the pitch shifts measured experimentally for the 100-Hz *F*0, −0.5-period phase-shift condition in experiments 1 and 3, a further analysis was conducted. D^2s were calculated between the SACFs for the +0.5 phase-shift stimulus and a no-shift 95-Hz *F*0 tone burst (this *F*0 is equal to the matched *F*0 of the comparison stimulus from experiment 1, nominal *F*0 =100 Hz), and between the SACFs for the +0.5 phase-shift stimulus and the 0 phase-shift stimulus. D^2s were calculated for SACFs covering autocorrelation delays from 0 up to 15, 25, 35, 45, 55, or 65 ms. D^2 for the former comparison should be smaller than D^2 for the latter comparison, if the model is to successfully predict the pitch shift. For delays up to 45 ms, D^2 was smaller for the latter comparison than for the former comparison. Meddis and O'Mard (1997) suggested that only delays that might reasonably influence pitch judgments should be considered. Pressnitzer *et al.* (2001) showed that humans are unable to derive a pitch from stimuli with *F*0s below approximately 32 Hz. This suggests it is unlikely that delays greater than 31 ms can be processed by the pitch mechanism. Therefore, the SACF for the +0.5 phase-shift stimulus was more similar to the SACF for the 0 phase-shift stimulus than the SACF for the tone burst it was matched to experimentally for delays that might be utilized in pitch discriminations.

It is important to emphasize that the experimental results presented in this paper provide evidence against two particular ways of using the information from the SACF (peak picking and Euclidean distance). The autocorrelation function contains a great deal of information about the stimulus (equivalent to the information in the magnitude spectrum), and it may well be possible to process the SACF to produce pitch estimates that are consistent with those measured here. There are also models that use processing that is similar in nature to autocorrelation, for example the neural timing nets model of Cariani (2001a; 2001b) and the auditory image

model of Patterson and colleagues (Patterson, 2000; Patterson *et al.*, 1995; Patterson and Holdsworth, 1996; Patterson *et al.*, 1992; Patterson *et al.*, 1992), that may have more success with the present data. The model of Meddis and O'Mard was chosen for scrutiny here because it is highly influential, and because of the success of the model in accounting for a wide range of pitch phenomena.

Autocorrelation, as implemented in Meddis and O'Mard's (1997) model, is an algorithm that detects regularity in the ISIs. Common-interval models in general (Moore, 2003; Patterson *et al.*, 1995) predict that two complexes that both have a large proportion of the same regular IPI will have the same pitch, regardless of the actual number of pulses in each. The results of experiments 1 and 2 contribute to the body of data (Carlyon, 1996; 1997; Carlyon *et al.*, 2002; Plack and White, 2000) suggesting that pitch models based on the isolation of a regular common interval do not provide a good description of the pitch of *some* unresolved complex tones. All the stimuli used in experiments 1 and 2 had a preponderance of IPIs at $1/F0$, $2/F0$, and $3/F0$, but frequently the pitch heard did not correspond to $F0$.

2. Mean-interval models

Carlyon *et al.* (2002) proposed a version of the mean-interval model that was based on a weighted average of the first-order IPIs in the stimulus. Longer intervals were weighted more heavily than shorter intervals to determine pitch

$$P = \frac{\sum_{t=1}^{n} W(\tau_t) p_t^2}{\sum_{t=1}^{n} W(\tau_t) p_t^2 \tau_t}, \tag{3}$$

where P is pitch, t is an index referring to each interval length present, τ_t is the duration of the interval (the IPI), $W(\tau_t)$ is the weighting applied to the interval, p_t is the relative proportion of the interval in the waveform (p_t is calculated as the number of intervals of length τ_t divided by the number of intervals in total), and n is the number of different intervals present. As described in the Introduction, the values of $W(\tau_t)$ were chosen so that larger IPIs (τ_ts) were weighted more heavily than shorter intervals. Specifically, the value of $W(\tau_t)$ was zero for τ_ts less than 1 ms, and increased logarithmically for τ_ts above 1 ms, converging on a value of 1.

In its original form, the Carlyon *et al.* (2002) model cannot account for the data from experiment 1. Figure 9 shows a comparison of the data from experiment 1 for the 100-Hz $F0$ tone bursts with a 0-period silent interval and the predictions of the Carlyon *et al.* (2002) model, based on Eq. (3). The model fails partly because the pitch shifts in the experimental data produced by a change in a single IPI were dependent upon the *proportion* of the waveform period by which the single IPI was shifted, not upon the *absolute* value of the IPI as suggested by the model. In other words, the model does not weight the IPIs appropriately for low $F0$'s. For example, all IPIs above 8 ms, corresponding to phase shifts of 0 period and above for an $F0$ of 100 Hz, are given a weight of 1. If the weights do not vary substantially with the duration of the central IPI, the model suffers because of the squaring of p_t in the numerator and the denominator of Eq. (3). By squaring the proportion of each IPI present, the model predicts that a change in just a single IPI should have a minimal effect on pitch. The present data suggest otherwise.

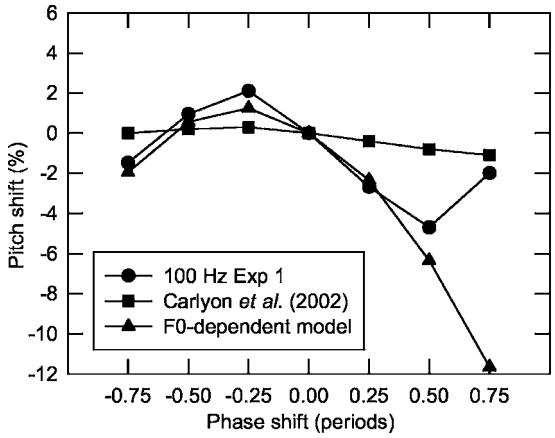


FIG. 9. The experimental data from experiment 1 (100-Hz $F0$, 0-period silent interval tone bursts only) plotted with the predictions of the weighted mean-interval model of Carlyon *et al.* (2002), and with the predictions of the $F0$-dependent, mean-interval model based upon data from experiment 1.

Examining the experimental data across $F0$s, it does not appear to be the case that longer intervals *per se* were weighted more heavily, but that intervals that were much longer or shorter *relative* to the most common interval were weighted less. Even taking this into account by transforming the weights for absolute time intervals into weights for *proportions* of the most common waveform period, the model of Carlyon *et al.* (2002) does not seem to capture adequately the resetting mechanism when the two consecutive pulses are too far apart in time or the more negative pitch shifts produced when two pulses occur too closely together in time. In order to attempt to rectify this latter difficulty, a modification of the model is proposed.

The new model also uses the mean IPI as the main predictor of pitch. However, the experimental results suggest that when two pulses occur close together in time then the mean IPI cannot predict the pitch. Plack and White (2000) suggested that this may be due to one of the pulses being "ignored." Plack and White (2000) proposed that this was due to a proportion of neurons being in refractory when the second pulse arrived. This possibility can be rejected as the more negative pitch shifts produced when two consecutive pulses are close together in time do not occur in response to a set time interval, like a neuronal refractory period. Nevertheless, it does appear to be the case that, when two pulses occur closely together in time, one of the pulses is ignored. [It is not clear whether this is because the rate of change of period (Gockel *et al.*, 2001) becomes too quick, or because the stimulus contains consecutive pulses that are temporally too close together. The experiments presented here are unable to differentiate between these two possibilities.] Therefore, the new model includes a function, $f(s)$, which is the probability that one of a pair of closely spaced pulses will be ignored. If a pulse is ignored then the predicted pitch becomes $(N-2)/D$, where N is the number of pulses and D is

the total duration of the tone (in seconds). If none of the pulses is ignored then the predicted pitch is simply $(N-1)/D$, and the probability of this occurring is $1-f(s)$.

Therefore, the model's equation for pitch is

$$P = f(s)\left(\frac{N-2}{D}\right) + [1-f(s)]\left(\frac{N-1}{D}\right), \tag{4}$$

where P is pitch in Hz. It was originally conceived that s would equal the product of the IPI_U and the ERB, where IPI_U is the single shifted IPI (the central IPI), and the ERB is the equivalent rectangular bandwidth. In other words, the chance that a pulse is ignored is dependent on the absolute duration of the IPI times the ERB. The ERB was calculated as a function of center frequency, F (in kHz; Glasberg and Moore, 1990)

$$\mathrm{ERB} = 24.7(4.37F + 1). \tag{5}$$

F was chosen to be the arithmetic mean of the cutoffs of the bandpass filter used to filter the harmonics. The duration of the impulse response of a filter is inversely proportional to its bandwidth. It is assumed that bandwidth measured psychophysically (ERB) is an approximation of the BM response at each place. As the IPI_U becomes shorter, it will become more likely that a pulse will be obscured by BM ringing. Therefore, it was anticipated that the product of the ERB and IPI_U would determine the effects of ringing on the pitch shift, regardless of the frequency region or the $F0$.

The value of $f(s)$ was calculated by substituting in Eq. (4) the known information (D, the total duration, N, the number of pulses, and P, the mean pitch matched experimentally) for the tone bursts, filtered into the lower frequency region (2200–3000 Hz) from experiment 3. Then, the values of $f(s)$ were plotted against s, and a quadratic regression line

$$f(s) = 0.04s^2 - 0.65s + 1.54, \tag{6}$$

was fitted ($R^2=0.90$). This ERB-dependent model was used to make predictions for the results of experiment 3 for the higher frequency region (3667–5000 Hz). The predictions did not match the experimental results (see Fig. 10). The effect of changing the frequency region in experiment 3 was less than that predicted on the assumption that the processing of a short IPI is limited by BM ringing. This suggests that BM ringing cannot be a complete explanation for the negative pitch shifts perceived when pulses are closely spaced in time.

This prompted a revision of the model. Further calculations were conducted using $s=\mathrm{IPI}_U/\mathrm{IPI}_C$, where IPI_U is the single, shifted IPI and IPI_C is the common IPI that produces the nominal $F0$. In this $F0$-dependent version of the model, the chance of a second pulse being ignored depends on the IPI *relative to* the common IPI. If the nominal $F0$ is decreased, a pulse occurring after a given IPI (say, 2 ms) is more likely to be ignored. This could be interpreted as the pitch mechanism ignoring pulses that are unlikely to be from the same source as the other pulses in the sequence. Again, $f(s)$ was calculated by substituting in Eq. (4) the known information (D, N, and P) for the low-frequency region tone bursts used in experiment 3. As before, the values of $f(s)$ were plotted against s, and a quadratic regression line

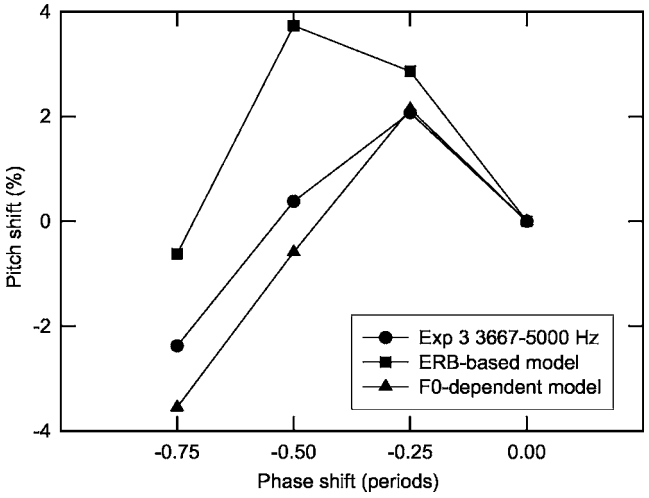


FIG. 10. The experimental data from experiment 3 (only the tone bursts filtered 3667–5000 Hz) plotted with the predictions of the ERB-dependent mean-interval model and the predictions of the $F0$-dependent mean-interval model. Both models are based upon the data from experiment 3 (only the tone bursts filtered 2200–3000 Hz).

$$f(s) = 0.04s^2 - 2.00s + 1.54, \tag{7}$$

was fitted ($R^2=0.90$). The predictions of this version of the model fit the experimental data for the high-frequency region tone bursts from experiment 3 better than the predictions of the ERB-based version of the model (see Fig. 10). This indicates that it may not be necessary to include a measure of the frequency region of the stimuli within the model. Therefore, an $F0$-dependent version of the model ($s=\mathrm{IPI}_U/\mathrm{IPI}_C$) was calculated from the larger dataset, that includes tone bursts with four different $F0$s, collected in experiment 1. As before, the known information (D, N, and P) was substituted in Eq. (4) for the tone bursts used in experiment 1 whose IPI_U was less than the period of the waveform. As before, the values of $f(s)$, for all $F0$s and phase shifts, were plotted against s, and a quadratic regression line

$$f(s) = 1.19s^2 - 2.73s + 1.52, \tag{8}$$

was fitted ($R^2=0.97$). Equations (7) and (8) produce similar values when $s<1$ (when $s>1$ it is assumed that the probability of a pulse being ignored is 0). The $\mathrm{rms}_{\mathrm{ERROR}}$ between the values produced by these equations, for values of $s<1$, is only about 0.12% (to produce the $\mathrm{rms}_{\mathrm{ERROR}}$, s was sampled at intervals of 0.1 from 0 to 1). The pitch shift predictions of this model for the 100-Hz $F0$, 0-period silent gap tone bursts from experiment 1 are shown in Fig. 9. The model closely fits the experimental data for all conditions, except the +0.75-period phase shift condition. This is because the +0.75-period phase shift produces a more positive pitch shift than would be expected on the basis of the mean rate. This may reflect the mechanism, that resets in response to gaps of around 1 period, beginning to act (see Sec. V A).

Carlyon *et al.*'s (2002) model could account for their finding that participants match a "4–6" pulse train (a pulse

train that has first-order intervals that alternate between 4 and 6 ms) to an isochronous pulse train with an interval of approximately 5.5 to 5.7 ms. The *F*0-dependent version of the model based upon the data in experiment 1 can also predict a pitch for a 4–6 pulse train if it is assumed (somewhat arbitrarily) that the longer interval, 6 ms, is IPI_C, and that the shorter interval, 4 ms, is IPI_U. The model predicts that a 4–6 pulse train will be matched to an isochronous pulse train that has an interval size of 5.6 ms. This is similar to the period of 5.5 to 5.7 ms matched experimentally by Carlyon *et al.* (2002).

To summarize the modeling section, it appears that neither the autocorrelation model nor previously published mean-interval models can account for the data. If the mean-interval model is modified to allow one of two closely spaced pulses to be ignored probabilistically, then the model provides a much better fit to the data. The probability of a pulse being ignored does not depend on the absolute time between the pulses, but on the interval between the pulses relative to the period of the complex. There is some evidence for an effect of frequency region, and hence possibly BM ringing, on whether a pulse is ignored, although this does not seem to be the most significant factor.

VI. SUMMARY

(i) Unresolved complex tone bursts of four *F*0s, 100, 125, 166.7, and 250 Hz, were presented. The tone bursts were five waveform periods long and two tone bursts were paired consecutively, separated by a 0-, 1-, or 2-period silent interval. The starting envelope phase of the second tone burst relative to the first tone burst was varied; in effect the center IPI of the pulse train was varied. Relative to the no-shift control, the variations in IPI produced substantial pitch shifts when there was no silent interval between the bursts, but little effect was seen for silent intervals of 1 or 2 periods.

(ii) The results are consistent with the hypothesis that the auditory system has an integration window for pitch that is reset in response to temporal discontinuities of greater than around one waveform period for unresolved complex tones.

(iii) The pitch shifts could not be explained on the basis of the Meddis and O'Mard (1997) autocorrelation model nor by the mean-interval model of Carlyon *et al.* (2002). They can be accounted for by a model based on the mean interpulse interval that includes a measure of the probability that one of two pulses closely spaced in time will be ignored. This probability is dependent on the separation of the pulses *relative* to the period of the complex.

ACKNOWLEDGMENTS

Thanks to Anastasios Sarampalis, Peter Cariani, and two anonymous reviewers for insightful comments on earlier versions of the manuscript; Ray Meddis, Lowel O'Mard, Steve Holmes, and Vit Drga for help in implementing the autocorrelation model; and Tim Rakow for useful statistical advice. Author R.W. was supported by an EPSRC doctoral training grant (GR/NO7219).

[1]The parameters of the Meddis inner hair cell are the same as those given in Meddis *et al.* (1990).

Bernstein, J. G., and Oxenham, A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?" J. Acoust. Soc. Am. **113**, 3323–3334.

Bregman, A. S., Ahad, P. A., and Kim, J. (**1994a**). "Resetting the pitch analysis system. II. Role of sudden onsets and offsets in the perception of individual components in a cluster of overlapping tones," J. Acoust. Soc. Am. **96**, 2694–2703.

Bregman, A. S., Ahad, P., Kim, J., and Melnerich, L. (**1994b**). "Resetting the pitch-analysis system. I. Effects of rise times of tones in noise backgrounds or of harmonics in a complex tone," Percept. Psychophys. **56**, 155–162.

Cariani, P. (**2001a**). "Neural timing nets," Neural Networks **14**, 737–753.

Cariani, P. (**2001b**). "Neural timing nets for auditory computation," in *Computational Models of Auditory Function* (IOS, Amsterdam), pp. 235–249.

Carlyon, R. P. (**1996**). "Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker," J. Acoust. Soc. Am. **99**, 517–524.

Carlyon, R. P. (**1997**). "The effects of two temporal cues on pitch judgments," J. Acoust. Soc. Am. **102**, 1097–1105.

Carlyon, R. P., Van Wieringen, A., Long, C. J., Deeks, J. M., and Wouters, J. (**2002**). "Temporal pitch mechanisms in acoustic and electric hearing," J. Acoust. Soc. Am. **112**, 621–633.

de Cheveigné, A. (**1998**). "Cancellation model of pitch perception," J. Acoust. Soc. Am. **103**, 1261–1271.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gockel, H., Moore, B. C. J., and Carlyon, R. P. (**2001**). "Influence of rate of change of frequency on the overall pitch of frequency-modulated tones," J. Acoust. Soc. Am. **109**, 701–712.

Jesteadt, W. (**1980**). "An adaptive procedure for subjective judgments," Percept. Psychophys. **28**, 85–88.

Joris, P. X., and Yin, T. C. T. (**1992**). "Responses to amplitude-modulated tones in the auditory nerve of the cat," J. Acoust. Soc. Am. **91**, 215–232.

Kiang, N. Y.-S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). *Discharge Patterns of Single Fibers in the Cat's Auditory Nerve* (MIT Press, Cambridge, MA).

Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128–134.

Licklider, J. C. R. (**1956**). "Auditory frequency analysis" in *Information Theory*, edited by C. Cherry (Academic, New York).

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., and Hewitt, M. J. (**1991**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification," J. Acoust. Soc. Am. **89**, 2883–2894.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.

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. (**2003**). *An Introduction to the Psychology of Hearing* (Academic, London).

Moore, G. A., and Moore, B. C. J. (**2003**). "Perception of the low pitch of frequency-shifted complexes." J. Acoust. Soc. Am. **113**, 977–985.

Nabelek, I. V. (**1996**). "Pitch of a sequence of two short tones and the critical pause duration," Acustica **82**, 531–539.

Patterson, R. D. (**2000**). "Auditory images: How complex sounds are represented in the auditory system." J. Acoust. Soc. Jpn. **21**, 183–190.

Patterson, R. D., and Holdsworth, J. (**1996**). "A functional model of neural activity patterns and auditory images," in *Advances in Speech, Hearing and Language Processing* (JAI, London), pp. 547–563.

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., Holdsworth, J., and Allerhand, M. (**1992a**). "Auditory models as preprocessors for speech recognition," in *The Auditory Processing of Speech: From the Auditory Periphery to Words* (Mouton de Gruyter, Berlin), pp. 67–83.
Patterson, R. D., Robinson, K., Holdsworth, J. W., McKeown, D., Zhang, C., and Allerhand, M. (**1992b**). "Complex sounds and auditory images," in *Auditory Physiology and Perception* (Pergamon, Oxford), pp. 429–446.
Plack, C. J., and White, L. J. (**2000**). "Pitch matches between unresolved complex tones differing by a single interpulse interval," J. Acoust. Soc. Am. **108**, 696–705.
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. Vol. **43**, 764–767.
Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "The lower limit of melodic pitch." J. Acoust. Soc. Am. **109**, 2074–2084.
Schouten, J. F., Ritsma, R. J., and Lopes Cardozo, B. (**1962**). "Pitch of the residue," J. Acoust. Soc. Am. **34**, 1418–1424.
Slaney, M., and Lyon, R. F. (**1990**). "A perceptual pitch detector," Proc. Int. Conf. Acoustic Speech Signal Process. **5**, 357–360.
Viemeister, N. F., and Wakefield, G. H. (**1991**). "Temporal integration and multiple looks," J. Acoust. Soc. Am. **90**, 858–865.
White, L. J., and Plack, C. J. (**1998**). "Temporal processing of the pitch of complex tones," J. Acoust. Soc. Am. **103**, 2051–2063.
Wilson, J. P., and Johnstone, J. R. (**1972**). "Capacitative probe measures of basilar membrane vibrations," in *Hearing Theory* (IPO, Eindhoven), pp. 172–181.

The effect of cross-channel synchrony on the perception of temporal regularity

Katrin Krumbholz[a)]
Institute of Medicine (IME), Research Center Jülich, D-52425 Jülich, Germany, and MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom

Stefan Bleeck and Roy D. Patterson
Centre for the Neural Basis of Hearing, Department of Physiology, University of Cambridge, Downing Street, Cambridge, CB2 3EG, United Kingdom

Maria Senokozlieva, Annemarie Seither-Preisler, and Bernd Lütkenhöner
Department of Experimental Audiology, ENT Clinic, Münster University Hospital, Kardinal-von-Galen-Ring 10, D-48149 Münster, Germany

(Received 11 August 2004; revised 20 April 2005; accepted 4 May 2005)

Temporal models of pitch are based on the assumption that the auditory system measures the time intervals between neural events, and that pitch corresponds to the most common time interval. The current experiments were designed to test whether time intervals are analyzed independently in each peripheral channel, or whether the time-interval analysis in one channel is affected by synchronous activity in other channels. Regular and irregular click trains were filtered into narrow frequency bands to produce target and flanker stimuli. The threshold for discriminating a regular target from an irregular distracter click train was measured in the presence of an irregular masker click train in the target band, as a function of the frequency separation between the target band and a flanker band. The flanker click train was either regular or irregular. The threshold for detecting the regular target was 5–7 dB lower when the flanker was regular. The data indicate that the detection of temporal regularity (and thus, pitch) involves cross-channel processes that can operate over widely separated channels. Model simulations suggest that these cross-channel processes occur after the time-interval extraction stage and that they depend on the similarity, or consistency, of the time-interval patterns in the relevant channels. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1941090]

PACS number(s): 43.66.Hg, 43.66.Mk, 43.66.Ba [RAL] Pages: 946–954

I. INTRODUCTION

The phase-locked temporal information produced by the cochlea seems likely to play a significant role in auditory perception. In time-domain models of perception, it is assumed that the auditory system analyses this temporal information by measuring the time intervals between the peaks in the phase-locked activity of auditory-nerve fibers. Time-domain models are important for explaining the pitch produced by iterated rippled noise and periodic stimuli limited to unresolved harmonics (Burns and Viemeister, 1976; Patterson *et al.*, 1996; Yost *et al.*, 1996). Typically, these models generate a multichannel simulation of the neural activity pattern (NAP) produced by the cochlea, and then perform a time-interval analysis separately within each NAP channel (Patterson *et al.*, 1995). The time-interval analysis is usually based on a form of autocorrelation (Licklider, 1951; Slaney and Lyon, 1990; Meddis and Hewitt, 1991a; Meddis and O'Mard, 1997). Autocorrelation models are now known to have several deficiencies, most of which are related to the fact that autocorrelation produces a representation analogous to an all-order time-interval histogram, whereas the auditory system appears to use only a subset of all possible time intervals (Kaernbach and Demany, 1998; Kaernbach and Bering, 2001). Patterson *et al.* (1992) proposed an alternative mechanism, referred to as strobed temporal integration, which restricts the analysis to time intervals measured from local activity peaks within a given NAP channel. Strobed temporal integration is similar to autocorrelation, but has the advantage of preserving short-term temporal asymmetry, which listeners are able to hear and which autocorrelation is insensitive to (Patterson and Irino, 1998). An example is the difference between time-reversed sounds such as damped and ramped sinusoids (Patterson, 1994a/b).

In both autocorrelation and strobed temporal integration, the time-interval analysis in one frequency channel of the NAP is entirely independent of the activity in other channels and any interaction between channels is relegated to a subsequent processing stage. In order to make quantitative predictions, it is often necessary to calculate a summary statistic, which involves combining the time-interval information across frequency channels (Meddis and Hewitt, 1991a/b; Meddis and O'Mard, 1997). In the neural response to broadband periodic click trains, for instance, the time-interval distribution within individual channels reflects not only the repetition period of the waveform but also the carrier period corresponding to the channel's characteristic frequency. A convenient way to explain the pitch elicited by periodic click

[a)]Corresponding author. Current address: MRC Institute of Hearing Research, University Park, Nottingham NG7 2RD, United Kingdom; electronic mail: katrin@ihr.mrc.ac.uk

 0001-4966/2005/118(2)/946/9/$22.50

trains is to simply sum the individual time-interval histograms across channels, which reinforces the repetition period and suppresses individual carrier periods. Typically, the cross-channel summation encompasses all channels in the simulated NAP. This approach, however, was guided more by computational simplicity rather than specific data about auditory processing, and will probably prove overly simplistic (see, e.g., Oxemham *et al.*, 2004). For example, the degree to which time-interval information from different channels is integrated might be expected to depend on the frequency separation between the channels, or cross-channel interactions may occur at the initial stage of time-interval analysis, rather than after this stage. The selection of the time intervals that are included in each channel's histogram, for instance, may be influenced by activity in flanking channels.

The current study used both psychophysics and modeling to investigate whether temporal regularity perception in one frequency channel is affected by synchronous activity in other channels, and how such cross-channel interactions are brought about. In particular, we wanted to examine whether cross-channel interactions occur at or after the initial stage of time-interval extraction, and how they depend on the channels' frequency separation.

II. METHOD

Regular and irregular click trains were filtered into relatively narrow frequency bands to produce target and flanker stimuli. Listeners were required to discriminate a regular target from an irregular distracter click train in the presence of an irregular masker click train in the target band. The target and distracter click trains had the same level, so that listeners had to base their decisions on sound-quality differences produced by the difference in regularity. In all but one condition, the target-band stimuli were presented together with a spectrally remote flanker click train, which was either regular or irregular. When the flanker was regular, it was either synchronous with, or delayed relative to, the regular target click train in the target band. The delayed regular flanker condition was included in order to distinguish between cross-channel interactions at or after the initial stage of time-interval analysis; cross-channel interactions at the time-interval extraction stage would be expected to depend on the relative delay between the interacting stimuli, while cross-channel interactions after the time-interval extraction stage would not. The flanker band was either below or above the target band and the frequency separation between the target and flanker bands was varied from half an octave to two octaves in both directions.

A. Stimuli and equipment

Trains of monophasic 40 μs clicks were generated digitally and filtered into 200 Hz wide frequency bands using 150th-order, Blackmann-windowed FIR filters. The target band was centered at 1.6 kHz and it contained two click trains, the regular target and the irregular masker click train in one of the two observation intervals that constituted each trial, and the irregular distracter and the masker click train in the other interval. The task of the listener was to choose the interval that contained the regular target click train. The level of the masker was fixed at 55 dB SPL; the levels of the target and distracter were identical and were varied together in an adaptive way to determine the threshold target/distracter-to-masker ratio for discriminating the regularity difference between the target and the distracter. As the target and distracter click trains had the same level, listeners could not base their decisions on loudness differences.

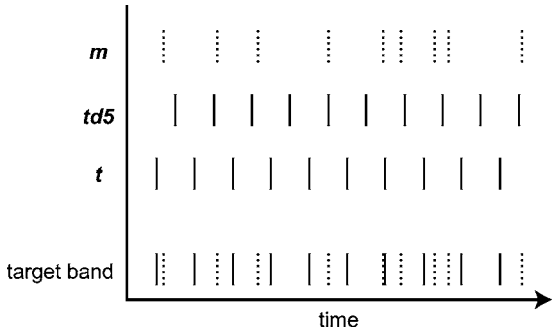


FIG. 1. Schematic representation of the stimuli: In this example, the target band (bottom trace) contains the regular target click train (solid lines) and the irregular masker click train (dotted lines). The three different flanker conditions are illustrated separately above the target band; **t**: flanker-like target; **td5**: flanker-like target, but with a 5 ms delay; **m**: flanker-like masker. Only one flanker was presented with the target band in the experiment.

In all irregular click trains (masker and distracter), the interclick intervals (ICIs) were randomly chosen from a uniform distribution between ST and 20 ms, where ST is the sampling period (40 μs), yielding a mean ICI of 10 ms and a mean click rate of 100 Hz. The irregular masker and distracter click trains were uncorrelated and created anew for each trial. The regular target click train had a fixed ICI of 10 ms, corresponding to a repetition rate of 100 Hz. The center frequency of the target band (1.6 kHz) corresponds to the 16th harmonic of 100 Hz, which means that the regular target click train was spectrally unresolved (see, e.g., Bernstein and Oxenham, 2003).

In all but one condition, the stimuli also contained a flanker band (200 Hz wide) centered at one of eight frequencies. Four of the flanker frequencies were below, and four above the target band, and within each set, the frequencies were spaced at half-octave intervals. The frequencies were 0.40, 0.57, 0.80, 1.13, 2.26, 3.20, 4.53, and 6.40 kHz. The frequency separation between the target- and flanker-band center frequencies ranged from 0.5 to 2 octaves in both directions. The flanker band contained only one click train with a fixed overall level of 55 dB SPL, like the masker click train in the target band. In the "flanker-like target" condition, designated **t**, the flanker click train was regular and synchronous with the regular target click train in the target band. In the "flanker-like target, but with 5 ms delay" condition, designated **td5**, the flanker click train was regular, but delayed relative to the target click train by half the repetition period (5 ms). Finally, in the "flanker-like masker" condition, designated **m**, the flanker was irregular and the flanker clicks were synchronous with the masker clicks in the target band. Figure 1 illustrates the target and flanker click trains for the three flanker modes. In this example, the target band contains the regular target click train (rather than the irregular dis-

tracter click train). In the "no flanker" condition, the target band was presented without any flanker band. In total, the experiment contained $3 \times 8 + 1 = 25$ conditions.

All of the click trains had a duration of 800 ms and were gated on and off with 25 ms cosine ramps. To mask any aural distortion products at the target repetition rate (100 Hz), a continuous low pass (second-order Butterworth) noise was presented with an overall level of 45 dB SPL. The low pass cutoff of the noise was at 0.25 kHz, which is 50 Hz below the lower edge of the lowest flanker band (centered at 0.4 kHz). The stimuli were generated digitally with Matlab in conjunction with TDT System 3 and digital to analog converted (TDT RP2.1). They were amplified (TDT HB7) and presented binaurally through AKG K240 DF headphones to the listener, who was seated in an echo-free, sound-insulated room (ENT Clinic of the University Hospital Münster, Germany).

B. Procedure

Each experimental trial consisted of two 800 ms observation intervals, during which the click trains were presented. The observation intervals were separated by a silent gap of 700 ms. In one of the two intervals, the target band contained the regular target click train, which the listeners were asked to detect. The other interval contained an irregular distracter click train with the same level as the target click train. Listeners were asked to judge which of the two stimuli contained the regular target, rather than the irregular distracter, by exploiting the fact that the target interval sounded more regular, or less crackly, than the distracter interval. The target interval was chosen randomly in each trial, and there was an equal probability for the target to occur in the first or second interval. The target/distracter level was varied adaptively using a three-down, one-up rule to track the signal level corresponding to 79% correct performance (Levitt, 1971). Visual feedback was provided. The level increments and decrements were 5 dB up to the first reversal of level, 3 dB from there to the second reversal, and 2 dB for the rest of the 10 reversals that made up each threshold run. The threshold was taken to be the average of the target/distracter level (in dB) at the last eight reversals. Each listener completed at least three threshold runs for each condition. Where the standard error of these three threshold estimates exceeded 2 dB, one or two more threshold estimates were obtained. Thus, the data points presented in the figures are the mean of three to five threshold estimates and the error bars show their standard error. The order in which conditions were tested was counterbalanced between the three threshold runs.

C. Listeners

Five listeners (two male, three female) participated in the experiment; they were authors KK and MS, and three students, who were paid for their services at an hourly rate. The listeners were between 21 and 32 years of age and had no reported history of hearing impairment or neurological disease.

III. RESULTS

The regularity thresholds for five listeners are presented as a function of the center frequency of the flanker band in Fig. 2; the average data are in the top left panel, the data for the individuals are in the remaining five panels. Threshold is expressed as the level (in dB) of the target click train relative to the masker click train in the target band. The masker had a fixed level of 55 dB SPL. The vertical dotted line marks the center frequency of the target band. The parameter is flanker mode (**t**, **td5**, and **m**, as noted in the legend of the top left panel). The thresholds for the conditions in which the flanker was regular (**t** and **td5**) are connected by solid lines; the thresholds for the irregular-flanker condition (**m**) are connected by dashed lines; the horizontal solid line shows the threshold for the no-flanker condition.

Figure 2 shows that regularity threshold was by about 5–7 dB lower in the regular flanker conditions (**t** and **td5**) than in the irregular flanker condition (**m**). This effect of flanker regularity was mainly due to a threshold increase for the irregular flanker (**m**) compared to the no-flanker condition, rather than a threshold decrease for the regular flanker conditions. The threshold difference between the regular and irregular flanker conditions, in the following referred to as the flanker regularity effect (FRE), was largely independent of the frequency separation between the target and flanker bands, but there was a tendency for both the absolute thresholds and the FRE to be larger for flankers below than above the target band. This effect of flanker frequency was quite pronounced in some listeners (e.g., KK), but almost absent in others (e.g., TM).

To determine the statistical significance of these results, the individual thresholds were submitted to a two-way repeated-measures ANOVA with flanker mode (**t**, **td5**, and **m**) and flanker frequency (eight levels) as independent within-subject factors. The analysis revealed significant main effects of flanker mode [$F(2,8)=53.588$, $p<0.0001$] and flanker frequency [$F(7,28)=4.773$, $p=0.0012$] as well as a significant interaction between these two factors [$F(14,56)=4.314$, $p<0.0001$].

Fischer PLSD post-hoc tests showed that the main effect of the flanker mode was due to significant differences between the regular flanker conditions (**t** and **td5**) on the one hand and the irregular flanker condition (**m**) on the other hand ($p<0.0001$ for both **t** vs **m** and **td5** vs **m**); the difference between the regular flanker conditions (**t** versus **td5**) did not reach significance ($p=0.0584$). The main effect of flanker frequency was mainly due to significant differences between thresholds for flanker frequencies below and above the target band; threshold differences within these two groups were not significant apart from the difference between 2.26 and 4.53 kHz ($p=0.0362$). Figure 3 shows that the interaction between flanker mode and flanker frequency was due to the fact that the difference between flanker frequencies below (squares) and above the target band (circles) was larger for the **m** condition than for the **t** and **td5** conditions, resulting in a larger FRE for flanker frequencies below than above the target band.

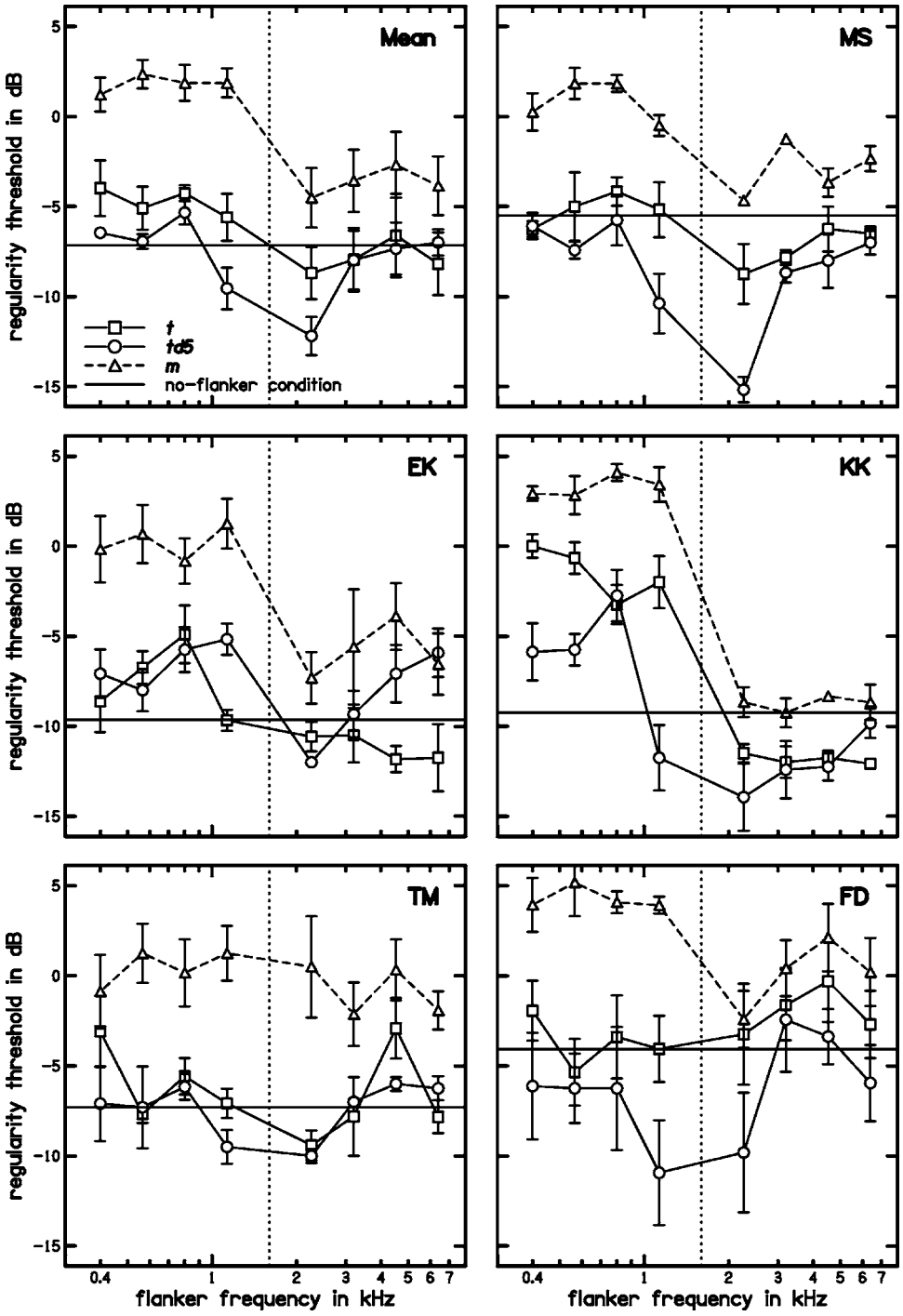


FIG. 2. Mean (top left panel) and individual (other panels) regularity thresholds as a function of flanker frequency; the parameter is the flanker mode (see the legend in the top left panel). The horizontal solid lines show the threshold for the no-flanker condition. The vertical dotted lines mark the center frequency of the target band.

While the Fischer PLSD post-hoc tests had revealed no significant difference between the **t** and **td5** conditions (see above), Fig. 2 suggests that the **td5** condition yielded lower thresholds than the **t** condition for flanker frequencies of 1.13 and 2.26 kHz, i.e., immediately below and above the target band. Paired *t* tests, however, showed that these differences did not reach significance, although the difference at 2.26 kHz was close to significant ($p=0.051$). The difference between the **td5** condition for 2.26 kHz and the no-flanker condition, however, did reach significance, indicating that there was a significant masking release for the delayed regular flanker immediately above the maker band.

IV. DISCUSSION

The current data indicate that temporal regularity, or pitch, is not always perceived independently within each peripheral channel. Rather, there appears to be a possibility for interference, or cross-talk, even between widely separated frequency regions. The frequency separation between the target and flanker bands ranged from 0.5 to 2 octaves. For all but possibly the smallest frequency separation (0.5 octaves), the target- and flanker-band click trains would not be expected to interact in the periphery, and so the observed effect of flanker regularity on regularity detection in the target band (the FRE) must have been generated more centrally. The fact

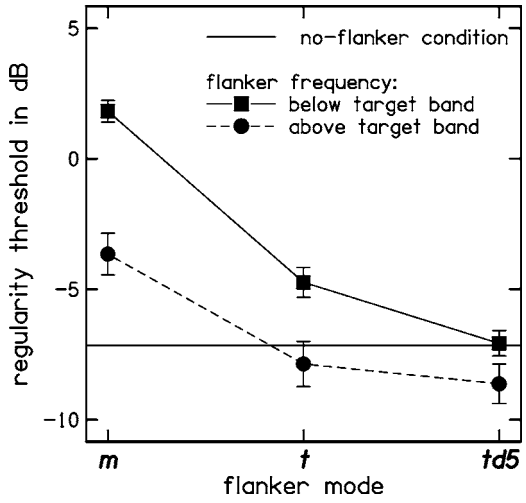


FIG. 3. Mean threshold, averaged over flanker frequencies below (squares) or above (circles) the target band, for each of the three flanker modes. As in Fig. 2, the horizontal line shows the threshold for the no-flanker condition.

that the FRE was largely independent of whether the flanker was synchronous with or delayed relative to the regular target suggests that the effect is based on cross-channel interactions after rather than at the time-interval extraction stage. This notion is corroborated by model simulations presented in the next section.

The FRE observed in the current study appears to be related to the phenomenon of pitch discrimination interference (PDI), which was recently discovered by Gockel *et al.* (2004). PDI refers to the fact that pitch discrimination in a filtered complex tone (target) can be severely impaired in the presence of another, spectrally remote, complex tone (interferer). Gockel *et al.* showed that the amount of interference depends on the salience of the pitch elicited by the interferer tone; a spectrally resolved interferer with a strong pitch produced a larger amount of interference than a spectrally unresolved interferer with a weaker pitch. This finding may be related to the current finding that the FRE tended to be larger when the flanker was below than above the target band. In the current study, the effect of flanker frequency did not appear to be strictly correlated with the spectral resolvability of harmonics; the limit of spectral resolution is generally assumed to be around the tenth harmonic (see, e.g., Berstein and Oxenham, 2003), which means that the regular click trains in all but the three lowest frequency bands (0.4, 0.57, and 0.8 kHz) were unresolved. In particular, the 1.13 kHz flanker click trains were probably unresolved, and yet, according to the statistical tests, the 1.13 kHz flanker band yielded similar regularity thresholds as the three lowest (resolved) flanker bands. The absence of any clear effect of harmonic resolvability may be due to the fact that the FRE was mainly due to a threshold increase for the irregular flanker condition, for which harmonic resolvability is undefined, and thus probably irrelevant. This suggests that, even for the three lowest, spectrally resolved flanker bands, the FRE was brought about by interactions in the temporal processing of the target- and flanker-band stimuli. The fact that the FRE was larger for flanker bands below than above the target band may be due to pitch-related temporal information being represented more accurately, and thus being more salient, in low- than in high-frequency channels (Krumbholz *et al.*, 2000; Pressnitzer *et al.*, 2001). The notion that the FRE was brought about by interactions in temporal processing is in accordance with the conclusion of Gockel *et al.* that their results on PDI are consistent with the existence of a single, temporal, pitch mechanism for the analysis of both resolved and unresolved complex tones. In the next section we explore how cross-channel interactions can be implemented in a temporal model of pitch perception.

The masking release for the **td5** flankers immediately below and above the target band just barely reached significance at 2.26 kHz (see Sec. III); future experiments will have to reveal whether it is a truly robust effect. If it were to be robust, it may be produced by peripheral interactions between the flanker and target stimuli. Alternatively, it may indicate the existence of two different kinds of cross-channel processes: one that operates over wide frequency separations and is independent of the relative phase of the interacting signals, and another one that operates over narrow frequency separations and depends on the relative phase, and may thus be part of the pitch extraction process. The model simulations presented in the next section suggest that the masking release in the **td5** condition at 2.26 kHz, i.e., immediately above the target band, may indeed be a true cross-channel effect.

V. MODELING

A. Cross-channel interactions after the time-interval analysis stage

1. Model architecture

The simulations were all based on a modified version of the auditory image model (AIM) presented in Patterson *et al.* (1995). The modeling was performed with the DSAM/AMS software package (http://www.essex.ac.uk/psychology/hearinglab/dsam) and MATLAB. The current implementation of the model consisted of (i) a second-order bandpass filter with cutoffs of 0.45 and 8.5 kHz, to simulate the action of the middle ear, (ii) a gammatone filterbank (**gtf**), to simulate the spectral analysis performed by the cochlea [the gammatone filter frequencies were evenly distributed on the ERB scale (Glasberg and Moore, 1990) with a density of 8.5544 filters per ERB], (iii) half-wave rectification, square-root compression, and fourth-order lowpass filtering with a 0.8 kHz cutoff frequency (**hcl**) in each channel, to simulate the neural activity pattern (NAP) flowing up the auditory nerve, and (iv) a channel-by-channel time-interval analysis performed by strobed temporal integration (**sti**; Patterson, 1994b; Patterson and Irino, 1998), which is similar to autocorrelation. Together, these operations produce a tonotopic array of time-interval histograms, which is referred to as an auditory image (AI). Similar models have been used by Krumbholz *et al.* (2001, 2003) to simulate the masking properties of noise and iterated rippled noise.

In the first set of simulations, described in this section, the time-interval information was analyzed independently within each frequency channel, and the cross-channel processing was performed *after* the time-interval analysis stage.

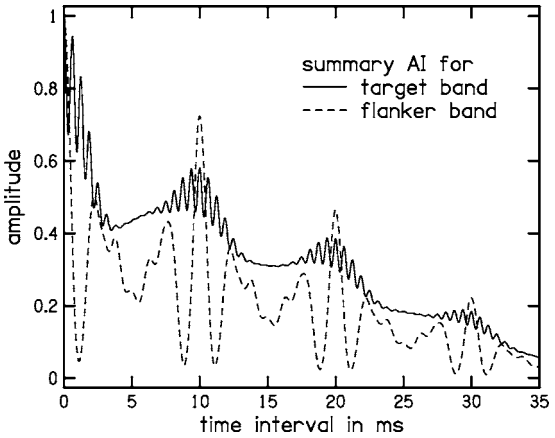


FIG. 4. Summary AIs for the target (solid line) and flanker bands (dashed line). In this example, the target band contained the regular target click train, and the flanker click train was centered at 0.4 kHz and was regular and synchronous with the regular target click train (**t**).

For each stimulus condition, 75 AIs were computed in successive 25 ms steps and averaged to produce a representative AI for that condition. Two "summary" AIs were derived from each of these AIs: one for the target band and one for the flanker band. They were generated by averaging the AI across 350 Hz wide bands around the target- and flanker-band center frequencies, and normalizing the resulting images to the value at 0 ms. Figure 4 shows the summary AIs of the target and flanker bands (solid and dashed lines, respectively) for the condition in which the target band contained the regular target click train with a relative level of 0 dB, and the flanker click train at 0.4 kHz was regular and synchronous with the target (**t**). Finally, the summary AIs for the target and flanker bands were combined in different ways (see later) to produce a combined summary AI, which was used to calculate the decision measure. The decision measure was simply the squared Euclidian distance, D^2 (Meddis and O'Mard, 1997), between the combined summary AIs for corresponding target and distracter stimuli for a single target/distracter level of 0 dB. D^2 is a measure of the *dissimilarity* between the stimulus representations. A large dissimilarity predicts a good discriminability and thus a low threshold. Thus, in case of a good match between model and data, D^2 would be inversely related to the measured threshold. D^2 rather than any more specific decision measure (see, e.g., Krumbholz *et al.*, 2001) was used, to account for any sound-quality differences between the target and distracter stimuli that listeners may have been using for their decisions.

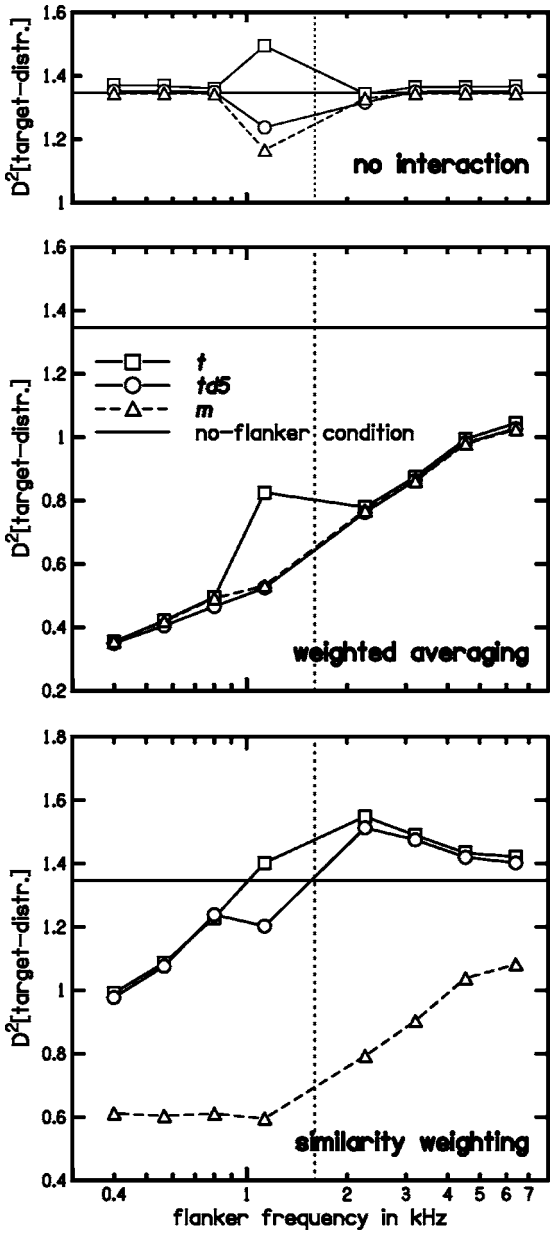


FIG. 5. Results of the first set of model simulations, in which the cross-channel interactions were implemented after the time-interval analysis stage. In all cases, the abscissa represents the flanker frequency and the ordinate shows the squared difference (D^2) between the stimuli in which the target band contained the regular target or the irregular distracter click train at a signal-to-masker level of 0 dB (see the text). The upper panel shows the results when D^2 is calculated from the summary AI of the target band alone, without any consideration of the flanker image (no interaction). The middle panel depicts the results of the simulation in which D^2 was calculated from a weighted average of the target and flanker images; the weighting factors corresponded to the number of channels encompassed by the target and flanker band, respectively. The lower panel shows the results of the similarity weighting model, in which D^2 was calculated from the target image after weighting that image with an image representing the similarity between the target and flanker images.

2. Simulation results

a. No interactions. In the first simulation, D^2 was calculated for the summary AI of the target band alone, without any consideration of the flanker image (upper panel of Fig. 5). In that case, D^2 was little affected by the presence of a flanker for most of the flanker frequencies, as would be expected. Only for the flanker frequency directly below the target band (1.13 kHz) did D^2 noticeably depend on the flanker mode. This suggests that, only for this flanker frequency, was there a possibility for a direct mechanical interaction between the target- and flanker-band click trains within a single auditory filter.

b. Weighted averaging. In the second simulation, cross-channel processes were implemented by a simple weighted averaging of the summary AIs for the target and

flanker bands. In order to accommodate for the fact that the flanker band encompassed varying numbers of frequency channels depending on the flanker frequency, the target and flanker images were weighted by the number of channels for which the respective image was calculated (corresponding to 350 Hz bands around the target- and flanker-band center frequencies). The images were then added and the resulting image divided by the total number of channels. The middle panel of Fig. 5 shows that this weighted averaging of the target and flanker images had the effect of reducing D^2 for all conditions with a flanker relative to the no-flanker condition. Due to the weighting by the number of channels, the reduction in D^2 decreased with increasing flanker frequency. Thus, the weighted averaging model correctly predicted the fact that most of the conditions with a flanker yielded a larger threshold than the no-flanker condition and the tendency for the threshold to be larger for the lower than for the higher flanker frequencies (see Fig. 2). However, the most important aspect of the data, namely, the threshold difference between the regular and irregular flanker conditions, was not predicted by the model; as for the model without any cross-channel interactions, the only flanker frequency for which D^2 diverged between the three flanker modes was the one directly below the target band.

c. Similarity weighting. In the third simulation, the effect of the flanker was modeled by multiplying the target image with an image representing the similarity between the target and flanker images as a function of time interval. The similarity image was calculated by (a) taking the absolute value of the difference between the target and flanker images, (b) multiplying it by the ratio of the number of channels in the target and flanker bands, and (c) subtracting the resulting image from unity. The similarity image was given by $\mathrm{Sim}(TI)=1-\min(1,N_{\mathrm{target}}/N_{\mathrm{flanker}})\cdot|sAI_{\mathrm{target}}(TI)-sAI_{\mathrm{flanker}}(TI)|$, where TI denotes time interval, N_{target} and N_{flanker} denote the number of frequency channels in the 350 Hz bands around the target- and flanker-band center frequencies, and $sAI_{\mathrm{target}}(TI)$ and $sAI_{\mathrm{flanker}}(TI)$ are the summary AIs for the target and flanker bands, respectively. In order to prevent the similarity image from becoming negative, the ratio of the number of channels was truncated at unity. Consequently, the range of the similarity image was between zero and unity.

In contrast to the weighted averaging model, this similarity weighting model yielded a clear D^2 difference between the regular and irregular flanker conditions, and the direction of the D^2 difference mirrored the difference between the measured thresholds for these conditions (the lower panel of Fig. 5). Like the weighted averaging model, the similarity weighting model also predicted the observed decrease in threshold for increasing flanker frequencies. Moreover, the similarity weighting model correctly predicted the fact that threshold for the regular flanker conditions (**t** and **td5**) was larger than the no-flanker threshold for flanker frequencies below the target band, and slightly smaller for flanker frequencies above the target band, while the threshold for the irregular flanker condition (**m**) was larger than the threshold for the no-flanker condition for all flanker frequencies. Thus, the similarity weighting model provides a reasonable account of most aspects of the current data.

B. Cross-channel effects at the level of time-interval extraction

1. Model architecture

In the second set of simulations, the cross-channel interactions were implemented at, rather than after, the time-interval analysis stage by allowing the activity in the flanker band to influence the strobed temporal integration (**sti**) process in the target band. In **sti**, local peaks in the activity within each channel elicit strobe pulses, which cause a copy of the preceding 35 ms of the NAP in the channel to be added into the corresponding channel of the auditory image (AI; Patterson, 1994b). In the original version of **sti**, the determination of strobe times and the adding of NAP segments in one AI channel was entirely independent of the activity in the other channels. In the current model, cross-channel interactions were implemented by multiplying the NAP segments in a given channel by a weighting factor that depended on the strobe times in the other channels. The simulation was limited to the 350 Hz band around the target band center frequency. The weighting factor for a given strobe in a given channel in this 350 Hz band was equal to the proportion of all channels in the 350 Hz bands around the target- and flanker-band center frequencies that contained a strobe within a 2 ms window around the time point of the relevant strobe. Thus, time intervals measured from events that elicited synchronous strobing in a large proportion of both the target- and flanker-band channels were weighted more strongly than time intervals from events that elicited strobes in only a few channels. The resulting AI was averaged across the 350 Hz band around the target-band center frequency to produce a summary AI for the target band. In all other respects, the current simulation was similar to the simulations described in the previous section.

2. Simulation results

a. Time-interval weighting. The upper panel of Fig. 6 shows that the cross-channel weighting had the effect of increasing the D^2 for the regular and synchronous flanker condition (**t**; open squares) at the medium flanker frequencies (1.13, 2.26 and, to a lesser extent, also 3.2 kHz), while decreasing the D^2 for the regular and delayed (**td5**; circles) as well as the irregular (**m**; triangles) flanker conditions relative to the no-flanker condition. The effect was reversed at the lowest flanker frequencies (0.4 and 0.57 kHz), and there was little effect at the highest flanker frequencies (4.53 and 6.4 kHz). The increase in D^2 for the **t** condition at the medium flanker frequencies was probably brought about by an increase in the weighting of time intervals measured from clicks in the regular target click train, due to the strobes elicited by the flanker clicks being synchronous with those elicited by the regular target clicks. In contrast, both a regular and delayed (**td5**) and an irregular (**m**) flanker click train at a frequency close to the target band would be expected to increase the weighting of time intervals measured from

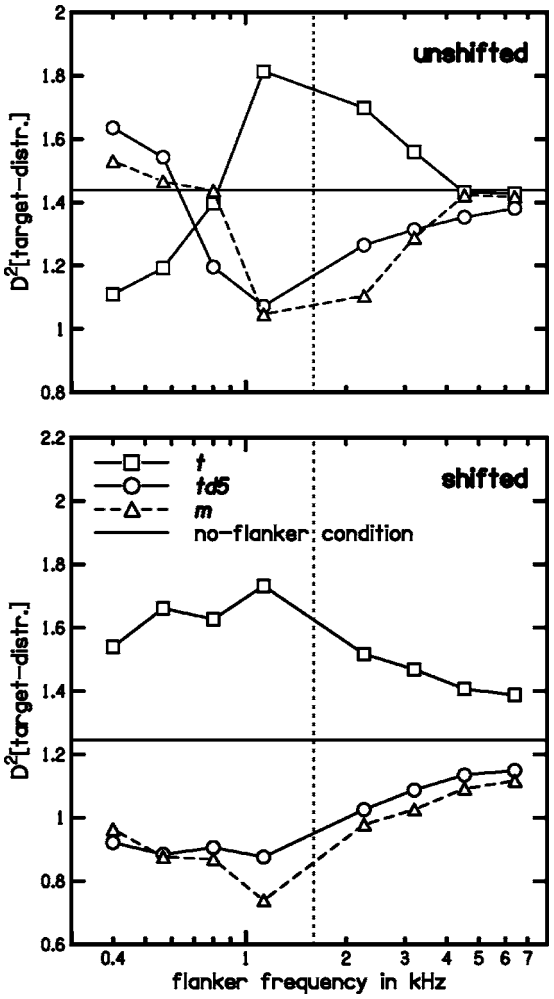


FIG. 6. Results of the second set of model simulations, in which cross-channel interactions were implemented at the time-interval analysis stage by weighting the transfer of time-interval information from a given NAP channel to the respective time-interval histogram with a factor that represented the amount of synchronous strobing activity in the other channels. The upper panel shows the results of the model in which the modified time-interval analysis was applied to the original NAP simulation, which had also been the front end of the previous set of simulations. The lower panel depicts the results of the model in which the NAP channels were shifted in time to compensate for the traveling-wave delay before the time-interval analysis was applied.

events other than the regular target clicks (e.g., the masker clicks). This may explain the relative decrease in D^2 for the **td5** and **m** conditions at the medium flanker frequencies. The waning of the D^2 differences toward higher flanker frequencies is probably due to the decrease in the number of frequency channels encompassed by the higher flanker bands and the resulting lessening of their relative effect on the target band. The reversal of the pattern at the lowest flanker frequencies, on the other hand, is most likely due to relative temporal shifts between the simulated neural responses in the target- and flanker-band channels, introduced by differences between the peak latencies of the respective gammatone-filter responses, which mimic the increase in the traveling-wave delay towards lower frequencies (Robles and Ruggero, 2001).

b. Time-interval weighting after realignment for traveling-wave delay. In order to test this assumption, we realigned the NAP by temporally shifting each channel's response by the peak latency of the respective gammatone filter's impulse response to neutralize the traveling-wave delay, and then repeated the above simulation on these realigned responses. The peak latency was defined as the time point at which the envelope of the filter's impulse response reaches its maximum, which is given by $\text{shift}(f_c)=(N-1)/[2\pi\cdot 1.019\cdot(0.0247+0.108\cdot f_c)]$ ms, where f_c is the filter center frequency in kHz (see Patterson, 1994a).

As expected, the temporal realignment of the NAP channels eliminated the reversal in the pattern of D^2 between the lower and medium flanker frequencies, indicating that this reversal was indeed related to the traveling-wave delay. The size of the D^2 differences still decreased towards higher flanker frequencies, but to a lesser degree than in the previous simulation, in which no temporal shifting had been applied. This indicates that the high-frequency behavior of the previous model was, at least to a certain degree, also determined by the traveling-wave delay.

VI. SUMMARY

The current results suggest that the perception of temporal regularity, or pitch, involves interactions between even widely separated frequency channels. The detection of regularity in the target band click trains was easier when the target band was presented together with a regular than an irregular flanker click train. Regularity detection performance was little affected, or in some cases even slightly enhanced, when the regular flanker click train was delayed relative to the regular target click train, suggesting that the cross-channel interactions underlying the observed effect of flanker regularity occur after rather than at the pitch extraction stage. This notion was corroborated by the modeling results: implementing cross-channel interactions at the time-interval analysis stage of a multichannel time-interval model of pitch predicted the effect of the flanker to be strongly dependent on the relative delay between the target and flanker click trains, which is inconsistent with the data; in contrast, implementing the interactions after the time-interval analysis stage correctly predicted the effect to be independent of the delay.

The threshold difference between the regular and the irregular flanker conditions (the flanker regularity effect, or FRE) was roughly independent of the frequency separation between the target and flanker bands, but there was a tendency for the threshold difference as well as the absolute threshold values to be larger for flanker frequencies below than above the target band. Recent results by Gockel *et al.* (2004) suggest that this effect of flanker frequency was due to the fact that the lower-frequency flankers mediated more salient pitch information than the higher-frequency flankers. The modeling showed that, in the current study, the effect could be simulated by taking account of the greater number of channels encompassed by the lower- than the higher-frequency flankers.

The modeling also showed that the FRE could not be explained by simply adding the time-interval distributions from the target and flanker bands; the only model of cross-

channel interactions that could correctly account for the difference between the regular and irregular flankers was the similarity weighting model, in which the time-interval distribution for the target band was weighted by the similarity between the time-interval distributions for the target and flanker bands. This suggests that cross-channel interactions in pitch perception involve processes that measure the similarity, or consistency, of pitch-related temporal information across channels. Similar processes have also been suggested to be involved in cross-channel interactions in the modulation domain (Buus, 1985; Hall, 1986; Richards, 1987).

Bernstein, J. G., and Oxenham A. J. (**2003**). "Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? J. Acoust. Soc. Am. **113**, 3323–3334.

Burns, E. M., and Viemeister, N. F. (**1976**). "Non-spectral pitch," J. Acoust. Soc. Am. **60**, 863–869.

Buus, S. (**1985**). "Release from masking caused by envelope fluctuations,"J. Acoust. Soc. Am. **78**, 1958–1965.

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Gockel, H., Carlyon, R. P., and Plack, C. J. (**2004**). "Across-frequency interference effects in fundamental frequency discrimination: questioning evidence for two pitch mechanisms," J. Acoust. Soc. Am. **116**, 1092–1104.

Hall, J. W. (**1986**). "The effect of across-frequency differences in masking level on spectro-temporal pattern analysis," J. Acoust. Soc. Am. **79**, 781–787.

Kaernbach, C., and Bering, C. (**2001**). "Exploring the temporal mechanism involved in the pitch of unresolved harmonics," J. Acoust. Soc. Am. **110**, 1039–1048.

Kaernbach, C., and Demany, L. (**1998**). "Psychophysical evidence against the autocorrelation theory of auditory temporal processing," J. Acoust. Soc. Am. **104**, 2298–2306.

Krumbholz, K., Patterson, R. D., and Nobbe, A. (**2001**). "Asymmetry of masking between noise and iterated rippled noise: evidence for time-interval processing in the auditory system," J. Acoust. Soc. Am. **110**, 2096–2107.

Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (**2000**). "The lower limit of pitch as determined by rate discrimination," J. Acoust. Soc. Am. **108**, 1170–1180.

Krumbholz, K., Patterson, R. D., Nobbe, A., and Fastl, H. (**2003**). "Microsecond temporal resolution in monaural hearing without spectral cues?" J. Acoust. Soc. Am. **113**, 2790–2800.

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

Licklider, J. C. R. (**1951**). "A duplex theory of pitch perception," Experientia **7**, 128–133.

Meddis, R., and Hewitt, M. J. (**1991a**). "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I. Pitch identification," J. Acoust. Soc. Am. **89**, 2866–2882.

Meddis, R., and Hewitt, M. J. (**1991b**) "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II. Phase sensitivity," J. Acoust. Soc. Am. **89**, 2883–2894.

Meddis, R., and O'Mard, L. (**1997**). "A unitary model of pitch perception," J. Acoust. Soc. Am. **102**, 1811–1820.

Oxenham, A. J., Bernstein, J. G., and Penagos, H. (**2004**). "Correct tonotopic representation is necessary for complex pitch perception," Proc. Natl. Acad. Sci. U.S.A. **101**, 1421–1425.

Patterson, R. D. (**1994a**). "The sound of a sinusoid: Spectral models," J. Acoust. Soc. Am. **96**, 1409–1418.

Patterson, R. D. (**1994b**). "The sound of a sinusoid: Time-interval models," J. Acoust. Soc. Am. **96**, 1419–1428.

Patterson, R. D., and Irino, T. (**1998**). "Modeling temporal asymmetry in the auditory system," J. Acoust. Soc. Am. **104**, 2967–2979.

Patterson, R. D., Allerhand, M., 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., Handel, S., Yost, W. A., and Datta, A. J. (**1996**). "The relative strength of tone and noise components of iterated rippled noise," J. Acoust. Soc. Am. **100**, 3286–3294.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1992**). "Complex sounds and auditory images," in *Auditory Physiology and Perception, Proceedings of the 9th International Symposium on Hearing*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 429–446.

Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (**2001**). "Lower limit of melodic pitch," J. Acoust. Soc. Am. **109**, 2074–2084.

Richards, V. M. (**1987**). "Monaural envelope correlation perception," J. Acoust. Soc. Am. **82**, 1621–1630.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Slaney, M., and Lyon, R. F. (**1990**). "A perceptual pitch detector," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Albuquerque, New Mexico* (IEEE, New York), pp. 357–360.

Yost, W. A., Patterson, R. D., and Sheft, S. (**1996**). "A time domain description for the pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **99**, 1066–1078.

Perception of dissonance by people with normal hearing and sensorineural hearing loss[a)]

Jennifer B. Tufts,[b)] Michelle R. Molis, and Marjorie R. Leek
Army Audiology and Speech Center, Walter Reed Army Medical Center, 6900 Georgia Avenue NW, Washington, DC 20307

(Received 21 July 2004; revised 21 April 2005; accepted 4 May 2005)

The purpose of this study was to determine whether the perceived sensory dissonance of pairs of pure tones (PT dyads) or pairs of harmonic complex tones (HC dyads) is altered due to sensorineural hearing loss. Four normal-hearing (NH) and four hearing-impaired (HI) listeners judged the sensory dissonance of PT dyads geometrically centered at 500 and 2000 Hz, and of HC dyads with fundamental frequencies geometrically centered at 500 Hz. The frequency separation of the members of the dyads varied from 0 Hz to just over an octave. In addition, frequency selectivity was assessed at 500 and 2000 Hz for each listener. Maximum dissonance was perceived at frequency separations smaller than the auditory filter bandwidth for both groups of listners, but maximum dissonance for HI listeners occurred at a greater proportion of their bandwidths at 500 Hz than at 2000 Hz. Further, their auditory filter bandwidths at 500 Hz were significantly wider than those of the NH listeners. For both the PT and HC dyads, curves displaying dissonance as a function of frequency separation were more compressed for the HI listeners, possibly reflecting less contrast between their perceptions of consonance and dissonance compared with the NH listeners. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1942347]

PACS number(s): 43.66.Jh, 43.66.Sr, 43.75.Cd [RAL] Pages: 955–967

I. INTRODUCTION

Research and anecdotal evidence suggest that many listeners with sensorineural hearing loss (SNHL) do not perceive music normally, even with the use of hearing aids and/or cochlear implants (e.g., Gfeller *et al.*, 2000; Chasin, 2003). An understanding of how SNHL affects the perception of music, besides providing greater insight into the nature of SNHL, may suggest approaches to improving hearing aids and cochlear implants for both music and speech listening.

Several aspects of the musical signal are affected by SNHL. deLaat and Plomp (1985) found that people with SNHL had greater difficulty recognizing a melody presented simultaneously with two other melodies than did normal-hearing people. Other reports have demonstrated poor complex pitch perception and inaccuracies in pure-tone octave judgments (Arehart and Burns, 1999), abnormal pitch-intensity shifts, frequency difference limens, and pitch-matching variability (Burns and Turner, 1986), and reduced pitch strength (Leek and Summers, 2001) in people with SNHL. To the extent that pitch perception and the identification and segregation of melodic lines are important in music listening, deficits in these areas may alter the perception of music by hearing-impaired listeners.

Another important aspect of musical expression that may be affected by SNHL is the perception of consonance and dissonance. Consonance and dissonance are perceptual attributes of musical intervals that convey variation in musical tension. Generally speaking, a musical interval is described as consonant if it sounds harmonious and restful, while an interval is described as dissonant if it sounds discordant and tense. Music theorists have classified the thirteen musical intervals of the Western tradition in terms of consonance and dissonance. The so-called "perfect" intervals—the unison, octave, fifth, and fourth—are considered to be highly consonant. The major and minor thirds and sixths are considered to be imperfect consonances, that is, less consonant than the perfect intervals; the major and minor seconds and sevenths, and the tritone, are categorized as dissonant intervals (Hutchinson and Knopoff, 1978; Huron, 2001). The classification of musical intervals in this way suggests that the ability to distinguish sensations of consonance and dissonance is important for music perception.

Consonance and dissonance have been investigated as phenomena not only of musical relevance, but also of psychoacoustic interest. In a purely psychoacoustic context, *sensory dissonance* refers to the degree to which a tone complex, presented in isolation, sounds dissonant, i.e., tense, rough, or unpleasant, while *sensory consonance* refers to a restful, smooth, or pleasant quality (Terhardt, 1984). Sensory consonance and dissonance provide the basis for the classification of intervals as *musically* consonant or dissonant. However, this relationship is not completely straightforward. For example, a minor third, which is considered a musically consonant interval, sounds dissonant when played in a very low register.

Sensory dissonance is generally regarded to be a consequence of the imperfect frequency resolution of the basilar membrane (Greenwood, 1991). According to von Helmholtz

[a)]Portions of this work were presented at the 147th meeting of the Acoustical Society of America, New York, NY, May 2004, and at the International Conference for Music Perception and Cognition, Evanston, IL, August 2004.

[b)]Current address: Department of Communication Sciences, University of Connecticut, Storrs, CT 06269; electronic mail: jennifer.tufts@uconn.edu

(1877/1954), sensory dissonance results from the perception of fast beats between two tones that are closely spaced in frequency. The beating tones, which are unresolved on the basilar membrane, produce amplitude fluctuations within a single auditory channel. These amplitude fluctuations create sensations of roughness and dissonance (Terhardt, 1978). Conversely, two tones that are spaced further apart in frequency do not interact within a single auditory filter to produce dissonance.

Several researchers have investigated the relationship between sensory dissonance and the frequency separation of two simultaneous pure tones in normal-hearing listeners (Plomp and Levelt, 1965; Plomp and Steeneken, 1968; Kameoka and Kuriyagawa, 1969a). In these studies, listeners evaluated the sensory consonance or dissonance of pure-tone pairs that were created either by fixing the frequency of one pure tone and varying the frequency of another, higher-frequency pure tone, or by varying the frequencies of two pure tones around a common geometric mean. These studies indicated that, across a wide frequency range, the sensory dissonance of two simultaneous pure tones is a relatively smooth function of their frequency separation in Hz. At a separation of 0 Hz, the two pure tones exactly coincide in frequency and therefore produce no dissonance. As the separation of the pure tones increases slightly from 0 Hz, beats are heard, first as fluctuations in loudness, then as a sensation of roughness. As a result, sensory dissonance increases rapidly with increasing frequency separation from 0 Hz. The frequency separation at which maximal dissonance is reached occurs at about 25%–40% of the critical bandwidth in the frequency region of the tones (Plomp and Levelt, 1965; Greenwood, 1991). As the separation between the pure tones widens further, the sensation of roughness subsides and sensory dissonance decreases. At approximately one critical bandwidth (Plomp and Levelt, 1965; Plomp and Steeneken, 1968; Greenwood, 1991), the two tones are resolved on the basilar membrane and produce a smoother sensation, with minimal dissonance. Further increases in frequency separation produce relatively little change in perceived sensory dissonance. This pattern has been observed for pure-tone pairs with lower-frequency components ranging from approximately 125 to 7000 Hz (Plomp and Levelt, 1965; Kameoka and Kuriyagawa, 1969a).

Unlike pure tone pairs, most musical sounds have many components that may interact to produce sensory dissonance. The sensory dissonance created by simultaneous harmonic complex tones is of particular interest, because the singing voice and the sounds of many musical instruments have harmonic complex spectra. While the sensory dissonance of pure-tone pairs is a relatively smooth function of their frequency separation, the dissonance of harmonic complex tone pairs varies markedly depending on their fundamental frequency (F0) ratios and their amplitude spectra (Plomp and Levelt, 1965). Specifically, intervals with small-integer F0 ratios (such as the octave, with an F0 ratio of 2:1) sound consonant, while intervals with larger-integer F0 ratios (such as the minor second, with an F0 ratio of 16:15) sound dissonant. Small-integer ratios sound consonant ostensibly because several of the partials of the two tones are either identical in frequency, or are spaced sufficiently far apart that they do not create roughness. On the other hand, larger-integer F0 ratios sound dissonant because many of the partials are noncoinciding and closely spaced. Given this explanation of the relationship among consonance, dissonance, and F0 ratio, it follows that the use of consonance and dissonance in music is predicated on normal or near-normal peripheral frequency resolution. For this reason, the perception of sensory consonance and dissonance by people with SNHL is of interest from a psychoacoustic as well as a music-perception standpoint.

Deficits in frequency resolution, which often coexist with SNHL, may cause the components of a tone complex to interact over a wider frequency range than in normal-hearing listeners. Therefore, listeners with SNHL may perceive an increase in sensory dissonance between some components of a tone complex that would ordinarily sound consonant for normal-hearing listeners. This in turn may have the effect of changing the relative dissonance of musical intervals, or reducing the perceptual contrast between consonant and dissonant intervals. To the best of our knowledge, no previous studies have examined the perception of consonance and dissonance by people with SNHL.

The present study addressed two questions. First, do people with SNHL judge the sensory dissonance of musical intervals differently than do normal-hearing people? Second, are judgments of sensory dissonance by normal-hearing and hearing-impaired listeners consistent with a relationship between peripheral frequency selectivity and dissonance perception? To answer these questions, normal-hearing (NH) and hearing-impaired (HI) subjects were recruited to complete two tasks. In the first task, subjects judged the relative sensory dissonance of musical intervals created with pure tones and with harmonic complex tones. In the second task, subjects' thresholds for pure tones at 500 and 2000 Hz were measured in the presence of a notched noise. The latter task was designed to assess their peripheral frequency resolution.

II. METHOD

A. Subjects, test environment, and order of procedures

Eight subjects participated in the study. Four subjects (1 M, 3 F; mean age=50 years, range=31–63) had normal hearing in the test ear (i.e., air-conduction thresholds $\leqslant$20 dB HL from 0.25 to 4 kHz re: ANSI, 1996). Each of the other four subjects (2 M, 2 F; mean age=69 years, range=61–80) had a mild to moderate sensorineural hearing loss in the test ear (i.e., air-conduction thresholds between 30 and 60 dB HL at audiometric frequencies from 0.25 to 4 kHz, air-bone gaps of $\leqslant$10 dB from 0.5 to 4 kHz, and a normal tympanogram). Figure 1 shows the mean air conduction thresholds of the subjects. According to self-report, none of the subjects had had training in music theory, none had perfect pitch, and none was able to recognize musical intervals by ear. Thus, it was assumed that the subjects' performance was not influenced by specialized knowledge of musical intervals.

All testing took place in a double-walled sound-treated

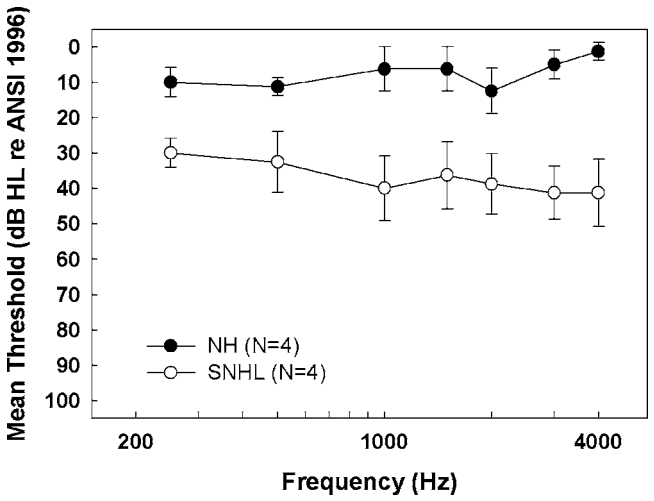


FIG. 1. Mean air-conduction thresholds (in dB HL re: ANSI 1996) for the subjects with normal hearing and the subjects with SNHL. Error bars represent standard deviations around the means.

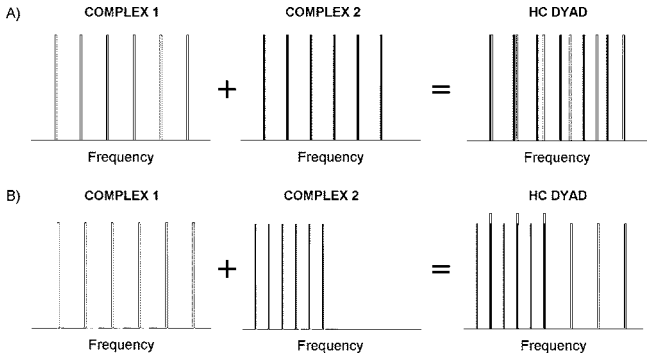


FIG. 2. Schematic example of the construction of two harmonic complex (HC) dyads. (A) HC dyad composed of two harmonic complex tones with a fundamental frequency ratio of 16:15 (a minor second). (B) HC dyad composed of two harmonic complex tones with a fundamental frequency ratio of 2:1 (an octave).

booth. Headphone output levels were calibrated with a 500 Hz pure tone each day before data collection. Prior to enrollment in the study, each subject signed an informed consent form and a privacy statement outlining the potential uses of his or her identifying information. The subject's audiogram and tympanogram were obtained upon enrollment. Next, the subject completed the sensory dissonance judgment task, followed by the threshold-in-noise task. All stimuli were presented to the same ear. Total participation time was approximately 8 h per subject, spread over four or five sessions.

B. Sensory dissonance judgments

Subjects judged the relative sensory dissonance of pairs of tones, called *dyads*, formed by combining either two pure tones or two harmonic complex tones. The pure-tone dyads fell within two frequency regions, one centered around 500 Hz and the other centered around 2000 Hz, while the F0's of the harmonic complex dyads were centered around 500 Hz. The intervals formed by the dyads included all of the equal-tempered musical intervals of the Western tradition.

1. Stimuli

All dyads were generated digitally and played through a 16 bit digital-to-analog converter (TDT DD1) at a rate of 40 000 samples/s. The stimuli were passed through an attenuator (TDT PA4) and a headphone buffer (TDT HB6) to one channel of a set of calibrated circumaural earphones (Sennheiser, HD540). *Pure-tone* dyads (PT dyads) were composed of two simultaneous pure tones with equal amplitudes and phases drawn randomly from a uniform distribution. *Harmonic complex* dyads (HC dyads) were composed of two simultaneous harmonic complex tones, each having six components (F0 and five harmonics). All twelve components were of equal amplitude and random phase. Each dyad was 750 ms in total duration, including 50 ms raised-cosine onset and offset ramps.

The overall presentation level of each dyad was approximately 83 dB SPL.[1] This level was chosen so that each dyad component would have a sensation level of at least 10 dB for each HI listener, without creating discomfort for either the NH or HI listeners. The levels of the individual components of the dyads were 80 and 72 dB SPL for the PT and HC dyads, respectively.

Two sets of 26 PT dyads were created. The dyads within a set were centered at a geometric mean frequency of either 500 Hz (PT 500 Hz dyads) or 2000 Hz (PT 2000 Hz dyads). In each set, the frequency separation between the dyad components was varied in quartertone steps (i.e., equal logarithmic steps of $2^{1/24}$), beginning at 0 Hz. With this step size, the thirteen equal-tempered musical intervals from the unison to the octave were represented in each set, as well as the thirteen interposed intervals ranging from the quartertone above the unison to the quartertone above the octave. One set of 26 HC dyads was also created (HC 500 Hz dyads). The F0's of the HC 500 Hz dyad components were identical to the frequencies of the PT 500 Hz dyad components. Figure 2 depicts a schematic example of the construction of two HC dyads whose F0's form a ratio of either (A) 16:15 (a minor second) or (B) 2:1 (an octave). Notice that the resulting dyads may have fewer than 12 components if the two added complexes have some partials in common.

Table I lists the PT dyads' musical interval names (where applicable), frequency ratios, and component frequencies. It should be noted that all of the equal-tempered musical intervals except the unison and the octave deviate slightly from simple integer ratios. For example, the equal-tempered fifth deviates from a ratio of 3:2 by two cents (1/600 of an octave). However, such small deviations do not reduce the acceptability of these intervals in musical practice (Vos, 1988).

2. Procedure

Subjects judged the relative sensory dissonance of the dyads within each set via a paired comparison task. On each trial, the subject heard two dyads from the same set (either the PT 500 Hz, PT 2000 Hz, or HC 500 Hz dyads), separated by a 500 ms silent interval, and chose the dyad that sounded more unpleasant. The term "unpleasant" was de-

TABLE I. The pure-tone dyads are shown in order of increasing interval width. Included are the musical interval names of the dyads (where applicable), their equal-tempered frequency ratios, and their component frequencies. The just integer ratios of the musical intervals are shown in parentheses. The fundamental frequencies of the components of the harmonic complex dyads are identical to the frequencies of the pure-tone dyads centered at 500 Hz.

Dyad No.	Musical interval name	Frequency ratio	Pure-tone frequencies in Hz for dyads centered at 500 Hz	Pure-tone frequencies in Hz for dyads centered at 2000 Hz
1	Unison	1.000 (1:1)	500–500	2000–2000
2		1.029	493–507	1971–2029
3	Minor second	1.059 (~16:15)	486–515	1943–2059
4		1.091	479–522	1915–2089
5	Major second	1.122 (~9:8)	472–530	1888–2119
6		1.155	465–537	1861–2150
7	Minor third	1.189 (~6:5)	459–545	1834–2181
8		1.224	452–553	1808–2213
9	Major third	1.260 (~5:4)	446–561	1782–2245
10		1.297	439–569	1756–2278
11	Perfect fourth	1.335 (~4:3)	433–578	1731–2311
12		1.374	427–586	1706–2344
13	Tritone	1.414 (~45:32)	420–595	1682–2378
14		1.456	414–603	1658–2413
15	Perfect fifth	1.498 (~3:2)	409–612	1634–2448
16		1.542	403–621	1611–2484
17	Minor sixth	1.587 (~8:5)	397–630	1587–2520
18		1.634	391–639	1565–2557
19	Major sixth	1.682 (~5:3)	386–648	1542–2594
20		1.731	380–658	1520–2631
21	Minor seventh	1.782 (~9:5)	375–667	1498–2670
22		1.834	369–677	1477–2709
23	Major seventh	1.888 (~15:8)	364–687	1458–2748
24		1.943	359–697	1435–2788
25	Octave	2.000 (2:1)	354–707	1414–2828
26		2.059	349–717	1394–2870

scribed as synonymous with "rough, dissonant, or discordant," as opposed to "pleasant, smooth, pure, or harmonious." Each dyad was paired twice with every other dyad in its set, once in each order, for a total of 650 trials. For each listener, the order of these 650 trials was randomized and then split to create two blocks of 325 trials each. This was done for each dyad set (PT 500 Hz, PT 2000 Hz, and HC 500 Hz dyads), giving a total of six blocks. The order of these six blocks was randomized for each listener.

A running score was kept for each dyad during testing. Initially, each dyad was assigned a score of zero. Each time a particular dyad was chosen as more unpleasant in a trial, 0.5 was subtracted from its score. For example, if a particular dyad was always chosen as more unpleasant in all of its pairings with the other dyads in its set, it would obtain a score of $-0.5*50=-25$, the lowest possible score. If a particular dyad was never chosen as more unpleasant in any of its pairings, it score would remain at zero, the highest possible score. At the completion of all six blocks, the subject had three sets of scores, each one representing the relative dissonance of the dyads within a stimulus set.

C. Estimation of peripheral frequency selectivity

Auditory filter bandwidths for each subject were estimated for signal frequencies of 500 and 2000 Hz using a notched-noise masking task (see Patterson and Moore, 1986, for a discussion of this method of estimating frequency resolution). This procedure involved obtaining several masked threshold estimates, with masking provided by two bands of noise, one above and one below the signal in frequency. As the frequency separation between the signal and the masking bands increased, the noise level required to mask the constant-level signal increased. The rate of increase of the masker level was used to estimate the shape of the auditory filter centered at the signal frequency.

1. Stimuli

The signals were tones of either 500 or 2000 Hz, selected to correspond to the geometric means of the dyads in the sensory dissonance judgment task. Signal duration was 300 ms, with 50 ms cosine-squared onsets and offsets. Notched-noise maskers consisted of two bands of noise positioned on either side of the signal frequency. The bandwidths of the noises were 0.4 times the signal frequency, with steep skirts (falling more than 75 dB per 100 Hz). The duration of the noise maskers was 400 ms, with the signal temporally centered in the noise. Eight notched-noise maskers were generated for each signal frequency, six with the notch centered on the signal frequency (i.e., symmetric

notches) and two with either the upper or lower noise band placed closer to the signal frequency (asymmetric notches).

Signals and notched-noise maskers were generated digitally, and played out at 40 000 samples/s through separate channels of a digital-to-analog converter (TDT DD1). The signal and noise channels were passed through separate programmable attenuators (TDT PA4), and added together (TDT SM3) before being passed through a headphone buffer (TDT HB6) to one channel of a set of TDH-49P earphones. Signal levels were fixed at 60 dB SPL for the NH listeners, and either 70 or 80 dB SPL for the HI listeners. Noise level was varied adaptively to determine the level that just masked the signal.

2. Procedure

Each masked threshold was measured using a modification of the single-interval yes-no maximum-likelihood adaptive procedure described by Green (1993). This procedure has been shown to produce reliable threshold estimates in 12 to 25 trials, depending on the number of catch trials included in the procedure (Gu and Green, 1994; Leek *et al.*, 2000). On each trial, the signal and noise were presented together, with the signal level fixed across trials and the notched-noise level determined by an adaptive track. Each block of trials began with the noise level low enough so that the signal was clearly heard. After each presentation, the subject indicated by a button press on a response box whether the signal was heard or not heard. A light on the box indicated that the response had been registered, but no other feedback was given. Catch trials, in which no signal was present, occurred on 20% of the presentations, and the responses to those trials were used to estimate a false alarm rate. After each presentation and response, a set of candidate psychometric functions was consulted. All previous responses in the block were used to determine which of the candidate functions was the most likely to represent the data collected up to that point. The selected function was then entered at 70% to determine the masker level for the next trial. Presentations continued until the confidence interval for the current threshold value was less than 1 dB. At the end of the threshold track, the level of noise necessary to produce 70% correct detections of the signal was extracted from the final estimated psychometric function. The average of three such measurements was taken as the threshold for that block of trials, and the average of two blocks for each notched-noise condition constituted the final threshold value. In cases where thresholds from the two blocks differed by more than 3 dB, an additional block was run and all threshold estimates were averaged.

While this adaptive procedure is not unbiased like a forced choice procedure, the implementation here has a number of characteristics that increase its reliability and validity. The starting level for each adaptive run was selected randomly over a range of values that produced a clearly audible signal; each track continued until a criterion variability measure was reached; and thresholds were estimated from a minimum of six adaptive thresholds (two blocks of three tracks each), with additional threshold runs if the two block estimates differed by more than 3 dB. The reliability of this procedure has been discussed extensively by Green and his colleagues and others (e.g., Gu and Green 1994; He *et al.*, 1998; Florentine *et al.*, 2000; Leek *et al.*, 2000).

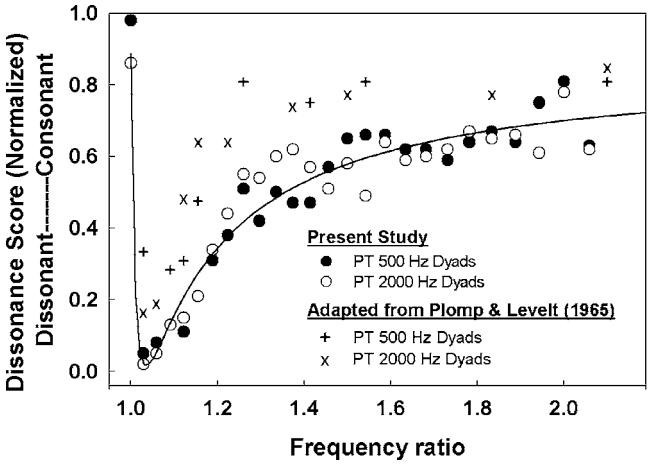


FIG. 3. Median normalized sensory dissonance scores of the NH group (N=4) for each of the two pure-tone dyad sets, plotted as a function of the frequency ratio of the dyad components. Closed circles represent the data for the pure-tone dyads centered at 500 Hz; open circles represent the data for the pure-tone dyads centered at 2000 Hz. The two sets of data are fitted by a single lognormal curve (see the text). Selected data from Plomp and Levelt (1965), translated to these axes, are shown for comparison. Crosses represent data for the pure-tone dyads centered at 500 Hz; Xs represent data for the pure-tone dyads centered at 2000 Hz.

III. RESULTS

A. Sensory dissonance judgments

Each subject produced three sets of dissonance scores, each one representing the relative sensory dissonance of the dyads within a stimulus set (PT 500 Hz, PT 2000 Hz, or HC 500 Hz dyads). Within each subject group (NH versus HI), the median dissonance score of each dyad, and its associated interquartile range, were calculated. Scores could range from 0, indicating maximum consonance, to −25, indicating maximum dissonance. For the NH subjects, the average interquartile ranges were 1.6 for the PT 500 Hz dyads, 4.7 for the PT 2000 Hz dyads, and 3.4 for the HC 500 Hz dyads. HI subjects had average interquartile ranges of 4.4, 2.3, and 3.8 for the corresponding dyad sets.

The median scores for each dyad were normalized between 0 and 1, with 1 representing maximal sensory consonance and 0 representing maximal sensory dissonance. The normalized scores of the PT 500 Hz and PT 2000 Hz dyads were plotted as a function of the frequency ratio of the dyad components; the normalized scores of the HC 500 Hz dyads were plotted as a function of the F0 ratio of the dyad components.

Figure 3 shows the scores of the PT 500 Hz and PT 2000 Hz dyads for the NH subjects. Plotted as a function of frequency ratio, the two sets of scores are nearly superimposed. The pattern of the data is similar to that observed in previous studies of the sensory dissonance of PT dyads. Specifically, as the frequency separation of the dyad components widens from 0 Hz, dissonance rapidly increases to a maximum, then decreases somewhat more gradually, eventually reaching a plateau. Selected points from Plomp and Levelt

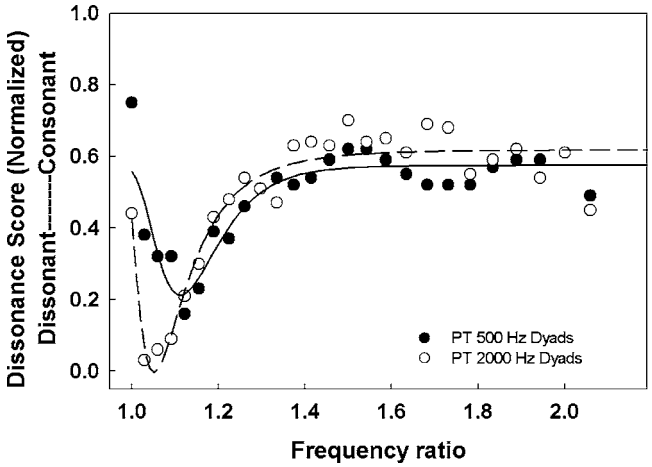


FIG. 4. Median normalized sensory dissonance scores of the HI group ($N=4$) for each of the two pure-tone dyad sets, plotted as a function of the frequency ratio of the dyad components. Closed circles represent the data for the pure-tone dyads centered at 500 Hz; open circles represent the data for the pure-tone dyads centered at 2000 Hz. Each of these two sets of data is fitted by a lognormal curve (see the text).

(1965) are included in Fig. 3 for comparison. These points represent approximate translations of Plomp and Levelt's (1965) data, which were obtained with a different methodology and different instrumentation than were used in the present study. Nevertheless, they show agreement with the current data in their overall pattern.

The two sets of dissonance scores of the NH group were initially fit by separate lognormal functions.[2] Comparison of these functions using Akaike's Information Criterion (AIC; Akaike, 1974) indicated that the two sets of scores could be represented by a single function fit to the combined data ($R^2=0.93$; corrected AIC values for one versus two functions, respectively: 58.25 and 65.93). The AIC compares maximum likelihood estimates of competing functions adjusted for the number of free parameters. The function associated with the smaller AIC value provides the better fit to the data. This function is shown as a solid line in Fig. 3. Additionally, lognormal functions were fit to the pure-tone dissonance scores of each individual NH subject, and are used in the analysis of the relationship of auditory filter bandwidth to sensory dissonance perception, described in Sec. III B.

Figure 4 shows the normalized scores of the PT 500 Hz and PT 2000 Hz dyads for the HI subjects. As is the case for the NH group, maximal dissonance occurs at a narrow frequency separation, followed by a decrease in dissonance to a plateau as the frequency separation increases. The two sets of dissonance scores of the HI group were fit by separate lognormal functions, shown as solid (PT 500 Hz dyads) and dashed (PT 2000 Hz dyads) lines in Fig. 4. Statistical evaluation using the AIC indicated that the two sets of scores were better fit by these two functions rather than one function for the combined data ($R^2=0.81$ and 0.91 for the PT 500 Hz and PT 2000 Hz curves, respectively; corrected AIC values for one versus two functions, respectively: 86.69 and 65.62). As was done for the NH subjects, lognormal functions were also fit to the pure-tone dissonance scores of each individual HI subject, with the exception of one subject's PT 500 Hz scores, which could not be adequately represented by a lognormal function.

Visual inspection of the pure-tone dissonance curves shown in Fig. 3 (NH) and Fig. 4 (HI) suggest several differences between the listener groups. First, the overall range of median dissonance scores for the HI listeners is compressed, relative to the range of scores for the NH listeners. While the normalized dissonance scores of the NH group span nearly the entire possible range, from 0.02 to 0.98, the normalized scores of the HI group span a smaller range from 0.03 to 0.75.

One aspect of the reduced range may be observed in the scores for the unison (the dyad with a 0 Hz frequency separation). The HI listeners did not judge the unison to be as consonant as did the NH listeners. For the NH group, the unison has median scores of 0.98 and 0.86, respectively, for the PT 500 Hz and PT 2000 Hz dyad sets, versus 0.75 and 0.44, respectively, for the HI group. A two-way analysis of variance with one repeated factor (frequency region) was carried out on the log-transformed individual unison scores across the two groups of listeners. The difference between the groups in the dissonance scores at the unison was statistically significant $[F(1,6)=7.20, p<0.04]$. In addition, the main effect of frequency region was statistically significant $[F(1,6)=15.89, p<0.01]$. The interaction between listener group and frequency region was not significant ($p>0.20$). These analyses indicate that, in general, the 500 Hz unison was more consonant than the 2000 Hz unison for both groups, but the unison at each frequency was perceived as less consonant by the HI listeners compared with the NH listeners.

Finally, maximal sensory dissonance (i.e., the minima of the dissonance curves) occurs at a larger frequency ratio on the PT 500 Hz curve of the HI listeners than on the other three pure-tone dissonance curves. Specifically, on the PT 500 Hz curve of the HI group, the point of maximal dissonance falls near the major second; on the PT 2000 Hz curve of the HI group, and both pure-tone dissonance curves of the NH group, the points of maximal dissonance occur at intervals smaller than the minor second. The frequency ratios at which maximal sensory dissonance occurred for each individual and frequency region were entered into a two-way analysis of variance, with frequency region as a repeated factor. The differences noted above were confirmed statistically: there was no main effect of group $[F(1,6)=3.008, p>0.10]$, but there was a significant main effect of frequency region $[F(1,6)=13.293, p=0.01]$. This main effect was primarily due to the significant interaction between group and frequency $[F(1,6)=7.78, p=0.03]$.

The closed symbols in panels (A) and (B) of Fig. 5 show the median scores of the HC 500 Hz dyads of the NH and HI listeners, respectively (open symbols will be discussed later). As the separation between the F0's of the dyad components widens from 0 Hz, the dissonance of the HC dyads rapidly increases to a maximum at an interval of one quartertone and then decreases somewhat more gradually for both groups of listeners. However, unlike the pure-tone sensory dissonance curves, the harmonic complex sensory dissonance curves ex-

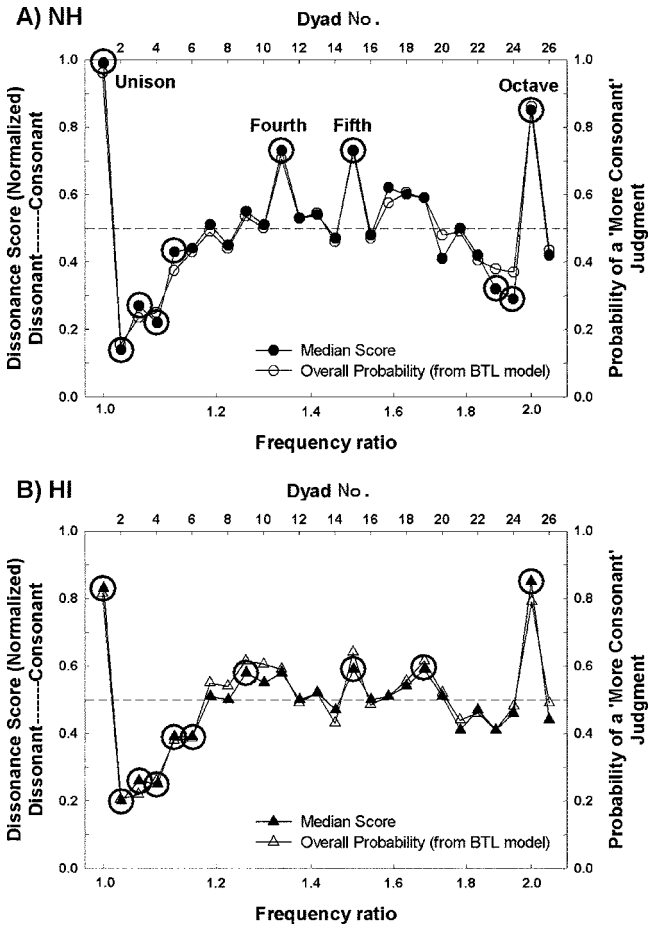


FIG. 5. Median normalized sensory dissonance scores of the NH (A) and HI (B) groups for the harmonic complex dyad set, shown as closed symbols and plotted as a function of the fundamental frequency ratio of the dyad components. Those scores that are circled are significantly more or less dissonant than predicted by chance ($p<0.05$), which is signified by the horizontal dashed line on each panel. [Note that the tests of difference from chance were conducted on pooled raw scores rather than the median normalized scores depicted here.] The open symbols indicate the overall probability of each dyad being judged more consonant than the other dyads in the set, based on the BTL model of paired comparisons. These data should be referred to the right axis.

hibit several sharp peaks associated with small-integer F0 ratios. This observation is consistent with previous research on the dissonance of HC dyads (e.g., Kameoka and Kuriyagawa, 1969b).

These data will be evaluated in two ways. First, one may ask which of these dyads were perceived as clearly dissonant or clearly consonant. This question will be addressed using a chi-squared analysis of differences of the scores from chance performance. Next, an analysis will be provided that takes into account the influence of all of the dyads in a set on individual judgments in determining the relative consonance or dissonance of the dyads.

Tests for paired comparison data (David, 1988) were used to determine which dyads were more consonant or dissonant than would be expected based on chance performance. The null hypothesis of no difference from chance was evaluated for each dyad using a normal approximation to the chi-square distribution. For each group of four listeners (with two replications each), the pooled scores were tested against a criterion value calculated for a significance level of 0.05, corrected for the 26 different significance tests to be performed, and divided by two for a two-tailed test of difference from chance (represented on the panels of Fig. 5 as a horizontal line at a score of 0.5). The dyads whose scores were significantly different from chance are circled in Fig. 5. These dyads were perceived as either highly dissonant or highly consonant, relative to the more ambiguous perceptions that are reflected by scores near chance.

As shown in Fig. 5, the NH group identified the intervals of the unison, octave, fifth, and fourth as highly consonant. Further, they judged the intervals equivalent to or smaller than a major second (dyads 2–5), and intervals near the octave (dyads 23 and 24), to be significantly dissonant. The HI group identified as highly consonant the unison, the octave, and the fifth, as well as the major third and major sixth, but not the fourth. HI listeners judged intervals smaller than a minor third (dyads 2–6) to be significantly dissonant, but did not judge dyads near the octave to be significantly dissonant.

Analyzing differences from chance performance does not directly take into account the influence of all the other dyads on individual comparisons between two dyads. For a comparison between a given pair of dyads, the probability of a particular outcome will be influenced by the other dyads in the stimulus set. A model of paired comparisons developed by Bradley and Terry (1952), further modified by Luce, and described in David (1988), was implemented to generate these probabilities. A similar analysis was conducted by Pressnitzer and McAdams (1999) to study the perception of roughness, as well as by Uppenkamp *et al.* (2001) to evaluate paired comparisons of "compactness" (an aspect of timbre).

The Bradley-Terry-Luce (BTL) model is usually described with reference to rankings of teams in a baseball league (see Agresti, 1990, for an intuitive description of this model of paired comparisons). The number of wins is not only dependent on a team's own ability, but also on the abilities of each of the other teams (see David, 1988). The BTL model takes these dependencies into account to establish the probability of a win for each team when playing every other team, as well as an overall probability of winning. This model was used here to establish the probabilities of the possible outcomes of all individual head-to-head comparisons between the dyads. These individual probabilities were combined so that an overall probability of "winning" (i.e., being judged more consonant in relation to all other dyads) was generated for each dyad. In effect, the analysis converted the individual judgments into a ranking of all of the dyads along the consonance/dissonance dimension. These overall probabilities were normalized to sum to 1.0. They are shown as the open symbols in Fig. 5, referred to the right-hand ordinate. These constructed probabilities, which take into account the scores of all of the dyads, closely follow the pattern of median scores.

For each dyad, a probability function was constructed that displays its probability of "winning" when paired with every other dyad. The 26 probability functions are shown in Fig. 6(A) (NH) and Fig. 6(B) (HI). Each curve on these panels represents the probability (shown along the ordinate)

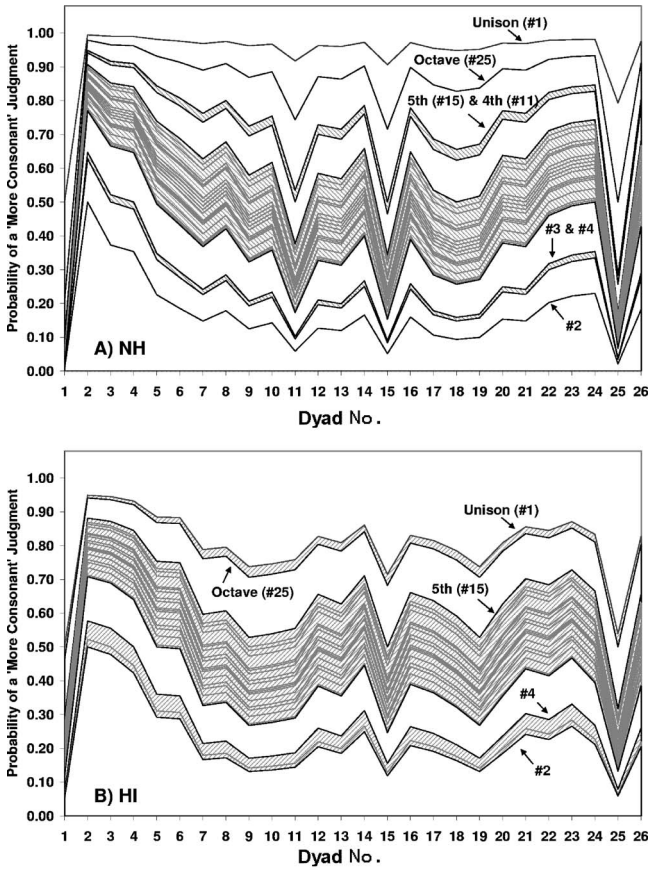


FIG. 6. Probability functions for each dyad, based on the BTL model of paired comparisons. Each curve shows the individual probabilities of that dyad being judged more consonant in relation to each of the other 25 dyads, shown on the abscissa. Curves whose average probabilities are less than one standard deviation different from neighboring curves are grouped together by shading (see the text for a more complete explanation). (A) and (B) Data from NH and HI listeners, respectively. There are six groupings of the probability functions in (A) and three groupings in (B) indicating that NH listeners more clearly separated the dyads in terms of their relative consonance/dissonance than did the HI listeners.

that the given dyad was perceived as more consonant when compared against every other dyad (shown along the abscissa), according to the BTL model. For example, in Fig. 6(A), the probability of the perfect fifth (dyad 15) being chosen as more consonant is 0.80 in its pairing with dyad 8, and 0.28 in its pairing with the octave (dyad 25).

The curves are arranged in Fig. 6 in order from high to low average probability, where the average probability for each dyad was computed across all of the values along its curve. It is apparent from Fig. 6 that the curves are not equally distributed along the probability axis. Instead, some appear to group together, reflecting those dyads that share similar consonance patterns. Groupings were defined on an ad hoc basis by taking the mean of all of the differences in average probabilities between adjacent curves, and grouping together those curves whose average probabilities did not differ from neighboring curves by more than one standard deviation of the mean. Groupings are demarcated in the figure by shading.

There are six distinct groupings for the NH listeners and three groupings for the HI listeners. For the NH listeners [Fig. 6(A)], the unison and the octave are not grouped with any other intervals or with each other. The fifth and fourth are grouped together. These dyads are very similar to one another in having lower average probabilities than the octave, but higher average probabilities than the rest of the dyads. The large shaded area in the middle of Fig. 6(A) groups together all but three of the remaining dyads. Two of these, the minor second and its upper neighboring dyad (dyads 3 and 4), are grouped together; the quartertone interval (dyad 2) has the lowest average probability and is separate from the rest, indicating that it is least consonant overall.

As shown in Fig. 6(A), the probability functions of the unison and the octave are fairly flat in comparison with the probability functions of the other dyads. The consistently high probabilities of the unison and octave indicate that these two dyads are very resistant to being chosen as more dissonant, no matter which dyad they are paired with. The only intervals that appear to affect the unison and octave, aside from each other, are the fifth and fourth. This is shown by the dips in the probability functions at the fifth and fourth (dyads 15 and 11, respectively, on the abscissa).

Compared with the NH group, the HI subjects produced fewer distinct groups of dyads. As shown in Fig. 6(B), the highly consonant unison and octave form a group, as do the highly dissonant minor second and its two neighbors (dyads 2–4). All of the 21 remaining dyads are grouped together. The probability functions of the unison and octave are not as flat as those of the NH group. Dips in the functions occur at the fifth (dyad No. 15) [and the octave (dyad No. 25) and unison (dyad No. 1), respectively], but not the fourth (dyad No. 11). The HI listeners apparently perceived the fifth as a relatively consonant interval, although neither it nor the fourth were grouped separately from the other dyads.

Overall, these results suggest that the HI listeners do not distinguish intervals in terms of consonance and dissonance as clearly as do the NH listeners. This is consistent with the more compressed puretone dissonance curves of the HI listeners, noted earlier.

B. Relationship of dissonance to auditory filter bandwidths

The thresholds generated by the notched-noise procedure for each individual were used to estimate auditory filter shapes. The auditory filters were derived using the polyfit procedure described by Rosen and Baker (1994). A least-squares fitting procedure was implemented to find the filter weighting function that best predicted the set of thresholds, given the assumptions of the power spectrum model of masking. The general weighting function has the form

$$W(g) = (1 - pg)e^{(-pg)}, \tag{1}$$

where g is a normalized frequency variable, representing the difference between the center frequency and a given point on the filter skirt, and p determines the passband and the slope of the filter skirt (Patterson and Moore, 1986). The skirt parameter, p, was fit separately on either side of the filter. Equivalent rectangular bandwidths (ERBs) of the individual auditory filters, given as a proportion of the center frequency, were computed as $[2/p(\text{lower skirt})]$

+[2/p(upper skirt)], as described by Glasberg and Moore (1990).

The mean relative ERBs of the auditory filters at 500 Hz were 0.20 and 0.26 for the NH and HI listeners, respectively. At the center frequency of 2000 Hz, the mean relative ERBs were 0.21 and 0.29 for the two groups, respectively. A t-test at each center frequency indicated that the difference in ERB between groups was statistically significant at 500 Hz ($p=0.002$), but not at 2000 Hz ($p>0.3$), where there was considerable variability among the HI listeners. Large variability in estimates of auditory filter bandwidths of HI listeners has been noted frequently before (e.g., Glasberg and Moore, 1986; Peters and Moore, 1992). Other factors that could account for the similar auditory filter bandwidths for NH and HI listeners at 2000 Hz are the high levels at which the measurements were made, and the relatively mild hearing losses of most of the HI subjects. Auditory filters of NH listeners measured at high stimulus levels are typically broader than those measured at lower levels (Leek and Summers, 1993). Given listeners with normal hearing and a signal level of 60 dB SPL, Rosen and Baker (1994) calculated an ERB of 0.22 at 2000 Hz, nearly identical to the mean ERB measured here. For HI listeners, Baker and Rosen (2002) reported an average ERB at 2000 Hz of about 0.25, measured at similar signal levels to those used here (the value estimated from their Fig. 5).

The significant difference in auditory bandwidths between the two groups of listeners at 500 Hz is noteworthy, given the differences observed in the dissonance curves of the PT 500 Hz dyads. Recall that previous investigators (Plomp and Levelt, 1965; Greenwood, 1991) estimated that maximum sensory dissonance occurs at frequency separations of approximately 25%–40% of the critical bandwidth, and that two tones separated by approximately one critical bandwidth are consonant. Figure 7 illustrates the strength of these two relationships in the data reported here.

Two points were extracted from each of the individual subjects' pure-tone dissonance curves for comparison with auditory filter bandwidth values. The curves fitted to the dissonance data have an exponential shape as the frequency difference between dyad components becomes large. Representing the exponential growth of the approach to more consonant responses with increasing frequency separation, a "growth constant" was extracted from each individual dissonance curve. The growth constant was defined as the abscissa value at which the curve approached an asymptotic value on the ordinate, i.e., at approximately two-thirds of the difference between the minimum and maximum ordinate values. The growth constant was used here as an estimate of the point beyond which increases in frequency separation did not result in appreciable increases in consonance. This point may be thought of as analogous to the "shoulder" in Plomp and Levelt's (1965) data. Figure 7(A) plots these growth constants as a function of the ERB in Hz for each subject at the two center frequencies. (Note that only three hearing impaired data points are shown for 500 Hz; this is because one subject's data could not be adequately fit by the lognormal function, and therefore the growth constant could not be adequately estimated for that condition and subject.) If the return to consonance occurs at approximately one critical bandwidth, and if the growth constant and ERB are reasonable approximations of these two phenomena, then the data points should fall on the main diagonal, labeled 100% ERB, in Fig. 7(A). For the NH listeners, the growth constant occurred at an average of 149% of the ERB at 500 Hz, and 100% of the ERB at 2000 Hz. For the HI listeners, the percentages were 93% and 75%, respectively. Thus, to a first approximation, the relationship between critical bandwidth and the return to consonance is observed here.

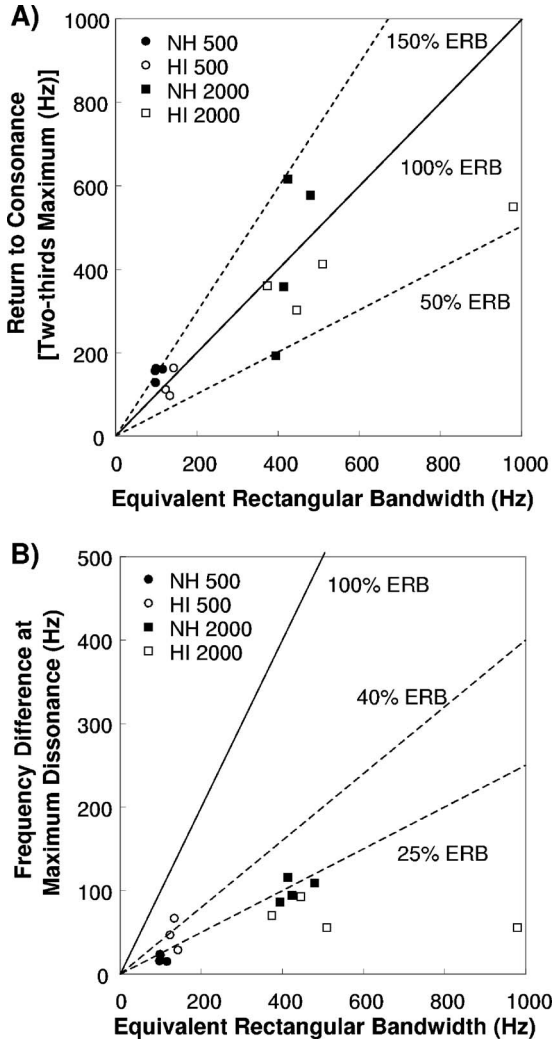


FIG. 7. (A) A measure of the return to consonance of the pure-tone dissonance curves as a function of the ERB of the auditory filters for individual subjects. NH data are shown with closed symbols; HI data are shown with open symbols. The solid line is included for reference, and represents frequency separations corresponding to 100% of the ERB. (B) Frequency separation at maximum sensory dissonance as a function of ERB for each individual subject. The solid and dashed lines are included for reference, and represent frequency separations corresponding to 100%, 40%, and 25% of the ERB, respectively.

The other noted relationship between auditory filter bandwidth and sensory dissonance perception is that maximum dissonance will occur when the two components of a dyad are separated by 25%–40% of a critical band. For the current data, this relationship was evaluated by first finding the minimum of each of the individual fitted curves for the PT 500 Hz and PT 2000 Hz dyad sets. Next, as shown in

Fig. 7(B), the frequency separation corresponding to the minimum of each curve was plotted as a function of the individual subject's measured ERB at the appropriate center frequency. The solid line on the panel, with a slope of one, represents the hypothetical case that maximal dissonance occurs exactly at the ERB value. The dashed lines on the panel indicate 25% and 40% of the measured ERBs. As expected, all the data fall below the 100% ERB line, indicating that maximum sensory dissonance is perceived at frequency separations that fall well within a single auditory filter for both NH and HI listeners. Most data are near or below the 25% ERB line. For the NH subjects, maximum dissonance occurred at an average of 19% of the ERB at 500 Hz, and 24% of the ERB at 2000 Hz. For the HI subjects, maximum dissonance occurred at an average of 36% of the ERB at 500 Hz, and 14% of the ERB at 2000 Hz. Interestingly, for NH listeners, the point of maximal dissonance is a more nearly constant proportion of the ERB across frequency regions than it is for the HI listeners.

IV. DISCUSSION

A. Sensory dissonance of pure-tone dyads

The pure-tone dissonance curves measured in this study were similar to one another in their general characteristics: a relatively consonant score for the unison (with the exception of the PT 2000 Hz curve of the HI group), followed by a dip to maximum dissonance and a subsequent return to consonance as the frequency difference between dyad components increased. Furthermore, the predicted relationships between critical bandwidth on the one hand, and maximum dissonance and the return to consonance on the other hand, were roughly upheld by the data (although the small number of subjects and the subjective nature of the dissonance judgment task probably contributed to the imprecise nature of the relationship seen here).

Some important differences were evident in the pure-tone dissonance curves of the NH and HI groups. First, the HI group did not judge the unison to be as consonant as did the NH group. In theory, the unison has zero dissonance, since it comprises two pure tones of identical frequency. Therefore, it was expected that the unison would be judged to be the most consonant interval in each stimulus set. In fact, this is what occurred for the NH listeners; however, the results of the HI listeners did not show the expected pattern. Although the HI listeners judged the unison to be the most consonant interval in the PT 500 Hz stimulus set, its score was lower than that given to the unison by the NH group. Even less expected was the finding that HI listeners judged the unison in the PT 2000 Hz stimulus set to be the seventh most *dissonant* interval in that set. It is possible that the 2000 Hz unison sounded sharper, shriller, or less "pure" for the HI subjects than for the NH listeners. Moore (2001) reported preliminary data in which subjects with hearing loss were asked to rate pure tones for their "distinctness" or "noisiness." The subjects produced scores indicating that tones in regions of hearing loss, particularly in the higher frequencies, were not as distinct as tones in regions of more normal hearing. This preliminary report provides some support for the notion that the 2000 Hz unison may have sounded slightly noisy or distorted to the HI listeners, which may account for its unexpectedly low consonance. Other evidence of distorted pitch perception in the presence of hearing loss has been reported by Larkin (1983) and by Leek and Summers (2001).

Maximal sensory dissonance fell at a larger interval on the PT 500 Hz dissonance curve of the HI group than it did on their PT 2000 Hz curve and on the two pure-tone dissonance curves of the NH listeners. Specifically, the major second (dyad 5) was maximally dissonant for the PT 500 Hz stimuli of the HI group, whereas the quartertone (dyad 2) was maximally dissonant for the PT 2000 Hz stimuli of the HI group and for both stimulus sets of the NH group (see Figs. 3 and 4). This finding is interesting in light of the two mechanisms of roughness perception proposed by Zwicker and Fastl (1990). [Recall that roughness and dissonance, while not synonymous, are closely related (Terhardt, 1974)]. Zwicker and Fastl (1990) provided data showing that maximal roughness of amplitude-modulated (AM) pure tones is frequency-dependent below about 1000 Hz. They argued, therefore, that at low frequencies, frequency selectivity is the limiting factor in roughness perception. Zwicker and Fastl (1990) also showed that for tones above about 1000 Hz, maximal roughness occurred at AM rates of approximately 75 Hz, independent of stimulus frequency. This finding suggests that, for high-frequency stimuli, roughness is not linked so strongly to critical bandwidth, but is instead limited by the ability of the ear to follow fast amplitude modulations. Given these data, it is reasonable to propose that any effects of broader auditory filters in the present study would be seen more clearly in the 500 Hz data than in the 2000 Hz data. Indeed, the significant broadening of the auditory filters at 500 Hz for the HI listeners relative to the filters of the NH listeners may be related to the shift of the point of maximal dissonance for the 500 Hz dyad set of the HI group.

B. Sensory dissonance of harmonic complex dyads

Consistent with previous research on the dissonance of HC dyads, the harmonic complex dissonance curves exhibit several sharp peaks associated with small-integer F0 ratios. For the NH group, peaks occur at the "perfect" consonances, i.e., the unison, octave, fifth, and fourth. These intervals, having coinciding or very nearly coinciding partials, sound more consonant than the immediately neighboring intervals. Another distinguishing characteristic of a perfect interval is its tendency to be perceived as highly fused. Terhardt (1974, 1984) explained the affinity of tones forming an octave, fifth, or fourth on the basis of our repeated exposure to these intervals in the spectra of harmonic complex sounds, including, especially, voiced speech. The fact that musically naïve listeners judged these intervals to be the most consonant of all the intervals in the set indicates that their special status in music is due to fundamental perceptual qualities of these intervals and is not merely a learned convention of music theory.

The harmonic complex dissonance curve of the NH group showed regions of marked sensory dissonance near the

unison, the octave, and the fifth. The existence of these regions of dissonance underscores the importance placed on preserving octaves and fifths in various tunings and temperaments (e.g., Pythagorean tuning and equal temperament, among others).

Like the NH listeners, HI listeners judged the unison and octave to be very consonant. However, the peaks at the fifth, and especially at the fourth, were not as robust. In addition, the harmonic complex dissonance curve of the HI group had a region of marked dissonance near the unison only. Together with the analysis presented in Fig. 6, these findings provide evidence that HI listeners do not distinguish the relative sensory dissonance and consonance of intervals as clearly as do NH listeners. This loss of contrast among the intervals suggests that HI listeners would not fully experience the variations in musical tension supplied by dissonant and consonant intervals.

The loss of contrast may be related to a reduction in pitch strength, which often accompanies SNHL (Leek and Summers, 2001). A reduction in pitch strength may lessen the degree of fusion of highly consonant intervals to the point that they do not contrast clearly with the more dissonant neighboring intervals. The loss of contrast may also be related to poorer frequency selectivity. If the auditory filter bandwidths of the HI listeners were generally somewhat broader across the frequency range in which the HC dyads' components fell (approximately 350–4300 Hz), as is suggested in the current data especially for the lower frequency regions, then more extensive interactions may have occurred among the components, thereby blurring distinctions in dissonance among the intervals.

It is likely that sensitivity to amplitude modulation (AM) per se does not account for the differences in the NH and HI groups seen here. Bacon and Gleitman (1992) reported that the ability of their HI subjects with relatively flat SNHL to detect AM did not differ from NH subjects when audibility was accounted for. Further, the stimuli in the present study were maximally amplitude-modulated, and presumably this modulation was detectable by both groups of listeners.

C. Effects of level

An additional factor that may affect sensory dissonance perception is the level above threshold at which the dyad components are presented. If lowering the sensation level (SL) of a dyad lessens its perceived dissonance, then this might explain the loss of contrast seen in the harmonic complex dissonance curves of the HI group. To investigate this possibility, two of the four NH listeners repeated the sensory dissonance judgment task at lower sensation levels for both sets of PT dyads and the set of HC dyads. All dyads were presented at 53 dB SPL, 30 dB lower than the original presentation level of 83 dB SPL. The levels of the individual components of the dyads were 50 and 42 dB SPL for the PT and HC dyads, respectively. At these SPLs, the sensation levels of the stimuli for these NH listeners were approximately equal to the original sensation levels for the HI listeners, i.e., approximately 20–50 dB SL for the PT dyad components and approximately 10–40 dB SL for the HC dyad components. The resulting median pure-tone and harmonic complex sensory dissonance curves were very similar in shape and range of scores to those obtained by the NH group at the higher intensity level. This finding suggests that the loss of contrast in the HI group resulted from characteristics of the hearing impairment and not the lower sensation level of the stimuli.

D. Relationship to music perception and training

The data reported here were obtained in a laboratory setting using isolated, artificial stimuli lacking a broader musical context. Under such conditions, judgments of dissonance are made largely on the basis of perceived roughness (Terhardt, 1974). In evaluating the quality of music, however, the unpleasantness due to roughness may be mitigated by other factors. For example, typical musical sounds include an attack portion in which the amplitudes of the partials change rapidly over time. Often, the spectra are highly complex, as when several harmonic complex tones are sounded simultaneously in a chord. Certain pitches may be produced by more than one voice, giving a chorus effect, or sounds may be frequency- or amplitude-modulated, as in vibrato. The effects of these factors on dissonance perception were not assessed in this study. Each, however, would likely reduce the contribution of sensory dissonance to the overall quality of music listening for a NH person (Terhardt, 1978). It is not known how these factors would impact dissonance perception by people with SNHL.

None of the subjects was trained in the sensory dissonance judgment task prior to data collection. The need for practice was judged to be minimal for several reasons. The task was subjective, with no "correct" answer, and each dyad was heard 50 times throughout the experimental sessions. The subjects' judgments were reasonably consistent, producing interpretable patterns of the data. In addition, the two listeners who participated in the lower-level repetition of the dissonance task produced data nearly identical to the first data set.

Trained musicians may base judgments of dissonance primarily on their knowledge of musical intervals, rather than on purely sensory qualities (Plomp and Levelt, 1965). The present study was designed to investigate differences in the perception of sensory dissonance between NH and HI listeners. Musical training may have obscured these differences. Therefore, subjects were excluded from participation if they reported such training. However, all of the subjects had been exposed to Western music over their lifetimes. It is not known whether and how this informal exposure may have influenced their judgments.

With regard to music listening through hearing aids, a future goal of advanced signal-processing algorithms may be the restoration of the normal contrast between consonance and dissonance for HI listeners. Tramo *et al.* (2001) showed that consonant and dissonant intervals produce very distinctive patterns of activity in the auditory nerve. If such patterns of neural activity are dependent upon a normal or near-normal representation of the signal at the level of the cochlea, then the signal must be altered externally to compen-

sate for the effects of the hearing impairment on the internal representation. One possibility may lie in manipulating the phase spectra of musical signals. This approach may be justified in light of evidence that the phase characteristic of the basilar membrane is altered in sensorineural hearing impairment (e.g., Lentz and Leek, 1999; Oxenham and Dau, 2004).

V. CONCLUSIONS

Judgments of the sensory dissonance of PT and HC dyads by the HI listeners were consistent in some respects with those of the NH listeners in this and previous studies. However, several differences were observed. HI listeners did not judge the unison to be as consonant relative to other dyads as the NH listeners did. For the HC dyads, NH listeners judged the musically significant intervals of the unison, octave, fifth, and fourth to be very consonant; HI listeners also judged the unison, octave, and the fifth to be very consonant, but did not clearly judge the fourth to be consonant relative to neighboring intervals. NH listeners showed regions of marked dissonance near the unison, octave, and fifth; HI listeners had a region of marked dissonance near the unison only. These findings suggest that the HI listeners did not distinguish the relative sensory dissonance of intervals as clearly as the NH listeners did. By extension, they may not fully experience the variations in musical tension supplied by dissonant and consonant intervals. The loss of contrast may have resulted from distortions in the representation of pitch in the impaired auditory system (e.g., a reduction in pitch strength), or from more extensive interactions among the components of the dyads. Judgments of sensory dissonance by NH and HI listeners were roughly consistent with a relationship between peripheral frequency selectivity and dissonance perception. A future goal of advanced signal-processing algorithms might be the restoration of the normal contrast between consonant and dissonant intervals for HI listeners, perhaps by altering the phase spectra of musical signals.

ACKNOWLEDGMENTS

This research was supported by a grant from NIH-NIDCD (No. DC 00626). It was approved by the Clinical Investigation Committee and the Human Use Committee, Department of Clinical Investigation, Walter Reed Army Medical Center, under Work Unit No. 03-25012. All subjects participating in this research provided written informed consent prior to beginning the study. The authors would like to thank Robert Lutfi and two anonymous reviewers for their comments on an earlier version of this article, as well as the staff of the Research Section of the Army Audiology and Speech Center at Walter Reed Army Medical Center for their helpful suggestions and discussions. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

[1]HC dyads forming a unison or an octave had overlapping components at six and three frequencies, respectively. The two components of the PT dyad forming a unison were of identical frequency. Since the phases of the dyad components were randomly selected, it is possible that partial or complete cancellation of the overlapping frequency components could have occurred in some cases. The amplitudes of all of the dyads were normalized prior to D/A conversion, however, so that the randomly chosen phases did not result in changes in the levels of the dyads as presented to the subjects.

[2]In order to facilitate curve fits, the normalized scores were expressed as a function of the frequency separation between dyad components divided by their geometric mean frequency (either 500 or 2000 Hz), or $(f_2 - f_1)/\sqrt{(f_1 * f_2)}$, where f_1 and f_2 were the frequencies of the two pure-tone components, $f_1 < f_2$. Lognormal functions were then fit to the dissonance scores. The abscissa values were converted back to frequency ratios to allow for easier interpretation of the data with regard to musical intervals.

Agresti, A. (**1990**). *Categorical Data Analysis* (Wiley, New York).
Akaike, H. (**1974**). "A new look at the statistical model identification," IEEE Trans. Autom. Control **19**, 716–723.
American National Standards Institute (ANSI) (1996). "American National Standard: Specifications for audiometers," ANSI S3.6-1996.
Arehart, K. H., and Burns, E. M. (**1999**). "A comparison of monotic and dichotic complex-tone pitch perception in listeners with hearing loss," J. Acoust. Soc. Am. **106**, 993–997.
Bacon, S. P., and Gleitman, R. M. (**1992**). "Modulation detection in subjects with relatively flat hearing losses," J. Speech Hear. Res. **35**, 642–653.
Baker, R. J., and Rosen, S. (**2002**). "Auditory filter nonlinearity in mild/moderate hearing impairment," J. Acoust. Soc. Am. **111**, 1330–1339.
Bradley, R. A., and Terry, M. E. (**1952**). "The rank analysis of incomplete block designs. I. The method of paired comparisons," Biometrika **39**, 324–345.
Burns, E. M., and Turner, C. (**1986**). "Pure-tone pitch anomalies. II. Pitch-intensity effects and diplacusis in impaired ears," J. Acoust. Soc. Am. **79**, 1530–1540.
Chasin, M. (**2003**). "Music and hearing aids," Hearing J. **56**, 36–41.
David, H. A. (**1988**). *The Method of Paired Comparisons*, 2nd ed. (Griffin, London).
deLaat, J. A. P. M., and Plomp, R. (**1985**). "The effect of competing melodies on melody recognition by hearing-impaired and normal-hearing listeners," J. Acoust. Soc. Am. **78**, 1574–1577.
Florentine, M., Buus, S., and Geng, W. (**2000**). "Toward a clinical procedure for narrowband gap detection. I. A psychophysical procedure," Audiology **39**, 161–167.
Gfeller, K., Christ, A., Knutson, J. F., Witt, S., Murray, K. T., and Tyler, R. (**2000**). "Musical backgrounds, listening habits, and aesthetic enjoyment of adult cochlear implant recipients," J. Am. Acad. Audiol **11**, 390–406.
Glasberg, B. R., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.
Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.
Green, D. M. (**1993**). "A maximum-likelihood method for estimating thresholds in a yes-no task," J. Acoust. Soc. Am. **93**, 2096–2105.
Greenwood, D. D. (**1991**). "Critical bandwidth and consonance in relation to cochlear frequency-position coordinates," Hear. Res. **54**, 164–208.
Gu, X., and Green, D. M. (**1994**). "Further studies of a maximum-likelihood yes-no procedure," J. Acoust. Soc. Am. **96**, 93–101.
He, N., Dubno, J. R., and Mills, J. H. (**1998**). "Frequency and intensity discrimination measured in a maximum-likelihood procedure from young and aged normal-hearing subjects," J. Acoust. Soc. Am. **103**, 553–565.
Huron, D. (**2001**). "Tone and voice: A derivation of the rules of voice-leading from perceptual principles," Music Percept. **19**, 1–64.
Hutchinson, W., and Knopoff, L. (**1978**). "The acoustic component of Western consonance," Interface (USA) 7, 1–29.
Kameoka, A., and Kuriyagawa, M. (**1969a**). "Consonance theory. I. Consonance of dyads," J. Acoust. Soc. Am. **45**, 1451–1459.
Kameoka, A., and Kuriyagawa, M. (**1969b**). "Consonance theory. II. Consonance of complex tones and its calculation method," J. Acoust. Soc. Am. **45**, 1460–1469.
Larkin, W. D. (**1983**). "Pitch vulnerability in sensorineural hearing impairment," Audiology **22**, 480–493.
Leek, M. R., Dubno, J. R., He, N., and Ahlstrom, J. B. (**2000**). "Experience with a yes-no single-interval maximum-likelihood procedure," J. Acoust. Soc. Am. **107**, 2674–2684.
Leek, M. R., and Summers, V. (**1993**). "Auditory filter shapes of normal-hearing and hearing-impaired listeners in continuous broadband noise," J. Acoust. Soc. Am. **94**, 3127–3137.

Leek, M. R., and Summers, V. (**2001**). "Pitch strength and pitch dominance of iterated rippled noise in hearing-impaired listeners," J. Acoust. Soc. Am. **109**, 2944–2954.
Lentz, J. J., and Leek, M. R. (**1999**). "Masking by harmonic complexes with different phase spectra in hearing-impaired listeners," J. Acoust. Soc. Am. **106**, 2146.
Moore, B. C. J. (**2001**). "Dead regions in the cochlea: Diagnosis, perceptual consequences, and implications for the fitting of hearing aids," Trends in Amplification **5**, 1–34.
Oxenham, A. J., and Dau, T. (**2004**). "Masker phase effects in normal-hearing and hearing-impaired listeners: Evidence for peripheral compression at low signal frequencies," J. Acoust. Soc. Am. **116**, 2248–2257.
Patterson, R. D., and Moore, B. C. J. (**1986**). "Auditory filters and excitation patterns as representations of frequency resolution," in *Frequency Selectivity in Hearing*, edited by B. C. J. Moore (Academic, London), pp. 123–127.
Peters, R. W., and Moore, B. C. J. (**1992**). "Auditory filter shapes at low center frequencies in young and elderly hearing-impaired subjects," J. Acoust. Soc. Am. **91**, 256–266.
Plomp, R., and Levelt, W. J. M. (**1965**). "Tonal consonance and critical bandwidth," J. Acoust. Soc. Am. **38**, 548–560.
Plomp, R., and Steeneken, H. J. M. (**1968**). "Interference between two simple tones," J. Acoust. Soc. Am. **43**, 883–884.
Pressnitzer, D., and McAdams, S. (**1999**). "Two phase effects in roughness perception," J. Acoust. Soc. Am. **105**, 2773–2782.
Rosen, S., and Baker, R. J. (**1994**). "Characterizing auditory filter nonlinearity," Hear. Res. **73**, 231–243.
Terhardt, E. (**1974**). "Pitch, consonance, and harmony," J. Acoust. Soc. Am. **55**, 1061–1069.
Terhardt, E. (**1978**). "Psychoacoustic evaluation of musical sounds," Percept. Psychophys. **23**, 483–492.
Terhardt, E. (**1984**). "The concept of musical consonance: A link between music and psychoacoustics," Music Percept. **1**, 276–295.
Tramo, M. J., Cariani, P. A., Delgutte, B., and Braida, L. D. (**2001**). "Neurobiological foundations for the theory of harmony in Western tonal music," Ann. N.Y. Acad. Sci. **930**, 92–116.
Uppenkamp, S., Fobel, S., and Patterson, R. D. (**2001**). "The effects of temporal asymmetry on the detection and perception of short chirps," Hear. Res. **158**, 71–83.
von Helmholtz, H. (**1877/1954**). *On the Sensations of Tone as a Physiological Basis for the Theory of Music* (Dover, New York).
Vos, J. (**1988**). "Subjective acceptability of various regular twelve-tone tuning systems in two-part musical fragments," J. Acoust. Soc. Am. **83**, 2383–2392.
Zwicker, E., and Fastl, H. (**1990**). *Psychoacoustics: Facts and Models* (Springer, New York).

On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality

Francis Rumsey, Sławomir Zieliński, and Rafael Kassier
Institute of Sound Recording, University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom

Søren Bech
Bang & Olufsen, Peter Bangsvej 15, DK-7600 Struer, Denmark

(Received 13 September 2004; revised 14 February 2005; accepted 6 May 2005)

Mean opinion score ratings of reproduced sound quality typically pool all contributing perceptual factors into a single rating of basic audio quality. In order to improve understanding of the trade-offs between selected sound quality degradations that might arise in systems for the delivery of high quality multichannel audio, it was necessary to evaluate the influence of timbral and spatial fidelity changes on basic audio quality grades. The relationship between listener ratings of degraded multichannel audio quality on one timbral and two spatial fidelity scales was exploited to predict basic audio quality ratings of the same material using a regression model. It was found that timbral fidelity ratings dominated but that spatial fidelity predicted a substantial proportion of the basic audio quality. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1945368]

PACS number(s): 43.66.Lj, 43.38.Md, 43.38.Vk [NX] Pages: 968–976

I. INTRODUCTION

The work reported in this paper concerns the prediction of basic audio quality grades from listener ratings of spatial and timbral fidelity, for the purpose of evaluating degraded multichannel (surround sound) audio signals. Mean opinion score ratings of reproduced sound quality typically pool all contributing factors into a single basic audio quality score. In order to improve understanding of the trade-offs between specific sound quality factors that might be made in systems for the delivery of high-quality multichannel audio, it is instructive to investigate their relative perceptual importance. An attempt is made here to develop a regression equation that enables the prediction of the global attribute "basic audio quality" from ratings made on two spatial scales and one timbral fidelity scale. This is based on data gathered from experienced listeners during a series of surround sound quality evaluation experiments.

A number of examples exist in the reproduced sound quality literature of attempts to evaluate the validity of different quality attributes for describing overall sound quality (itself judged on various scales). These mainly studied two-channel stereophonic reproduction but some recent examples studied multichannel surround sound. Some experiments evaluated aspects of spatial quality in addition to timbral and other factors but analyses of relationships to overall sound quality have usually gone no further than observations about correlation between ratings. There is a notable absence of attempts to quantify the relative importance of spatial and timbral quality ratings.

Toole (1985), for example, found that loudspeaker fidelity ratings and spatial quality ratings were quite highly correlated ($r=0.7$) but did not quantify the relative contributions of the different quality factors he tested (many of which were highly correlated) to the overall fidelity ratings. Gabrielsson and Lyndstrom (1985) also evaluated a number of attributes in terms of their validity for describing perceived sound quality (PSQ). They found both spatial and timbral attributes to be valid and moderately highly correlated with PSQ, but did not attempt to quantify their relative importance. Huopaniemi *et al.* (1998) and Zacharov and Huopaniemi (1999) evaluated spatial and timbral qualities of binaural filters and virtual home theater systems but did not map these to basic audio quality ratings, finding the ratings on the two scales to be quite highly correlated. As these tests had used untrained listeners, and the two attributes were rated in the same session, they supposed that listeners might have been unable to distinguish sufficiently between the attributes and could have been providing a response close to a mean opinion score on both scales. Zacharov and Kuovuniemi (2001), however, found the correlation between the attribute "tone color" and most spatial attributes to be low (<0.1). This experiment used a range of different spatial microphone techniques and a wide range of different naturalistic source material. In this case trained listeners had been used to grade the various attributes, suggesting that, when trained and familiar with the stimuli, listeners were able to distinguish between spatial and timbral qualities. They also mapped direct attribute ratings to components of subjective preference, but did not conduct mappings to basic audio quality (although the two may be related as discussed below).

In Zielinski *et al.* (2005) we reported some observations about the relationships between basic audio quality and spatial and timbral fidelity, based on the data used to derive the regression models reported here. These showed that timbral fidelity and the two spatial fidelity scales were correlated at a relatively low level (0.33 for frontal spatial fidelity and 0.26 for surround spatial fidelity). This was argued to be due to characteristics of the stimulus degradations used. We observed that basic audio quality seemed to be more strongly influenced by timbral fidelity than by spatial fidelity but that

TABLE I. Algorithms used for bandwidth limitation.

Label	Cutoff frequency in individual channels in kHz: Front left and right	Center	Left and right surround	Comments
12 kHz	12.0	12.0	12.0	Used for all items
8 kHz	8.0	8.0	8.0	Used for all items
3.5 kHz	3.5	3.5	3.5	Used for all items
A	20.0	10.0	5.0	Used only for *F-B* program material
B	20.0	13.0	3.5	Used only for *F-B* program material
C	18.25	3.5	10	Used only for *F-F* program material
D	14.125	3.5	14.125	Used only for *F-F* program material
E	13.0	7.0	3.5	Used only for *F-B* program material
F	10.0	13.0	3.5	Used only for *F-B* program material
G	11.25	3.5	7.0	Used only for *F-F* program material
H	9.125	3.5	9.125	Used only for *F-F* program material

advanced statistical regression models would have to be used to quantify the relationship owing to the multicollinearities observed.

The experiments reported here were undertaken using the 3-2 stereo (5.1 surround) multichannel replay configuration (ITU-R, 1993) and were conducted in an ITU-R BS.1116 conformant listening room (ITU-R, 1994) at the University of Surrey. Sound quality was intentionally degraded in a controlled manner using bandwidth limitations and "down-mixing" algorithms. This was because the project from which these results are derived aimed to study the trade-offs between different types of quality degradation on surround sound reproduction. The experimental setup and design details, as well as detailed descriptions of the stimuli, may be found in Zielinski *et al.* (2003a, b, 2005) and will be summarized here. The main emphasis in the current paper is on a novel statistical analysis of already published results.

II. SUMMARY OF EXPERIMENT USED TO DERIVE THE DATA USED IN REGRESSION MODELING

A. Audio stimuli

Twelve multichannel audio excerpts were selected. They represented the following genres: classical music, pop music, movie, sport, TV show, and ambient sounds (applause and rain). Since some of the audio degradation algorithms involved low-pass filtering, care was taken to select critical program material—that is material having pronounced high-frequency content. The rationale for selection of audio material was described in more detail in Zielinski *et al.* (2003a, b, 2005). It was also important that the selection of audio excerpts had a suitable spatial characteristic. For example, half of the selected items contained an *F-B* (foreground-background) spatial characteristic and the other half contained an *F-F* (foreground-foreground) spatial characteristic in order to preserve a balance in the selection of program material. In the case of *F-B* recordings front channels reproduce predominantly foreground audio content (mainly close and clearly perceived audio sources), whereas rear channels contain only background audio content (room response, reverberant sounds, unclear, "foggy"). This situation may be compared to the typical sound impression perceived by a listener sitting in a concert hall (sound stage with musicians at the front, reflections from side and rear). In the case of the recordings exhibiting the *F-F* spatial characteristic, both front and rear channels contain predominantly foreground content. This category has similarities to the auditory impression encountered when a listener is surrounded by an orchestra. Rear channels contain clearly identifiable sound sources, often different from the instruments reproduced by front channels, for example percussion instruments, backing vocals, etc. [See Zielinski *et al.* (2003a, b) and Rumsey (2002) for a detailed discussion on the categorisation of audio program material according to spatial characteristics based on a scene-based paradigm.] A more detailed description of some of the selected material is given in Kirby *et al.* (1999). The average duration of the selected excerpts was 20 s. The stimuli were played back to the subjects in audio-only mode (without picture).

The quality of the stimuli was degraded using two types of signal processing: bandwidth limitation and down-mixing. Eleven algorithms were used for limitation of bandwidth (see Table I) and four algorithms for the down-mixing procedure (see Table II). The rationale for selection of these algorithms is discussed in detail in Zielinski *et al.* (2003a, b). (The effect of these algorithms on the overall information rate of the audio signal, considered in the PCM domain, was designed to be comparable between the bandwidth-limited versions

TABLE II. Algorithms used for down-mixing.

Label	Algorithm description	Comments
3/0	Down-mixing the rear channels to the front channels	Used for all items
1/2	Down-mixing front channels to the centre (mono) channel, leaving the rear channels intact	Used only for *F-F* program material
Stereo	Down-mixing to two-channel stereo	Used for all items
Mono	Down-mix to mono	Used for all items

and the down-mixed versions. For example, down-mixing the number of audio channels from five to three has the same effect on information rate as reducing the bandwidth of all five channels from 20 to 12 kHz. A range of values was used, extending from the equivalent of five full bandwidth channels, representing the original unprocessed recordings, down to one effective channel. After processing, the total set of stimuli to be evaluated consisted of 138 items (including unprocessed recordings). The reference sample against which the degraded samples were compared was always the undegraded (original) five-channel recording of the extract in question, representing the highest possible quality grade. The 138 audio stimuli used in the listening tests spanned the whole range of audio quality (ranging from bad to excellent) and excited a broad range of perceptual impressions regarding the employed fidelity attributes. For example, the low-quality anchors (3.5-kHz low-pass filtered items and mono down-mixed items) excited perceptual impressions that were on average graded using the bottom of the timbral fidelity and the spatial fidelity scales. In contrast, the top of the fidelity scales was used in order to evaluate the unprocessed items (hidden reference).

B. Grading attributes and scales

All stimuli were evaluated using four attributes:

(i) basic audio quality,
(ii) timbral fidelity,
(iii) frontal spatial fidelity, and
(iv) surround spatial fidelity.

The first attribute (basic audio quality) was described to the listeners according to the definition provided by the international recommendation (ITU-R, 1994), that is, as the single, global attribute describing any and all detected differences between the reference and the evaluated excerpt. (This attribute was evaluated in an earlier listening experiment to the three remaining attributes, so there should have been no bias in the basic audio quality ratings due to attention being drawn to specific attributes such as those evaluated in the later experiment.)

The remaining three attributes, evaluated in the second experiment, were concerned with audio fidelity (not quality). Fidelity scales were chosen, as opposed to descriptive or quality scales, because the task in hand concerned the comparison of stimuli to an unimpaired reference. "Fidelity" implies trueness of reproduction quality to that of the original. It was also clear from small-scale pilot elicitation studies (not reported here) that generalized fidelity scales were more suitable to describe the nature of the changes in the stimuli due to the degradations applied (which had a negligible effect on other factors such as noise and distortion characteristics) than a set of descriptive attribute scales. An illustrative example would be that spatial attributes such as "*ensemble width*" (Rumsey, 2002) could be ambiguously interpreted because the stimuli had a variety of different source or ensemble types, and sometimes (in the case of the ambient sounds) no clearly prominent sources or ensembles whatsoever. It would therefore have been necessary to specify which "source" or "ensemble" was intended in each of the stimuli, which could have led to additional confusion.

TABLE III. Grading scale used in the experiment, based on ITU-R 1534 (2001).

Quality / Fidelity	Grading range
Excellent	80–100
Good	60–80
Fair	40–60
Poor	20–40
Bad	0–20

In the case of the second attribute (timbral fidelity), listeners were asked to "grade each stimulus according to how similar it is to the reference, taking into account changes in timbre only." They were requested to ignore any spatial changes in the sound reproduction. (The timbral fidelity scale therefore provided a place for rating all aspects of sound quality that were not spatial.) As far as the frontal spatial fidelity is concerned (third attribute), listeners were asked to "grade each stimulus according to how similar it is to the reference, taking into account changes in spatial sound reproduction within the frontal arc (between the left and right loudspeaker)." They were instructed to ignore timbral changes, as well as any spatial changes outside the frontal arc. [This is similar to the attribute "front image quality," suggested as an additional evaluation scale for surround sound material in the ITU-R (1994) BS.1116 standard.]

The last graded attribute was surround spatial fidelity. In this case the listeners were asked to "grade each stimulus according to how similar it is to the reference, taking into account changes in spatial sound reproduction outside the frontal arc (not between the left and right loudspeaker)." They were instructed to ignore timbral changes, as well as any spatial changes inside the frontal arc. [This is similar to the attribute "impression of surround quality," suggested as an additional evaluation scale for surround sound material in the ITU-R (1994) BS.1116 standard.]

When taken together, frontal spatial fidelity and surround spatial fidelity were intended to provide a rating of the spatial trueness to the reference of a given stimulus for the complete horizontal soundfield.

For all attributes a 100-point continuous scale with labels was employed (see Table III). Listeners were instructed to use the maximum value of the scale when evaluating a hidden reference (an unprocessed recording).

C. Experimental design

The basic audio quality and the three fidelity attributes were graded on two separate occasions (basic audio quality was evaluated in the first experiment and the remaining three attributes were evaluated in the subsequent experiment). The experiments were designed on the basis of a double-blind multi-stimulus test paradigm with hidden reference and anchors (ITU-R, 2001). The same test paradigm was used to grade all attributes. Due to the large number of evaluations to be made by the listeners (138 items $\times$ 4 attributes) all audio

items were evaluated without any repetitions, except for the subset of listeners' error check items which were evaluated twice. The subset of error check items consisted of the original recordings (hidden reference), 3.5-kHz low-pass filtered items (low quality timbral anchor), and down-mixed to mono items (low quality spatial anchor).

The listening tests were organized using 16 30-min-long listening sessions. Each session consisted of six blocks, where each block contained seven or eight audio items to be evaluated. In each block the listener was presented with the original, unprocessed recording (labelled as a reference). The listener could switch at will between the stimuli. In order to reduce the bias and the risk of confusion due to the evaluation of several attributes simultaneously, each listener graded only one attribute during each listening session. The order of presentation of the audio stimuli and also the order in which subjects were asked to grade the fidelity attributes were randomized for each listener. More details about the experimental design can be found in Zielinski *et al.* (2003a, b, and 2005).

D. Listening panel

The listening panel was recruited from the staff and students of the Tonmeister course (Music and Sound Recording) at the University of Surrey. The procedure for recruitment, screening (including audiometric evaluation), and training of the listeners is described in Zielinski *et al.* (2003a, b,). According to the results of the audiometric examination, about 50% of the listeners can be characterized as having normal hearing [the threshold of 0–15 dB HL *re* ISO 389 (1991) from 125 Hz to 8 kHz], whereas the remaining listeners exhibited a slight loss of hearing (16–25 dB) HL. According to the results of the screening test it was found that the slight hearing losses did not have any adverse effect on evaluation of basic audio quality. It is not known whether these hearing losses had any detrimental effect on the evaluation of the timbral and spatial fidelities. After the training and screening a panel consisting of 21 experienced listeners graded the stimuli in terms of basic audio quality. In the next stage, which took place on a separate occasion, a subset of this panel, consisting of 16 of the listeners from the previous listening panel, was asked to evaluate the degraded items using the three remaining fidelity attributes. (Not all of the 21 listeners from the first experiment were available for the second experiment.)

All subjects had extensive experience with critical listening to traditional two-channel stereo recordings. Some of them had also experience in listening to 5.1 surround audio recordings. In order to reduce the bias related to the habits of listening to two-channel stereo recordings all subjects were given the opportunity to become familiar with a range of different surround audio recordings prior to the proper listening test.

Some further observations on the suitability of the listening panel are offered in Sec. IV.

E. Acoustical conditions

The listening tests were conducted in the Listening Room of the Institute of Sound Recording, University of Surrey. The acoustical parameters of this room conform to the requirements of the ITU-R Recommendation BS.1116 (ITU-R, 1994). All channels were aligned relative to each other with a tolerance less than ±0.5 dB SPL, measured at the reference listening position. The loudness of all stimuli (both original and processed) used in the experiment was equalized in order to minimize any experimental error due to loudness changes. Equalization was performed objectively using Moore's loudness model (Moore *et al.*, 1997) and corrected subjectively ("fine-equalized") at the center listening position by a small panel of listeners.

III. RESULTS

A. Listener consistency

In order to monitor the intralistener consistency the tests were designed in such a way that some items were evaluated twice (hidden references, down-mix to mono items, 3.5-kHz low-pass filtered items). Consistency was examined by inspecting the average grading error, estimated as the square root of the error variance obtained from the ANOVA model calculated separately for each subject. A small error indicates a high consistency of grading for a given listener. According to the obtained results, the average error of evaluation of basic audio quality was equal to only 1% for the most consistent listener and to 7% for the least consistent listener. A similar magnitude of grading error was observed for timbral fidelity scores. However, for the frontal and surround spatial fidelity scores the error was larger and ranged from 2% for the most consistent listener to 13% for the least consistent subject. The magnitude of the observed grading error is acceptable in the context of this experiment, taking into account the high complexity of the task undertaken by the listeners. (Subjective evaluation of spatial attributes is in general a challenging task and a grading error of the order of 10% is often encountered in listening tests of this nature.) These results confirmed that the recruited listening panel was able to use the provided scales with adequate consistency.

In order to assess the degree of interlistener consistency, initially a number of univariate tests were used. The consistency between the listeners was assessed by examining the correlation between the scores obtained for a given listener and the mean scores averaged across the listeners for each evaluated item. For the basic audio quality scores and for the timbral fidelity scores the correlation was higher than 0.9 for all the listeners, indicating a high consistency between subjects. However, for the frontal and surround spatial fidelities the correlation was not as high and for three subjects its values were slightly smaller than 0.8. Nevertheless, the observed correlation values can be still regarded as high, and therefore it could be concluded that the subjects were relatively consistent.

In order to check whether all the listeners were using the scale in a similar manner the histograms of the data were examined separately for each evaluated item. The histograms were inspected visually and also more formally using the

TABLE IV. Results of correlation analysis.

		Basic audio quality	Timbral fidelity	Frontal spatial fidelity	Surround spatial fidelity
Basic audio quality	Pearson correlation	1	0.925	0.628	0.427
	Significance (two-tailed)	⋯	0.000	0.000	0.000
	N	138	138	138	138
Timbral fidelity	Pearson correlation	0.925	1	0.329	0.192
	Significance (two-tailed)	0.000	⋯	0.000	0.024
	N	138	138	138	138
Frontal spatial fidelity	Pearson correlation	0.628	0.329	1	0.664
	Significance (two-tailed)	0.000	0.000	⋯	0.000
	N	138	138	138	138
Surround spatial fidelty	Pearson correlation	0.427	0.192	0.664	1
	Significance (two-tailed)	0.000	0.024	0.000	⋯
	N	138	138	138	138

Kolmogorov-Smirnov test for all four grading scales respectively. The scores exhibited normal (Gaussian) distribution for about 90% of the items with the standard deviation ranging up to 20 relative to the 100-point scale. This result indicated that the listeners were using the scales in a similar manner for most of the evaluated items. For the remaining items the distribution was asymmetrical due to the scale floor/ceiling effect. In a few cases the distribution obtained for the surround spatial fidelity resembled a uniform distribution with the standard deviation ranging up to 42, which could indicate that for this attribute and these items there was no consensus between the listeners. Therefore, it was decided to further investigate the consistency between subjects using a multivariate approach. Scatter plots from a PCA analysis, with the subjects plotted as the objects, were analyzed for each item separately. The scatter plots were studied for the first two dimensions only since they accounted for the most significant proportion of variance. The plots obtained for most of the evaluated items exhibited patterns with no groupings (subpopulations), indicating a high consistency between subjects. For three items (out of 138) some groupings of scores were observed, which showed that occasionally groups of listeners used the scales differently. Moreover, for six items it was observed that there was one listener who used the scale differently from other subjects. Therefore, in order to check to what extent this listener biased the results, the regression model was calculated with and without the scores obtained from this listener. The differences between the resultant regression coefficients were negligibly small and therefore it was decided not to screen the data from this listener. Since there was sufficient intersubject consistency, it was decided to average the results across all the listeners for each evaluated item and to use the mean values in subsequent analyses.

B. Correlation analysis

The results of a Pearson correlation analysis are presented in Table IV. The results of the rank order correlation were checked and are similar. According to the results, basic audio quality is most correlated with timbral fidelity ($r=0.93$). There is also a high correlation between frontal spatial fidelity and surround spatial fidelity ($r=0.66$). These results give rise to a question about whether a high correlation between the scores was caused by the fact that the listeners could not distinguish between the attributes or by other factors like properties of the sound field, etc. In order to answer this question it was decided to examine selected scatter plots between the scores obtained for different attributes. For example, Fig. 1 shows the scatter plot of basic audio quality scores as a function of the timbral fidelity scores. It can be seen that most of the scores are located on the diagonal ($y=x$), which accounts for the high correlation between the scores (0.93). However, there is also a group of scores departing from this pattern (outliers). These scores are attributed primarily to the mono down-mix items and 1/2 downmix items. According to this figure, the basic audio quality of these items is lower than the timbral fidelity (due to spatial degradation), which proves that the listeners could differentiate between the basic audio quality and the timbral fidelity. Similarly, it can be shown that the listeners could distinguish between the frontal spatial fidelity and the surround spatial fidelity. Figure 2 shows the scatter plot of frontal spatial fidelity scores as a function of the surround spatial fidelity scores. It is possible to notice that the scores are not only grouped along the diagonal ($y=x$) but are also scattered, to some extent, across the whole figure, which demonstrated

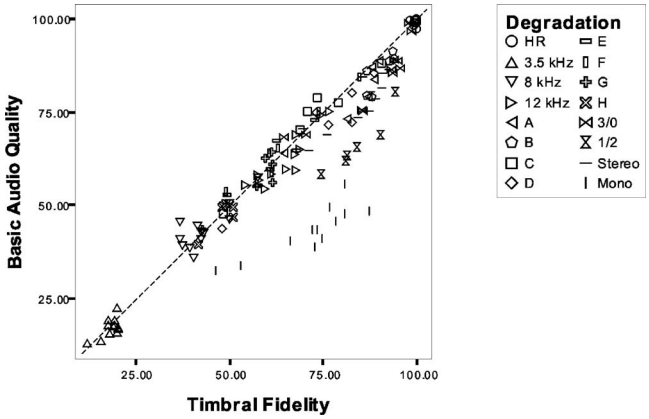


FIG. 1. Scatter plot of basic audio quality scores as a function of timbral fidelity scores. Scores averaged across subjects (dashed line $y=x$).

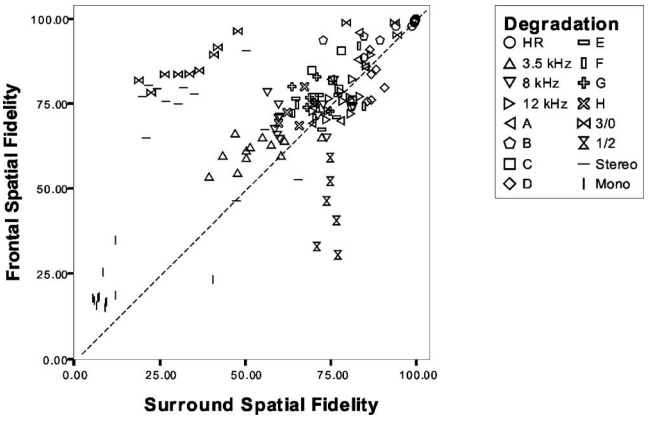


FIG. 2. Scatter plot of frontal spatial fidelity scores as a function of surround spatial fidelity scores. Scores averaged across subjects (dashed line $y=x$).

that the listeners could distinguish between the frontal spatial fidelity and the surround spatial fidelity. For example, the scores obtained for 1/2 down-mix items are scattered under the diagonal, which means that in this case, as one might expect, the frontal spatial fidelity was degraded more than the surround spatial fidelity. An opposite interaction can be observed for the scores obtained for down-mix to front channels 3/0 and for down-mix to stereo, where deterioration in frontal spatial fidelity was smaller than deterioration in surround spatial fidelity. An interesting grouping of scores can be also observed in Fig. 3. This figure depicts frontal spatial fidelity scores as a function of the timbral fidelity scores. It can be noted that for low-pass-filtered items (e.g., 3.5 kHz) the frontal spatial fidelity is much higher than the timbral fidelity. On the contrary, for the down-mixed items (e.g., mono) the frontal spatial fidelity scores are much lower than the timbral fidelity scores. According to these results, it can be concluded that the listeners could distinguish between the evaluated attributes. The high correlation between some of the scales (see Table IV) can be attributed to the fact that audio degradations used in this experiment were not perceptually "orthogonal." As shown in Zielinski *et al.* (2005), the low-pass filtering caused not only timbral distortion but also some spatial distortions; likewise down-mixing produced not only spatial changes but also timbral ones. This inevitably resulted in some correlation between the scores.

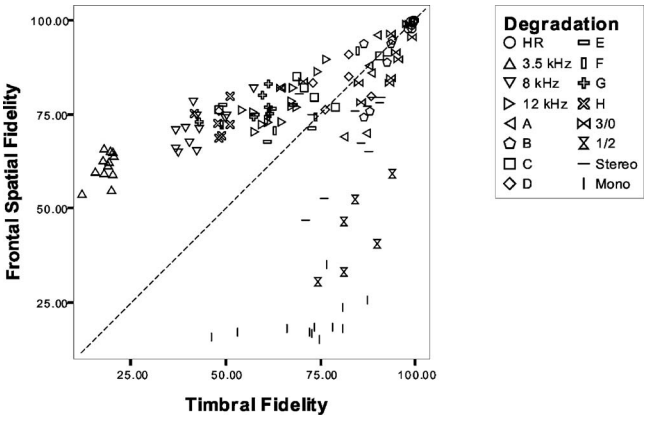


FIG. 3. Scatter plot of frontal spatial fidelity scores as a function of timbral fidelity scores. Scores averaged across subjects (dashed line $y=x$).

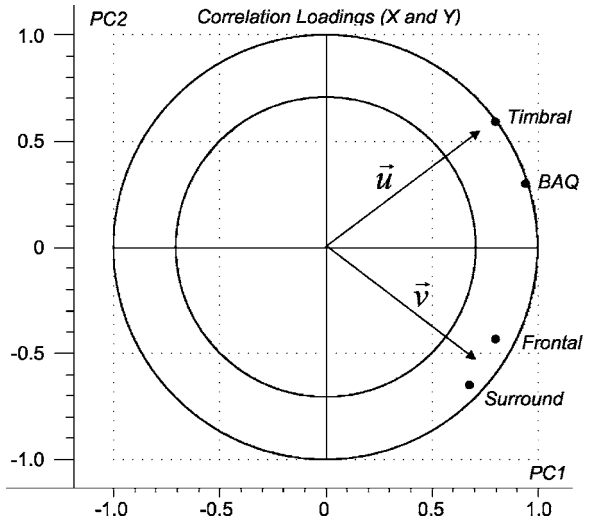


FIG. 4. Correlation loadings plot obtained from the PLS regression (PC1 and PC2 explain 88% and 9% of BAQ variance, respectively).

C. Regression model

The main research task in this study was to quantify the relationship between the basic audio quality ratings and the timbral and spatial fidelity ratings. Three different regression methods were employed at the initial stage of the study: a multiple linear regression, a principal component regression, and a partial least squares regression. These led to almost identical results. Due to space limitations, it was decided to report here only the results obtained using the partial least squares regression (PLS-R), since it is known that this method is superior to the previously mentioned methods in terms of its statistical properties (Esbensen, 2002).

The first task required when performing a PLS-R is to determine the optimum number of principal components to be taken into account in the final solution. The primary reason for developing the regression model is to explore the data, which would imply using a model consisting of three principal components (a full model). However, since some attempts will be made in subsequent sections to generalize the conclusions, it was decided to develop a regression model that also has good predictive properties. The task of finding the optimum number of principal components in the regression model can be undertaken either by finding the point of inflection in the screen plot or by analyzing the changes in the explained variance. According to the obtained results, the regression model based on the first principal component accounted for 88% of the variance. If two principal components were included in the model, the percentage of explained variance rose to 97%. However, the performance of the model improved only marginally (by 0.1%) if the third principal component was taken into account. The optimum number of principal components to be included in the regression model is therefore two, since this represents the best trade-off between the percentage of explained variance and the number of principal components used for the multidimensional decomposition of data. [It is known that "simpler" models generally have better prediction properties when used with new data (Esbensen, 2002).]

Figure 4 shows the correlation loading plots obtained

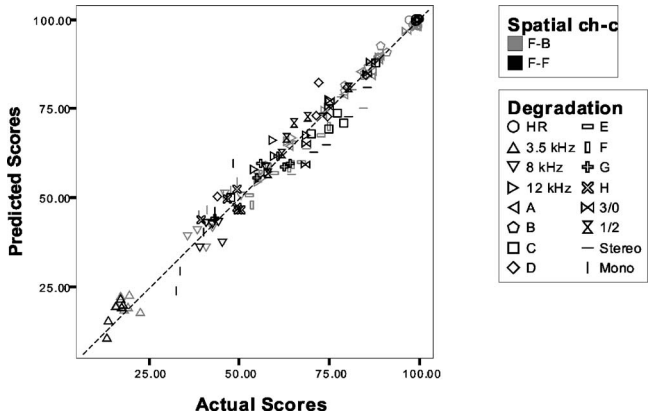


FIG. 5. Basic audio quality—scatter plot of the predicted scores as a function of the actual scores obtained during the listening tests (PLS-R model using two PCs) (dashed line $y=x$).

during the PLS regression. The correlation loading plots are a useful tool for exploration of the interrelationships between the X variables and the relationships between the X and Y variables (basic audio quality is the Y variable in our case whereas the remaining three fidelity scales constitute the X variables). As far as the X variables are concerned it is possible to note that frontal and surround spatial fidelity scales are grouped close to each other, which means that they are correlated. The correlation between these variables equals 0.66 (see Table IV). It can be hypothesized that these two scales taken together form a new combined scale describing the total spatial fidelity, represented by vector $\vec{v}$ in Fig. 4. If this hypothesis is true, it is interesting to note that the angle between the vector $\vec{v}$ and the vector $\vec{u}$ pointing at the timbral fidelity is equal to about 90°. This would mean, as one could expect, that from a perceptual point of view timbral fidelity and spatial fidelity are "orthogonal."

According to Fig. 4 it is also clear that the basic audio quality (BAQ) is more correlated with the timbral fidelity than with the frontal or surround fidelity. This observation is also supported by the regression equation derived during this analysis, which can be written in the following form:

$$\mathrm{BAQ} = 0.80\ \mathrm{Timbral} + 0.30\ \mathrm{Frontal} + 0.09\ \mathrm{Surround} - 18.7. \quad (1)$$

The main conclusion that can be drawn from this equation is that the changes in basic audio quality (BAQ) depend approximately two times more on timbral fidelity changes than they do on spatial fidelity changes (Frontal and Surround).

More information regarding the regression model can be obtained by analyzing the scatter plot between the actual and the predicted scores (see Fig. 5). It is clear that it fits the actual data obtained during the listening tests well. The correlation coefficient between the actual and predicted scores is very high (0.99). The slope of the regression line presented in this figure is close to unity (0.98) whereas its vertical offset is small (1.64 points relative to 100-point scale). It is also important to note that the root mean square error of the calibration (RMSEC) is small and approximately equals 4 points relative to the employed 100-point scale, which shows that the developed model describes the data with a high accuracy (4%). RMSEC is a useful measure for the model fit, calculated for the calibration objects only and expressed in original measuring units (Esbensen, 2002). It can be interpreted as the average error between the predicted and actual scores.

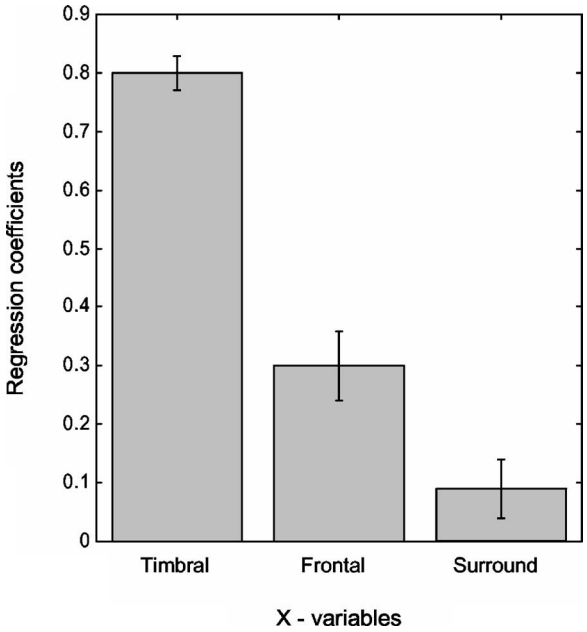


FIG. 6. Unstandardized regression coefficients from the PLS-R model with associated Martens' uncertainty limits.

The reliability of the obtained regression model was tested using the random cross-validation approach combined with the Martens' uncertainty test (Martens and Martens, 2000). The obtained results are summarized in Fig. 6, which shows in a graphical form the regression coefficients included in Eq. (1). The uncertainty limits obtained using the Martens' test are also presented in this figure. In the case of the first two variables (timbral and frontal fidelities) the size of the uncertainty intervals is small relative to the values of the regression coefficients, which indicates a high stability solution in those two cases. On the contrary, the relative stability of the coefficient obtained for the surround spatial fidelity is small. It is important to note that all regression coefficients presented in Fig. 6 are statistically significant and therefore should be included in the regression equation.

D. Dependency on program material

It is reasonable to ask whether or not the choice of program material, and in particular the surround sound mixing style, had a marked effect on the resulting regression equations shown above. For example, is it reasonable to pool all of the perceptual judgments when generating the regression equations because one might expect "*F-B*" program items to give rise to different results from "*F-F*" items (because the former makes more subtle use of the surround channels)? Furthermore, does the regression model hold true for all the quality degradation types used in the experiment? Certainly the high degree of fit seen in the regression models suggests that they are valid across all conditions, but some examples are given below to demonstrate this point.

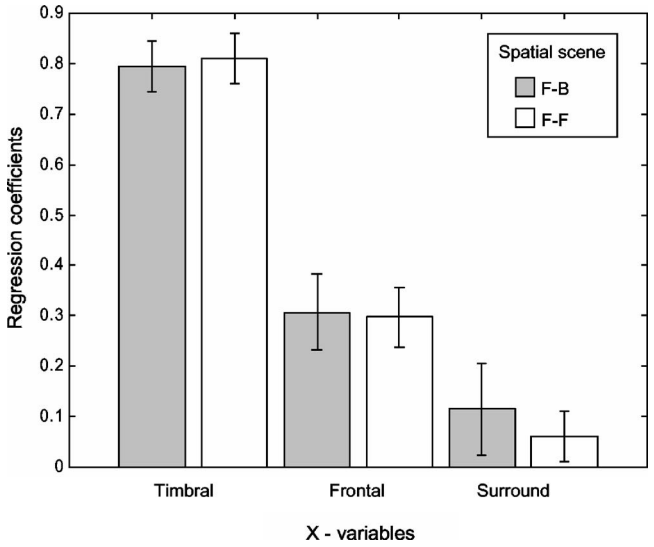


FIG. 7. Unstandardized regression coefficients and Martens' uncertainty limits for two regression models based on program exhibiting *F-B* and *F-F* scene characteristics, respectively.

Figure 7 shows the unstandardized regression coefficients for two different PLS regression models, one based only on the *F-B* program material and the other based on the *F-F* program material. Although the coefficients are not identical, they are not in fact significantly different from a statistical point of view because the Martens' uncertainty limits can be seen to overlap. This suggests that the perceptual importance of the different fidelity scales does not differ appreciably depending on the spatial mixing style of the program material.

Previously discussed Fig. 5 shows the scatter plot for *all* the data points used in the PLS regression model. These consisted of different mixing styles (spatial scenes—*F-F* and *F-B*) and different degradations covering wide range of quality. It can be seen that all the data points lie close to the diagonal line, suggesting that the model exhibits a good fit no matter what the nature of the program material or the quality level.

IV. DISCUSSION AND FUTURE WORK

Although the regression equations shown above demonstrate that, in the perceptual domain, spatial fidelity has less influence over global quality judgments than timbral fidelity, it is far from unimportant. In fact, it can be seen to account for approximately 30% of the overall quality rating. Accordingly it can be proposed that spatial attributes should be incorporated into future perceptual models of audio quality as an important factor. Current perceptual models that attempt to predict quality differences between a reference and a degraded version of an audio signal, such as that given in ITU-R BS. 1387 (1998), do not take into account changes in spatial factors, for example. It is recognized that the perceptual audio codecs for which the ITU model was developed did not primarily introduce spatial changes (although stereo imaging was sometimes affected in various subtle ways), giving rise primarily to differences in technical quality (noise, distortion, spurious components, etc.). Nonetheless, it is likely that the growing number of codecs that involve parametric, scalable, and scene-based representation of spatial audio using various forms of stereophony (e.g., multichannel, binaural, wavefield synthesis) may introduce substantial compromises in spatial quality, particularly at the lowest bit rates, giving rise to a need to implement trade-offs such as those implied by the regression models given in this paper. Furthermore, quality evaluation models will need to take spatial changes into account if they are to predict perceived quality accurately.

The choice of stimuli for the experiments on which these regression models were based is bound to have affected the results to some extent. Nonetheless, these items were chosen to be typical of a wide range of program material, including music, movie, broadcast, and effects, using a variety of recording and/or mixing styles, and so can be considered representative. The authors introduced quality changes designed to cover a wide range of perceptual values in the spatial and timbral domains. If smaller changes were to be introduced, it is anticipated that the overall magnitude of the perceived effect would be smaller but that the overall regression values would be likely to remain similar. As can be seen in Fig. 5, higher quality items seem to fit the model as well as lower quality items. The models, although pointing to an interesting trend in the general relationship between timbral and spatial quality in overall quality judgment, can only be said to apply accurately to the types of quality changes applied in this experiment.

Informal comments from some listeners may help to explain the way in which they originally rated "basic audio quality" and the reason for the emphasis on timbral fidelity in the models. Although basic audio quality is defined as incorporating "any and all differences" between the reference and the impaired item, comments from some listeners following the test suggested that they found it difficult to consider spatial changes as changes in "quality." Some appeared to be strongly wedded to the idea that "sound quality" is primarily to do with the technical attributes of the sound that they hear, such as distortion and spectral changes, which may have had a bearing on the smaller weighting given to spatial fidelity scales in the regression equations. Whatever the truth of this suggestion, it seems most likely that these experienced listeners would opt to preserve technical quality if they had to choose something to be degraded, rather than spatial quality.

It is interesting to notice that the regression equations contain a relatively small weighting with respect to surround spatial fidelity. One might argue that this is because half of the program items were mixed in the "*F*-B" mode, and that the rear channels contained primarily diffuse effects or reverberation. However, the evidence presented in Sec. III D does not bear this out. The small weighting with respect to this factor suggests that the listeners did not take much notice of changes in the spatial distribution of surround information when rating basic audio quality, a factor that may be due to the degree of familiarity they had with conventional two-channel stereo, and the fact that surround audio as a listening experience is a relatively new phenomenon. An alternative,

and probably more likely, explanation is that subjects noticed these changes but did not regard them as particularly detrimental to overall quality.

There is an ongoing debate as to whether experienced listeners should be used for both evaluation of basic audio quality and for scaling more precise attributes (e.g., timbral or spatial fidelity), as was the case in this study. So far, there is no evidence that this approach is invalid and there is general agreement within the scientific community and in international standards that basic audio quality should be evaluated by experienced listeners. The results of this experiment are therefore only representative of a group of experienced listeners. There is an argument for investigating the degree to which basic audio quality grades represent the *preference* of listeners and whether the preferences of naive listeners differ from the quality ratings of experienced listeners. This has been reported separately (Rumsey *et al.*, 2005). Results suggest that naive subject preference has a similar structure in relation to experienced listener fidelity ratings as that of BAQ ratings given by the experienced listeners, but that the weighting given to surround spatial fidelity is slightly higher.

V. CONCLUSIONS

A regression model mapping three audio fidelity attributes (timbral fidelity, frontal spatial fidelity, and surround spatial fidelity) onto basic audio quality was developed in the context of home cinema surround sound reproduction. This model fits the data acquired in the listening tests with high accuracy, since the correlation between the predicted and the actual basic audio quality scores is close to unity (0.99) and the average error is of the order of 4%. According to the regression model timbral fidelity ratings were dominant in the judgment of basic audio quality, and frontal spatial fidelity was regarded as more significant than surround spatial fidelity. Spatial fidelity can be seen to account for approximately 30% of the basic audio quality rating and must therefore be considered as an important factor in future perceptual models of sound quality.

It is hoped that the obtained results will contribute to the development of a cognitive model for audio quality. Moreover, the developed regression equation could be used by audio engineers during psychoacoustical optimization of audio systems where the trade-off between their timbral performance and spatial properties is sought.

ACKNOWLEDGMENTS

This project was carried out with the financial support of the Engineering and Physical Sciences Research Council, UK (GR/N24032). Some excerpts used in the project were kindly supplied by BBC, R&D Department (used with permission).

Esbensen, K. (**2002**). *Multivariate Data Analysis—in practice* (Camo, Oslo).

Gabrielsson, A., and Lyndstrom, B. (**1985**). "Perceived sound quality of high-fidelity loudspeakers," J. Audio Eng. Soc. **33**, 33–53.

Huopaniemi, J., Zacharov, N., and Karjalainen, M. (**1998**). "Objective and subjective evaluation of head-related transfer function filter design," presented at 105th AES Convention, San Francisco, 26–29 September, Paper 4805, Audio Engineering Society.

ISO (**1991**). *389: Acoustics—Standard reference zero for the calibration of pure-tone air conduction audiometers*, International Organization for Standardization.

ITU-R (**1993**). *Recommendation BS. 775: Multichannel stereophonic sound system with and without accompanying picture*, International Telecommunications Union.

ITU-R (**1994**). *Recommendation BS. 1116: Methods for subjective assessment of small impairments in audio systems including multichannel sound systems*, International Telecommunications Union.

ITU-R (**1998**). *Recommendation BS. 1387: Method for objective measurements of perceived audio quality*, International Telecommunications Union.

ITU-R (**2001**). *Recommendation BS. 1534: Method for the subjective assessment of intermediate quality level of coding systems*, International Telecommunicatons Union.

Kirby, D. G., Cutmore, N., and Fletcher, J. (**1999**). "Program origination of five-channel surround sound," J. Audio Eng. Soc. **46**, 323–330.

Martens, H., and Martens, M. (**2000**). "Modified jack-knife estimation of parameter uncertainty in bilinear modelling by partial least squares regression (PLSR)," Food Quality and Preference **11**, 5–16.

Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of thresholds, loudness and partial loudness," J. Audio Eng. Soc. **45**, 224–240.

Rumsey, F. (**2002**). "Spatial quality evaluation for reproduced sound: terminology, meaning and a scene-based paradigm," J. Audio Eng. Soc. **50**, 651–666.

Rumsey, F., Zielinski, S., Kassier, R., and Bech, S. (**2005**). "Relationships between experienced listener ratings of multichannel audio quality and naïve listener preferences," J. Acoust. Soc. Am. **117**, 3832–3840.

Toole, F. (**1985**). "Subjective measurements of loudspeaker sound quality and listener performance," J. Audio Eng. Soc. **33**, 2–32.

Zacharov, N., and Huopaniemi, J. (**1999**). "Results of a round robin subjective evaluation of virtual home theatre sound systems." presented at the 107th AES Convention, New York, 24–27, September, Paper 5067.

Zacharov, N., and Koivuniemi, K. (**2001**). "Unravelling the perception of spatial sound reproduction: analysis & external preference mapping," presented at 111th AES Convention, New York, 30 November–3 December, Paper 5423.

Zielinski, S. K., Rumsey, F., and Bech, S. (**2003**). "Comparison of quality degradation effects caused by limitation of bandwidth and by down-mix algorithms in consumer multichannel audio delivery systems," presented at 114th AES Convention, Amsterdam, 22–25 March, Paper 5802.

Zielinski, S. K., Rumsey, F., and Bech, S. (**2003b**). "Effects of down-mix algorithms on quality of surround sound," J. Audio Eng. Soc. **51**, 780–798.

Zielinski, S. K., Rumsey, F., Kassier, R., and Bech, S. (**2005**). "Comparison of basic audio quality, timbral and spatial fidelity changes caused by limitation of bandwidth and by down-mix algorithms in 5.1 surround audio systems," J. Audio Eng. Soc. **53**, 174–192.

Can dichotic pitches form two streams?

Michael A. Akeroyd[a)]
MRC Institute of Hearing Research (Scottish Section), Glasgow Royal Infirmary, Alexandra Parade, Glasgow, G31 2ER, United Kingdom

Robert P. Carlyon and John M. Deeks
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 2EF, United Kingdom

(Received 18 November 2004; revised 9 May 2005; accepted 12 May 2005)

The phenomenon of auditory streaming reflects the perceptual organization of sounds over time. A series of "A" and "B" tones, presented in a repeating "ABA-ABA" sequence, may be perceived as one "galloping" stream or as two separate streams, depending on the presentation rate and the A-B frequency separation. The present experiment examined whether streaming occurs for sequences of "Huggins pitches," for which the percepts of pitch are derived from the binaural processing of a sharp transition in interaural phase in an otherwise diotic noise. Ten-second "ABA" sequences were presented to eight normal-hearing listeners for two types of stimuli: Huggins-pitch stimuli with interaural phase transitions centered on frequencies between 400 and 800 Hz, or partially-masked diotic tones-in-noise, acting as controls. Listeners indicated, throughout the sequence, the number of streams perceived. The results showed that, for both Huggins-pitch stimuli and tones-in-noise, two streams were often reported. In both cases, the amount of streaming built up over time, and depended on the frequency separation between the A and B tones. These results provide evidence that streaming can occur between stimuli whose pitch percept is derived binaurally. They are inconsistent with models of streaming based solely on differences in the monaural excitation pattern. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1945566]

PACS number(s): 43.66.Pn, 43.66.Ba [AK] Pages: 977–981

I. INTRODUCTION

If short bursts of two tones, A and B, separated by only a small frequency difference (ΔF), are concatenated into an ABA-ABA-ABA-$\cdots$ sequence, then a listener will hear one, "galloping" stream of tones, varying in pitch. In contrast, at wider frequency separations, the listener will hear two separate streams, each with its own, steady pitch, and the tendency to hear two streams increases with presentation rate. Furthermore, for a given ΔF and presentation rate, the percept will tend to "build-up" from a single stream near the beginning of the sequence to two streams at the end (Anstis and Saida, 1985; Carlyon *et al.*, 2001; Cusack *et al.*, 2004). This phenomenon reflects the process of "auditory streaming," and is important both for our ability to separate one speaker from a background of others and for following the melody of one instrument in an orchestra (van Noorden, 1975; Bregman, 1990).

According to an influential computational model (Beauvois and Meddis, 1991, 1996), the effect of frequency separation on streaming results from its effect on the overlap between the peripheral excitation patterns produced by the A and B tones. However, it appears that streaming based on pitch differences also occurs when produced by complex tones from which the lower (resolved) harmonics have been removed, and where the peripheral excitation patterns produced by the A and B tones do not differ systematically (Vliegen and Oxenham, 1999; Vliegen *et al.*, 1999; Grimault *et al.*, 2000; for a review see Moore and Gockel, 2002). This suggests that streaming can take place at neural sites which do not receive, or at least do not require, a peripheral tonotopic representation.

In an effort to further constrain the sites at which pitch-based streaming may occur, we investigated streaming produced by "dichotic pitches." Cramer and Huggins (1958) discovered that pitch sensations could be created by the binaural interaction of noise stimuli. A typical stimulus is a white noise, diotic apart from a transition in interaural phase across a narrow band of frequencies around 500 Hz. The waveforms at the two ears differ *only* in the phases of these frequencies. When played monaurally, either of the left and right waveforms sound like white noise, but when played together, a percept of a faint 500-Hz tone is also heard, lateralized to one side or the other (e.g., Raatgever and Bilsen, 1986; Akeroyd and Summerfield, 2000; Zhang and Hartmann, 2004). In many ways this "Huggins pitch" behaves as an ordinary tone, for example, the "octave enlargement" effect occurs for both (Hartmann, 1993), and Huggins pitches are strong enough to form melodies and be easily heard by untrained listeners (Akeroyd *et al.*, 2001). Because a Huggins pitch can *only* be heard when both waveforms are played, its percept must be derived from auditory processing at the brainstem or higher. Modern theories of the creation of the perception of dichotic pitch all postulate an internal spectrum in which there is a peak at the frequency of the center of the transition in interaural phase, although there is as yet no consensus as to quite how the spectrum is generated (e.g., Raatgever and Bilsen, 1986; Culling *et al.*, 1998; Hartmann and Zhang, 2003).

[a)]Electronic mail: maa@ihr.gla.ac.uk

The primary motivation of the present study was to ascertain if stream segregation occurred with ABA-ABA-ABA-$\cdots$ sequences of dichotic pitches. A secondary aspect of the design tested if two of the factors that influence the streaming of pure tones—namely, ΔF and the build-up over time—also affected the streaming of dichotic pitches. If stream segregation can indeed operate on pitch information that has been derived from binaural interactions, then listeners should report one stream at the beginning of the sequence, but, over the course of about 10 s, would report two. They should also report two streams more often for larger than for smaller ΔFs.

II. METHODS

A. Stimuli

Four conditions were tested. The stimuli were 10-s sequences of either Huggins-pitch stimuli or tone-in-noise control stimuli (see below), with an A frequency of 400 or 800 Hz. For the 400-Hz A tones, the B frequencies were, in different subconditions, either 4, 6, or 8 semitones higher, whereas for the 800-Hz A tones the B frequencies were 4, 6, or 8 semitones lower. Each Huggins-pitch stimulus in a sequence was constructed in the frequency domain, as described by Akeroyd *et al.* (2001). Two matched spectral buffers, representing the left and right channels of a diotic Gaussian noise sampled at 22050 Hz, were created and then rectangular filtered (0–4000 Hz passband) in the spectral domain. The interaural phase shift was implemented by modifying the phases of the frequency components in the spectral buffer representing one channel: a linear shift of 0 to 2π radians was added to the phases for frequency components from 10% below to 10% above the frequency of the note. Subsequently the signal waveforms for the left and right channels were created by applying an inverse discrete Fourier transform to the two spectral buffers, giving waveforms of 125-ms duration. They were then concatenated into a 10-s sequence of "ABA" triplets, each separated by 125 ms of diotic noise. Each sequence was bandpass filtered between 100 and 2000 Hz, in order to remove the transients from the concatenation of each 125-ms segment, and finally given a smoothed onset and offset of 30-ms duration. The spectrum level of the noise used in this and in the following control condition was 40 dB SPL.

In order to compare the results obtained with Huggins-pitch stimuli to those occurring when monaural excitation-pattern cues are available, we included control conditions with sequences of pure tones. To produce a pitch percept that was similar to the Huggins-pitch sequences, the tones were presented diotically against a continuous background noise (i.e., NoSo), lowpass filtered at 2000 Hz. The NoSo configuration was chosen so that any streaming could only have been attributable to *monaural* processing; although other configurations (e.g., NoSπ) would have created localizations similar to those of the Huggins-pitch stimuli, these would have been strongly influenced by binaural processes. The following procedure was adopted in an attempt to match, roughly, the strength of the pure tones in these diotic (NoSo) stimuli to those of the Huggins-pitch stimuli. First, we computed the mean interaural correlation at the output of a gammatone filter placed at the center frequency of the Huggins-pitch transition band; for example, at 400 Hz it was approximately 0.02. Second, we found the level of a 125-ms duration NoSπ tone which gave the same amount of interaural decorrelation; at 400 Hz and for a noise spectrum level of 40 dB, it was found to be 57 dB. Third, from the data of Blodgett *et al.* (1958), we estimated the detection threshold for such an NoSπ tone to be 47 dB.[1] Accordingly, we estimated the sensation level of the Huggins pitches to be $57-47=10$ dB. Finally, we noted from Blodgett *et al.*'s data that the threshold of a 125-ms NoSo tone was about 63 dB, and so we set the level of the pure tones in the NoSo control stimuli to be $63+10=73$ dB.

B. Procedure

In the main part of the experiment we presented 10-s sequences of ABA stimuli, and asked our listeners to report throughout how many streams they heard. They did so by clicking with a mouse on one of two virtual buttons on a computer screen, marked "one stream" and "two streams." They were told to press one of these buttons whenever their percept changed, and their responses therefore map out what each listener perceived at each point during each sequence. For statistical convenience, we quantized the responses into nonoverlapping, 1-s bins. The first two bins were excluded from the analysis because subjects did not always make their first response within the first 2 s of each sequence (Carlyon *et al.*, 2001; Cusack *et al.*, 2004).

Prior to the main test, subjects were first played eight simple Huggins-pitch melodies, to confirm that they could indeed perceive a dichotic pitch (Akeroyd *et al.*, 2001). All eight listeners reported hearing the appropriate melodies, with six listeners hearing the melodies to the left of the center of the head, and two to the right. Next, they were shown a diagram illustrating the two possible perceptual organizations, told that these could change during a sequence, and performed some training runs with demonstration versions of the pure-tone sequences. The demonstration versions used ΔFs of 3 and 12 semitones, to illustrate percepts of one and two streams, respectively. Their frequencies were increased by a factor of 2.25 relative to those used in the main experiment, and there was no background noise. Listeners were encouraged to concentrate on the rhythm instead of the overall pitch of the stimuli. Third, they practiced making streaming judgements for about 10 min on the experimental stimuli. They were told that the pitches of these stimuli would be lower in frequency and fainter than in the demonstration stimuli, and so they should continue making their judgments on the rhythm. The main test followed, in which each experimental sequence was presented 20 times per listener. The listeners were not instructed to try to keep the stimuli into one "galloping" stream nor to try to separate them into two streams; instead, they were encouraged to listen naturally and to report what they perceived (Carlyon *et al.*, 2001).

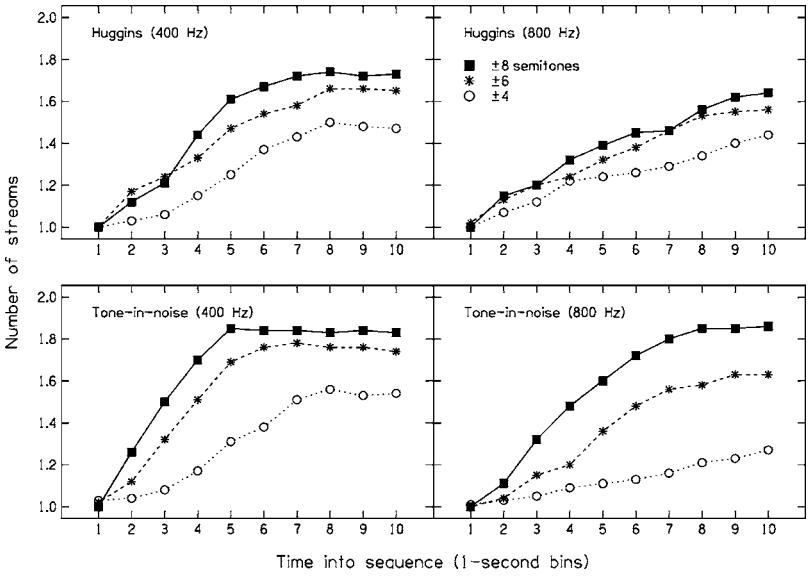


FIG. 1. Mean number of streams perceived by the listeners as a function of time into each 10-s sequence of stimuli. The top panels are for the Huggins-pitch stimuli, and the bottom panels are for the tone-in-noise control stimuli. The left panels are for an A frequency of 400 Hz, and the right panels are for an A frequency of 800 Hz. The parameter is the frequency separation (ΔF) between the A and B tones; 4 semitones (circles), 6 semitones (asterisks), and 8 semitones (squares). The results were quantized into 1-s bins and then averaged across the eight listeners.

C. Listeners

Eight normal-hearing listeners participated. Four of them completed all the Huggins-pitch conditions before starting any of the tones-in-noise conditions, while the other four did the tones-in-noise conditions first.

III. RESULTS

The results are plotted in Fig. 1. Each panel shows the number of streams reported, averaged across the listeners, for $\Delta F = \pm 4$, ± 6, or ± 8 semitones (shown by circles, asterisks, and squares, respectively). The upper panels show the results for the Huggins-pitch stimuli, and the lower panels show the results for the tone-in-noise control stimuli; the left panels are for the A=400 Hz stimuli, and the right panels are for the A=800 Hz stimuli.

The primary result is that listeners did indeed report stream segregation for the Huggins-pitch stimuli. Furthermore, the build-up and ΔF effects occurred for Huggins-pitch stimuli almost as much as for the tone-in-noise control stimuli; the number of reported streams increased towards the end of the sequence, and they were more likely to report two streams in the higher ΔF sequences than in the lower ΔF sequences.

To assess the significance of these effects, we conducted a three-way within-subjects ANOVA, contrasting the effect of frequency of the A tone, ΔF, and time-in-sequence upon the number of reported streams.[2] For the Huggins-pitch stimuli, there was a significant effect of time-in-sequence $[F(7,49)=13.9, p<0.001]$ and ΔF $[F(2,14)=6.5, p=0.01]$, but not of A frequency $[F(1,7)=0.8, p>0.1]$. The A-frequency by time-in-sequence interaction was found to be marginally significant $[F(7,49)=2.7, p=0.07]$. The other interactions were found to be insignificant. A separate ANOVA was conducted for the tones-in-noise stimuli. It showed that the three factors all gave significant effects: time-in-sequence $[F(7,49)=30.6, p<0.001]$, ΔF $[F(2,14)=59.0, p<0.001]$, and A-frequency $[F(1,7)=21.8, p=0.002]$. Two of the three two-way interactions were, at least, marginally significant: time-in-sequence by ΔF $[F(14,98)=2.5, p=0.03]$, and ΔF by A-frequency $[F(2,14)=3.5, p=0.07]$, but time-in-sequence by A frequency was not significant $[F(7,49)=2.1, p=0.1]$. Finally, the three-way interaction was also significant: time-in-sequence by ΔF by A-frequency $[F(14,98)=10.7, p<0.001]$. The interaction between ΔF and time-in-sequence occurred because listeners always reported one stream at the beginning of each sequence, but the number of "two stream" judgements at the end was lower for small ΔFs. The interaction between ΔF and A-frequency occurred because, although fewer two-stream responses were made for the 800-Hz than for the 400-Hz A tones at most ΔFs, this difference was smaller at the largest ΔF due to ceiling effects. This was especially true later in the stream, so accounting for the 3-way interaction between ΔF, A-frequency, and time-in-sequence.

We conducted a multiple-regression analysis to rank the importance of the various factors in determining the data. The quantitative factors of "time-in-sequence" and ΔF were coded as $1, 2, 3, \ldots 10$ s, and 4, 6, or 8 semitones, respectively, whilst the binary factors of A-frequency and stimulus type were coded as 1=400 Hz, 2=800 Hz, or 1=tones-in-noise, 2=Huggins-pitch stimulus. The analysis was applied to the mean data plotted in Fig. 1. It showed that the most-important factor was time-in-sequence ($r^2=0.58$), followed, at some remove, by ΔF ($r^2=0.19$). The factors of the A-frequency ($r^2=0.05$) and type-of-stimulus ($r^2=0.02$) were the least important predictors of the data.

The small effect of type-of-stimulus is shown in the figure by the curves for different ΔFs being lower for the Huggins-pitch stimuli than for the tone-in-noise stimuli. This was confirmed by a four-way ANOVA with factors of stimulus type, A-tone frequency, ΔF, and time-in-sequence, which revealed a main effect of stimulus type $[F(1,7)=6.2, p=0.04]$. There was also a significant interaction between stimulus type and ΔF $[F(2,14)=7.8, p=0.005]$; this is reflected in Fig. 1 by the fact that the separation between the curves for different ΔFs differ for the Huggins than for the tone-in-noise stimuli. The reason for the smaller effect of ΔF for the Huggins-pitch stimuli is not certain. It may be a result of the "sluggish" response of the binaural system in response

to dynamic changes (e.g., Grantham and Wightman, 1978; Culling and Summerfield, 1998, Akeroyd and Summerfield, 1999), which could impair the ability to follow the frequency changes between successive Huggins pitches (although we note that not all monaural analyses of pitch are fast: some, such as the pitch of unresolved harmonics, are sluggish; e.g., White and Plack, 2003). A second possibility is that the internal representation of dichotic pitches may be less accurate in frequency than for typical, monaural pitches. Henning and Wartini (1990) have shown that the frequency difference limen is larger for a tone presented dichotically in a noise ($NoS\pi$) than diotically (NoSo) at an equal sensation level, and Hartmann's (1993) direct measurements of the accuracy of pitch matching of a Huggins-pitch stimulus found an average value of 0.5%, whilst the value for a diotic pure-tone stimulus in silence is approximately 0.1% (Kohlrausch and Houtsma, 1992). This conjecture is consistent with Grimault *et al.*'s (2000) study of the effect of harmonic resolvability on the stream segregation of complex tones. They observed fewer "two-stream" responses when the harmonics were highly unresolvable than when they were highly resolvable (see also Vliegen *et al.*, 1999). As other data (e.g., Houtsma and Smurzynski, 1990; Carlyon and Shackleton, 1994) shows that the detectability of changes in fundamental frequency is considerably worse for a set of unresolved harmonics than for a set of resolved harmonics; it may well be the case that a larger ΔF is needed to induce streaming for stimuli with a relatively indistinct, imprecise, representation of pitch or pitch strength.

A final analysis was performed to test for a potential alternative explanation for the build-up observed in our mean data for the Huggins-pitch stimuli. As all listeners received initial practice with high-frequency tones in quiet before the experiment started, it is possible that they learned to expect a switch from one stream to two as the sequences progressed. This may have caused them to adopt a similar strategy when listening to the Huggins-pitch stimuli. Furthermore, four of our listeners were tested with the Huggins-pitch sequences only *after* being tested on the tones-in-noise sequences. We reasoned that if the build-up observed with the Huggins-pitch stimuli were due to subjects having learnt "what to expect" from the diotic stimuli, it should be greater in those subjects tested with the Huggins-pitch stimuli last, compared to those tested with Huggins-pitch stimuli first. We therefore conducted another ANOVA, with the order in which subjects were tested entered as a between-subjects factor. This factor was not significant [$F(1,6)=0.8, p>0.1$], and did not interact with A-frequency, ΔF, or time-in-sequence [respectively, $F(1,6)=1.0, p>0.1$; $F(2,12)=0.0, p>0.1$; $F(7,42)=0.4, p>0.1$]. We conclude that the data were not compromised by a learning effect.[3]

IV. DISCUSSION

Overall, the results demonstrate that Huggins-pitch stimuli can form two streams, like partially-masked tones-in-noise do. A build-up of streaming was observed in the Huggins-pitch condition; listeners were often reporting two streams at the end of the ABA-ABA-$\cdots$ sequences. The A-B frequency difference ΔF had a similar, albeit slightly smaller, effect to that which it has for pure tones; the larger the frequency difference between the A and B tones, the more two-stream reports were made. Thus, pitch information derived from binaural processing is sufficient for streaming to occur.

The results are inconsistent with the predictions of the model of Beauvois and Meddis (1991, 1996), according to which streaming arises solely from monaural peripheral processes, and they add to others that show that binaurally-derived lateralization information—from ear-of-presentation or interaural-time-differences—can help in the segregation of pure-tone melodies (Hartmann and Johnson, 1991). Our results are consistent with Moore and Gockel's (2002) hypothesis that streaming can stem from a variety of cues, including both spectral and purely temporal differences, and that the amount of streaming depends on the strength of the perceptual differences between stimuli. A complete account of streaming would have to include what those loci are, how the operations interact, and, given recent evidence for a strong effect of attention on streaming (Carlyon *et al.*, 2001; Carlyon *et al.*, 2003; Cusack *et al.*, 2004), how they are modified by attentional input. The results described here make a small contribution towards this endeavor by demonstrating that streaming based on pitch differences can occur solely as the result of binaural interactions.

ACKNOWLEDGMENTS

We wish to thank Dr. Armin Kohlrausch and three anonymous reviewers for their insightful comments on the manuscript. The Scottish Section of the IHR is co-funded by the Medical Research Council and the Chief Scientist's Office of the Scottish Executive Health Department.

[1]Although Blodgett *et al.*'s (1958) data were obtained for a signal frequency of 500 Hz, we assumed that they would not have differed substantially for the 400- or 800-Hz frequencies of our A tones.

[2]For this and the following statistical analyses, we applied the Huyhn-Feldt sphericity correction to account for the fact that the response in any bin will be unlikely to be independent of the response in the preceding bin. The effect of the Huyhn-Feldt correction is to reduce the effective degrees of freedom in the F-ratio test and to increase the p value for any given F; we report the corrected p values and the uncorrected degrees of freedom.

[3]Furthermore, it is worth noting that the amount of build-up was not significantly smaller for the Huggins-pitch stimuli than for the tones-in-noise stimuli, as one might expect if the former were simply a side-effect of the latter. The earlier four-way ANOVA showed that the interaction between stimulus type and time-in-sequence was not significant [$F(7,49)=0.66, p>0.5$].

Akeroyd, M. A., and Summerfield, A. Q. (**1999**). "A binaural analog of gap detection," J. Acoust. Soc. Am. **105**, 2807–2820.

Akeroyd, M. A., and Summerfield, A. Q. (**2000**). "The lateralization of simple dichotic pitches," J. Acoust. Soc. Am. **108**, 316–334.

Akeroyd, M. A., Moore, B. C. J., and Moore, G. A. (**2001**). "Melody recognition using three types of dichotic-pitch stimulus," J. Acoust. Soc. Am. **110**, 1498–1504.

Anstis, S., and Saida, S. (**1985**). "Adaptation to auditory streaming of frequency-modulated tones," J. Exp. Psychol. Hum. Percept. Perform. **11**, 257–272.

Beauvois, M. W., and Meddis, R. (**1991**). "A computer model of auditory stream segregation," Q. J. Exp. Psychol. A **43A**, 517–542.

Beauvois, M. W., and Meddis, R. (**1996**). "Computer simulation of auditory stream segregation in alternating-tone sequences," J. Acoust. Soc. Am. **99**, 2270–2280.

Blodgett, H. C., Jeffress, L. A., and Taylor, R. W. (**1958**). "Relation of masked threshold to signal-duration for various interaural phase-combinations," Am. J. Psychol. **71**, 283–290.
Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge).
Carlyon, R. P., Cusack, R., Foxton, J. M., and Robertson, I. H. (**2001**). "Effects of attention and unilateral neglect on auditory stream segregation," J. Exp. Psychol. Hum. Percept. Perform. **27**, 115–127.
Carlyon, R. P., Plack, C. J., Fantini, D. A., and Cusack, R. (**2003**). "Cross-modal and nonsensory influences on auditory streaming," Perception **32**, 1393–1402.
Carlyon, R. P., and Shackleton, T. M. (**1994**). "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms," J. Acoust. Soc. Am. **95**, 3541–3554.
Cramer, E. M., and Huggins, W. H. (**1958**). "Creation of pitch through binaural interaction," J. Acoust. Soc. Am. **30**, 413–417.
Culling, J. G., and Summerfield, A. Q. (**1998**). "Measurements of the binaural temporal window using a detection task," J. Acoust. Soc. Am. **103**, 3540–3553.
Culling, J. F., Summerfield, A. Q., and Marshall, D. H. (**1998a**). "Dichotic pitches as illusions of binaural unmasking. I. Huggins pitch and the binaural edge pitch," J. Acoust. Soc. Am. **103**, 3509–3526.
Cusack, R. P., Deeks, J., Aikman, G., and Carlyon, R. P. (**2004**). "Effects of location, frequency region, and time course of selective attention on auditory scene analysis," J. Exp. Psychol. Hum. Percept. Perform. **30**, 643–656.
Grantham, D. W., and Wightman, F. L. (**1978**). "Detectability of varying interaural temporal difference," J. Acoust. Soc. Am. **63**, 511–523.
Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., and Collet, L. (**2000**). "Influence of peripheral resolvability on the perceptual segregation of harmonic complex tones differing in fundamental frequency," J. Acoust. Soc. Am. **108**, 263–271.
Hartmann, W. M. (**1993**). "On the origin of the enlarged melodic octave," J. Acoust. Soc. Am. **93**, 3400–3409.
Hartmann, W. M., and Johnson, D. (**1991**). "Stream segregation and peripheral channelling," Music Percept. **9**, 155–184.
Hartmann, W. M., and Zhang, P. X. (**2003**). "Binaural models and the strength of dichotic pitches," J. Acoust. Soc. Am. **114**, 3317–3326.
Henning, G. B., and Wartini, S. (**1990**). "The effect of signal duration on frequency discrimination at low signal-to-noise ratios in different conditions of interaural phase,"Hear. Res. **48**, 201–207.
Houtsma, A. J. M., and Smurzynski, J. (**1990**). "Pitch identification and discrimination for complex tones with many harmonics," J. Acoust. Soc. Am. **87**, 304–310.
Kohlrausch, A., and Houtsma, A. J. M. (**1992**). "Pitch related to spectral edges of broadband signals," in *Processing of Complex Sounds by the Auditory System*, edited by R. P. Carlyon, C. J. Darwin, and I. J. Russell (Oxford University Press, Oxford).
Moore, B. C. J., and Gockel, H. (**2002**). "Factors influencing sequential stream segregation," Acust. Acta Acust. 88, 320–332.
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.
van Noorden, L. P. A. S. (1975). "Temporal coherence in the perception of tone sequences," Ph.D. thesis, Eindhoven University of Technology.
Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (**1999**). "The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task," J. Acoust. Soc. Am. **106**, 938–945.
Vliegen, J., and Oxenham, A. J. (**1999**). "Sequential stream segregation in the absence of spectral cues," J. Acoust. Soc. Am. **105**, 339–346.
White, L. J., and Plack, C. J. (**2003**). "Factors affecting the duration effect in pitch perception for unresolved complex tones," J. Acoust. Soc. Am. **114**, 3309–3316.
Zhang, P. X., and Hartmann, W. M. (**2004**). "Lateralization of the Huggins pitch," J. Acoust. Soc. Am. **115**, 2534.

Combining energetic and informational masking for speech identification

Gerald Kidd, Jr.,[a] Christine R. Mason, and Frederick J. Gallun
Department of Speech, Language and Hearing Sciences and Hearing Research Center, Sargent College, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

(Received 14 June 2004; revised 12 May 2005; accepted 17 May 2005)

This study examined combinations of energetic and informational maskers in speech identification. Speech targets and maskers (speech or noise) were processed and filtered into sets of 15 narrow frequency bands. The target was the sum of eight randomly selected bands. More masking occurred for speech maskers than for spectrally matched noise maskers regardless of whether the masker bands overlapped the target bands. The greater effect of the speech maskers was interpreted as due to informational masking. When the masker was comprised of nonoverlapping bands of speech, the addition of bands of noise overlapping the speech masker, but not the speech target, reduced the overall amount of masking. Surprisingly, presenting the noise to the ear contralateral to the target and masker produced an even greater release from masking. The contralateral noise was apparently sufficient to cause a slight change in the image of the ipsilateral speech masker, possibly pulling it away from the target enough to allow the focus of attention on the target. This finding is consistent with the interpretation that in some conditions small binaural differences may be sufficient to cause, or significantly strengthen, the perceptual segregation of sounds. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1953167]

PACS number(s): 43.66.Rq, 43.66.Dc, 43.66.Lj [AK] Pages: 982–992

I. INTRODUCTION

The purpose of this study was to examine the effect of combining energetic and informational maskers. Both energetic and informational masking may be present in many everyday listening situations but it can be difficult to ascertain the proportion or strength of each component, especially for complex, time-varying stimuli like speech. Even in the laboratory it is challenging to devise measurement procedures in which the two types of masking may be varied independently or combined in a controlled manner. However, it is important to understand the conditions that produce each type of masking and how the two interact if there is any hope of successfully predicting masking in situations where both are present or change over time. For example, do the effects of energetic and informational masking simply add?

The experiments described here examined combinations of these two types of masking in different ways. To varying degrees, the common theme that emerges is that it is very difficult to predict accurately how energetic and informational maskers interact based on intuitions drawn solely from studies of energetic masking. In this study, energetic and informational maskers were combined in a speech identification task. It extended earlier unpublished work using pure-tone targets and randomized multitone maskers in a detection task (Kidd and Mason, 1997). The results of that study indicated that, in certain conditions, adding an energetic masker to an informational masker may have a greater effect on the informational masker than on the target, thus reducing the overall amount of masking observed.

There are a few previous studies that are relevant to the issue above although they were not explicitly intended to examine combinations of energetic and informational maskers. Neff and Green (1987) and Oh and Lutfi (1998) have found that increasing the number of masker components, over a certain range of values, in a random-frequency multitone masker decreased the overall amount of masking observed.[1] Those results occurred presumably because adding masker components, thereby increasing the density of the spectra, changed the proportion of energetic-to-informational masking. In effect, this manipulation caused the masker components to interact with each other within an auditory filter increasing energetic masking but reducing informational masking. Overall masker power was held constant, so that as the number of components in the masker increased the level per component decreased. Oh and Lutfi (1998) proposed that the energetic-informational masking distinction, in those experiments, could be accounted for by the component-relative entropy (CoRE) model which is based on the statistical summation of the outputs of auditory filters across trials. When few masker components are present, the variance in the output of any given filter is expected to be relatively large. As the number of masker components increases, the likelihood of one or more components falling in a filter increases such that the expected across-trial variance in the output of the filter decreases. If it is assumed that informational masking is proportional to this variance, then it is possible to change the relative amounts of energetic and informational masking by changing the density of the random-frequency components in the masker. With respect to how energetic and informational masking interact, the underlying assumption was that the effects of the two types of masking simply add in decibels (see also Lutfi, 1990).

[a]Electronic mail: gkidd@bu.edu

Kidd *et al.* (2003b) measured the discriminability of multitone complexes based on how closely they matched an exact harmonic relationship. "Harmonicity" discrimination was masked by random-frequency multitone maskers in a manner similar to the pure-tone detection studies described above. As the number of multitone masker components increased, the ratio of energetic-to-informational masking (EM/IM) also increased. Kidd *et al.* found that the effectiveness of a manipulation intended to perceptually segregate target and masker (dichotic presentation) decreased as the EM/IM increased. Thus, in that study, changing EM/IM affected both the overall amount of masking observed and the usefulness of perceptual cues in overcoming the informational component.

In a study more directly related to the current experiments, Neff and Jesteadt (1996) reported conditions in which energetic and informational maskers were directly combined. The informational masker was a set of random-frequency tones (similar to the studies discussed above) presented outside of a protected region around the signal frequency. The energetic masker was a pure tone having the same frequency as the target, so that the in-phase addition of the target and pure-tone masker created an increase in level *re* masker alone. The unincremented masker component was presented in the nonsignal interval. Thus, this was a pure-tone intensity discrimination task in the presence of random-frequency multitone maskers. When the amount of masking was measured for each type of masker separately and compared to the masking produced when the two were combined, a model like that which is often applied to combinations of two or more energetic maskers (e.g., Penner, 1980; Humes *et al.*, 1992) could successfully predict the results. In these models, the prediction is that the amount of masking produced when maskers are combined is equal to, or greater than, the masking produced by the more effective masker in isolation.

In the current experiments, which employ stimuli and procedures very similar to those described by Arbogast *et al.* (2002), speech targets and speech and noise maskers were used that were processed into sets of very narrow frequency bands. Subsets of the bands for target and masker were presented on each stimulus interval. The frequency bands comprising the masker could overlap the target bands exactly (same-band masker), maximizing energetic masking, or could be mutually exclusive with the target bands (different-band masker), thereby minimizing energetic masking. The speech/noise contrast, in the context of the speech identification task employed, allowed for different strengths of informational masking to be added to the same-band versus different-band distinction. The results are organized into three sections emphasizing specific comparisons of interest. In the first section, all four combinations of same-band and different-band presentation for both speech and noise maskers were systematically examined. The goal was to determine how much additional masking may be produced by an informational masker than is produced by a spectrally matched predominantly energetic masker. Of particular interest was the increase in masking attributable to the informational component when the energetic component was very low (different band) versus very high (same band). In the second section, the effect of combining energetic and informational maskers was assessed for several relative levels of the two maskers. The intent was to examine the issue of additivity of masking directly by measuring the effect of each masker separately and then comparing that to the conditions where the two were combined. Finally, in the third section the combination of a different-band informational masker (speech) with a different-band energetic masker (noise) was examined—as above—but rather than presenting both in the same ear as the target, the energetic masker was presented to the contralateral ear. Our expectation was that the effect of the energetic masker on the informational masker would be eliminated by dichotic presentation, thereby restoring the full masking effectiveness of the ipsilateral informational masker. However, as discussed below, that expectation was not borne out by the results.

II. METHODS

A. Subjects

The listeners were three female graduate students ranging in age from 21 to 25 years. Routine audiometric examination indicated that all three listeners had normal hearing. The listeners were paid for their participation in the experiment. Although the time course of participation and number of sessions varied somewhat for the three listeners, they typically participated in three 2-h sessions per week over the course of approximately 9 weeks.

B. Stimuli

The stimuli and methods were similar to those used by Arbogast *et al.* (2002) and employed the coordinate response measure ("CRM;" Bolia *et al.*, 2000) materials for target and speech maskers. Each CRM sentence has the structure "Ready [callsign] go to [color] [number] now" with the task being to identify the color and number associated with a specified callsign. A detailed description of the stimulus generation procedure is given in Arbogast *et al.* (2002). The targets and maskers were processed into 15 narrow frequency bands using a modified version of cochlear implant simulation software (Shannon *et al.*, 1995). The bands had center frequencies that ranged from 215 to 4891 Hz spaced at a ratio of 1.25:1 and were each approximately 1/3 octave wide. The envelopes were then extracted from each band, low-pass filtered at 50 Hz, and used to modulate pure-tone carriers having frequencies equal to the center of each of the bands. This resulted in a set of 15 extremely narrow nonoverlapping bands spaced equally in logarithmic frequency.[2] For the targets, 8 of the 15 bands were randomly selected and summed for presentation on each observation interval. A recent study by Brungart *et al.* (2005) has found that as few as five bands of speech generated in this manner will yield identification performance near 100% correct in quiet. There were two types of maskers—speech and noise—that were also processed into the same 15 narrow bands. The speech maskers were processed in exactly the same way as the targets. Both speech and noise maskers were comprised of either the same 8 bands as the target on a given trial (called same-band speech, SBS, and same-band noise, SBN, respec-

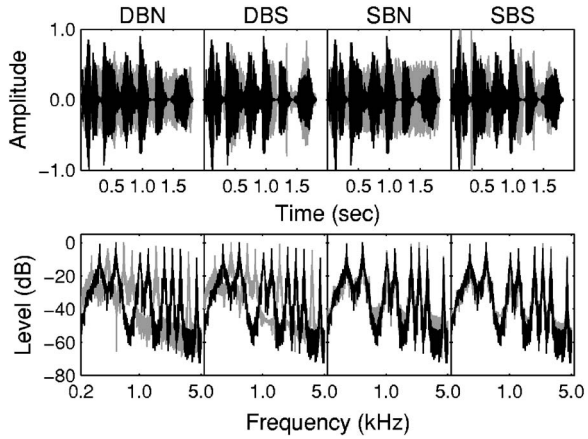


FIG. 1. Examples of the stimuli used in this study. The top row shows waveforms of a target sentence "Ready Baron go to White 8 now" (black) paired with four versions of a masker sentence "Ready Charlie go to Red 2 now" (gray). The bottom row plots the magnitude spectra of the target and maskers directly above each. The four types of maskers are different-band speech (DBS), same-band speech (SBS), different-band noise (DBN), and same-band noise (SBN). Details are provided in the text and in Arbogast *et al.* (2002).

tively) overlapping the target bands completely, or were six mutually exclusive bands chosen randomly from the remaining seven bands (called different-band speech, DBS, and different-band noise, DBN, respectively). The SBS and DBS maskers were CRM sentences from a different talker than the target uttering another phrase of the same structure, but having a different callsign (e.g., "Charlie" or "Ringo," etc.), color, and number than the target (which always had the callsign "Baron"). The targets and maskers were presented simultaneously so that the callsigns, colors, and numbers roughly coincided [the speaking rates for different sentences differed somewhat so the extent to which words overlapped in time varied; cf. Bolia *et al.*, (2000)]. This technique thus yielded two sources of speech—target and SBS/DBS maskers—that were each highly intelligible in isolation but consisted of spectrally overlapping (SBS) or nonoverlapping (DBS) bands.

To create the SBN and DBN maskers, the processed speech maskers described above were multiplied in the frequency domain with a Gaussian noise and inverse Fourier transformed prior to presentation. This yielded noise stimuli that had spectral shapes nearly identical to the multiband speech maskers from which they were derived but were completely unintelligible. Thus, each masker sample had a speech and noise version so that either could be added to the same target sentence (see Arbogast *et al.*, 2002, for details). Examples of the target and maskers are shown in Fig. 1 as waveforms and magnitude spectra.

C. Procedures

The task of the listener was to identify the color and number from the sentence having the callsign "Baron." A response was counted correct only if both the color and the number were reported accurately. Chance performance was thus about 3% (4 colors by 8 numbers). Response feedback was given after every trial.

All of the stimuli were stored on a computer and played through Tucker-Davis Technology (TDT) 16-bit digital-to-analog converters at a rate of 50 kHz, then low-pass filtered at 7.5 kHz. Stimulus level was controlled separately for targets and maskers by programmable attenuators (TDT PA4). The listeners were seated in individual double-walled IAC booths. The stimuli were presented through matched and calibrated TDH-50 earphones. Response feedback and interval timing information was displayed on LCD monitors and responses were entered on a computer keyboard. The target was always presented to the listener's right ear at a level of 60 dB SPL. When a masker was presented to the same ear as the target it is referred to as "ipsilateral" and when a masker was presented to the ear opposite the target it is referred to as "contralateral."

The listeners were initially tested in quiet to assure that identification performance was near 100% correct at the target level of 60 dB SPL. The various masked conditions tested were then intermixed and presented in a different random order for each subject. The results are presented as three types of comparisons, each focusing on a particular issue. Additional procedural details concerning the conditions upon which each comparison is based are given in the corresponding sections below. For the DBN masker conditions, where performance was nearly perfect, each data point reflects roughly 50 to 100 trials per subject. For the other conditions, each data point represents at least 100 trials per subject with most conditions averaging about 150 trials per subject per point.

III. RESULTS

A. Ipsilateral speech and noise maskers presented separately

The results described in this section are a comparison of same-band and different-band masking for noise and speech maskers presented ipsilateral to the target. There were four combinations of band placement (same or different) and masker type (speech or noise): DBN, DBS, SBN, and SBS. Three of these maskers were also used by Arbogast *et al.* (2002). The targets and maskers were all presented monaurally to the listener's right ear. The maskers were presented at 50, 60, and 70 dB SPL corresponding to 10-, 0-, and −10-dB target-to-masker ratios (T/M), respectively (note that the different-band maskers were comprised of only six bands and hence were approximately 1.25 dB higher in level per band than in the same-band maskers).

The results are shown in Fig. 2 for all three listeners (data points) along with group means (solid lines) and expected chance performance (horizontal dotted line at ~3%). Each panel contains the proportion of correct responses as masker level was increased for one masker type. The three subjects generally performed quite similarly, so the description of the results will focus on group-mean data. First, there was a decrease in identification performance with increasing masker level that was apparent for all four maskers. Also, there was a clear ordering of the amount of masking produced,[3] at least for the two higher masker levels, with the greatest amount of masking found for the SBS masker fol-

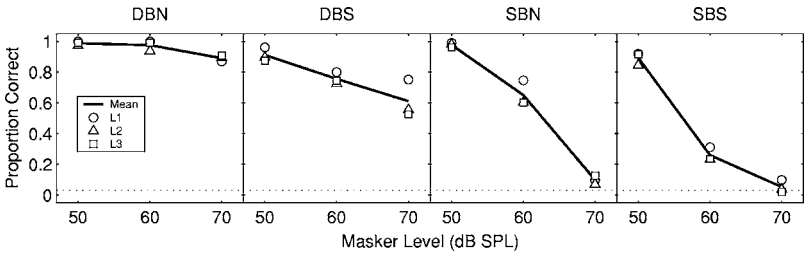


FIG. 2. Individual and group mean results from Sec. III A. The data points are the results from listeners 1–3 with the solid lines connecting group means. The dotted line at the bottom indicates chance performance. Each panel contains the results from a different masker (left to right): different-band noise (DBN), different-band speech (DBS), same-band noise (SBN), and same-band speech (SBS).

lowed by SBN, DBS, and DBN. A two-way within-subjects analysis of variance indicated that both main factors, masker type $[F(3,6)=227.2, p<0.001]$ and masker level $[F(2,4)=1327.42, p<0.001]$, were significant as was the two-way interaction $[F(6,12)=75.58, p<0.001]$.[4] Thus, as is apparent in Fig. 2, the effect of masker level depended on which masker was tested.

For the two different-band maskers, DBN and DBS (left two panels of Figure 2), performance was poorer when the masker was speech than when the masker was noise (a *post-hoc* simple contrast reveals a significant difference between the DBN and DBS maskers with $p=0.037$). The DBN masker produced very little masking at all, with group mean proportion correct identification performance at 0.89 even for the highest masker level. For the DBS masker, performance dropped from a proportion correct of about 0.91 at a masker level of 50 dB SPL to about 0.61 at a masker level of 70 dB. Because these two maskers were equated in spectral content and level, the expectation was that the energetic masking they produced should be about the same. This consideration does not take into account temporal differences between speech and noise maskers, which is discussed more fully below and in the Appendix. However, this difference in effectiveness between DBN and DBS maskers is taken as an indication of how much additional masking (presumably informational masking) was produced beyond the small amounts of energetic masking that may have occurred.

For the two same-band maskers, SBN and SBS, a similar comparison may be made. Here, though, both maskers likely caused large amounts of energetic masking because they both overlapped the target bands in the frequency domain almost exactly. For the SBN masker, it was assumed that the masking that was produced was primarily energetic in nature. It is clear from Fig. 2 that performance was significantly degraded as the SBN masker level was increased with the proportion correct identification decreasing from about 0.98 at a masker level of 50 dB (indicating no effect of the masker at a +10 dB T/M) to near chance performance of 0.10 at a masker level of 70 dB (-10 dB T/M). Arbogast *et al.* (2002) also found that the SBN masker produced significant amounts of masking. For the SBS masker, which was not tested in the Arbogast *et al.* study, the masking produced was only slightly greater than that produced by the SBN masker at the high (0.89) and low (0.05) ends of the functions. However, in the middle of the range (0 dB T/M) performance was about 0.39 lower for the SBS masker than for the SBN masker. As was the case for the different-band maskers, it seems likely that the greater masking produced by the speech masker than by the noise masker was due to the added informational masking it caused. At the 70-dB masker level, the energetic masking alone for either masker was probably sufficient to drive performance to near chance, so the additional informational masking effect of the SBS masker was not apparent, that is, a performance "floor" likely caused the masking to be about the same for SBN and SBS. Therefore, in this case, what we conclude is that the additional informational masking caused by the speech masker, compared to the noise masker, was only apparent at the 0 dB T/M and would not have been obvious had only the extreme values been tested.

The main difference between speech and noise maskers that is not accounted for by spectral matching and rms equalization is a difference in the amplitude envelopes. The speech bands have low-amplitude phonemes and pauses corresponding to boundaries between syllables and words, where the envelopes are at minima. Equating the maskers by rms produces speech envelopes that have higher peaks than the noise to compensate for these minima (cf. Fig. 1) and, of course, the distributions of envelope frequencies are slightly different as well. Ordinarily, the expectation about the difference in energetic masking between broadband speech and noise maskers has to do with the listener having a greater opportunity to hear the speech target during the envelope minima of the speech masker fluctuations, decreasing masking relative to the more constant noise envelope. That expected result is not always obtained, depending on the stimuli and how the task is structured, and speech may sometimes produce more masking than anticipated purely from spectrotemporal overlap [e.g., "perceptual masking" (Carhart *et al.*, 1968)].

In the current conditions, however, the target and speech masker have the same sentence structure and are, to some degree, temporally aligned so that the colors and numbers (which are the only words that count toward the identification score) from both sources tend to occur at about the same time reducing (but not eliminating) instances where crucial information from the target is available while the masker is at an envelope minimum. It should be pointed out, though, that the listener still must follow the target talker from the time that the callsign "Baron" is uttered until the test words occur. Therefore, even though the identification score is based on the color and number, the other words in the target sentence may be crucial, too. It is possible that the dips/peaks that occur for the nontest words in the sentences may be useful because they provide the continuity of cues that allows the listener to follow the target talker from the callsign to the test items. Thus, although the SBS and SBN maskers are spectrally matched, it is possible that they produce different amounts of energetic masking due to differences in the fluctuations of their temporal envelopes. This

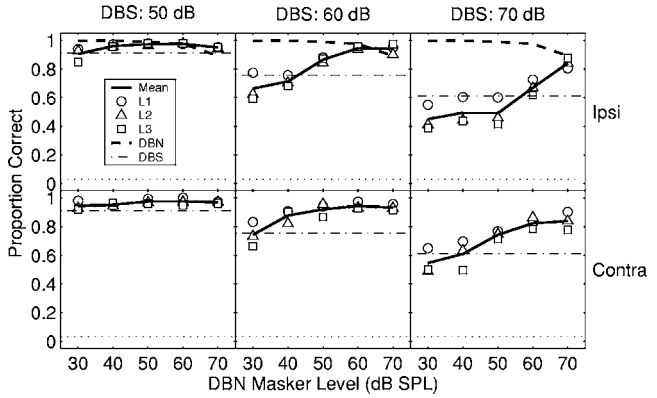


FIG. 3. Individual and group mean results from the comparisons in Sec. III B (upper row) and Sec. III C (lower row). The data points show results from individual subjects and the lines connect group averages. For each row, the panels contain data for 50, 60, and 70 dB (left to right) levels of the DBS masker. For each panel, the abscissa is the level of the added DBN masker. For the upper row, the DBN masker was added to the ipsilateral ear while the lower row is for the condition where the DBN masker was added to the contralateral ear. The dotted line at the bottom indicates chance performance. The dashed-dotted line in each panel indicates group mean performance for the DBS masker alone while the dashed lines along the top of each upper-row panel show group mean performance for the DBN masker alone in the ipsilateral ear.

issue is considered in more detail in the Appendix. The conclusion from these data and that analysis, though, is that the greater masking produced by the SBS masker than the SBN masker at 0 dB T/M was due to a difference in informational, rather than energetic, masking.

B. Ipsilateral speech and noise maskers presented in combination

This section provides a comparison of conditions in which the effects of two maskers—one highly informational (speech) and the other highly energetic (noise)—were measured separately and in combination. All of the maskers were presented ipsilateral to the target in the following combinations: target plus DBS masker only, target plus DBN masker only, and target plus *both* DBS and DBN maskers (denoted DBN+DBS). There were three levels of the DBS masker: 50, 60, and 70 dB SPL. The level of the DBS masker was randomly chosen from among those three levels on every trial. The level of the DBN masker was chosen randomly from a set of levels at 10-dB intervals from 30 to 70 dB SPL. Note that the DBS and DBN maskers in the combined DBN +DBS condition have different bands from the target but occupy the same bands as each other.

The results from individual subjects (data points) and group means (solid lines) are shown in the top three panels of Fig. 3. The abscissa is the level of the DBN masker while the three panels show identification performance at the 50-, 60-, and 70-dB levels of the DBS masker (left to right), respectively. The horizontal dashed-dotted lines indicate group-mean performance for the three levels (separate panels) of the DBS-only masker (also shown in Fig. 2). The dashed lines at the very top of the graph show performance for the different levels of the DBN-only masker (shown in Fig. 2) and replotted in each panel, with two lower levels included).

As noted from the previous section, and as may be seen again in the top three panels of Fig. 3, increasing the level of the DBN-only masker hardly affected identification performance at all whereas increasing the level of the DBS-only masker (dashed-dotted horizontal lines) caused discrimination performance to drop significantly. In addition, when only one masker was present, the DBS masker was always more effective than the DBN masker. The combination of the two might be expected to be at least as effective as the more effective of the pair. Performance for the DBN+DBS masker that falls below the horizontal DBS-only lines would indicate more masking than either alone (additive) whereas performance above that line would indicate less masking than the more effective of the pair (nonadditive).

The main finding apparent in Fig. 3 is that adding the DBN masker to the DBS masker had a complex effect on the overall amount of masking produced. A two-way repeated-measures analysis of variance on the DBN+DBS data found that both main factors, DBS level $[F(2,4)=129.89, p<0.001]$ and DBN level $[F(4,8)=22.67, p<0.001]$, were statistically significant as was the interaction $[F(8,16)=21.03, p<0.001]$. For lower levels of the DBN masker, performance was poorer than for the DBS masker alone—an additive masking effect. However, as the level of the DBN masker increased, performance improved such that, at the highest noise levels, percent correct identification was much better than for the DBS masker alone—a clearly *non*additive effect. For example, at the highest level of the DBN masker and the DBS masker (both at 70 dB, in top right panel), identification performance was about 23 percentage points better than for the 70-dB DBS masker alone. At the lowest DBS level, performance is too high to show much of an improvement due to adding the DBN masker (i.e., a ceiling effect). The explanation for the increase in identification performance as the level of the DBN masker increases seems straightforward: the largely energetic DBN masker is "masking the masker," thereby decreasing the effectiveness of the DBS informational masker without adversely affecting the target. This is not too different from what might be expected if the level of the DBS masker were simply decreased. It is also consistent with the earlier report by Kidd and Mason (1997) for detecting a tone masked by a random-frequency multitone masker combined with a notched-filtered Gaussian noise.

In fact, the more difficult result to explain is the initial decrease in overall performance at the lower levels of the DBN masker given that performance for those levels of the DBN masker was nearly perfect when presented alone. The largest "additive" effect thus occurred for the greatest level disparity between maskers. Consider, for example, the 30-dB level for the DBN masker. When the DBS masker was at 50 dB (top left panel), the addition of a 30-dB DBN masker decreased the proportion correct by 0.02, relative to the DBS-only condition. When the DBS masker was raised in level to 60 dB (top middle panel), the 30-dB DBN masker had its *greater* additive masking effect—identification performance decreased by about 0.09. Lastly, when the level of the DBS masker was raised even higher to 70 dB SPL (top right panel), the 30-dB DBN masker had its greatest "addi-

tive" masking effect, decreasing proportion correct identification performance by an average of 0.16. Higher levels of the DBN masker produced smaller additive masking effects. So, paradoxically, within the range of values tested here, the greater the difference in level between the two maskers the more they interacted to increase masking (keeping in mind the combinations causing *non*additive effects discussed above).[5] Inspection of the individual data reveals that this effect occurred for all three subjects although it was smaller for L1 than for L2 and L3. Paired samples t tests comparing DBS alone to DBS+DBN for the lower levels of DBN were not significant when DBS was 50 or 60 dB but were significant when the DBS masker was 70 dB and the DBN masker was 30, 40, or 50 dB [$t(2)=10.6$, 9.7, and 7.5 with $p=0.009$, 0.012, and 0.017, respectively]. At present, we have no satisfactory explanation for this small but consistent finding.

C. Combinations of ipsilateral and contralateral speech and noise maskers

The results discussed above indicated that, under appropriate conditions, an energetic masker may interact with an informational masker such that the overall amount of masking is reduced relative to the informational masker alone. In this section, the results of the ipsilateral combination of energetic and informational maskers described above are compared to a condition in which the DBN masker was presented in the ear contralateral to both the target and the DBS masker. Because the explanation for the decrease in the effectiveness of the DBS masker when the DBN masker was added at sufficiently high levels was that the DBN masker energetically masked the DBS masker, it was anticipated that contralateral presentation of the DBN masker would restore the effectiveness of the DBS masker.

All 15 combinations of the three DBS masker levels (50, 60, and 70 dB SPL) and five DBN masker levels (30 to 70 dB SPL in 10-dB steps) were tested with the DBS masker presented to the ipsilateral ear and the DBN masker presented to the contralateral ear. These conditions are therefore identical to those shown in the top three panels of Fig. 3 with the exception of moving the noise masker to the opposite ear. The results are shown in the lower panels of Fig. 3. Surprisingly, the functions plotted there are qualitatively quite similar to those obtained when the DBN masker was presented ipsilaterally (upper panels). If the DBN masker presented to the opposite ear had no effect (as was the case when it was presented alone), the functions should fall on the horizontal lines representing performance for DBS only. In most cases, presenting the DBN masker in the contralateral ear improved performance relative to the reference condition of ipsilateral DBS masker alone.

Comparison across the upper and lower rows of Fig. 3 indicates that group mean performance was usually *better* when the DBN masker was presented contralaterally than when it was presented ipsilaterally. This trend was apparent to varying degrees for all three subjects. Figure 4 illustrates this point. The group mean differences in performance between ipsilateral and contralateral presentation of the DBN masker are plotted for all of the combinations of DBS and DBN levels. Inspection of Fig. 4 reveals that contralateral presentation of the DBN masker did indeed result in a greater release from masking than did ipsilateral presentation at most level combinations (i.e., identification performance improved for contralateral versus ipsilateral presentation shown as a positive difference in proportion correct). There was no convincing evidence for a low-level additive masking effect for contralateral presentation. As the level of the DBS masker increased, the size of the difference increased as well and tended to shift toward the higher DBN levels with a maximum difference greater than 0.25 found for DBS=70 dB SPL and DBN=50 dB SPL (note that the largest advantage of adding the DBN masker was always at the highest DBN level of 70 dB SPL, as shown in Fig. 3, but the greatest asymmetry between ipsilateral and contralateral advantage occurred at lower DBN levels). A three-way repeated-measures ANOVA was performed on these data.[6] The three factors were (ipsilateral) DBS level, DBN level, and DBN ear of presentation (ipsilateral or contralateral). All three main effects were significant: DBS level [$F(2,4)=96.6, p<0.001$], DBN level [$F(4,8)=34.8, p<0.001$], and ear of presentation of the DBN masker [$F(1,2)=66.4, p=0.015$]. Furthermore, all of the two-way interactions, and the three-way interaction, were reliable as well. These results are consistent with the interpretation that, while the addition of the DBN masker in either ear significantly affected performance, and did so in a level-dependent manner, the greater effect occurred for contralateral presentation.

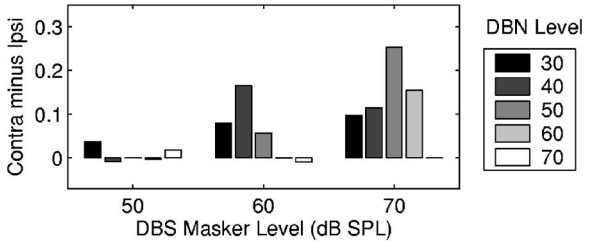


FIG. 4. The difference in speech identification score between ipsilateral and contralateral presentation of the DBN masker. The level of the DBS masker (50, 60, and 70 dB) forms the major divisions along the abscissa while the levels of the added DBN masker (30–70 dB SPL) are indicated for each DBS level. A positive difference means that speech identification performance was better when the DBN masker was presented to the contralateral ear than when it was presented to the ipsilateral ear (see text).

IV. DISCUSSION

The first point concerns the degree to which the various maskers produced energetic or informational masking, or both. Consistent with a large body of prior work on informational masking, and specifically in agreement with the interpretation of Arbogast *et al.* (2002) using many of these same stimuli and conditions, it seems very likely that the two noise maskers (DBN and SBN) produced primarily energetic masking. The large difference in the amount of masking caused by these two maskers—greater than 80 percentage points in identification scores at the highest level—suggests that the DBN masker produced relatively little energetic masking, whereas the SBN masker produced considerable

energetic masking. However, the conclusion about the nature of the masking caused by the noise maskers must be tempered by two observations. First, the large difference in the effectiveness of the DBN and SBN maskers does not mean that there are no interactions among bands in the peripheral filters between DBN and the target; quite likely there are. However, the DBN bands are sufficiently separated from the target bands that the interactions must occur many decibels down from the peaks of the bands (cf. endnote 2) and were generally not strong enough to interfere with speech reception. The difference between the two maskers essentially is dependent on the frequency selectivity of the auditory system (cf. Arbogast *et al.*, 2005). And, second, even for the DBN masker, there was a great deal of frequency uncertainty in the stimulus because the bands comprising the speech target and the masker were chosen randomly on every trial. If one listens only to the target or only to the masker, there are clearly large sample-to-sample variations in sound quality. However, here, as in past studies using this masker (e.g., Arbogast *et al.*, 2002, 2005; Kidd *et al.*, 2005), little masking occurs despite the high masker uncertainty. Therefore, trial by trial frequency uncertainty *per se* is not the critical variable.

For the two speech maskers, both presumably produce large amounts of informational masking while the SBS masker also produces substantial amounts of energetic masking. To the extent that the noise maskers serve as controls for the amount of energetic masking produced by the speech maskers, the additional masking produced by the speech maskers may be attributed to informational masking. The maximum group mean difference in performance between the DBN and DBS maskers was about 30 percentage points (for the 70-dB masker level, −10 dB T/M) and between the SBN and SBS maskers was about 40 percentage points (60-dB masker level, 0 dB T/M). The difference in T/M in the middle portions of the functions was about 8 dB for SBS vs. SBN. The same computation could not be made for DBS and DBN because the DBN function, in particular, only declined to about 90% correct at the highest masker level. However, Arbogast *et al.* (2002) reported about 22 dB less masking for the DBN masker than for the DBS masker computed in the middle portion of the performance-level functions.

In a recent study that bears on the current work, Qin and Oxenham (2003) measured speech recognition for natural speech and noise-excited cochlear-implant simulation speech (which is similar to our stimuli in many respects[7]) as the number of channels was varied. Their targets were masked by speech-shaped noise or by speech from other talkers that was either natural speech or was processed into *N* contiguous bands of cochlear implant simulation speech where *N* varied from 4 to 24. They found that, for the natural speech targets, masked "thresholds" (T/Ms at 50% correct performance obtained from psychometric functions) were lower in the speech maskers than in equal rms noise. However, the reverse was true for the cochlear implant simulation speech targets and maskers. Their interpretation of this result was that the degree to which fundamental frequency/intonation information was available determined the effectiveness of the speech masker. In natural speech, differences in fundamental frequency and intonation contours can provide a means for perceptually segregating two talkers. For cochlear implant simulation speech, most, if not all, of that information is lost making the segregation task more difficult and requiring a higher T/M to achieve criterion performance. The same appears to be true in the current conditions. The lack of intonation patterns for use in segregating talkers, as well as the semantic content of the speech masker, caused much greater masking in many corresponding conditions for speech maskers than for noise maskers. The situation is somewhat more complicated in the current conditions, though. For the target combined with a DBS masker, the listener can potentially use differences in sound quality or timbre created by the two sets of nonoverlapping frequency bands in a way analogous to using the natural differences in source characteristics to segregate talkers. This assumes that the listener is able to extract some aspect of sound quality (i.e., from the set of bands chosen for the target on that trial) when the callsign is presented and use that cue to follow the target over time until the test words occur. The listener may also try to ignore the DBS masker which is comprised of a different set of bands and presumably has a distinctively different quality. If the bands overlap exactly, as with the current SBS maskers and the stimuli used by Qin and Oxenham (2003), that qualitative difference would be reduced or eliminated in addition to the loss of intonation cues.

Next, with respect to the interaction among simultaneously presented ipsilateral maskers, it is quite clear that, in certain conditions, combining energetic and informational maskers may actually reduce the overall amount of masking observed relative to the informational masker alone. Also, it is clear that performance can improve as the level of the (energetic) masker increases. This nonadditivity of masking is possible because informational masking is often produced at frequencies remote from the frequencies comprising the target. Thus, interactions among energetic and informational maskers may occur without directly affecting the target. Note that the greater masking produced by SBS than by SBN indicates that substantial informational masking may *also* occur when the target and masker exactly overlap in frequency. Adding an energetic masker to an informational masker thus may have a much greater effect on the informational masker than it does on the target.

It should also be acknowledged, as discussed in the Introduction, that other conditions in which energetic and informational maskers are combined may have a different effect and masking may increase relative to either masker alone, as demonstrated by Neff and Jesteadt (1996). These findings reveal the difficulty in attempting to arrive at a comprehensive model of masking that incorporates both energetic and informational components because the interactions among the maskers, as well as the effects on the target, must be taken into account. Furthermore, very little masking of any sort occurred for the DBN masker even though there is a high degree of uncertainty—in the form of sample-to-sample spectral variability—in that masker condition. Presumably, if the task were structured differently—for example, if the listener were required to detect a target that was highly similar

to the DBN masker, such as a narrow-band noise presented at a known frequency, substantial informational masking almost certainly would occur. That hypothetical stimulus set and task are somewhat similar to those used in the multitone masking paradigm by Neff and Green (1987) and several others (e.g., Oh and Lutfi, 1998; Wright and Saberi, 1999; Richards *et al.*, 2002; Durlach *et al.* 2003a), which is known to produce large amounts of informational masking in many listeners. Thus, the informational masking produced by a given masker depends on its inherent statistical variability *and* its similarity to the target in the context of the task required of the listener (cf. Kidd *et al.*, 2002; Durlach *et al.*, 2003b).

The explanation for the release from masking that occurred when the DBN masker was added contralateral to the DBS masker is very different than that advanced above to account for the reduction in masking that occurred when both maskers were presented to the same ear. When the maskers and target are in the same (ipsilateral) ear, the DBN masker energetically masks the DBS masker. In that case the intelligibility—or perhaps simply the saliency—of the DBS masker was reduced, rendering it a less effective informational masker. When the DBN masker was moved to the contralateral ear, a similar, but even stronger, influence on the ipsilateral DBS masker was observed. Despite the similarity of the effect of the DBN masker in ipsilateral and contralateral ears, the mechanisms seem likely to be different because energetic masking is generally thought of as occurring in the cochlea and auditory nerve. Dichotic presentation of stimuli would usually be considered sufficient to eliminate interactions based on overlapping patterns of excitation. It is possible, of course, that the inputs from the two ears are combined at the first site of binaural interaction, or at a higher site, in a way that is analogous to overlapping patterns of excitation in the cochlea. Durlach *et al.* (2003a) point out that the energetic-informational distinction may change or be different at different physiological locations from periphery to cortex and, to the extent possible, the "physiological vantage point" at which the distinction is made must be considered. Generally, though, the expectation about informational masking is that it occurs despite there being a physiological representation of the target that is sufficient for an ideal observer to solve the task and that presenting target and masker to separate ears satisfies that criterion. This condition is somewhat unusual in the context of discussions about energetic/informational masking because the "masking" is of one masker by another, making it disadvantageous for the listener to ignore the masker.

A recent study by Brungart and Simpson (2005) has also found that, in certain conditions, a masker may be more effective when presented contralateral to the target than when it is presented in the same ear. As in the current experiment, the task was speech recognition using the CRM materials and test procedures, and there were two maskers present in addition to the target. However, their "reverse cocktail party effect" occurred when both maskers were speech and only when one masker was much lower in level than the target and other masker. In that case, moving the low-level speech masker to the contralateral ear increased the total amount of masking observed, probably because the low-level masker was much easier to hear in the contralateral ear and its informational masking effect was thus stronger. Although we cannot rule out a similar effect here, it seems unlikely given that the DBN masker produces primarily energetic masking. Also, for the reasons discussed below, the binaural interaction between maskers appears to be crucial in producing the contralateral effect found in this study whereas little binaural interaction would be expected for the two independent speech sources used in Brungart and Simpson's (2005) experiment.

The stimulus configuration examined in Sec. III C. (signal plus DBS masker in one ear and DBN masker in the opposite ear) is somewhat like the masking-level difference (MLD) condition of signal-monaural masker-monaural (S_mM_m) versus signal-monaural masker-uncorrelated (S_mM_u, or partially correlated, S_mM_r). It has long been known that the magnitude of the binaural advantage for that stimulus configuration depends on the interaural correlation of the masker (e.g., Wilbanks and Whitmore, 1968). However, for a "typical" MLD condition, the binaural release from masking is largely a release from energetic masking as for the case of detecting a tone in noise. In the current study, the overlap between the masker bands and the target bands is minimized, so the masking is, we believe, primarily informational masking and the binaural condition tested presumably provides a release from informational masking.

The finding that contralateral presentation of the DBN masker improved performance (relative to ipsilateral presentation), rather than degrading it, was completely opposite of what we had expected. Instead of the DBN masker losing its ability to reduce the effectiveness of the DBS masker, it appears to have exerted a even greater reduction in the effectiveness of the DBS masker when presented to the opposite ear. One possible explanation for this unexpected finding has to do with the inherent difference between the truly monotic target and the dichotic combination of the ipsilateral DBS masker and the contralateral DBN masker. It seems possible that contralateral presentation of the DBN masker influenced the quality of the ipsilateral DBS masker image, essentially "binauralizing" it sufficiently to sound perceptually distinct from the truly monaural target. Thus, the advantage of adding the contralateral noise in narrow frequency bands corresponding to the DBS bands may be due to a qualitative difference between a target that stimulates only monaural channels and a binaural masker that, despite the uncorrelated nature of the narrow bands in the two ears, nonetheless stimulates binaural channels.

Several studies have reported that the contralateral presentation of noise may influence the perceived location of speech presented in the opposite ear. Warren and Bashford (1976) refer to this phenomenon as "contralateral induction" (see also Warren, 1999). They demonstrated that alternately switching the simultaneous presentation of speech to one ear and noise to the other ear influenced the perceived location of the speech—drawing it away from the ear of presentation—more than the noise. Contralateral induction occurred, they hypothesized, if the inducing (contralateral) sound overlapped the spectrum of the target so that the target

could plausibly have been present, but masked, in the contralateral ear. Kidd *et al.* (1994) found that the informational masking produced in an ipsilateral detection task could be reduced significantly simply by adding an exact copy of the masker to the opposite ear (masker diotic). Their interpretation of that result was that moving the masker image away from the monotic target toward the middle of the head improved the ability of the observer to focus attention on the target, thereby greatly reducing the informational masking caused by the masker. Similar findings have been reported in other informational masking tasks (e.g., Kidd *et al.*, 2003a; Durlach *et al.*, 2003b).

With respect to the present findings, consider first what happens to any single band of the two maskers. Because the envelopes of the bands in the two ears are uncorrelated and time varying, the interaural time and level differences also vary over time, sometimes favoring one ear and sometimes favoring the other ear. The very narrow-band nature of the sounds, though, may foster some degree of fusion across ears. It is well known that differences in interaural envelopes can provide a basis for lateralization even when the fine structure is not correlated between ears or occurs at high frequencies where the ability to code fine structure diminishes (e.g., Henning, 1974; McFadden and Pasanen, 1976; Nuetzel and Hafter, 1977; Bernstein and Trahiotis, 1992). Here, the interaural envelope difference function would be different for each frequency band. Summed across bands, the combined ipsilateral DBS-contralateral DBN masker creates a perceptually complex image having a binaural quality and an indistinct location that may be different from the target. Thus, both a qualitative difference and a shift in masker image location may underlie the observed contralateral advantage.

One final point concerns the possible relation between the current findings and several recent studies examining "contralateral masking effects" (CME). Brungart and Simpson (2002) reported that listeners were unable to ignore a speech masker in the ear contralateral to a speech target if a second, unrelated speech masker was also present in the target ear. This CME was interpreted as resulting from a limitation on attentional resources: the listener was unable to ignore a contralateral speech signal when engaged in a speech recognition task in the ipsilateral ear that required segregating two talkers and attending to one of them. The contralateral talker placed too great a load on the capacity of the observer to both attend to the target and ignore two similar competing talkers. Thus, the normally high degree of binaural channel separation broke down and an increase in masking occurred. A similar conclusion regarding the effect of limited attentional capacity in monitoring multiple sources presented dichotically was reached by Kidd *et al.* (2003a) for nonspeech sounds. In that study, a significant increase in masking was found when the listener had to segregate a target tone from a masker in one ear and an informational masker was added to the opposite ear. In both of these cases, the CME that was observed indicated that the listeners were unable to ignore the irrelevant stimulus in the nontest ear even though it adversely affected performance. Several other studies have also demonstrated that listeners are often unable to ignore irrelevant information presented contralaterally to a target for detection (e.g., Langhans and Kohlrausch, 1992) or discrimination (e.g., Heller and Trahiotis, 1995) tasks. In the present conditions, the subjective impression is also that the maskers in the two ears must be attended to, or combined in, an obligatory way. It happens that in this particular task, combining the two ears benefits the listener. However, preliminary work from our laboratory has indicated that presenting the SBN masker (instead of the DBN masker as in Sec. III C) contralateral to the target and DBS masker *degrades* identification performance—very much like the CMEs above. This suggests that the listener is not able to selectively combine the masker bands—either DBN or SBN—presented contralaterally but, like the CMEs described above, is obliged to do so by virtue of the nature of the stimuli and the way they are presented.

V. SUMMARY

Psychophysical experiments were conducted that examined combinations of energetic and informational maskers. Greater masking for audible speech maskers was found than for spectrally matched noise masker controls. The additional masking caused by the speech was thought to reflect informational masking. The informational masking caused by a speech masker could be significantly reduced by the simultaneous presentation of an energetic masker that overlapped the masker speech bands. However, an equal, or often greater, reduction in the effectiveness of the speech masker was found when the noise masker occupying the same frequency bands was presented to the contralateral ear. In that case, it appears that a binaural image was created that was sufficiently distinct from the target to improve the listener's ability to focus attention on the target, thereby reducing informational masking significantly.

ACKNOWLEDGMENTS

This work was supported by Grant Nos. DC00100, DC04545, DC04663, and F32 DC006526 from NIH/NIDCD and by the Boston University Hearing Research Center. The authors are grateful to their listeners, to Kelly Egan for her assistance, and to Nathaniel Durlach for many helpful discussions.

APPENDIX: ANALYSIS OF ENVELOPE SPECTRA

In Sec. III A, the noise maskers—either different band (DBN) or same band (SBN)—were spectrally matched with the corresponding speech maskers. The intent was to equate the noise and speech maskers in terms of their energetic masking values so that any difference could be attributed to an additional effect of informational masking. However, the spectral matching and rms equalization were based on computations performed over the entire duration of each stimulus. There are differences, therefore, in the time-varying nature of the two types of maskers that could cause differences in the amount of energetic masking each produces. In comparing the DBN masker to the DBS masker it is not much of an issue because the DBN masker produced almost no masking of any sort, so the conclusion that the DBS masker was

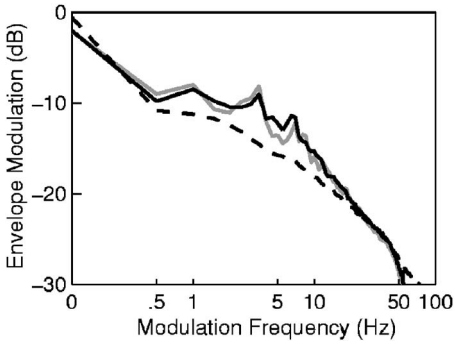


FIG. 5. Envelope magnitude spectra computed and averaged across all bands for all targets (gray), SBS maskers (black), and SBN maskers (dashed line). The abscissa is modulation frequency and the ordinate is envelope magnitude.

primarily informational seems well supported (see also Arbogast *et al.*, 2002). However, the same cannot be said for the comparison of the SBN and SBS maskers because both presumably produce large amounts of energetic masking. In this Appendix, we consider the time-varying properties of the two types of maskers, speech and noise (see also Buss *et al.*, 2003).

First, both types of maskers are composed of a set of very narrow bands. Because the stimuli are equated for rms values, the speech tends to have more peaks and valleys than the noise (this is apparent in a crude way simply from the plots in Fig. 1). Crest factors (ratio of peak to rms amplitudes), computed on the entire set of speech and noise waveforms, support this qualitative assessment and were, on average approximately 6.5 for the speech waveforms and 4.0 for the noise waveforms. Probably more importantly, though, are the differences in the envelopes of the two types of maskers. It has been speculated, for example, that the envelope characteristics of speech are essential in producing informational masking of speech targets (cf. Brungart *et al.*, 2005) in part because, in certain conditions, time-reversed speech can also produce significantly more informational masking than speech-shaped noise. Figure 5 plots the envelope spectra averaged across all bands of the speech maskers and the noise maskers computed over the duration of each stimulus.

Inspection of the envelope spectra reveals some noticeable differences. The long-term average envelope spectrum of the noise bands is smoother than for the speech and has a slightly higher dc component. The speech spectra have peaks at about 1, 3.5, and 7 Hz, corresponding roughly to 1, 0.33, and 0.14 s, respectively, likely due to boundaries between words or syllables. However, noting differences in the long-term average envelope spectra does not really inform us about possible differences in energetic masking between speech and noise. In order to more directly address this question, the following computation was made: For every target and masker pair, the rms amplitudes of the digital waveforms were equated and a band-by-band calculation of the sample-by-sample differences between target and masker envelopes was computed. Only positive values—meaning that the target envelope was greater than the masker envelope at a given point in time—were saved. Then the sum of the values was calculated for the entire duration of the target. Thus, an estimate was obtained on a band-by-band basis of the degree to which the target envelope exceeded the masker envelope—a quantity which presumably is (inversely) related to the amount of energetic masking produced in that band (at the relative level, i.e., 0 dB T/M, used in the computation; higher/lower T/Ms would of course produce less/more overlap). This was done for all bands for both speech maskers and noise maskers paired with the targets actually forming the speech corpus. The result of this computation indicated that, for the overwhelming (greater than 95%) number of comparisons, the target overlapped the speech masker envelope to a greater degree than it did the corresponding noise envelope (i.e., more of the target was "available" when the masker was speech than when the masker was noise computed on a band-by-band basis). Thus, there appears to be no evidence suggesting that the SBS masker produced greater energetic masking than the SBN masker when presented at equal T/Ms and, in fact, the evidence here suggests the opposite. This again agrees with the interpretation above that the greater effectiveness of the SBS masker than the SBN masker was attributable to informational masking.

One final point about the envelopes is that there was no significant correlation between the envelopes of the speech and noise maskers (accounted for less than 1% of the variance, on average, computed over corresponding frequency bands). Therefore, as intended, the masker samples for the noise and speech stimuli were spectrally matched but were not correlated in the time domain. This is consistent with the observation of Arbogast *et al.* (2002) that the noise maskers did not retain the intelligibility of the speech from which they were created.

[1]The decrease in masking as the number of masker components increased followed an initial increase in masking as the number of masker components increased for very few components (i.e., 2–10). Because masker power was held constant and little energetic masking would be expected for these few components outside of the "protected region" around the signal frequency, this initial increase was thought to reflect increased informational masking, which then was reduced when more components were added beyond an intermediate value (i.e., 10–20 components).

[2]The bands of speech—and the bands of noise which were derived from the bands of speech—were very narrow in frequency. Initially, the bandpass filtering (prior to extracting speech envelopes) used constant-Q filters having bandwidths about 22% of the center frequency. The envelope functions subsequently obtained from each narrow band of speech, as noted in the text, were low-pass filtered at 50 Hz and used to modulate sinusoidal carriers located at the center frequencies of the bands. Spectral analysis of the resulting bands of speech indicated that each band was characterized by a sharp peak corresponding to the carrier frequency. The bandwidths at the −3-dB points were thus extremely narrow and difficult to compute accurately. Better estimates were obtained at −10- and −20-dB points *re* the spectral peak in each band. Those estimates increased in proportion to center frequency ranging from about 4–16 Hz at the −10-dB bandwidths to 40–100 Hz at the −20-dB bandwidths. The points where adjacent bands overlapped in frequency occurred approximately at the geometric mean of the center frequencies and also varied in proportion to frequency ranging from about −20 dB *re* adjacent band peaks for the lowest-frequency bands to about −60 dB for the highest bands.

[3]By "amount of masking" we mean reduction in proportion correct identification performance due to the presence of the masker. Computations of masking, or target-to-maker ratios, in terms of decibel differences were not possible (without extrapolation) in many cases. Further, because the slopes of the psychometric functions varied across conditions, changes in T/M

across conditions depended on the performance level at which the comparison is made.

[4]The statistical analyses were performed on the proportion correct values. A separate analysis was done on the data transformed into rationalized arcsine units (cf. Studebaker, 1985). Results of the analyses agreed in all cases so only the one analysis is presented.

[5]As pointed out during the review process, we do not know whether a given change in percent correct produces an equivalent perceptual change throughout the entire range of values tested.

[6]This ANOVA includes data (DBS and DBN) analyzed in the comparison discussed in Sec. III B.

[7]Our stimuli were comprised of pure-tone carriers modulated by the envelopes of narrow bands of speech or noise with the envelope frequencies low-passed at 50 Hz. In contrast to noise-excited vocoder speech in which the bands are contiguous, the present stimuli differ not only in fine structure but also in the reduced spectral overlap between bands (cf. Qin and Oxenham, 2003; Shannon *et al.*, 1995; Brungart *et al.*, 2005).

Arbogast, T. L., Mason, C. R. and Kidd, G., Jr. (**2002**). "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098.

Arbogast, T. L., Mason, C. R., and Kidd, G. Jr. (**2005**). "The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **117**, 2169–2180.

Bernstein, L. R., and Trahiotis, C. (**1992**). "Discrimination of interaural envelope correlation and its relation to binaural unmasking at high frequencies," J. Acoust. Soc. Am. **91**, 306–316.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Brungart, D. S., and Simpson, B. D. (**2002**). "Within-ear and across-ear interference in a cocktail-party listening task," J. Acoust. Soc. Am. **112**, 2985–2995.

Brungart, D. S., and Simpson, B. D. (**2005**). "Evidence for a 'reverse cocktail-party' effect in a three-talker dichotic listening task," Acta. Acust. Acust. **91**, 564–566.

Brungart, D. S., Simpson, B. D., Darwin, C. J., Arbogast, T. L., and Kidd, G., Jr. (**2005**). "Across-ear interference from parametrically-degraded synthetic speech signals in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **117**, 292–304.

Buss, E., Hall, J. W., III, and Grose, J. H. (**2003**). "Effect of modulation coherence for masked speech signals filtered into narrow bands," J. Acoust. Soc. Am. **113**, 462–467.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1968**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

Durlach, N. I., Mason, C. R., Kidd, G., Jr., Arbogast, T. L., Colburn, H. S., and Shinn-Cunningham, B. G. (**2003a**). "Note on informational masking," J. Acoust. Soc. Am. **113**, 2984–2987.

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S. and Kidd, G., Jr. (**2003b**). "Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.

Heller, L. M., and Trahiotis, C. (**1995**). "The discrimination of samples of noise in monotic, diotic and dichotic conditions," J. Acoust. Soc. Am. **97**, 3775–3781.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Humes, L. E., Jesteadt, W., and Lee, L. W. (**1992**). "Modeling the effects of sensorineural hearing loss on auditory perception," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford).

Kidd, G., Jr., and Mason, C. R. (**1997**). "Combining energetic and informational maskers," presented at the annual meeting of the American Speech-Language-Hearing Association.

Kidd, G., Jr., Mason, C. R., and Arbogast, T. L. (**2002**). "Similarity, uncertainty and masking in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. **111**, 1367–1376.

Kidd, G., Jr., Mason, C. R., Brughera, A., and Chiu, C. Y. P., (**2003b**). "Discriminating harmonicity," J. Acoust. Soc. Am. **114**, 967–977.

Kidd, G., Jr., Mason, C. R., Brughera, A., and Hartmann, W. M. (**2005**). "The role of reverberation in release from masking due to spatial separation of source for speech identification," Acta. Acust. Acust. **91**, 526–536.

Kidd, G., Jr., Mason, C. R., Arbogast, T. L., Brungart, D., and Simpson, B. (**2003a**). "Informational masking caused by contralateral stimulation," J. Acoust. Soc. Am. **113**, 1594–1603.

Kidd, G., Jr., Mason, C. R., Deliwala, P. S., Woods, W. S., and Colburn, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.

Langhans, A., and Kohlrausch, A. (**1992**). "Differences in auditory performance between monaural and diotic conditions. I. Masked thresholds in frozen noise," J. Acoust. Soc. Am. **91**, 3456–3470.

Lutfi, R. A. (**1990**). "How much masking is informational masking?" J. Acoust. Soc. Am. **80**, 2607–2610.

McFadden, D., and Pasanen, E. G. (**1976**). "Lateralization at high frequencies based on interaural time differences," J. Acoust. Soc. Am. **59**, 634–639.

Neff, D. L., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 409–415.

Neff, D. L., and Jesteadt, W. (**1996**). "Intensity discrimination in the presence of random-frequency, multicomponent maskers and broadband noise," J. Acoust. Soc. Am. **100**, 2289–2298.

Nuetzel, J. M., and Hafter, E. R. (**1977**). "Lateralization of complex waveforms: Effects of fine structure, amplitude, and duration," J. Acoust. Soc. Am. **60**, 1339–1346.

Oh, E., and Lutfi, R. A. (**1998**). "Nonmonotonicity of informational masking," J. Acoust. Soc. Am. **104**, 3489–3499.

Penner, M. J. (**1980**). "The coding of intensity and the interaction of forward and backward masking," J. Acoust. Soc. Am. **67**, 608–616.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Richards, V. M., Tang, Z., and Kidd, G. Jr. (**2002**). "Informational masking with small set sizes," J. Acoust. Soc. Am. **111**, 1359–1366.

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

Studebaker, G. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.

Warren, R. M., (**1999**) *Auditory Perception: A New Analysis and Synthesis* (Cambridge U. P. Cambridge, UK).

Warren, R. M., and Bashford, J. A., Jr. (**1976**). "Auditory contralateral induction: An early stage in binaural processing," Percept. Psychophys. **20**, 380–386.

Wilbanks, W. A., and Whitmore, J. K. (**1968**). "Detection of monaural signals as a function of interaural noise correlation and signal frequency," J. Acoust. Soc. Am. **43**, 785–797.

Wright, B. A., and Saberi, K. (**1999**). "Strategies used to detect auditory signals in small sets of random maskers," J. Acoust. Soc. Am. **105**, 1765–1775.

Noise improves modulation detection by cochlear implant listeners at moderate carrier levels[a)]

Monita Chatterjee[b)] and Sandra I. Oba[c)]
Department of Auditory Implants and Perception, House Ear Institute, 2100 W. Third Street, Los Angeles, California 90057

(Received 25 August 2004; revised 30 March 2005; accepted 15 April 2005)

Envelope detection and processing are very important for cochlear implant (CI) listeners, who must rely on obtaining significant amounts of acoustic information from the time-varying envelopes of stimuli. In previous work, Chatterjee and Robert [JARO **2**(2), 159–171 (2001)] reported on a stochastic-resonance-type effect in modulation detection by CI listeners: optimum levels of noise in the envelope enhanced modulation detection under certain conditions, particularly when the carrier level was low. The results of that study suggested that a low carrier level was *sufficient* to evoke the observed stochastic resonance effect, but did not clarify whether a low carrier level was *necessary* to evoke the effect. Modulation thresholds in CI listeners generally decrease with increasing carrier level. The experiments in this study were designed to investigate whether the observed noise-induced enhancement is related to the low carrier level *per se*, or to the poor modulation sensitivity that accompanies it. This was done by keeping the carrier amplitude fixed at a moderate level and increasing modulation frequency so that modulation sensitivity could be reduced without lowering carrier level. The results suggest that modulation sensitivity, not carrier level, is the primary factor determining the effect of the noise. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1929258]

PACS number(s): 43.66.Ts, 43.66.Dc [GDK] Pages: 993–1002

I. INTRODUCTION

The phenomenon dubbed "stochastic resonance" (SR), which has excited interest in various fields including neuroscience in recent years, consists of an enhancement in signal transmission through a nonlinear system with the addition of an optimum noise (see Moss *et al.*, 2004, for a review). In sensory systems, SR has been shown in behavioral, psychophysical, and physiological experiments as well as in modeling studies of single-neuron and distributed neuronal systems (Douglas *et al.*, 1993; Moss *et al.*, 2004; Ward *et al.*, 2002; Simonotto *et al.*, 1997, 1999; Kitajo *et al.*, 2003; Collins *et al.*, 1995; Stocks, 2001; Bahar and Moss, 2003; Morse *et al.*, 2003). Optimum levels of noise have produced SR-like gains in the responses of bullfrog saccular hair cells and primary neurons (Jaramillo and Wiesenfeld, 1998; Indresano *et al.*, 2003), as well as in gerbil auditory-nerve neurons (Henry, 1999). In the present paper, we explore SR in cochlear-implant (CI) listeners performing a modulation detection task.

Electrical stimulation of the cochlear-damaged auditory system results in a remarkably low-noise neural response in the periphery, with strong phase-locking to the stimulus and large across-fiber synchrony (Hartmann *et al.*, 1984). This pattern of response contrasts with that observed in the normal auditory nerve, which shows random activity both in quiet and in the presence of sound (Kiang, 1965). It has been thought for some time that in the normal cochlea, auditory-nerve fibers fire not only randomly but also independently of each other (Johnson and Kiang, 1976): one potential benefit of this across-fiber desynchrony may be an increase the number of independent sources of information. In contrast, the electrically stimulated, deafened auditory system, with its synchronized across-fiber responses, would seem to have a lower information capacity: hence, the appeal of using external noise to improve thresholds and information capacity in cochlear implants.

Using an electrically stimulated frog sciatic nerve as a model, Morse and Evans (1996; also see Moss *et al.*, 1996) were the first to show that the representation of fine time structures of vowels by a single neuron could be enhanced by using optimal external noise. Following Morse and Evans' initial findings, various investigators have explored the idea that external noise may be of benefit to cochlear implant listeners, to the extent that it can restore the "normal" stochasticity of the peripheral neural response (White *et al.*, 2000).

Under some conditions, noise-induced benefits have been demonstrated in CI listeners (Zeng *et al.*, 2000; Behnam and Zeng, 2003). In guinea pigs, Matusuoka *et al.* (2000) have shown that the addition of Gaussian noise to an electrical pulse train has the effect of desynchronizing the neural responses to electrical stimulation. Rubinstein *et al.* (1999) proposed to achieve this using very high-rate "conditioning" pulse trains, which would drive the neurons into a state of relative refractoriness. There is, in fact, evidence to

[a)]Portions of this work were published in the Proceedings of the 2003 Meeting of the Association for Research in Otolaryngology, Daytona Beach, FL, the 2003 Conference on Implantable Auditory Prostheses, Asilomar, CA, and the 2003 SPIE Conference on Fluctuations and Noise, Santa Fe, NM.

[b)]Electronic mail: mchatterjee@hesp.umd.edu Present address: Department of Hearing and Speech Sciences, 0100 LeFrak Hall, University of Maryland at College Park, College Park, MD 20742.

[c)]Present address: Children's Auditory Research and Evaluation Center, House Ear Institute, 2100 W. Third St., Los Angeles, CA 90057.

TABLE I. Potentially relevant information about subjects.

Subject	Age (years)	Gender	Age at onset of profound deafness	Age at implantation	Etiology	Device type	Electrode pair used for stimulation
S1	57	F	26	44	Ototoxic medication	N24	10,13
S2	47	M	35	35	Trauma	N22	10,13
S3	61	F	40	60	Hereditary	N24	10,12
S4	53	M	43	43	Unknown	N22	10,13
S5	63	M	44	53	Trauma/unknown	N22	10,13
S6	76	F	60	65	Unknown	N22	8,10
S7	71	F	58	66	Unknown	N24	14,16
S8	70	F	45	56	Cochlear otosclerosis	N22	6,8

suggest that neuronal responses do become more stochastic under high-rate stimulation conditions, and that both intensity and temporal coding might improve under these conditions (Miller *et al.*, 2001; Runge-Samuelson *et al.*, 2004; Litvak *et al.*, 2001, 2003a, 2003b; Hong *et al.*, 2003).

Present-day CI devices utilize a speech-processing strategy that involves the stimulation of tonotopically appropriate electrodes with carrier pulse trains that are amplitude modulated by envelopes extracted from the frequency regions of interest (see Loizou, 1998, for review). The ability to resolve these temporal fluctuations has been shown to be an important predictor of speech perception with the CI (Fu, 2002). When the acoustic input is a mix of signal and noise, the envelope fluctuations of the noise are mixed in with those of the signal. It is of some interest, therefore, to understand the processing of signal-related envelope modulations by CI listeners in the presence of competing noise fluctuations.

In a modulation detection task performed by CI listeners, Chatterjee and Robert (2001) found that in most instances, noise introduced into the envelope acted as an interferer, making it more difficult to detect the modulation in the signal. Under some conditions, however, they also found that the noise *enhanced* the detection of the target modulation. This enhancement was observed when the carrier level was low, the modulation frequency was relatively high, and the noise applied at an optimum level. The function relating modulation sensitivity to the noise level had an inverted-U-shaped appearance. This is the signature shape of stochastic resonance.

The findings of Chatterjee and Robert were intriguing but not conclusive as to underlying mechanisms. Modulation sensitivity in cochlear implant listeners varies monotonically with carrier level. It was not clear from the results of that study whether the observed SR phenomenon was tied to the lower carrier level or the ensuing poorer modulation sensitivity. If the lower carrier level was both necessary and sufficient to produce SR, it might be easier to explain the phenomenon as arising at a stage where the *carrier* is being detected. However, if it was observed that SR was produced at higher carrier levels under conditions that compromised modulation sensitivity, we might conclude that it arose at a stage where *modulations* were detected. In this paper, we examine a scenario in which modulation sensitivity is compromised but carrier level remains moderate—thus, audibility or detectability of the carrier can be eliminated as an issue. One way to achieve this is to keep the carrier fixed at a moderate level, and increase modulation frequency. As in normal-hearing listeners, the modulation transfer function (MTF) in cochlear implant listeners has a low-pass filter shape, and modulation sensitivity worsens as modulation frequency increases. Thus, modulation sensitivity can be compromised without lowering carrier level.

Results from two experiments are presented here. In experiment 1, modulation sensitivity was manipulated by varying carrier level for a fixed modulation frequency. This essentially replicated the original findings of Chatterjee and Robert in a larger group of subjects. In experiment 2, modulation sensitivity was manipulated by fixing the carrier at a moderate level (well above threshold) and varying modulation frequency.

II. EXPERIMENTAL METHODS

A. Subjects

Subjects were eight adult, postlingually deafened users of the Nucleus-22/Nucleus-24 cochlear implant device. Subjects S2, S4, S5, S6, and S8 had considerable experience in psychophysical experiments in our laboratory, particularly in modulation detection experiments; subjects S1, S3, and S7 had little or no experience with psychophysical experiments. Relevant information about individual subjects is provided in Table I.

B. Stimuli

Stimuli were presented through a custom-built research interface (Shannon *et al.*, 1990; Robert, 2002). Each carrier stimulus consisted of a 200-ms-long train of 200-μs/phase biphasic current pulses, either 500 or 1000 pulses/s in rate, presented to a single electrode pair, usually a centrally located pair in the array (see Table I). For some subjects, some apical electrodes were not usable and the chosen electrode

pair was located more basally. Pulse amplitudes were obtained for each subject's device from the calibration information provided by the manufacturer.

Sinusoidal amplitude modulation (SAM) was applied to the charge/phase of each pulse: either the pulse amplitude or the pulse phase duration was modulated, depending on the subject's resolution. In the Nucleus device, finer resolution is available in allowable pulse phase duration increments than in pulse amplitude increments. In pulse duration, the smallest increment is 0.4 μs, translating to a modulation index of 0.002 (-53.97 dB) for a reference pulse phase duration of 200 μs/phase. Cochlear implant listeners are often very sensitive to modulation, and in the best case, their sensitivity can exceed the resolution available in amplitude. Therefore, when measuring modulation sensitivity, pulse phase duration is modulated. If a subject shows very poor sensitivity to modulation, the required increments in pulse phase duration may become too large for a fixed pulse rate, and pulse amplitude will have to be modulated instead. In this study, for subject S7, modulation sensitivity was very poor: therefore, pulse amplitude was modulated instead of pulse phase duration. Generally, modulation sensitivity is greater when pulse amplitude is modulated than when pulse phase duration is modulated. This likely stems from the nonlinear relation between pulse duration and pulse amplitude for constant loudness (Chatterjee *et al.*, 2000; Zeng *et al.*, 1998). Thus, if a subject exhibits very poor sensitivity with pulse phase duration modulation, s/he is generally more sensitive to pulse amplitude modulation. However, in our experience, the shape of the MTFs does not depend on whether pulse amplitude or pulse phase duration is being modulated. In all cases, modulation was applied in linear microampere units. For pulse phase duration modulation, the equation used was

$$D(n) = D_{\text{ref}}[1 + m\cos(2\pi f_m n/fc)],$$

where $D(n)$ is the pulse phase duration of the nth pulse, D_{ref} is the reference pulse phase duration (200-μs/phase in this case), m is the modulation index, f_m is the modulation frequency, n is the pulse number, and f_c is the carrier pulse rate (pulses/s). For pulse amplitude modulation, the equation was identical except that pulse amplitudes replaced pulse phase durations.

Noise was generated as an array of pseudorandom numbers drawn from a uniform distribution within the range $(I \pm r^* I)$, where I was the reference carrier amplitude and r determined the level of the noise $(0 < r < 1.0)$. This noise was added to the amplitude of each pulse of the train. The noise had a flat envelope spectrum for frequencies up to half the carrier pulse rate. Figure 1(a) shows an example of the amplitude series generated with 20% noise added to a pulse train with an amplitude of 427 μA. Figure 1(b) shows the mean DFT (discrete Fourier transform) of 20 such pulse trains. Error bars show ±1 standard deviation. The amplitudes correspond to subject S4's calibration table (see Chatterjee, 2003 for a detailed description of a similar noise). The levels of noise were generally high enough that concerns about the step size/resolution did not arise. The noise was also applied in linear microampere units.

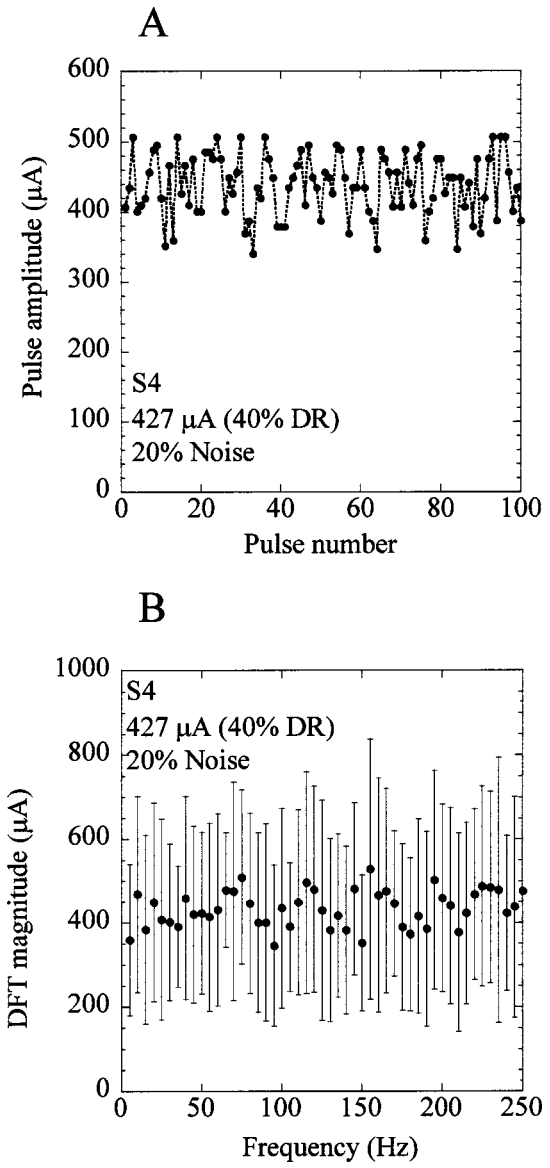


FIG. 1. (A). An example of a pulse amplitude series generated when a 20% noise is added to the 100 pulses of a 200-ms long, 500-Hz pulse train with a 427-μA reference amplitude. The amplitude calibration table corresponds to that of subject S4. (B). The mean envelope spectrum of 20 amplitude series such as the one in (A). Error bars show ±1 standard deviation.

In all cases, care was taken to ensure that the levels of the sounds did not exceed comfortable loudness limits. Experimental protocols were approved by the Institutional Review Board of House Ear Institute and St. Vincent's Hospital, Los Angeles.

C. Procedures

1. Modulation detection thresholds

Modulation detection thresholds were measured using an adaptive, 3-down, 1-up, 3-interval forced-choice task, in which the noise was present in all three intervals, and one of the intervals (random order) carried the noise plus the modulation to be detected. The subject was asked to indicate which of the intervals contained the modulation. Correct/incorrect feedback was provided. The run continued until a maximum of 60 trials was completed, or a total of 12 rever-

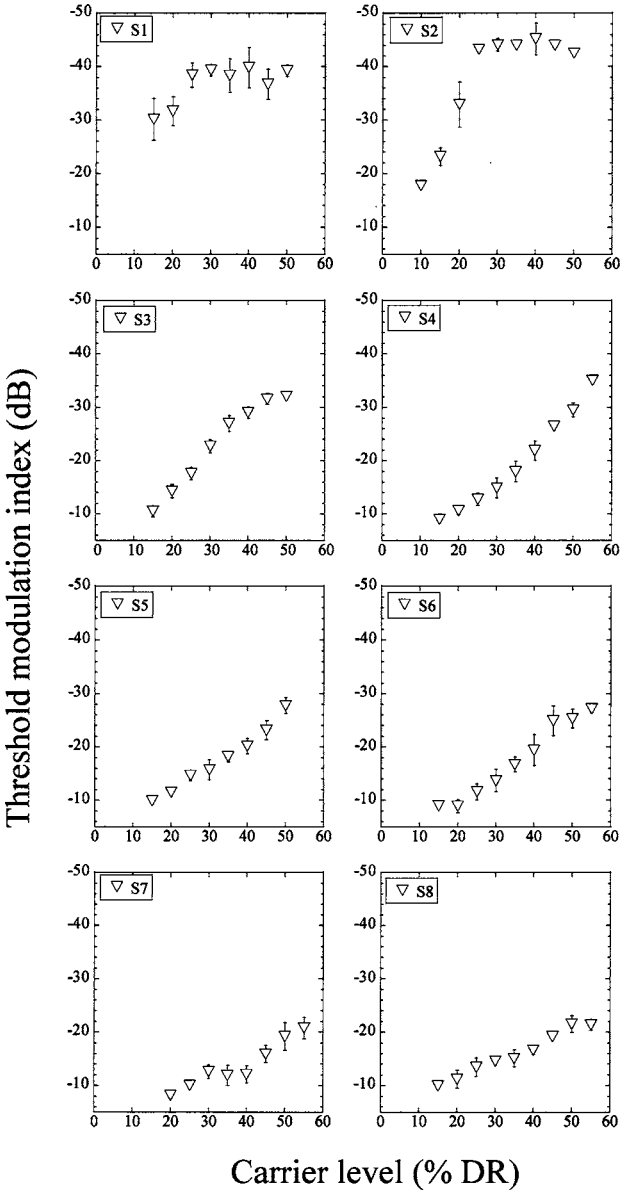


FIG. 2. Modulation detection thresholds ($20^*\log m$, where m is the modulation index) plotted as a function of signal carrier level, expressed as percent dynamic range (dynamic range was calculated as linear microamps). The carrier pulse rate was 500 pulses/s, and the modulation frequency was 50 Hz. Each panel corresponds to a different subject.

sals was completed. The run was discarded if fewer than ten reversals had been achieved. The initial step size was decreased after the first four reversals. The mean of all but the first four reversals was calculated to obtain the final threshold for the run. With the exception of the data shown in Fig. 2, the mean and standard deviation of at least four such runs was calculated for each measurement. For a particular experimental condition, the presentation sequence of the stimuli was randomized across noise levels.

2. Dynamic range

The dynamic range was determined for each carrier as follows. Detection threshold was measured using a 3-down, 1-up, 2-interval forced-choice procedure. Initial and final steps were 1 and 0.5 dB, respectively. Each measurement was repeated three times; if two of the measures differed by more than 30 μA, a fourth measurement was made. The mean of all measurements was taken as detection threshold. To obtain "maximum acceptable level" (MAL), the subject was asked to adjust the current level of the pulse train (by pressing the "up" or "down" arrow keys on the keyboard) until the sound reached the upper limit of the comfortable loudness range. The mean of three repetitions of the MAL procedure was calculated for each subject. Dynamic range was defined as the difference between MAL and detection threshold in microamperes.

3. Noise detection thresholds

Noise detection thresholds were measured using methods identical to the modulation detection experiments described before. Noise was applied to the pulse amplitudes by multiplying successive pulses by a number drawn from a uniform distribution, so that the amplitudes were uniformly distributed about the mean (i.e., the reference carrier level), with a range that was adaptively varied from trial to trial. This method of applying the noise is different from the additive method used in the modulation detection experiment. However, the two methods result in identical distributions of pulse amplitudes.

III. RESULTS

A. Experiment 1: Effects of noise on modulation sensitivity; carrier level as parameter

Detection thresholds for 50-Hz modulation were measured as a function of carrier level (carrier pulse rate was 500 pulses/s). Carrier levels spanned a range from 10% to 55% of DR. Initially, only one or two repetitions of the measurement were obtained at each carrier level. The modulation threshold vs carrier level functions are shown in Fig. 2 for each of the eight subjects. For each subject, a number of carrier levels was selected that would yield a range of modulation thresholds. For the selected carrier levels, the data in Fig. 2 represent the mean and standard deviation of four or more repetitions: the remaining data represent the mean of only one or two repetitions. In all such figures, the vertical axes show modulation threshold decreasing (i.e., modulation sensitivity increasing) in the upward direction.

Effects of noise on the 50-Hz modulation thresholds are shown in Fig. 3 for each subject. Within each panel, carrier level is the parameter. The solid lines indicate the conditions in which the improvement in modulation sensitivity over the 0% noise condition was statistically significant (Student's t-test, $p<0.05$). The subjects showed a wide range of variation in baseline sensitivity (i.e., the sensitivity to modulation at 0% noise level). For subjects S1 and S2, who had the best baseline sensitivity, the axes have been extended relative to the other subjects. For these two subjects, the noise worsened modulation thresholds at most carrier levels. At the lowest level, however, S2's thresholds showed a gain in modulation sensitivity with increasing noise level for a limited range of noise levels. Beyond this range, sensitivity dropped again, resulting in a shallow, inverted-U shape.

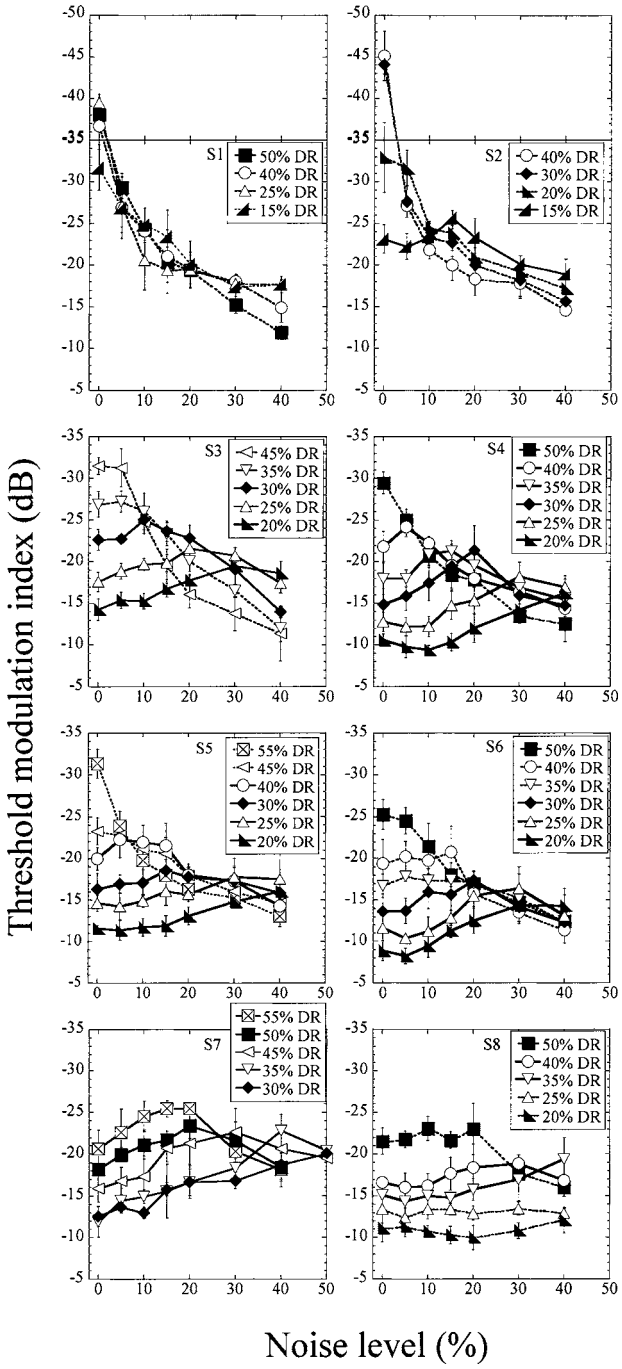


FIG. 3. Modulation detection thresholds plotted as a function of noise level (expressed as percent of reference carrier amplitude), for all subjects. The parameter is carrier level. The carrier pulse rate was 500 pulses/s, and the modulation frequency was fixed at 50 Hz. Error bars show ±1 standard deviation. Solid lines indicate conditions which yielded a statistically significant improvement in modulation sensitivity over the baseline sensitivity (zero-noise condition).

Subjects S3, S4, S5, and S6, who were all less sensitive to modulation than S1 and S2, showed a continuum of effects of noise. At higher carrier levels, when they were more sensitive to modulation, the effect of noise was similar to that observed in S1 and S2: increasing noise resulted in poorer modulation thresholds. At moderate carrier levels, the function showed a plateau for low noise levels before dropping down toward higher thresholds at higher noise levels. At low carrier levels, noise improved modulation thresholds over a limited range of noise levels: as observed with subject S2, modulation thresholds worsened at high noise levels, often resulting in an inverted-U-shaped function.

Subjects S7 and S8 had the poorest baseline modulation thresholds overall, and showed the portion of the continuum observed at lower carrier levels in subjects S3, S4, S5, and S6. In fact, the continuum observed in individual subjects could be observed across subjects as well—when modulation thresholds were high, noise generally worsened performance: when modulation thresholds were low, noise improved performance over a range of optimal noise levels, and when modulation thresholds were moderate, a plateau was observed, followed by a decline with increasing noise levels. Thus, the effect of the noise seemed to be closely related to baseline modulation thresholds. In subject S7's case, for instance, modulation threshold was poor enough at the highest carrier level used, to yield the U-shaped function at the higher levels. Subjects S3, S4, S5, and S6, who were moderately sensitive to 50-Hz modulation, showed the full range of effects. At high carrier levels, noise was a masker; at low carrier levels, optimum noise enhanced modulation thresholds, and at medium levels, there was a prolonged plateau at low noise levels, followed by the masking effect. A floor effect was observed in subject S8 at the lowest carrier levels.

For the cases in which statistically significant improvement in modulation sensitivity was observed ($p<0.05$, Student's t-test), the size of the peak improvement in modulation threshold observed for each subject is listed in Table II, along with the noise level at which the peak occurred.

The results obtained with subjects S1 and S2 did not show the full continuum observed with the remaining subjects. This was possibly because these subjects were very sensitive to modulation at 50 Hz down to the lowest carrier levels. To test this possibility, we repeated the experiment using a high carrier pulse rate (1000 pulses/s) and a high modulation frequency (300 Hz) for these two subjects. Similar data were also collected with subject S3. Because of the low-pass characteristic of the MTF, all three subjects had reduced sensitivity at the higher modulation frequency. Figure 4 shows the effects of noise on their modulation detection thresholds for the 300-Hz modulation detection task. Decreasing the carrier level had a significant effect on modulation sensitivity at this modulation frequency, and we are now able to observe the full continuum of effects for all three subjects. Again, Table III lists the size of the peak improvement in modulation threshold for each condition and subject ($p<0.05$).

These results suggest that modulation sensitivity, and not carrier level itself, dictates the response to noise. This idea will be explored further in the next section.

B. Experiment 2: Effects of noise on modulation sensitivity; modulation frequency as parameter

Our goal in this experiment was to select a carrier level that would yield a range of modulation sensitivities across modulation frequency. Thus, the carrier level should not be so high that modulation sensitivity did not decline enough at high modulation frequencies, nor so low that a floor effect

TABLE II. Statistically significant (t-test, $p<0.05$) peak enhancement in 50-Hz modulation sensitivity shown as the absolute decrease in modulation threshold (dB). The noise level (r value) at which the peak occurred is also listed. S1 showed no improvement with added noise.

	Carrier level (% dynamic range)								
Subject	15%	20%	25%	30%	35%	40%	45%	50%	55%
	2.469 dB								
S2	$r=0.15$								
S3		5.219 dB	4.040 dB	2.478 dB					
		$r=0.30$	$r=0.20$	$r=0.10$					
S4		5.629 dB	5.432 dB	6.515 dB	3.328 dB	2.387 dB			
		$r=0.40$	$r=0.30$	$r=0.20$	$r=0.15$	$r=0.05$			
S5		4.463 dB	3.127 dB	2.269 dB		2.299 dB			
		$r=0.40$	$r=0.30$	$r=0.15$		$r=0.05$			
S6		5.492 dB	4.774 dB	3.657 dB					
		$r=0.30$	$r=0.30$	$r=0.20$					
S7				7.556 dB	10.976 dB		6.648 dB	5.236 dB	4.916 dB
				$r=0.50$	$r=0.40$		$r=0.30$	$r=0.20$	$r=0.20$
S8					4.075 dB	2.240 dB			
					$r=0.40$	$r=0.30$			

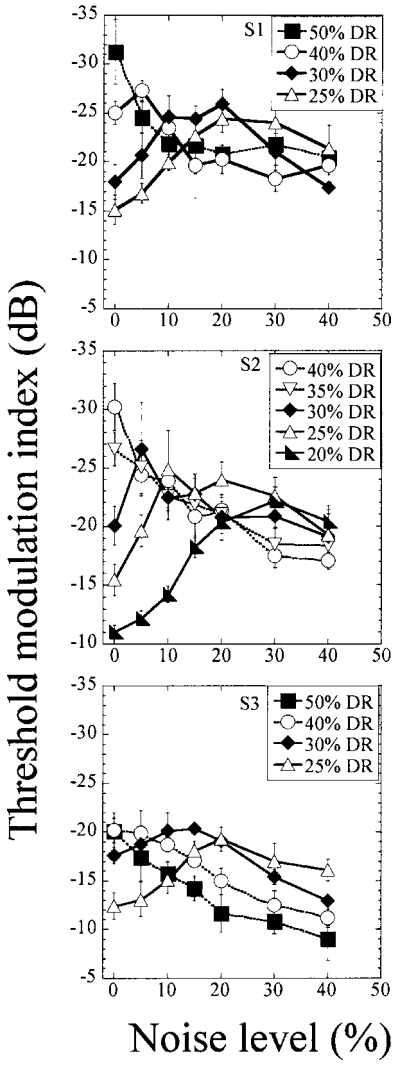


FIG. 4. Modulation detection thresholds as a function of noise level, for a 1000-pulses/s carrier and 300-Hz modulation frequency, for subjects S1–S3. The parameter is carrier level. As in previous figures, solid lines indicate conditions which yielded statistically significant improvements over baseline sensitivity.

was reached. We therefore selected moderate carrier levels for each subject, and measured the effects of noise, with modulation frequency as the parameter. If the baseline modulation sensitivity determines the effect of the noise, then we should observe a continuum of functions similar to those observed in the previous experiment as baseline modulation sensitivity declines with increasing modulation frequency. In Fig. 5, we show results with subjects S3, S4, S5, S6, S7, and S8.

The carrier pulse rate was 500 pulses/s for each subject, and the level was one that produced a moderate modulation sensitivity at low modulation frequencies. We observe that modulation sensitivity did decline with increasing modulation frequency for these subjects, albeit by different amounts. Subjects S4, S5, S7, and S8 do show the kind of continuum observed in the previous section—i.e., when modulation sensitivity was good, noise generally made it poorer, and when modulation sensitivity was poor, the familiar inverted-U-shaped function was observed. In between, a range of effects was observed. In Subject S3's case, the beneficial effect of the noise was small. It is possible that for S3, modulation sensitivity did not decline to the point at which the noise-induced enhancement could be observed. The same, however, could not be said for subject S6, who always showed a monotonic decline in modulation sensitivity with increasing

TABLE III. Statistically significant (t-test, $p<0.05$) peak enhancement ($p<0.05$) in 300-Hz modulation sensitivity shown as the absolute decrease in modulation threshold (dB). The noise level (r value) at which the peak occurred is also listed.

	Carrier level (% dynamic range)			
Subject	20%	25%	30%	40%
		9.313 dB	7.948 dB	2.238 dB
S1		$r=0.20$	$r=0.20$	$r=0.05$
S2	11.175 dB	9.441 dB	6.536 dB	
	$r=0.30$	$r=0.10$	$r=0.05$	
S3		6.826 dB	2.736 dB	
		$r=0.20$	$r=0.15$	

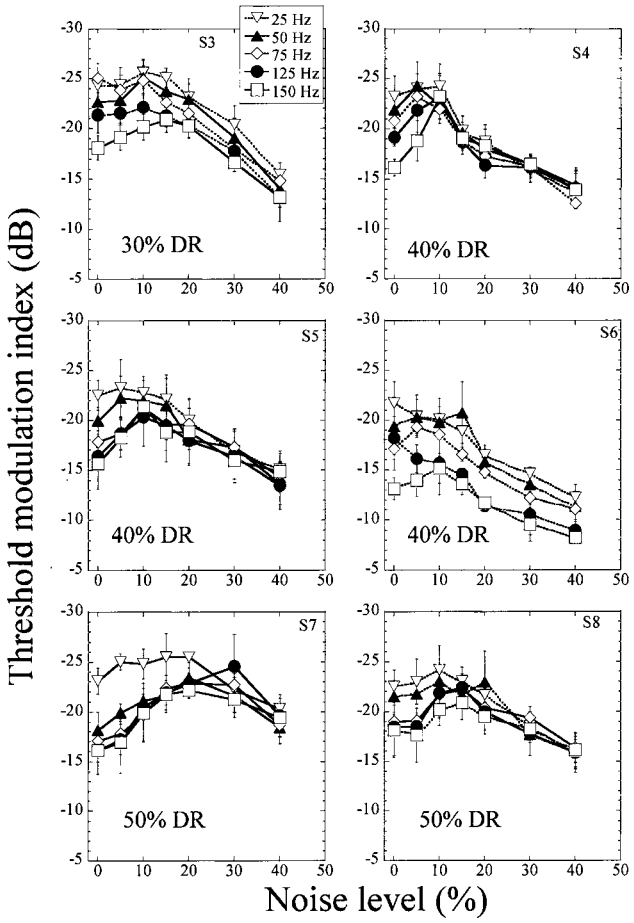


FIG. 5. Modulation detection thresholds as a function of noise level, for subjects S3–S8. The carrier level was fixed at a moderate level (indicated in each panel). The carrier frequency was fixed at 500 Hz, and the parameter is modulation frequency (shown in inset).

noise, even when the baseline modulation sensitivity was very poor. Table IV lists the size of the peak improvement in modulation threshold for each subject ($p<0.05$).

For subjects S1 and S2, modulation sensitivity could not be lowered sufficiently by increasing modulation frequency up to 150 Hz. Therefore, we increased the carrier pulse rate to 1000 pulses/s and used a wider range of modulation frequencies in their cases. This was also done for subject S3, to test the idea that if we can reduce modulation sensitivity sufficiently, we might see the full range of effects observed with carrier level. Results are shown in Fig. 6. It is apparent that the continuum of effects can indeed be observed with all three subjects. Table V lists the size of the peak improvement in each subject ($p<0.05$).

We note here that, for a particular noise level, the net energy in the signal was identical across conditions in this experiment. As the carrier level was held constant, the different pattern of results obtained were due only to changes in modulation frequency. The net energy in the signal did not change with changing modulation frequency. A particular noise level resulted in an identical energy increment across conditions. Thus, the differential effects of the noise observed at different modulation frequencies were not due to differences in energy.

C. The role of sensitivity to the noise

Sensitivity to the noise, like sensitivity to modulation, depends strongly on carrier level. The decrease in noise threshold (increase in noise sensitivity) with increasing carrier level is shown in Fig. 7(a) for all the subjects. This level dependence of noise threshold explains the difference between the effects observed by lowering carrier level at a fixed modulation frequency and those observed by increasing modulation frequency at a fixed carrier level. In the first case, as discussed previously, sensitivity to noise decreased with decreasing carrier level: as a result, higher levels of noise were required to reach the peak of the U-shaped function at low carrier levels. This is shown in a scatterplot in Fig. 7(b), showing the noise level at the peak modulation sensitivity vs noise detection threshold. In contrast, when carrier level was fixed and modulation frequency altered, noise threshold remained constant, and the peak of the U-shaped function remained at approximately the same noise level at all modulation frequencies. This is observed in the results of experiment 2: given a fixed carrier level, for all modulation frequencies, the peak enhancement in modulation sensitivity generally occurred at the same noise level.

TABLE IV. Statistically significant (t-test, $p<0.05$) peak enhancement ($p<0.05$) in 25,50,75,125, and/or 150-Hz modulation sensitivity shown as the absolute decrease in modulation threshold (dB). The noise level (r value) at which the peak occurred is also listed. S6 showed no significant improvement with added noise.

	Modulation frequency				
Subject	25 Hz	50 Hz	75 Hz	125 Hz	150 Hz
		2.478 dB			2.762 dB
S3		$r=0.10$			$r=0.15$
(30% DR)					
S4		2.387 dB		3.959 dB	7.054 dB
(40% DR)		$r=0.05$		$r=0.10$	$r=0.10$
S5		2.299 dB	3.421 dB	3.910 dB	5.706 dB
(40% DR)		$r=0.05$	$r=0.10$	$r=0.10$	$r=0.10$
S7	2.693 dB	5.236 dB	5.816 dB	7.409 dB	5.498 dB
(50% DR)	$r=0.15$	$r=0.20$	$r=0.20$	$r=0.30$	$r=0.20$
S8			3.253 dB	3.977 dB	
(50% DR)			$r=0.15$	$r=0.15$	

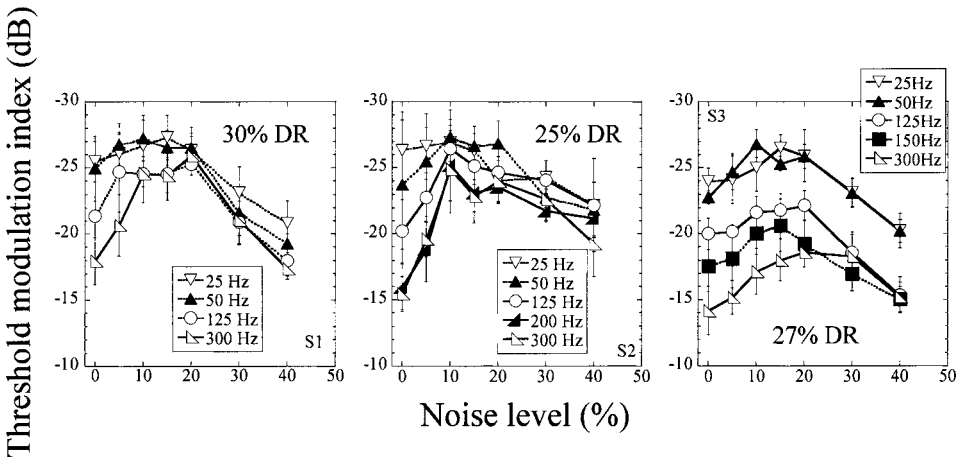


FIG. 6. Modulation detection thresholds as a function of noise level, for subjects S1–S3. The carrier pulse rate was fixed at 1000 pulses/s. The parameter is modulation frequency. Carrier level was fixed at moderate (indicated in each panel).

IV. DISCUSSION

The results presented here reinforce the original finding by Chatterjee and Robert (2001) that noise added to the envelope of a modulated pulse train can have a continuum of effects in cochlear implant listeners, ranging from masking the modulation to enhancing it. The question raised in the Introduction has been answered at least in part by the results. It appears that a low carrier level is sufficient but not necessary to produce the noise-induced gain in modulation sensitivity observed earlier by Chatterjee and Robert (2001). At higher carrier levels, when modulation sensitivity was compromised by increasing the frequency of the modulation being detected, the noise produced a similar enhancement at the higher modulation frequency.

The results suggest that the effect of the noise depends on modulation sensitivity in the no-noise, baseline condition. When modulation sensitivity is compromised—whether by lowering carrier level or by increasing modulation frequency, or by the presence of a fluctuating masker—optimal noise in the envelope has a beneficial effect on modulation thresholds. Further increases in noise beyond this point result in poorer modulation sensitivity. The nonmonotonic dependence of modulation thresholds on noise has the signature shape of stochastic resonance. The data indicate that sensitivity to the noise itself also plays a role in the shape of the function. Thus, at low carrier levels, detection thresholds of the noise are higher, and higher levels of noise are required to produce the maximum enhancement effect. On the other hand, when carrier level is fixed (thus fixing noise threshold) and the parameter being varied is modulation frequency, the peak of the U-shaped function occurs at a constant noise level, as would be expected if the location of the peak depended only on the salience of the noise (i.e., on the noise threshold).

Thus, the results show that the continuum of effects of noise is not based on the detectability of the carrier, but rather on the detectability of the modulation. Let us consider this finding in the light of the literature on MTFs of the auditory-nerve fiber (ANF). It is known that, in electrical stimulation, ANFs show a remarkably flat MTF, even flatter than the MTF of ANFs responding to acoustic stimuli, with a low-pass filter roll-off occurring only above 1000–2000 Hz (Dynes and Delgutte, 1992). In contrast, the psychophysical MTFs measured in cochlear implant listeners here and elsewhere (Shannon, 1992; Chatterjee and Robert, 2001) show a roll-off at about 125 Hz. The difference in the peripheral and psychophysical MTFs would suggest that the shape of the psychophysical MTF is limited by central mechanisms. Given that our primary finding involves the decline in modulation thresholds at modulation frequencies exceeding 100 Hz, it seems that the ANF cannot be implicated as a major player in these experiments. Is it still possible that the noise enhances modulation sensitivity in the periphery under certain conditions? At low carrier levels, it is certainly possible

TABLE V. Statistically significant (t-test, $p<0.05$) peak enhancement ($p<0.05$) in 25,50,75,200, and/or 300-Hz modulation sensitivity shown as the absolute decrease in modulation threshold (dB). The noise level (r value) at which the peak occurred is also listed.

	Modulation frequency				
Subject	25 Hz	50 Hz	125 Hz	200 Hz	300 Hz
					7.948 dB
S1					$r=0.20$
(30% DR)					
S2			4.939 dB	9.168 dB	9.441 dB
(25% DR)			$r=0.15$	$r=0.10$	$r=0.10$
S3	2.512 dB	4.033 dB	2.116 dB		4.430 dB
(27% DR)	$r=0.15$	$r=0.10$	$r=0.20$		$r=0.20$

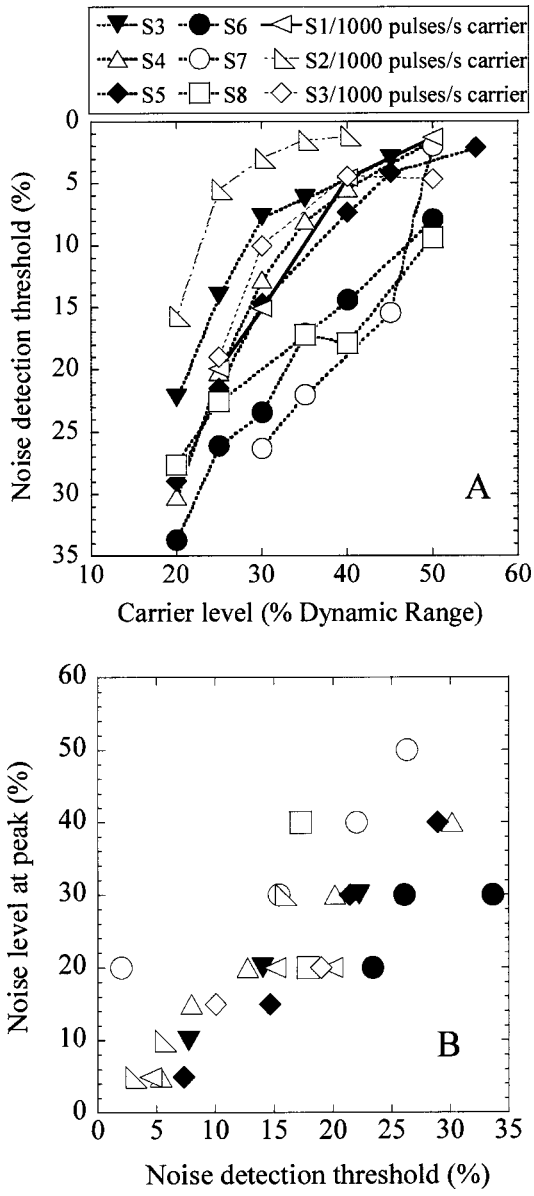


FIG. 7. (A). Noise detection thresholds (in % of reference carrier amplitude) as a function of carrier level for each of the eight subjects. (B). The noise level at which peak improvement in modulation sensitivity was reached, as a function of the noise detection threshold at that carrier level.

that this occurs. However, at *moderate* carrier levels, the psychophysical results observed in experiment 2 depend entirely on modulation frequency. This dependence is observed within a range of modulation frequencies that correspond to a flat gain in the ANF (Dynes and Delgutte, 1992): therefore, this part of the results is difficult to explain based on a peripheral SR effect in the ANF. It is still possible that the two sets of effects (the continuum observed with changing carrier level and its counterpart observed by changing modulation frequency) are determined by two mechanisms—one peripheral and one central. In this case, parsimony dictates that the similarity between the two continua is too strong to require two mechanisms.

Although some recent work suggests that SR in systems of neurons may improve information transmission rather than simply signal detection, SR has been largely modeled and described as a *threshold* phenomenon. Although conducted above carrier threshold, our experiment can be considered a threshold experiment, where the signal being detected is the *modulation* in the carrier. There is considerable evidence that envelope features of sounds are processed by specific pathways in the brainstem nuclei (Frisina, 2001). It is possible that extracted envelope information is passed to an "envelope detector" neuron residing in the cochlear nucleus or inferior colliculus. The input to such a neuron would consist of both the envelope noise and the signal modulation. For weak modulation signals just below detection threshold, such a neuron would be likely to show SR at an optimal level of envelope noise. Independent of the carrier level, SR would only be observed at the output of this neuron when the modulation signal was weak, but not when the modulation signal was strong. On the other hand, if a single auditory-nerve neuron shows SR, it would only occur at the lowest carrier levels, near the threshold for the *carrier*. The data presented here indicate that the effect occurs at higher carrier levels as well, and only when *modulation* threshold is compromised. Based on these considerations, it seems reasonable to hypothesize that the continuum of effects of noise observed in these experiments arises at a relatively central stage of processing in the brain.

Cochlear implant listeners have a limited dynamic range of hearing. The level-and modulation-frequency dependence of their modulation sensitivity means that their range of useful levels is even narrower when it comes to processing dynamic aspects of stimuli. By improving modulation sensitivity under these conditions, external noise is likely to extend the range of amplitudes that can carry useful information.

ACKNOWLEDGMENTS

We acknowledge Dr. Robert P. Morse for many helpful discussions about stochastic resonance. We thank Mark E. Robert for software support. We also thank Cochlear Corporation for providing us with the amplitude calibration information for the individual subjects. We are grateful to our subjects for their enthusiastic support of our research over the years. This work was funded by NIDCD Award No. R01DC004786.

Bahar, S., and Moss, F. (**2003**). "The nonlinear dynamics of the crayfish mechanoreceptor system," Int. J. Bifurcation Chaos Appl. Sci. Eng. **13** (8), 2013–2034.

Behnam, S. E., and Zeng, F. G. (**2003**). "Noise improves suprathreshold discrimination in cochlear implant listeners," Hear. Res. **186**(1–2), 91–93.

Chatterjee, M. (**2003**). "Modulation masking in cochlear implant listeners: Envelope vs tonotopic components," J. Acoust. Soc. Am. **113** (4), 2042–2053.

Chatterjee, M., and Robert, M. E. (**2001**). "Noise enhances modulation sensitivity in cochlear implant listeners: Stochastic resonance in a prosthetic sensory system?," J. Assoc. Res. Otolaryngol. **2**(2), 159–171.

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**(3), 1637–1644.

Collins, J. J., Chow, C. C., and Imhoff, T. T. (**1995**). "Stochastic resonance without tuning," Nature (London) **378** (6555), 341–342.

Douglas, J. K., Wilkens, L., Pantazelou, E., and Moss, F. (**1993**). "Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance," Nature (London) **365**, 337–340.

Dynes, S. B. C., and Delgutte, B. (**1992**). "Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea," Hear. Res.

58, 79–90.
Frisina, R. D. (**2001**) "Subcortical neural coding mechanisms for auditory temporal processing," Hear. Res. **158**(1–2) 1–27.
Fu, Q. J. (**2002**). "Temporal processing and speech recognition in cochlear implant users," NeuroReport **13**(13), 1635–1639.
Hartmann, R., Topp, G., and Klinke, R. (**1984**) "Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea," Hear. Res. **13**(1) 47–62.
Henry, K. R. (**1999**). "Noise improves transfer of near-threshold, phase-locked activity of the cochlear nerve: Evidence for stochastic resonance?" J. Comp. Physiol. [A] **184**(6), 577–584.
Hong, R. S., Rubinstein, J. T., Wehner, D., and Horn, D. (**2003**). "Dynamic range enhancement for cochlear implants," Otol. Neurotol. **24**(4), 590–595.
Indresano, A. A., Frank, J. E., Middleton, P., and Jaramillo, F. (**2003**). "Mechanical noise enhances signal transmission in the bullfrog sacculus," J. Assoc. Res. Otolaryngol. **4**(3), 363–370.
Jaramillo, F., and Wiesenfeld, K. (**1998**). "Mechanoelectrical transduction assisted by Brownian motion: A role for noise in the auditory system," Nat. Neurosci. **1**(5), 384–388.
Johnson, D. H., and Kiang, N. Y. (**1976**). "Analysis of discharges recorded simultaneously from pairs of auditory nerve fibers," Biophys. J. **16**(7), 719–734.
Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (**1965**). "Discharge patterns of single fibers in the cat's auditory nerve," *Research Monographs 35* (MIT Press, Cambridge, MA).
Kitajo, K., Nozaki, D., Ward, L. M., and Yamamoto, Y. (**2003**). "Behavioral stochastic resonance within the human brain," Phys. Rev. Lett. **90**(21), 218103.
Litvak, L., Delgutte, B., and Eddington, D. K. (**2001**). "Auditory nerve fiber responses to electrical stimulation: Modulated and unmodulated pulse trains," J. Acoust. Soc. Am. **110**(1), 368–379.
Litvak, L., Delgutte, B., and Eddington, D. K. (**2003a**). "Improved temporal coding of sinusoids in electric stimulation of the auditory nerve using desynchronizing pulse trains," J. Acoust. Soc. Am. **114**(4), 2079–2098.
Litvak, L., Delgutte, B., and Eddington, D. K. (**2003b**). "Improved neural representation of vowels in electric stimulation using desynchronizing pulse trains," J. Acoust. Soc. Am. **114**(4), 2099–2111.
Loizou, P. (**1998**). "Mimicking the human ear," IEEE Signal Process. Mag. **15**(5), 101–130.
Matsuoka, A. J., Abbas, P. J., Rubinstein, J. T., and Miller, C. A. (**2000**). "The neuronal response to electrical constant-amplitude pulse train stimulation: Additive Gaussian noise," Hear. Res. **149**, 129–137.
Miller, C. A., Abbas, P. J., and Robinson, B. K. (**2001**). "Response properties of the refractory auditory nerve fiber," J. Assoc. Res. Otolaryngol. **2**, 216–232.
Morse R. P., and Evans E. F. (**1996**). "Enhancement of vowel coding for cochlear implants by addition of noise," Nat. Med. **2**(8), 928–932.
Morse, R. P., Allingham, D., and Stocks, N. G. (**2003**). "An information-theoretic approach to cochlear implant coding," in *Unsolved Problems of Noise*, edited by S. Bezrukov (AIP, Melville, NY), pp. 125–132.
Moss, F., Chiou-Tan, F., and Klinke, R. (**1996**). "Will there be noise in their ears?," Nat. Med. **2**(8), 860–862.
Moss, F., Ward, L. M., and Sannita, W. G. (**2004**). "Stochastic resonance and sensory information processing: A tutorial and review of application," Clin. Neurophysiol. **115**(2), 267–281.
Robert, M. E. (**2002**) "House Ear Institute Nucleus Research Interface User's Guide," House Ear Institute, Los Angeles.
Rubinstein, J. T., Wilson, B. S., Finley, C. C., and Abbas, P. J. (**1999**). "Pseudospontaneous activity: Stochastic independence of auditory nerve fibers with electrical stimulation," Hear. Res. **127**, 108–118.
Runge-Samuelson, C. L., Abbas, P. J., Rubinstein, J. T., Miller, C. A., and Robinson, B. K. (**2004**). "Response of the auditory nerve to sinusoidal electrical stimulation: Effects of high-rate pulse trains," Hear. Res. **194** (1–2), 1–13.
Shannon, R. V. (**1992**). "Temporal MTFs in patients with cochlear implants," J. Acoust. Soc. Am. **91**(4, Pt 1), 2156–2164.
Shannon, R. V., Adams, D. D., Ferrel, R. L., Palumbo, R. L., and Grandgenett M. (**1990**) "A computer interface for psychophysical and speech research with the Nucleus cochlear implant," J. Acoust. Soc. Am. **87**, 905–907
Simonotto, E., Riani, M., Seife, C., Roberts, M., Twitty, J., and Moss, F. (**1997**). "Visual perception of stochastic resonance," Phys. Rev. Lett. **78**(6), 1186–1189.
Simonotto, E., Spano, F., Riani, M., Ferrari, A., Levrero, F., Pilot, A., Renzetti, P., Parodi, R. C., Sardanelli, F., Vitali, P., Twitty, J., Chiou-Tan, F., and Moss, F. (**1999**). "fMRI studies of visual cortical activity during noise stimulation," Neurocomputing **26–27**, 511–516.
Stocks, N. G. (**2001**). "Information transmission in parallel arrays of threshold elements: Suprathreshold stochastic resonance," Phys. Rev. E **63**, 041114: 1-9.
Ward, L. M., Neiman, A., and Moss, F. (**2002**) "Stochastic resonance in psychophysics and in animal behavior," Biol. Cybern. **87**, 91–101.
White, J. A., Rubinstein, J. T., and Kay, A. R. (**2000**). "Channel noise in neurons," Trends Neurosci. **23**(3), 131–137.
Zeng, F. G., Fu, Q. J., and Morse, R. (**2000**). "Human hearing enhanced by noise," Brain Res. **869**(1–2), 251–255.
Zeng, F. G., Galvin, J. J. III, and Zhang, C. Y. (**1998**). "Encoding loudness by electric stimulation of the auditory nerve," NeuroReport **9**(8), 1845–1848.

Tactual display of consonant voicing as a supplement to lipreading

Hanfeng Yuan,[a] Charlotte M. Reed, and Nathaniel I. Durlach
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 7 June 2004; revised 23 April 2005; accepted 12 May 2005)

This research is concerned with the development and evaluation of a tactual display of consonant voicing to supplement the information available through lipreading for persons with profound hearing impairment. The voicing cue selected is based on the envelope onset asynchrony derived from two different filtered bands (a low-pass band and a high-pass band) of speech. The amplitude envelope of each of the two bands was used to modulate a different carrier frequency which in turn was delivered to one of the two fingers of a tactual stimulating device. Perceptual evaluations of speech reception through this tactual display included the pairwise discrimination of consonants contrasting voicing and identification of a set of 16 consonants under conditions of the tactual cue alone (T), lipreading alone (L), and the combined condition (L+T). The tactual display was highly effective for discriminating voicing at the segmental level and provided a substantial benefit to lipreading on the consonant-identification task. No such benefits of the tactual cue were observed, however, for lipreading of words in sentences due perhaps to difficulties in integrating the tactual and visual cues and to insufficient training on the more difficult task of connected-speech reception. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1945787]

PACS number(s): 43.66.Ts, 43.66.Sr, 43.66.Wv [KWG] Pages: 1003–1015

I. INTRODUCTION

One class of speech-communication aids for persons with profound hearing impairment or deafness is based on the transformation of information about speech (e.g., its acoustic or articulatory properties) into vibratory or electrical patterns presented to the skin. A variety of different tactual devices have been designed and evaluated over the years. Gault (1924, 1926) employed the diaphragm of a special telephone receiver to deliver unprocessed sound vibrations to the fingers or hands. Since then, various artificial tactile displays have been developed as aids for speech communication (e.g., see reviews by Kirman, 1973; Reed *et al.*, 1982, 1989; Bernstein, 1992). These devices have included different types of stimulation (either mechanical or electrocutaneous), have been applied to a variety of body sites (e.g., finger, hand, forearm, abdomen, thigh), and have employed various numbers and configurations of stimulators (e.g., single-channel or multi-channel stimulation, linear or two-dimensional arrays).

Many of these tactual displays have been evaluated in terms of their ability to supplement information that is available through lipreading. Lipreading, which is an important source of speech information for many hard-of-hearing and deaf individuals, is largely based on the visibility of various lip features that arise during the articulation of speech (Jeffers and Barley, 1971). Perceptual studies of vowel and consonant reception through lipreading alone (e.g., Jackson *et al.*, 1976; Owens and Blazak, 1985) indicate that phonemes are classified into groups of stimuli with identical or highly similar articulatory shapes and movements (referred to as visemes). This impoverished segmental information contributes to the difficulty experienced by most lipreaders in the reception of speech through the lipreading signal alone. Among the various important features of speech, voicing is poorly perceived through lipreading alone (see Heider and Heider, 1940; Erber, 1974; Walden *et al.*, 1977; Bratakos *et al.*, 2001). By providing supplemental information for voicing with lipreading, the viseme groups can be reduced to smaller groups, thereby leading to improved lipreading ability (see Auer and Bernstein, 1996; Grant *et al.*, 1998).

A number of previous studies have been concerned with the tactual display of fundamental frequency (F0) to provide information about voicing as a supplement to speechreading (Grant *et al.*, 1986; Hanin *et al.*, 1988; Eberhardt *et al.*, 1990; Summers *et al.*, 1994; Waldstein and Boothroyd, 1995a, b; Auer *et al.*, 1998). At the segmental speech level, F0 provides an acoustic manifestation of the presence or absence of vocal-fold activity which is highly related to voicing. Auditory displays of F0 have been shown to provide substantial benefits to performance through lipreading alone, with improvements of roughly 30 to 40 percentage points for the reception of words in sentences (e.g., see Breeuwer and Plomp, 1986; Boothroyd *et al.*, 1988). Information about F0 has been encoded in a variety of ways for presentation through the tactual sense, including (a) rate of vibration to a single transducer (Hanin *et al.*, 1988), (b) frequency and/or amplitude modulation of pulse-train signals (Summers *et al.*, 1994), (c) place of stimulation in a multi-channel array of transducers (Hanin *et al.*, 1988; Waldstein and Boothroyd, 1995a, b), and (d) vertical displacement of a finger rest (Waldstein and Boothroyd, 1995b). Generally, the results of speech-reception studies with these tactual displays indicate

[a]Electronic mail: hfyuan@mit.edu

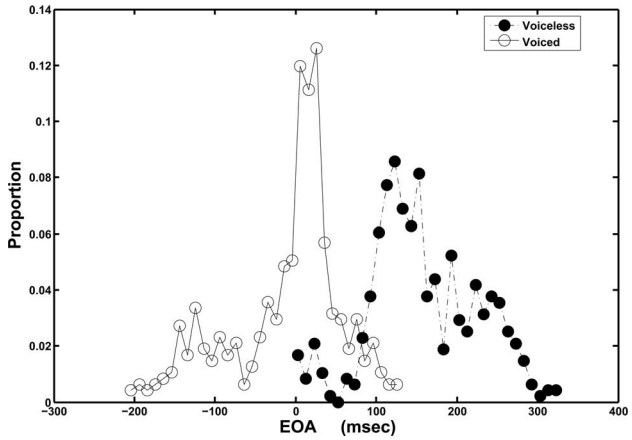


FIG. 1. Probability distribution of EOA for two categories of consonants: voiceless versus voiced. Within the voiceless category, measurements were made on 480 tokens pooled across the eight voiceless consonants. Likewise, 480 tokens were used in measurements within the voiced category.

modest improvements to speechreading at both the segmental and connected-speech levels, although not to the same extent as has been observed with auditory F0 supplements. The percentage-point improvements to lipreading typically observed with tactual displays of F0 are roughly one-fifth to one-third of those observed with auditory displays of similar information (e.g., see Hanin *et al.*, 1988; Kishon-Rabin *et al.*, 1996).

The current research was concerned with developing an improved tactual display of voicing that would lead to a higher level of voicing discriminability at the segmental level than that achieved previously. Improvements at the segmental level may be expected to translate into improvements in aided lipreading of connected speech. In particular, the current work was concerned with the development and evaluation of a tactual display based on a temporal cue to voicing. In previous work we demonstrated that a simply-derived, real-time cue based on the envelope-onset asynchrony (EOA) of two different filtered bands of speech serves as a reliable source of information about voicing for initial-position obstruents (Yuan *et al.*, 2004). Acoustic measurements of the onset-timing difference between a low-frequency envelope (350-Hz low-pass) and a high-frequency envelope (3000-Hz high-pass) derived from speech syllables indicated a different pattern of EOA values for voiced versus voiceless consonants. The probability distributions of EOA (defined as the onset time of the low-pass envelope minus the onset time of the high-pass envelope) are shown in Fig. 1 for two categories of consonants: voiced versus voiceless. Measurements are shown for initial consonants in 960 C_1VC_2 syllables (2 talkers $\times$ 16 C_1 $\times$ 3 V $\times$ 10 repetitions). The two distributions are well separated. For voiceless consonants, the high-frequency envelope tends to precede the low-frequency envelope (with a mean EOA of roughly 154 ms and s.d. of 66 ms). For voiced consonants, on the other hand, the two envelopes tend to occur roughly simultaneously or with the low-frequency envelope preceding the high-frequency envelope (with a mean EOA of roughly -14 ms and s.d. of 70 ms). Distributions of EOA values for pairs of voiced-voiceless consonant contrasts were used to estimate the sensitivity of an ideal observer for voicing discrimination using the same set of syllables whose EOA measurements are shown in Fig. 1. Resulting d' values for each of eight pairs of voicing contrasts ranged from 4.0 (for /f-v/) to 13.0 for /s-z/. These results indicate excellent sensitivity, corresponding to percent-correct scores for unbiased performance in the range of roughly 98%–100% correct. Therefore, the onset asynchrony of the two envelopes is expected to be a good indicator of the voicing distinction. This cue, which is independent of manner of production, can be implemented in real time and synchronized with the lipreading signal.

The feasibility of the EOA cue as a tactual supplement to lipreading is obviously dependent on the temporal resolution of the tactual sense. In a previous study (Yuan *et al.*, 2005), tactual temporal onset-order thresholds were measured for two sinusoidal vibrations of different frequencies delivered to two channels (thumb and index finger) of a multi-finger tactual stimulating device. The frequency delivered to the thumb was fixed at 50 Hz and that to the index finger at 250 Hz. The amplitudes and durations of the two sinusoidal vibrations were selected to simulate those of the envelopes derived from the speech waveforms. The temporal-onset-order threshold averaged 34 ms across the four subjects who participated in the study. This temporal resolution appears to be sufficient to distinguish different patterns of envelope-onset asynchrony associated with the feature voicing. This result is encouraging for the development of signal-processing and display schemes for encoding this temporal speech property (EOA) through the tactual sense.

The current study is concerned with the development and evaluation of a tactual display of the EOA cue to supplement lipreading. This EOA cue was presented through a two-finger tactual display such that the envelope of the high-frequency band was used to modulate a 250-Hz carrier signal delivered to the index finger and the envelope of the low-frequency band was used to modulate a 50-Hz carrier delivered to the thumb. Three perceptual evaluations of speech reception through this tactual display were examined under three presentation conditions: tactual display alone (T), lipreading alone (L), and the combined condition (L+T). These evaluations included training and testing on the following tasks: discrimination of pairs of consonants contrasting voicing (experiment 1), identification of a set of 16 consonants (experiment 2), and recognition of CUNY sentences (experiment 3). Training is an important component of the evaluation procedure and is necessary to familiarize subjects with the novel tactual cue. At the segmental level (experiment 1 and 2), training was provided in the form of correct-answer feedback on the discrimination and identification tasks. Sentence testing was also conducted to determine the extent to which segmental-level training might transfer to the task of connected-speech recognition.

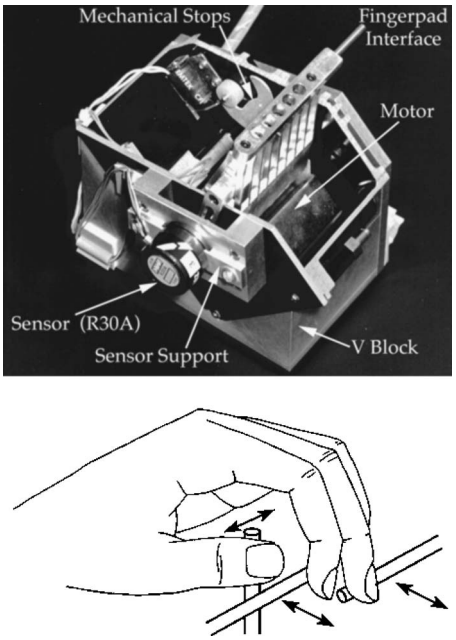


FIG. 2. Illustration of the Tactuator stimulating device. The upper panel is a photograph of one of the three motor assemblies with labeled components. The lower panel is a schematic drawing illustrating finger placement on three vibrating rods of the Tactuator device (taken from Tan, 1996).

II. GENERAL METHODOLOGY

A. Tactual stimulating device

The tactual stimulating device used in the experiments (referred to as the Tactuator) was initially developed by Tan (1996) for research with multidimensional tactual stimulation. A complete discussion of this system is provided by Tan (1996) and Tan and Rabinowitz (1996), which includes a detailed description of the hardware components, controller components, and performance characteristics of the device. For the current research, the original system was upgraded with a new computer, DSP system, and electronic control system to improve its performance capabilities. A complete description of the characteristics of the modified system is available in Yuan (2003).

The device is a three-finger display capable of presenting a broad range of tactual movement to the passive human fingers. It consists of three mutually perpendicular rods that interface with the thumb, index finger, and middle finger in a manner that allows for a natural hand configuration (see bottom panel of Fig. 2). A photograph of the motor assembly (with labeled components) associated with one of the rods is provided in the upper panel of Fig. 2. Each rod is driven independently by an actuator that is a head-positioning motor from a hard-disk drive. The position of the rod is controlled by an external voltage source to the actuator and is measured by an angular position sensor that is attached to the moving part of each of the three motor assemblies. The rods are capable of moving the fingers, which rest lightly on the rods, in an outward (extension) and inward (flexion) direction relative to a neutral resting position.

The overall performance of the device is well suited for psychophysical studies of the tactual sensory system. First, the device is capable of delivering frequencies along a continuum from dc to 300 Hz, allowing for stimulation in the kinesthetic (low-frequency) and cutaneous (high-frequency) regions, as well as in the mid-frequency range. Second, the range of motion provided by the display for each digit is roughly 26 mm. This range allows delivery of stimulation at levels from threshold to approximately 50 dB SL throughout the frequency range from dc to 300 Hz. Third, each channel is highly linear, with low harmonic distortion and negligible interchannel crosstalk. Fourth, loading (resulting from resting a finger lightly on a moving bar of the actuator) does not have a significant effect on the magnitude of the stimulation.

B. Speech materials

The speech materials used in this study include audiovisual recordings of C_1VC_2 nonsense syllables and CUNY sentences.

All speech materials were digitized for real-time speech signal processing and for rapid-access control in the experiments. The audio-visual stimuli were originally available on videotapes. These analog recordings were digitized using the Pinnacle DV500 Plus system and then stored in individual files on the host computer. The video was digitized in NTSC standard, with frame size 720×480, frame rate of 29.97 frames per second, and pixel depth of 24 bits. To reduce their size, the segmented files were compressed using the Sorenson Video compressor and converted to QuickTime format. The audio was set to 22 050 samples/second, with 16-bit resolution. The files were played by QuickTime.

1. Nonsense syllables

The nonsense syllables were balanced for the initial consonant C_1. The initial consonant C_1 was selected from 16 values of C_1=/p t k b d g f θ s ʃ v ð z ʒ tʃ dʒ/ and the middle vowel V was selected from three values of V =/i a u/. The final consonant C_2 was selected randomly for each syllable from a set of 21 consonants: /p t k b d g f θ s ʃ v ð z ʒ tʃ dʒ m n ŋ r l/. The speakers for these recordings were two females, SR and RK, both of whom were teachers of the deaf and were approximately 30 years of age at the time of the recordings. Each talker recorded ten lists of syllables containing one representation of each C_1V combination. The total corpus consists of 960 C_1VC_2 syllables representing 60 tokens of each of the 16 values of C_1 (2 speakers $\times$ 10 lists/speaker $\times$ 3 vowels). Each value of C_2 was roughly equally distributed among the 16 values of C_1, that is, there were approximately three representations of each C_2 among the 60 syllables containing each C_1.

2. CUNY sentences

The CUNY sentences were recorded by a female talker onto laser videodisc (Boothroyd *et al.*, 1985) and consist of 60 lists of 12 sentences each. The length of the sentences ranges from 3 to 14 words and each list contains 102 words (all of which are used in scoring). The CUNY sentences are considered to be of easy-to-medium difficulty because of

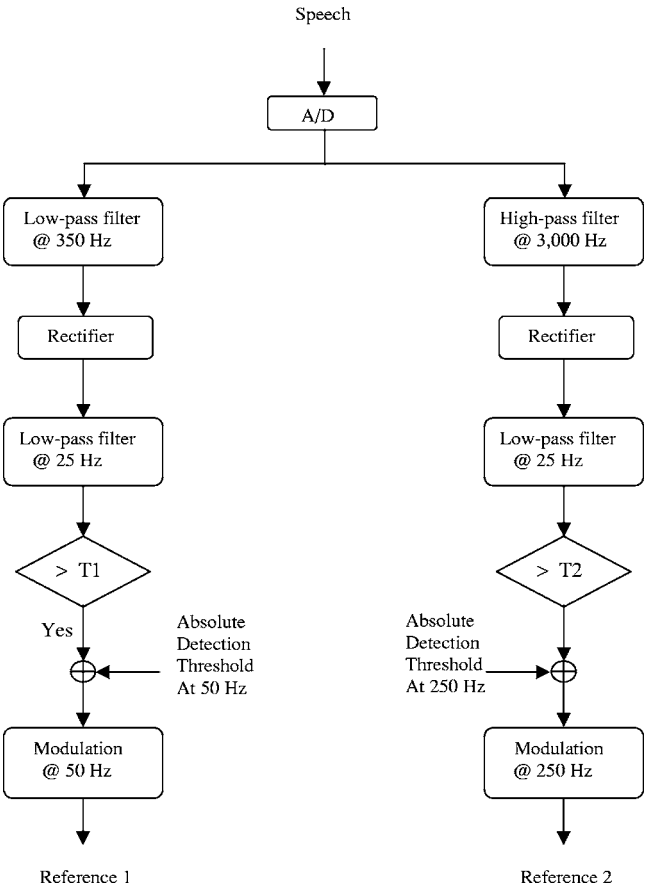


FIG. 3. A flowchart of the envelope-extraction algorithm.

their conversational style. Although the sentences in each list are arranged by topic, these topics were not made known to the subjects in this study.

C. Derivation of envelope-onset asynchrony (EOA) cue

Envelope-onset asynchrony (EOA) is believed to carry reliable information about voicing for English obstruent consonants in initial position (Yuan *et al.*, 2004). The derivation of EOA from the speech signal and its presentation through a multi-finger tactual display is shown in Fig. 3. The multimedia file of the token was played through QuickTime for Java. The speech sound from the audio channel of the host computer was routed to the input of the Audio PMC IO card. The acoustic input was sampled at 48 000 samples/second by an A/D converter. The speech samples were then processed through two parallel branches of the DSP board.

In branch 1, speech samples were passed through a discrete second-order Butterworth low-pass filter with a cutoff frequency of 350 Hz. This band was selected to monitor the presence or absence of low-frequency energy that accompanies glottal vibration. The low-pass-filtered speech samples were then rectified and smoothed by a discrete second-order Butterworth low-pass filter with a cutoff frequency of 25 Hz. A threshold (T1) was established for the level of the smoothed amplitude envelope (A1) in order to eliminate envelope signals arising primarily from random noise fluctuations in the passband, but yet sufficient for passing signals driven by the acoustic speech waveform. Values of amplitude envelope below this threshold were set to zero. Values of the amplitude envelope samples above this threshold were added to the average threshold for a 50-Hz signal at the left thumb (1 dB *re* 1μ peak).[1] The resulting amplitude envelope signals generally fell in the range of 21 to 51 dB *re* 1 μ peak and were well above threshold for each of the four subjects. Finally, the smoothed amplitude envelope modulated a 50-Hz sinusoid and was converted to an analog signal through a D/A converter. This modulated low-frequency signal was then routed to the reference signal of the PID controller to drive the thumb.

In branch 2, speech samples were passed through a discrete second-order Butterworth high-pass filter with a cutoff frequency of 3000 Hz. This band was selected to monitor the presence or absence of high-frequency energy that accompanies aspiration, frication, and burst characteristics of consonants. The high-pass-filtered speech samples were then rectified and smoothed by a discrete second-order Butterworth low-pass filter with a cutoff frequency of 25 Hz. A threshold (T2) was established for the level of the smoothed amplitude envelope (A2) to eliminate envelope signals arising from random-noise fluctuations in the passband, but yet sufficiently low to pass signals arising from speech energy in the passband. Values of the amplitude envelope below this threshold were set to zero. Values of the amplitude envelope above this threshold were added to the average absolute-detection threshold for a 250-Hz signal at the left index finger (−19 dB *re* 1 μ peak).[1] The resulting amplitude envelopes generally fell in the range of +1 to 31 dB *re* 1μ peak, again well above threshold of each of the four subjects. The smoothed envelope signal modulated a 250-Hz sinusoid and was passed through a D/A converter. The resulting signal was then routed to the reference signal of the PID controller to drive the index finger.

The derivation of EOA shown in the flowchart in Fig. 3 is simple and can be implemented in real time, such that the tactual display of EOA can be synchronized with lipreading. The specific scheme for the presentation of the EOA cue was designed to optimize the tactual delivery of EOA information: (1) the two modulated envelopes were presented to two different fingers with the assumption that the amount of cross-finger masking is less than that of same-finger masking (Verrillo *et al.*, 1983) and (2) the two amplitude envelopes were modulated by two different frequencies with the belief that the farther the distance between the two modulating frequencies the less the amount of masking between them (Gescheider *et al.*, 2001; Tan *et al.*, 2003).

The selection of the two modulating frequencies of 50 and 250 Hz for the low- and high-pass bands, respectively, was based on the following criteria: (1) The two modulating envelopes have the same order as the two frequency regions from which the envelopes are extracted, i.e., the carrier with higher frequency was modulated by the amplitude envelope from the higher frequency band, while the carrier with lower frequency was modulated by the amplitude envelope from the lower frequency band; (2) the two frequencies are perceptually distinct (see Tan, 1996); and (3) in general, the

duration of the amplitude envelope signals exceeds 20 ms, allowing each modulating envelope to contain at least one cycle.

D. Experimental conditions

Experimental conditions included testing under three different conditions: tactual cue alone (T), lipreading alone (L), and the combined condition (L+T). For conditions of L and L+T, the video image of the talker was displayed on a 19-in. color video monitor. The subject was seated roughly 0.8 m in front of the video monitor. For T and L+T conditions, subjects were seated 0.5 m from the Tactuator and placed the thumb and index finger of the left hand on the device. The tactual cue for aided lipreading in the L+T condition was presented simultaneously with the video signal. To eliminate any auditory cues from the vibration of the Tactuator, subjects wore foam earplugs that were designed to provide 30-dB attenuation and also wore earphones that delivered pink masking noise at an overall level of roughly 90 dB SPL.

E. Subjects

The same four individuals ranging in age from 21 to 32 years (three male and one female) served as subjects in all three experiments in this study. Subjects were screened for absolute-threshold detection before participating in the experiments. Hearing tests were conducted through the Audiology Department at the MIT Medical Department before the start of the experiment (to provide a baseline for each subject's hearing level prior to exposure to masking noise) and at the completion of the study (for comparison with the baseline results). No threshold shifts were noted for any of the subjects. None of the subjects had previous experience in experiments concerned with lipreading or tactual perception. Subjects S1, S2, and S3 require corrective lenses for normal vision and wore either glasses or contact lenses for the lipreading experiments. Subjects were paid for their participation in the study and received periodic bonuses (that were unrelated to performance) throughout the course of the study.

III. EXPERIMENT 1: PAIRWISE DISCRIMINATION

The ability to discriminate consonant voicing was examined for each of eight pairs of English consonants that contrast the feature of voicing: /p-b/, /t-d/, /k-g/, /f-v/, /θ-ð/, /s-z/, /ʃ-ʒ/, and /tʃ-dʒ/.

A. Stimuli

The stimuli used in this experiment were the C_1VC_2 nonsense syllables described in Sec. II B. The tokens were subdivided into two different sets for use in this experiment: a "training" set and a "testing" set. The training set consisted of 576 tokens (6 tokens per talker for each of the 48 C_1V combinations). The testing set consisted of the remaining 384 tokens (4 tokens per talker for each of the 48 C_1V combinations). In other words, 36 tokens (2 speakers $\times 3$ vowels $\times 6$ repetitions) of each initial consonant were used in the "training" set, and 24 tokens (2 speakers $\times 3$ vowels $\times 4$ repetitions) of each initial consonant were used in the "testing" set.

B. Procedure

The ability to discriminate consonant voicing for each of the eight pairs of consonants was tested separately under the three conditions of L, T, and L+T. The tests were conducted using a two-interval, two-alternative, forced-choice (2I-2AFC) procedure. Some conditions of the experiment employed trial-by-trial correct-answer feedback while others did not. In each trial, one of the tokens representing each of the two consonants in a given pair was selected at random from either the "training" set or the "testing" set with replacement. The two tokens were presented in one of two possible stimulus orders (voiced followed by voiceless consonant or voiceless followed by voiced consonant) selected at random with equal *a priori* probability. The subject was instructed to report the order of the presentation by clicking an icon labeled "voiced-voiceless" if he/she perceived that the token with voiced initial consonant was presented in the first interval, or clicking an icon labeled "voiceless-voiced" for the other alternative.

Each interval of a trial had a 2-s duration during which a randomly selected audiovisual file was played. The duration of the files ranged from roughly 300 to 930 ms, averaging 800 ms. The effective interstimulus interval (ISI) was the duration between the closing of the lips in the file in interval 1 and the opening of the lips in the file in interval 2. It was thus a random variable (with mean=530 ms and s.d. =315 ms). A blue background appeared on the screen in the period between the offset of the file in the first interval and the onset of the file in the second interval.

Data were collected in five replications in which the eight pairs of consonants were tested separately under each of the three conditions. The order in which L and T were tested was chosen randomly for each consonant contrast; the L+T condition was always tested last. Each run consisted of 60 trials. The purpose of this design was to provide subjects with training on the task and then to test posttraining performance. In the first replication, the eight stimulus pairs were tested in order of increasing difficulty based on the results of an informal pilot study (i.e., /s-z/, /ʃ-ʒ/, /θ-ð/, /p-b/, /t-d/, /k-g/, /f-v/ and /tʃ-dʒ/). In the remaining four replications, the stimulus pairs were tested in a randomly determined order. The speech tokens from the "training" set were employed in replications 1–3. In the first two replications, training was provided through the use of trial-by-trial correct-answer feedback. (Pilot data obtained on the pairwise discrimination task indicated that performance typically reached asymptotic levels within the first 50–100 trials.) In replication 3, subjects performed the task without feedback using the tokens from the training set. The final two replications of the experiment tested subjects' ability to transfer their training to a fresh set of stimuli (the "testing" tokens) without feedback. Each run took roughly 10 min; on average each subject required ten 2-h sessions over ten different days to complete the experiment.

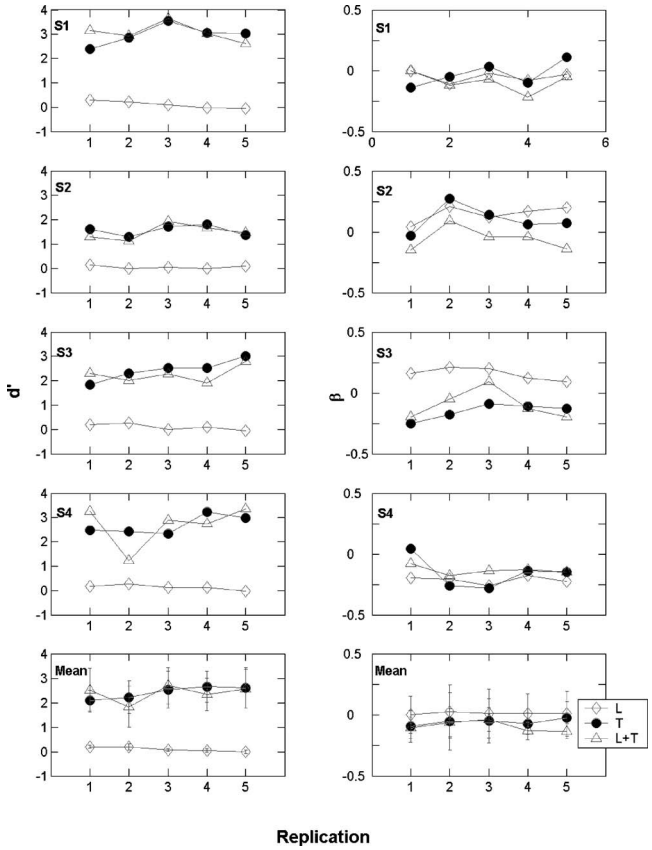


FIG. 4. In the five panels on the left, d' scores for 2I-2AFC discrimination are plotted as a function of replication under three conditions for each individual subject and for means across subjects. In the five panels on the right, β is plotted as a function of replication. Each data point represents the average across eight pairs; error bars represent ±1 s.d. Unfilled diamonds represent lipreading alone (L), filled circles represent tactual condition (T), and unfilled triangles represent the combined condition (L+T).

C. Data analysis

Results for each experimental run (4 subjects $\times 8$ consonant pairs $\times 3$ conditions $\times 5$ replications) were summarized in terms of a 2×2 stimulus-response confusion matrix. Signal-detection measures of sensitivity (d') and bias (β) (Green and Swets, 1966; Durlach, 1968) for each matrix were computed assuming equal variance Gaussian distributions. Values of d' and β were computed for each of the eight consonant pairs for each subject under each of the five replications.

D. Results

Mean and individual performance in d' (left panels) and β (right panels) are provided in Fig. 4 for each of the five replications of the experiment. Each of the top four rows of Fig. 4 represents the results for one of the four subjects and the bottom row represents the results averaged across the four subjects. For each subject, values of d' and β were averaged across the eight consonant pairs for each of the three conditions (L, T, L+T) under each replication.

A clear and consistent effect was observed for condition. The d' for L was near 0 for each subject under each replication, indicating that performance was at chance level. This result is consistent with the well-known observation that lipreading carries little if any information about voicing for consonants in initial position. Performance under the conditions of T and L+T was similar within a given subject and ranged from d' of roughly 1.5 for S2 to roughly 3.0 for S1. Averaged over subjects and replications, d' values were roughly 0.09 for L, 2.4 for T, and 2.4 for L+T.

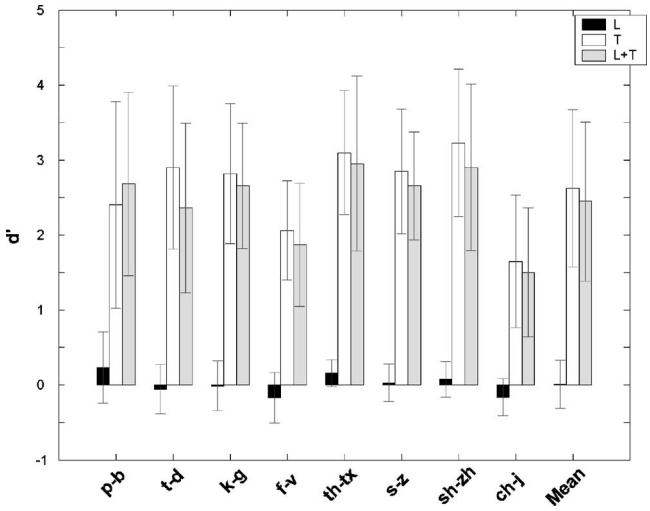


FIG. 5. Values of d' averaged across the four subjects for conditions of no-feedback with the test tokens (replications 4 and 5) as a function of consonant pair and condition in the 2I-2AFC discrimination experiment. Filled bars represent lipreading (L), unfilled bars represent tactual condition (T), and shaded bars represent the combined condition (L+T). In the labels of the plot, th stands for $|\theta|$, tx stands for $|\eth|$, sh stands for $|\int|$, zh stands for $|ʒ|$, ch stands for $|t\int|$, and j stands for $|dʒ|$.

The effects of token set and feedback on average performance can also be examined from the data shown in Fig. 4. There seems to be a slight improvement in performance in replication 3 (with "training" set and no feedback) compared to performance in replications 1 and 2 (with "training" set and feedback). The improvement due to training seems to saturate quickly. No apparent difference is found for no-feedback performance between the "test" set (replications 4 and 5 without feedback) and the "training" set (replication 3 without feedback). This result indicates that training conducted with a large set of tokens (multiple repetitions in three vowel contexts from 2 speakers) was sufficient for generalization to a "fresh" set of utterances by the same speakers in the same vowel contexts in a simple discrimination task.

Minimal effects of response bias were observed. For individual subjects, β was in the range of −0.3 to +0.3. Averaged across subjects, β ranged from −0.15 to +0.05 across replications.

Values of d' averaged over replications 4 and 5, and across the four subjects, are shown in Fig. 5 for each of the eight pairs of consonant contrasts under each of the three conditions. A clear and consistent effect was observed for condition. The d' for L was near 0 for each of the eight pairs, indicating that performance was at chance level. Performance under the conditions of T and L+T was similar within each pair: difference in d' never exceeded 0.5. A one-way ANOVA indicated no significant difference in d' scores between T and L+T conditions $[F(1,126)=0.92; p=0.3391]$.

Interpair variability was observed on T and L+T. Performance ranged from d' of roughly 3.3 for the pair /ʃ-ʒ/ to roughly 1.6 for the pair /tʃ-dʒ/.

E. Discussion

Although it is difficult to make direct comparisons between the current results and those obtained in previous studies, the voicing-discriminability performance achieved with the tactual display of an EOA cue appears to compare quite favorably with that reported in other studies of tactual displays. Reed *et al.* (1992) reported results for pairwise discrimination of initial consonants through two different tactual aids in which spectral information is encoded through site of stimulation on a two-channel (Tactaid 2) or seven-channel (Tactaid 7) array of vibrators. Performance for pairs contrasting voicing was roughly 65% correct under a 1I-2AFC procedure for both of these devices. Waldstein and Boothroyd (1995a) obtained a similar raw score of roughly 60% correct for initial-consonant voicing through the Tactaid 7 alone (which improved to roughly 70% in conjunction with lipreading). Hnath-Chisolm and Kishon-Rabin (1988) compared the performance of two different tactual displays of fundamental-frequency (F0) information: a temporal, single-channel display and a spatial 16-channel display. Performance on initial-consonant voicing, which was similar for the two displays, averaged roughly 60% correct for the tactual cue alone and 68% for aided lipreading. Waldstein and Boothroyd (1995b) also reported results for the 16-channel F0 display, which averaged roughly 80% correct for the tactual cue alone or combined with lipreading. On average, the discriminability of roughly 90% observed in the current study represents an improvement over that obtained with previous tactual devices.

The four subjects tested in the voicing-discrimination task also participated in a psychophysical study of tactual temporal onset-order discrimination for sinusoidal signals presented to the index finger and thumb (see Yuan *et al.*, 2005). Across individual subjects, temporal onset-order thresholds ranged from roughly 18 to 43 ms. There appears to be a rough correspondence between subjects' abilities on the two tasks. The subject with the highest sensitivity in the onset-order discrimination task (i.e., lowest threshold in ms) was also the subject with the best performance in the voicing-discrimination task. Correlation coefficients computed under conditions T and L+T (Pearson's $r=0.66$ and 0.77, respectively) did not reach significance, however.

Yuan *et al.* (2005) computed the sensitivity of an ideal observer on a voicing-detection task using the EOA cue that was employed in the tactual display studied here and with limited temporal-order resolution of 34 ms (corresponding to the average threshold obtained in the temporal onset-order task). The sensitivity index d' of the ideal observer was computed for each of the eight pairs of voiced-voiceless contrasts using a corpus of speech materials that included 64 tokens of each initial consonant in C_1VC_2 syllables (2 speakers $\times$16 vowels$\times$2 repetitions). Averaged across the eight pairwise contrasts, the performance of the ideal observer led to d' values that were roughly 1.6 times those obtained by the human subjects. The performance of the ideal observer exceeded that of the human observers even though there was greater variability in the speech corpus arising from the use of 16 vowels as opposed to 3 vowels in the human experiments. This comparison indicates that the EOA cue is not optimally employed by the human subjects in performing the voicing-discrimination task. Limitations that may play a role in degrading human performance include transformations of the signal at the peripheral level of the tactual sensory system as well as limitations in memory and attention arising at a more central level of processing.

IV. EXPERIMENT 2: 16-CONSONANT IDENTIFICATON

The results of experiment 1 indicate that the tactual presentation of the EOA cue was highly effective for pairwise discrimination of initial voicing contrasts. The purpose of experiment 2 was to examine the contribution of this tactual voicing cue to the task of consonant identification. The ability to identify the initial consonant in C_1VC_2 syllables was tested using a one-interval, 16-alternative forced-choice (1I-16AFC) paradigm.

A. Speech stimuli

Tokens of the C_1VC_2 nonsense syllables were subdivided into two different sets for use in the two experiments: a "training" set and a "testing" set as described in Sec. III A.

B. Procedure

Consonant identification ability was studied under the three conditions of L, T, and L+T using a 1I-16AFC procedure. On each trial, one of the tokens from either the "training" set or "testing" set was selected at random with replacement. Each consonant was assigned a response code and the 16 response alternatives were displayed as a 4×4 array of labeled buttons on the computer screen following each trial. The subject's task was to select a response by using a computer mouse to click on the button corresponding to his/her response. No time limit was imposed on the subject's response. Trial-by-trial correct-answer feedback (used only in replications 1 and 2) was provided in the form of computer-generated text.

Data were collected in seven replications. Each replication involved the presentation of four consecutive runs of each of the three conditions. The order in which L and T were presented was chosen randomly. The L+T condition was always run last. Each run consisted of 80 trials. Subjects received training in replications 1 and 2 using the "training" set with trial-by-trial correct-answer feedback. In replication 3, subjects performed the task without feedback, using the "training" set of tokens. Testing was conducted in replications 4–7 using the "testing" set of tokens without feedback. On each trial, one of the tokens was selected randomly with replacement from the appropriate set (either "training set" or "testing set") with equal probability. Each run took roughly 12 min; the entire experiment required an average of six 2-h sessions over six different days for each subject.

TABLE I. Classification of the 16 consonants along features of voicing, place, and manner. "+" indicates that the consonant owns that feature.

		p	b	t	d	k	g	f	v	θ	ð	s	z	ʃ	ʒ	tʃ	dʒ
Voicing	Voiced		+		+		+		+		+		+		+		+
	Voiceless	+		+		+		+		+		+		+		+	
Place	Labial	+	+					+	+								
	Dental									+	+						
	Alveolar			+	+							+	+				
	Velar					+	+							+	+	+	+
Manner	Plosives	+	+	+	+	+	+										
	Fricatives							+	+	+	+	+	+	+	+		
	Affricates															+	+

C. Data analysis

Individual performance in percent-correct score was calculated for each of the seven replications for each condition. The results from the final four replications for each subject and each condition were used to construct 16×16 stimulus-response confusion matrices. Confusion matrices were also constructed for each of three features (voicing, manner, and place) by grouping the consonants into classes according to their definitions along each feature (see Table I for the feature classifications). Consonant confusions were analyzed in terms of information transfer (IT) (Miller and Nicely, 1955). For the full 16×16 matrices, results are presented in terms of the percentage of overall IT (%-IT). For each of the three features, results are presented in terms of the percentage of unconditional feature IT (%-feature IT). Percent IT measurements are preferred over simple percent-correct measures for their ability to quantify the covariance between stimuli and responses (see Miller and Nicely, 1955) and are widely used in evaluations of speech-reception aids for the deaf (e.g., see Dorman *et al.*, 1990; Rabinowitz *et al.*, 1992; Fu and Shannon, 1998).

D. Results

Mean and individual performance in percent-correct score is shown in Fig. 6 for each of the seven replications of the experiment. Each of the top four panels of Fig. 6 represents results for one of the four subjects. For each subject, data are shown for L (diamonds), T (circles), and L+T (triangles). Each data point represents the percent-correct identification score across the four runs (320 trials) collected at that particular replication. Results averaged across the four subjects are also shown in the bottom panel of Fig. 6. Chance performance on this task is 6.25%.

Performance followed the same order across subjects: T<L<L+T. All subjects demonstrated some type of learning effect from replication 1 to 2 (with feedback) for the two conditions L and L+T except S4. The improvement from replication 1 to 2 averaged roughly 7 percentage points for these two conditions. The learning effect was negligible for condition T. Performance on the "training" set did not decrease when feedback was eliminated (i.e., results of replication 3 were similar to those of replication 2). The performance using the "testing" set without feedback (replications 4–7) decreased by roughly 12 percentage points from scores using the "training" set without feedback (replication 3). This result suggests that the subjects may have learned to take advantage of some artifactual types of cues following repeated exposure to the "training" set both with and without feedback. Intersubject variance is most obvious under the condition L+T: performance varies from roughly 40% correct for S2 to roughly 60% correct for S3.

Confusion matrices for the set of 16 consonants were compiled for each subject and each condition using the no-feedback replications with the "test" tokens (i.e., replications

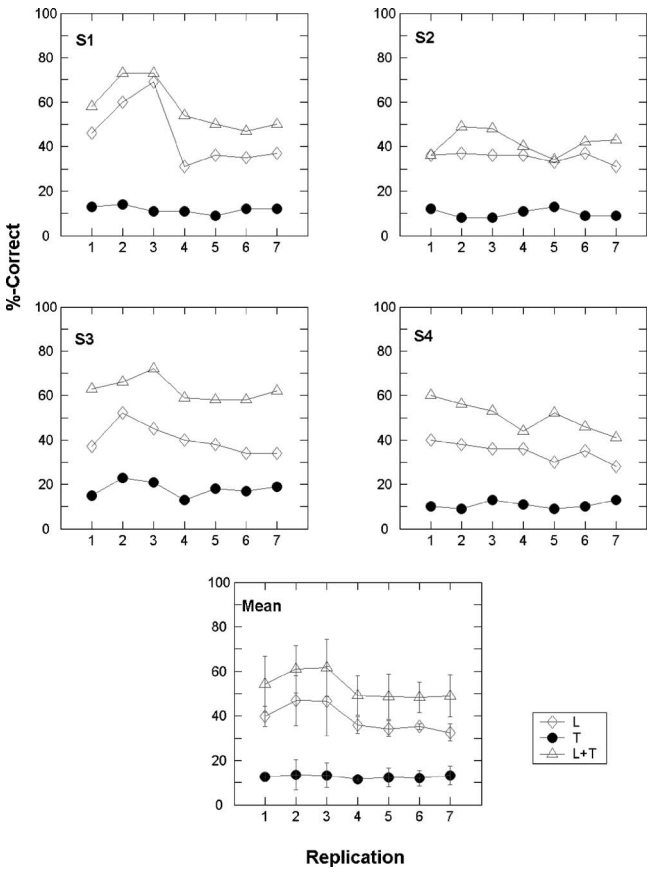


FIG. 6. Scores in %-correct versus replication number for the 16-consonant identification experiment are shown in separate panels for individual subjects and for means across subjects. Unfilled diamonds represent lipreading alone (L), filled circles represent tactual condition (T), and unfilled triangles represent the combined condition (L+T).

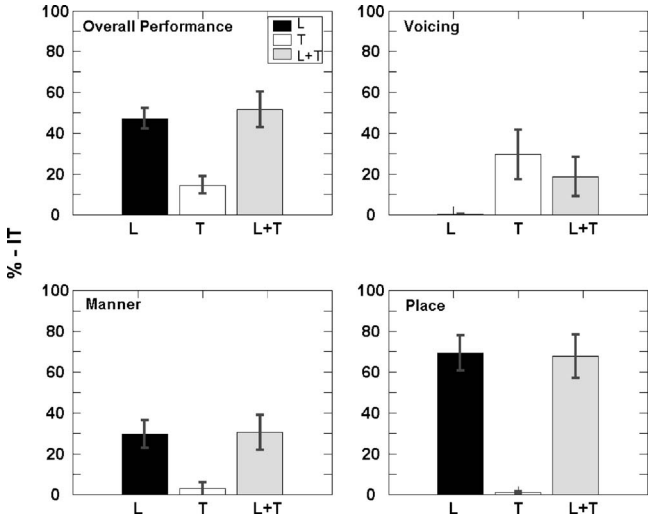


FIG. 7. Average results over four subjects and the final four replications in %-IT for the 16-consonant-identification experiment under three conditions for overall performance (top left panel), voicing (top right), manner (bottom left), and place (bottom right). Filled bars represent lipreading (L), unfilled bars represent the tactual condition (T), and shaded bars represent the combined condition (L+T). Error bars represent ±1 s.d.

4–7). Averaged across subjects, the percent-correct scores were roughly 34% under condition L, 12% under condition T (slightly better than chance), and 49% under condition L+T. The improvement in L+T over L alone averaged roughly 15 percentage points. Measurements of %-IT for overall performance are provided in the top left panel of Fig. 7. The %-IT scores were 47%, 14%, and 52% for conditions L, T, and L+T, respectively. The improvement in %-IT performance observed for L+T compared to L alone averaged roughly 5 percentage points. The primary difference between results for %-correct and %-IT is that the %-IT scores for L and L+T are more similar than are the %-correct scores.

Measurements of %-feature IT are provided in Fig. 7 for voicing (top right panel), manner (bottom left panel), and place (bottom right panel). For the feature voicing, the %-feature IT was ordered as $L<L+T<T$. For condition L, %-feature IT was close to 0 (i.e., no transfer of information), consistent with the performance observed in pair discrimination and with the knowledge that the initial-consonant voicing, characterized by the activities of the vocal folds, is almost invisible through lipreading. The %-feature IT under condition T averaged roughly 30%, representing an improvement of 30 percentage points in the delivery of voicing relative to lipreading alone despite the fact that the overall performance in the T condition was near chance. The %-feature IT under L+T averaged roughly 19% and was 11 percentage points lower than that for T alone. For the features of manner and place, on the other hand, performance on %-feature IT was ordered as $T<L=L+T$. For manner, %-IT under T averaged 3% (indicating essentially no information transfer) and was 30% for L and L+T. For the feature place, %-feature IT averaged 1% on T (i.e., nearly no transfer of information), 69% on L, and 67% on L+T.

E. Discussion

For individual features, performance under the bimodal condition L+T appears to be a simple combination of the information transmitted through each of the two separate modalities (L alone and T alone). The contribution towards performance on the features of manner and place arises entirely from lipreading, and performance on voicing appears to arise entirely from the tactual cue. For manner and place, the %-feature IT through lipreading alone is nearly identical to that observed through L+T (i.e., 30% for manner and 67% for place). For voicing, performance through T alone is roughly 10 percentage points higher than for the combined condition L+T. Although it is clear that voicing information is being contributed by the tactual cue, it also appears that subjects may experience some difficulty attending to the tactual cue in the presence of lipreading.

Performance on bimodal conditions may be assessed using models of bimodal integration, defined by Grant (2002) as the processes employed by individual receivers to combine information extracted from two separate sources. Based on the assumptions of a given model regarding the bimodal integration process, mathematical operations are employed to compute predictions of bimodal performance from observed performance in each unimodal condition. The assumptions of the postlabelling integration model (POST) proposed by Braida (1991) appear to provide a reasonable fit to the bimodal feature-processing strategies employed by the subjects in the current study. The POST model assumes that subjects process cues from each modality separately and combine these judgments in selecting a response to the bimodal stimulus (as appears to be the case for the three features shown in Fig. 7). Using the observed confusion matrix from each separate modality, joint probabilities are calculated for each possible pair of response labels that can be generated for each stimulus. A maximum-likelihood rule is used to assign a response to each label pair. A predicted stimulus-response confusion matrix for the bimodal condition is then generated by summing over the probabilities associated with those label pairs that lead to a particular response to each stimulus [see Braida (1991, 1995) for a detailed description].

Observed overall percent-correct scores (for L, T, and L+T) and predicted overall percent-correct scores (for L+T) are provided in Table II for each of the four subjects and for means across subjects. The integration efficiency (IE) ratio, shown in the last column of Table II, is defined as the ratio between observed and predicted performance for L+T. The range of observed performance across subjects under each of the two individual modalities is fairly limited, ranging from 10.7% to 16.6% correct in condition T (with a standard deviation of 2.9) and from 32.3% to 36.8% correct in condition L (with a standard deviation of 2.0). For each of the subjects, observed L+T scores were higher than the score under each modality alone (by an average of 16 percentage points for L and 37 percentage points for T). The intersubject variability for L+T (40.4% to 59.3% correct with a standard deviation of 8.2), however, is larger than that observed under the single modalities, suggesting individual differences in bimodal integration ability. The predictions of Braida's (1991) POST model for performance under L+T ranged from 47.9% to 66.4% correct (with a standard deviation of 7.6% correct) across subjects and averaged roughly 9 percentage points higher than the observed performance. The

TABLE II. Observed performance in %-correct scores in the L, T, and L+T conditions and predicted scores under L+T using the post-labelling integration model of Braida (1991). The final column is a measure of integration efficiency (IE), defined as the ratio between the observed and predicted score for L+T.

Subject	L (Observed)	T (Observed)	L+T (Observed)	L+T (Predicted)	IE ratio
S1	34.49	11.03	51.34	57.62	0.89
S2	32.33	10.70	40.39	47.85	0.84
S3	36.75	16.64	59.25	66.36	0.89
S4	32.88	10.73	44.64	58.81	0.76
Mean	34.11	12.28	48.91	57.66	0.85

substantial improvement predicted for the L+T condition over either L or T alone supports the notion that the two single modalities supply complementary rather than redundant cues for identifying consonants (see Braida, 1995).

Although the subjects performed similarly on the unimodal conditions, the relatively wide range of individual predicted scores for L+T arises from differences in the structure of the confusion matrices for individual subjects under each of the two separate modalities. The integration efficiency ratios shown in the final column of Table II (which ranged from 0.76 to 0.89 across subjects) indicate that some subjects were better able to integrate information across the two modalities than others but that none of the subjects made optimal use of the available information. Values of IE lower than 1.0 (representing suboptimal bimodal integration ability relative to the predictions of the POST model) may arise in part from the tendency shown in Fig. 7 for lower scores on L+T compared to T alone for the feature voicing. The lower observed scores for the bimodal condition compared to T alone on voicing may arise from cross-channel interference or from a degradation of the information in each channel under the combined condition relative to unimodal performance. Blamey *et al.* (1989) have also reported inefficient integration of tactual information with lipreading. Using a simple probabilistic model to predict performance in the bimodal case, Blamey *et al.* (1989) observed less efficient integration of lipreading with tactual cues than with a simple auditory supplemement. The less effective integration of tactual cues (as opposed to auditory cues) with lipreading is likely related to subjects' lack of previous experience with this modality.

Overall performance followed the pattern of T<L<L+T for each subject (see Fig. 6). To normalize for different levels of ability on lipreading alone, a relative gain measure was used to assess the benefit provided for aided lipreading. Relative gain is calculated as [(L+T)-L]/(100-L), thus providing a measure of the proportion of the total possible improvement to lipreading that was actually carried by the tactual supplement. Across subjects in the current study, relative gain ranged from 0.11 to 0.36 and averaged 0.23.

The average relative gain of the current study appears to fall into the high end of the range of such measurements obtained in previous studies of consonant identification employing tactual displays as a supplement to lipreading. The scores reported by Carney (1988) using a 24-channel tactual vocoder to supplement speechreading translate into an average of roughly 0.31. Bratakos *et al.* (2001), using a single-channel vibrator that delivered the envelope of a frequency band of speech centered at 500 Hz modulating a 200-Hz carrier to supplement lipreading, reported a gain of 0.23 for L+T relative to L alone. Blamey *et al.* (1988) measured phonemic transmission using a multichannel electrotactile display of three acoustic-based speech parameters (F0, F1, and amplitude) and achieved a gain of roughly 0.21. Slightly lower values of gain were calculated from consonant identification scores reported by Weisenberger and Percy (1995) using the Tactaid 7 device, where gain ranged from 0.07 to 0.16.

V. EXPERIMENT 3: CUNY SENTENCE RECOGNITION

Sentence testing was conducted to determine whether the benefits observed for the tactual cue at the segmental level would have immediate carryover to the task of connected speech reception. Performance in sentence reception was examined under two conditions: lipreading alone (L) and lipreading combined with the tactual supplement (L+T). Data were not obtained with touch alone (T) because no significant information was expected for delivery through touch alone without extended training.

A. Stimuli

The speech stimuli used in this experiment were audiovisual recordings of the CUNY sentences.

B. Procedure

Each subject viewed 3 lists/condition for training and 27 lists/condition for testing. Lists were alternated between conditions of L and L+T. After each sentence presentation, the subjects were given as much time as necessary to type their responses into the computer. The subjects were instructed to write down any part of the sentence that they understood. They were encouraged to guess even when they were not sure of the answer. Prior to testing, subjects received a brief period of training on the sentence-reception task with three lists of sentences on each of the two conditions (L and L+T) in which multiple repetitions of stimuli were permitted and correct-answer feedback was provided. The testing employed only one presentation of each sentence with no correct-answer feedback. Responses were recorded by the host computer and scored later. The response was compared to the stimulus sentence. Any words in the response that corresponded exactly to words in the stimulus sentence were

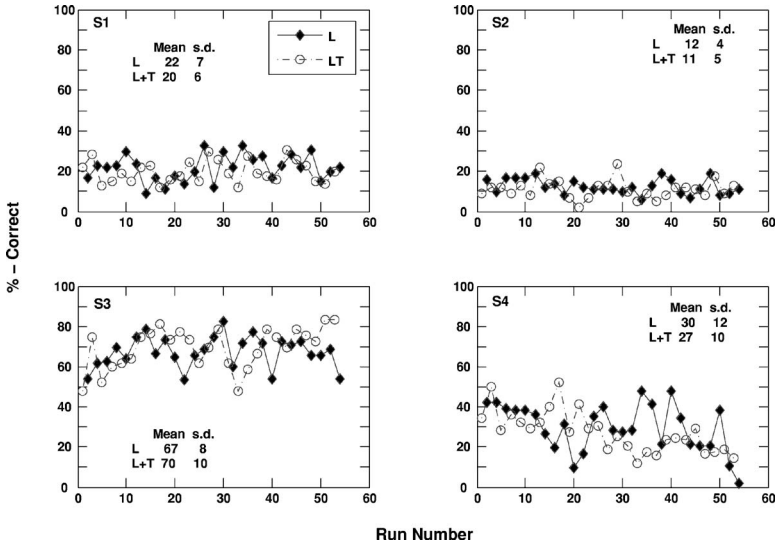


FIG. 8. Percent-correct performance on CUNY sentences as a function of run number for each of two conditions: lipreading (L—filled diamonds) and lipreading+tactual cue (L+T—unfilled circles). Results are shown in separate panels for each individual subject (S1, S2, S3, and S4).

scored as correct. The total number of correct words was computed across the 12 sentences in each list and converted to a percent-correct score.

C. Results

Results of the CUNY sentence test under the two conditions of L and L+T are presented in Fig. 8 for each subject. Each data point represents the percent-correct score for each list presented under each condition. No learning over time was observed for any of the subjects. For each subject, performance appears to be similar for L and L+T and never differed by more than 3 percentage points. In addition, performance appears to be fairly stable over time. Interlist standard deviation ranged from 4 to 12 percentage points across subjects. Across subjects, mean lipreading scores ranged from 12% to 67% correct word recognition in sentences in contrast to the narrow range of performance (32%–35% correct) observed in the consonant-identification task. Thus, segmental performance does not appear to be a good predictor of sentence performance through lipreading in the current study. Bernstein *et al.* (2000) also observed a larger range of performance in lipreading of words in sentences compared to phonemes in normal-hearing subjects. In a group of 96 normal-hearing subjects, the percentage of correctly identified words in CID sentences (delivered by a female talker) ranged from roughly 5% to 75%, compared to scores in the range of 15%–40% correct on a consonant-identification task.

D. Discussion

Miller *et al.* (1951) investigated the effects of speech material (digits, sentences, and nonsense syllable) on the intelligibility of speech under varied signal-to-noise ratio (S/N). Three functions of percent-correct scores versus S/N ratio were obtained for the three types of speech material, respectively. These results can be used to estimate the potential percent-correct score for word recognition in sentences given the percent-correct score for nonsense syllables. The tactual cue is effective in improving consonant recognition at the segmental level from 34% correct for lipreading alone to 50% correct for the combined condition of L+T. According to Fig. 1 of Miller *et al.* (1951), such an improvement in performance at the segmental level translates into an improvement in word recognition in sentences from 60% to 83%. A 16-percentage-point improvement at the segmental level thus leads to a larger 23-percentage-point improvement in word recognition in sentences. This amplification is due to the steeper slope of the "words in sentences" curve than that of the "nonsense syllables" curve in the region of S/N roughly from −10 to 10 dB.

The results of the CUNY sentence testing, however, indicated no benefit to lipreading with the tactual cue studied here. Improvements to lipreading of connected speech have been observed in previous studies with various types of tactual displays of speech, including multi-channel tactual displays of F0 (Grant *et al.*, 1986; Hanin *et al.*, 1988; Waldstein and Brothroyd, 1995a, b), tactual vocoders (e.g., Weisenberger *et al.*, 1989; Bernstein *et al.*, 1991), the multi-channel formant display of the Tactaid 7 device (Reed and Delhorne, 1995; Waldstein and Boothroyd, 1995a), a multi-channel electrotactile display of speech parameters (Cowan *et al.*, 1991), and a single-channel amplitude-envelope cue (Bratakos *et al.*, 2001). For example, results with tactual displays of F0 (designed to convey voicing-related cues as is the display studied here) indicate mean improvements over lipreading in the range of roughly 5 to 20 percentage points across subjects and studies (e.g., see Hanin *et al.*, 1988; Waldstein and Brothroyd, 1995a, b; Kishon-Rabin *et al.*, 1996). This magnitude of improvement is typical of that observed with the other types of displays summarized above, with the exception of two cases where larger benefits have been observed. These cases include an aided lipreading benefit of 25 percentage points achieved with sentence materials presented through the Tickle Talker device (Cowan *et al.*, 1991) and a 40-word/min improvement in continuous-discourse tracking observed with the 16-channel tactile vocoder studied by Weisenberger *et al.* (1989).

The lack of benefit of the tactual EOA cue may be due to a variety of factors. First, the amount of training subjects received on this task was limited to roughly 9 h (5 h for pairwise discrimination, 4 h for 16-consonant identification,

and only roughly 10 min for continuous sentences). Training in the use of novel cues through the skin is critical in the evaluation of tactual aids. In cases of good performance through the tactual sense (such as Tadoma), subjects received intensive training over a long period of time (years). Second, the subjects may have experienced difficulty in integrating information from the tactual and visual cues (as appears to have occurred at the segmental level). Third, temporal masking of the tactual signals may play a greater role in continuous-sentence recognition than in identification of initial consonants in nonsense syllables. Fourth, the additional complexity of continuous speech compared to isolated syllables (such as consonant clusters, faster speaking rate, coarticulation, etc.) leads to increased variability in the acoustic cue provided by the tactual display. Fifth, as the current display is designed specifically for initial consonant voicing, its ability to resolve the ambiguity of consonant voicing in other syllable positions (final or middle) is unknown. Sixth, the nature of the cue itself, which requires discriminating an onset-time difference between the vibrotactile signals presented at two fingers, may make it difficult to integrate with continuous speech.

Finally, it may be possible that the phonetic information of consonant voicing does not play as big a role in continuous-sentence recognition due to the redundancy of language. In other words, even if the tactual cue of consonant voicing were to be perfectly perceived and perfectly integrated with lipreading, it may not provide much benefit to the recognition of continuous sentences, specifically the CUNY sentences. In continuous-sentence recognition, phonetic information is not the only information available. Other information such as lexical structure, syntax, and grammar all play a role in language processing. For example, lexical knowledge can help to solve the ambiguity for the word pool when /p/ and /b/ are perceptually ambiguous because bool is not a word. In this case, the voicing information is redundant with lexical knowledge. On the other hand, lexical knowledge is of no help in distinguishing between the words park and bark when /p/ and /b/ are perceptually ambiguous because both words are normal members of the lexicon. In this case, voicing information can be used to solve this ambiguity. In a sentence, even more redundant information in the form of grammatical cues will be available.

Although voicing information may not necessarily be required or may be redundant with lexical or other information, this does not imply that the voicing cue has no contribution to word recognition. Voicing information enables the listener to limit the range of alternatives from which a response can be selected (i.e., reduce the stimulus uncertainty), which, as Miller *et al.* (1951) suggested, can improve word recognition. Iverson *et al.* (1998) investigated the interaction of phonemic information and lexical structure with a computational approach. The phonemic information was obtained through experimental studies of the identification of nonsense syllables under conditions of lipreading alone, auditory alone, and audio-visual conditions. From the data, categories of perceptually equivalent phonemes were constructed under each condition. These phonemic equivalence classes were then used to retranscribe a lexicon. In general, the additional information under each condition dramatically increased the percent of unique words, and decreased the expected class size or alternatives to be selected from the lexicon.

VI. CONCLUDING REMARKS

A two-channel tactual display of speech to provide information about consonant voicing was implemented for presentation of two envelopes derived from two different bands of speech with a two-finger tactual stimulating device. The efficacy of this display was evaluated by three perceptual experiments. Results of a pairwise discrimination experiment indicate that voicing is well discriminated through the tactual display of the EOA cue for eight pairs of initial voicing contrasts. Consonant identification studies indicate that voicing information derived from the tactual display improved performance by 15 percentage points over lipreading alone. No significant improvement was observed over lipreading alone with the addition of the tactual cue for sentence reception.

These results indicate that the tactual cue for voicing was effective at the segmental level and led to levels of performance superior to those obtained with previous tactual displays. These results demonstrate that the approach taken here of selecting information to complement that available through lipreading was a judicious use of the capacity of the tactual channel. This strategy may likewise be applied to other features that are impoverished through lipreading, such as aspects of manner of consonant production. The tactual display, however, did not lead to improvements over lipreading at the continuous-speech level.

ACKNOWLEDGMENTS

This research was supported by Research Grant No. R01-DC00126 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. The authors wish to thank Professor L. D. Braida and Professor K. N. Stevens for their helpful comments on the research reported here, and Professor Braida for his assistance in deriving predictions of bimodal integration.

[1]Thresholds were measured using a 2I-2AFC procedure and averaged across the results of the first three subjects. Subject 4 entered the experiment at a later date. This subject's thresholds for 50T and 250I were within 3 dB of the average thresholds across S1, S2, and S3.

Auer, E. T., and Bernstein, L. E. (**1996**). "Lipreading supplemented by voice fundamental frequency: To what extent does the addition of voicing increase lexical uniqueness for the lipreader?" in ICSLP'96 Proc., Philadelphia, PA, 3–6 October, pp. 86–93.

Auer, E. T., Bernstein, L. E., and Coulter, D. C. (**1998**). "Temporal and spatio-temporal vibrotactile displays for voice fundamental frequency: An initial evaluation of a new vibrotactile speech perception aid with normal-hearing and hearing-impaired individuals," J. Acoust. Soc. Am. **104**, 2477–2489.

Bernstein, L. E. (**1992**). "The evaluation of tactile aids," in *Tactile Aids for the Hearing Impaired*, edited by I. R. Summers (Whurr, London), pp. 167–186.

Bernstein, L. E., Demorest, M. E., and Tucker, P. E. (**2000**). "Speech perception without hearing," Percept. Psychophys. **62**, 233–252.

Bernstein, L. E., Demorest, M. E., Coulter, D. C., and O'Connell, M. P. (**1991**). "Lipreading sentences with vibrotactile vocoders: Performance of

normal-hearing and hearing-impaired subjects," J. Acoust. Soc. Am. **90**, 2971–2984.
Blamey, P. J., Cowan, R. S. C., Alcantara, J. I., and Clark, G. M. (**1988**). "Phonemic information transmitted by a multichannel electrotactile speech processor," J. Speech Hear. Res. **31**, 620–629.
Blamey, P. J., Cowan, R. S. C., Alcantara, J. I., Whitford, L. A., and Clark, G. M. (**1989**). "Speech perception using combinations of auditory, visual, and tactile information," J. Rehabil. Res. Dev. **26**, 15-24.
Boothroyd, A., Hnath-Chisolm, T., and Hanin, L. (**1985**). "A sentence test of speech perception: Reliability, set-equivalence, and short-term learning," City University of New York, Rep. No. RCI10.
Boothroyd, A., Hnath-Chisolm, T., and Kishon-Rabin, L. (**1988**). "Voice fundamental frequency as an auditory supplement to the speechreading of sentences," Ear Hear. **9**, 306–312.
Braida, L. D. (**1991**). "Crossmodal integration in the identification of consonant segments,"Q. J. Exp. Psychol. **43**, 647–677.
Braida, L. D. (**1995**). "Integration models of speech intelligibility," Symposium on Speech Communication Metrics and Human Performance, AL/CF-SR-1995-0023, 129–144.
Bratakos, M. S., Reed, C. M., Delhorne, L. A., and Denesvich, G. (**2001**). "A single-band envelope cue as a supplement to speechreading of segmentals: A comparison of auditory versus tactual presentation," Ear Hear. **22**, 225–235.
Breeuwer, M., and Plomp, R. (**1986**). "Speechreading supplemented with auditorily presented speech parameters," J. Acoust. Soc. Am. **79**, 481–499.
Carney, A. E. (**1988**). "Vibrotactile perception of segmental features of speech—a comparison of single-channel and multichannel instruments," J. Speech Hear. Res. **31**, 438–448.
Cowan, R. S. C., Blamey, P. J., Sarant, J. Z., Galvin, K. L., Alcantara, J. I., Whitford, L. A., and Clark, G. M. (**1991**). "Role of a multichannel electrotactile speech processor in a cochlear implant program for profoundly hearing-impaired adults," Ear Hear. **12**, 39–46.
Dorman, M. F., Soli, S., Dankowski, K., Smith, L. M., McCandless, G., and Parkin, J. (**1990**). "Acoustic cues for consonant identification by patients who use the Ineraid cochlear implant," J. Acoust. Soc. Am. **88**, 2074–2079.
Durlach, N. I. (**1968**). "A decision model for psychophysics," Communication Biophysics Group, Research Laboratory of Electronics, MIT, MA.
Eberhardt, S. P., Bernstein, L. E., Demorest, M. E., and Goldstein, M. H. (**1990**). "Speechreading sentences with single-channel vibrotactile presentation of voice fundamental frequency," J. Acoust. Soc. Am. **88**, 1274–1285.
Erber, N. P. (**1974**). "Visual perception of speech by deaf children—recent developments and continuing needs," J. Acoust. Soc. Am. **39**, 178–185.
Fu, Q.-J., and Shannon, R. V. (**1998**). "Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **104**, 2570–2577.
Gault, R. H. (**1924**). "Progress in experiments on tactual interpretation of oral speech," J. Abnorm. Soc. Psychol. **14**, 155–159.
Gault, R. H. (**1926**). "On the interpretation of speech sounds by means of their tactual correlates," Ann. Otol. Rhinol. Laryngol. **35**, 1050–1063.
Gescheider, G. A., Bolanowski, S. J., and Hardick, K. R. (**2001**). "The frequency selectivity of information-processing channels in the tactile sensory system," Somatosens. Mot. Res. **18**, 191–201.
Grant, K. W. (**2002**). "Measures of auditory-visual integration for speech understanding: A theoretical perspective (L)," J. Acoust. Soc. Am. **112**, 30–33.
Grant, K. W., Ardell, L. A. H., and Kuhl, P. K. (**1986**). "The transmission of prosodic information via an electrotactile speechreading aid," Ear Hear. **7**, 328–335.
Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). "Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration," J. Acoust. Soc. Am. **103**, 2677–2690.
Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).
Hanin, L., Boothroyd, A., and Hnath-Chisolm, T. (**1988**). "Tactile presentation of voice fundamental frequency as an aid to the speechreading of sentences," Ear Hear. **9**, 335–341.
Heider, F., and Heider, G. M. (**1940**). "An experimental investigation of lipreading," Psychol. Monogr. **52**, 124–133.
Hnath-Chisolm, T., and Kishon-Rabin, L. (**1988**). "Tactile presentation of voice fundamental frequency as an aid to the perception of speech pattern contrasts," Ear Hear.. **9**, 329–334.
Iverson, P., Bernstein, L. E., and Auer, E. T. (**1998**). "Modeling the interaction of phonemic intelligibility and lexical structure in audiovisual word recognition," Speech Commun. **26**, 45–63.
Jackson, P. L., Montgomery, A. A., and Binnie, C. A. (**1976**). "Perceptual dimensions underlying vowel lipreading performance," J. Speech Hear. Res. **19**, 796–812.
Jeffers, J., and Barley, M. (**1971**). *Speechreading* (Thomas, Springfield, Ill.).
Kirman, J. H. (**1973**). "Tactile communication of speech," Psychol. Bull. **80**, 54–74.
Kishon-Rabin, L., Boothroyd, A., and Hanin, L. (**1996**). "Speechreading enhancement: A comparison of spatial-tactile display of voice fundamental frequency (F_0) with auditory (F_0)," J. Acoust. Soc. Am. **100**, 593–602.
Miller, G. A., and Nicely, P. E. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.
Miller, G. A., Heise, G. A., and Lichten, W. (**1951**). "The intelligibility of speech as a function of the context of the test materials," J. Exp. Psychol. **41**, 329–335.
Owens, E., and Blazak, B. (**1985**). "Visemes observed by hearing-impaired and normal-hearing adult viewers," J. Speech Hear. Res. **28**, 381–393.
Rabinowitz, W. M., Eddington, D. K., Delhorne, L. A., and Cuneo, P. A. (**1992**). "Relations among different measures of speech reception in subjects using a cochlear implant," J. Acoust. Soc. Am. **92**, 1869–1881.
Reed, C. M., and Delhorne, L. A. (**1995**). "Current results of a field study of adult users of tactile aids," Semin. Hear. **16**, 305–315.
Reed, C. M., Delhorne, L. A., and Durlach, N. I. (**1992**). "Results obtained with Tactaid II and Tactaid VII," in *Proceedings of the Second International Conference on Tactile Aids, Hearing Aids and Cochlear Implants*, edited by A. Risberg, S. Felicetti, G. Plant, and K.-E. Spens (Royal Institute of Technology, Stockholm, Sweden) pp. 149–155.
Reed, C. M., Durlach, N. I., and Braida, L. D. (**1982**). "Research on tactile communication of speech: A review," ASHA Monogr. No. 20.
Reed, C. M., Durlach, N. I., Delhorne, L. A., Rabinowitz, W. M., and Grant, K. W. (**1989**). "Research on tactual communication of speech: Ideas, issues, and findings," Volta Rev. (Monogr.) **91**, 65–78.
Summers, I. R., Dixon, P. R., Cooper, P. G., Gratton, D. A., Brown, B. H., and Stevens, J. C. (**1994**). "Vibrotactile and electrotactile perception of time-varying pulse trains," J. Acoust. Soc. Am. **95**, 1548–1558.
Tan, H. Z. (**1996**). "Information transmission with a multi-finger tactual display," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.
Tan, H. Z., and Rabinowitz, W. M. (**1996**). "A new multi-finger tactual display," in *Proceedings of the Dynamic Systems and Control Division*, DSC-Vol. **58**, pp. 515–522.
Tan, H. Z., Reed, C. M., Delhorne, L. A., Durlach, N. I., and Wan, N. (**2003**). "Temporal masking of multidimensional tactual stimuli," J. Acoust. Soc. Am. **114**, 3295–3308.
Verrillo, R. T., Gescheider, G. A., Calman, B. G., and Van Doren, C. L. (**1983**). "Vibrotactile masking—effects of one-site and 2-site stimulation," Percept. Psychophys. **33**, 379–387.
Walden, B. E., Prosek, R. A., Montgomery, A. A., Scherr, C. K., and Jones, C. J. (**1977**). "Effects of training on visual recognition of consonants," J. Speech Hear. Res. **20**, 130–145.
Waldstein, R. S., and Boothroyd, A. (**1995a**). "Comparison of two multichannel tactile devices as supplements to speechreading in a postlingually deafened adult," Ear Hear. **16**, 198–208.
Waldstein, R. S., and Boothroyd, A. (**1995b**). "Speechreading supplemented by single-channel and multichannel tactile displays of voice fundamental-frequency," J. Speech Hear. Res. **38**, 690–705.
Weisenberger, J. M., and Percy, M. E. (**1995**). "The transmission of phoneme-level information by multichannel tactile speech-perception aids," Ear Hear. **16**, 392–406.
Weisenberger, J. M., Broadstone, S. M., and Saunders, F. A. (**1989**). "Evaluation of two multichannel tactile aids for the hearing impaired," J. Acoust. Soc. Am. **86**, 1764–1775.
Yuan, H. F. (2003). "Tactual display of consonant voicing to supplement lipreading," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.
Yuan, H. F., Reed, C. M., and Durlach, N. I. (**2004**). "Envelope-onset asynchrony as a cue to voicing in initial English consonants," J. Acoust. Soc. Am. **116**, 3156–3167.
Yuan, H. F., Reed, C. M., and Durlach, N. I. (**2005**). "Temporal onset-order discrimination through the tactual sense," J. Acoust. Soc. Am. **117**, 3139–3148.

Acoustic characteristics of Mandarin esophageal speech

Hanjun Liu, Mingxi Wan,[a] Supin Wang, and Xiaodong Wang
The Key Laboratory of Biomedical Information Engineering of Ministry of Education, Department of Biomedical Engineering, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an, 710049, People's Republic of China

Chunmei Lu
Cancer Institute & Cancer Hospital, Chinese Academy Medical Science, Beijing Union Medical College, Beijing, 100021, People's Republic of China

(Received 20 June 2004; revised 22 April 2005; accepted 5 May 2005)

The present study attempted to investigate the acoustic characteristics of Mandarin laryngeal and esophageal speech. Eight normal laryngeal and seven esophageal speakers participated in the acoustic experiments. Results from acoustic analyses of syllables /ma/and /ba/ indicated that, F_0, intensity, and signal-to-noise ratio of laryngeal speech were significantly higher than those of esophageal speech. However, opposite results were found for vowel duration, jitter, and shimmer. Mean F_0, intensity, and word per minute in reading were greater but number of pauses was smaller in laryngeal speech than those in esophageal speech. Similar patterns of F_0 contours and vowel duration as a function of tone were found between laryngeal and esophageal speakers. Long-time spectra analysis indicated that higher first and second formant frequencies were associated with esophageal speech than that with normal laryngeal speech. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1942349]

PACS number(s): 43.70.Bk, 43.70.Gr [AL] Pages: 1016–1025

I. INTRODUCTION

As a common treatment of laryngeal cancer, total laryngectomy involves the removal of the entire larynx, which results in the loss of ability to produce voice and speech. Currently, three main methods for vocal rehabilitation is available to laryngectomees: electrolarynx (EL), tracheoesophageal (TE), and standard esophageal (SE) phonation. EL is a hand-held, battery-powered transducer that transmits sounds through the neck tissues to the vocal tract (Barney *et al.*, 1959). Despite the fact that EL is relatively easier to use, the intrinsic limitations of the device make it difficult to produce phonetic information such as intonation and stress, and EL speech sounds monotonic and machinelike. TE speech involves a surgical-prosthetic restoration procedure that creates a fistula between the trachea and the esophagus, enabling the patients to power esophageal phonation with pulmonary air. It is treated as a standard technique for vocal rehabilitation in developed countries, despite nearly one-third of laryngectomees being found unsuitable for anatomical or personal considerations (Karen and Joel, 2000). As the major method of voice rehabilitation after laryngectomy, SE speech is produced by using the esophagus as an air supply and the pharyngo-esophageal (PE) segment situated on the superior aspect of the esophagus as a voicing source. As compared to TE or EL speech, SE speech requires no dependence on mechanical instrument, and both hands are free during speech. Therefore, SE speech has been one of the favorite methods of voice restoration.

Previous studies of SE speech have focused on American English, many of which have been conducted to compare the acoustic characteristics of SE and normal laryngeal (NL) phonation (Snidecor and Curry, 1959; Shipp, 1967; Hoops and Noll, 1969; Filter and Hyman, 1975; Christensen and Weinberg, 1976; Baggs and Pine, 1983; Blood, 1984; Robbins, 1984; Robbins *et al.*, 1984a, b; Sedory *et al.*, 1989; Bellandese, 1998; Bellandese *et al.*, 2001). In the study of SE and NL speech, Robbins *et al.* (1984a) described acoustic variables in the frequency, intensity, and time domains, mainly including mean F_0 (reading and vowel), jitter ratio, directional jitter, median intensity (reading), mean shimmer, maximum phonation time, percent pause time, words per minute (WPM), syllable duration, etc. Mean F_0 value obtained for SE speakers was 77.1 Hz, much lower than that from NL speakers (102.8 Hz). Slower reading rates (99 WPM), lower vocal intensity (60 dB), but greater mean F_0 range (118.1 Hz) were demonstrated for SE speakers than those for NL speakers (172.8 WPM, 60 dB, and 85.9 Hz). SE speech was also characterized by greater maximum phonation time (MPT), jitter, shimmer values, less periodicity, and more frequent short pauses when compared to NL speech. In the report by Robbins *et al.* (1984b) with the same data from the above study, median intensity (reading), mean MPT, WPM, and mean F_0 (reading) were found to be significantly different between SE and NL groups.

Blood (1984) compared SE and NL speakers on the acoustic parameters of F_0 and vocal intensity. Mean F_0 (102 Hz) and vocal intensity (84 dB) for NL speech were significantly higher than those (63 Hz and 72 dB for SE speech. Baggs and Pine (1983) noted that relative intensity, sentence duration, percent silence time, and maximum vowel duration were all significantly different between NL and SE speakers. Maximum vowel duration for SE speech was reported to be 4.6 s, much longer than previous reports (Snide-

[a]Electronic mail: mxwan@mail.xjtu.edu.cn

 0001-4966/2005/118(2)/1016/10/$22.50

cor and Curry, 1959; Hoops and Noll, 1969; Shipp, 1967; Robbins *et al.*, 1984a, b). Bellandese *et al.* (2001) compared six acoustic characteristics of excellent female SE and NL speakers, including mean F_0 values in the vowel /a/ and reading, signal-to-noise ratio (SNR), syllables per minute, number of pauses, and total duration of the first paragraph of the *Rainbow Passage*. Results indicated significant differences between NL and SE groups for all variables, in which number of pauses and total duration times were significantly higher but the other variables were significantly less for SE group as compared to NL group.

Although considerable research has focused on acoustic characteristics of SE speech, the literature presents conflicting information according to the comparisons between SE and NL speakers (Weinberg and Bennett, 1972; Robbins, 1984; Robbins *et al.*, 1984b; Blood, 1984; Bellandese, 1998; Bellandese *et al.*, 2001). For example, Bellandese *et al.* (2001) reported higher mean F_0 /a/ (107 Hz) but lower F_0 range (97 Hz) as compared to those (87 and 167 Hz, respectively) reported by Weinberg and Bennett (1972). Significant differences were found in mean F_0 (reading) between SE and NL groups (Robbins, 1984a), which disagreed with the results reported by Robbins *et al.* (1984b) and Blood (1984). Discrepancies in the above studies may be due to differences in subject sampling, recordings methods, methods of analysis, or speech samples used.

A few studies have investigated acoustic and perceptual characteristics of SE speech of tone languages, which are mainly focused on Thai and Cantonese (Gandour *et al.*, 1986, 1987a, b, 1988; Ching *et al.*, 1994; Yiu *et al.*, 1994; Ng *et al.*, 1997, 1998, 2001). A tone language is defined as a language that uses lexical tone to signify meaning, and lexical tone is defined as the use of fundamental frequency to distinguish minimal word pairs that are not differentiated by segmental information (Yiu *et al.*, 1994). Compared with the intonation patterns of American English that is defined at the sentence level, tonal patterns in tone languages are defined at the word or syllable level. In the acoustic studies of Thai and Cantonese alaryngeal speech, parameters were fundamental frequency, intensity and vowel duration, mean pause time, syllable per phrase, voice onset time, etc. (Gandour *et al.*, 1986, 1987a, b, 1988; Ng *et al.*, 2001). Gandour *et al.* (1988) reported that SE speakers of Thai were unable to produce the five phonemic tones at a level of proficiency comparable to that of NL speakers, and acoustic analysis revealed that it was attributed to their inability to consistently produce F_0 contours comparable to those of NL speakers. Durational measurements associated with rhythmic aspects of speech produced by one Thai SE speaker were investigated, and results indicated that SE speaker was able to maintain normal relative temporal relations among syllables within phrases (Gandour *et al.*, 1986). With regard to Cantonese, Ching *et al.* (1994) found that SE speakers were not able to proficiently produce the six different tones. Ng *et al.* (1998) reported that listeners' identification of the six tones for each syllable produced by SE speaker was similar in pattern to those produced by NL speakers, which suggested that listeners were able to identify accurate words spoken by Cantonese SE speakers. Acoustic characteristics of Cantonese SE and NL speakers were compared in the study of Ng *et al.* (2001), in which SE speakers exhibited significantly higher mean F_0 reading values and vowel duration but lower vocal intensity levels than NL speakers. And no significant differences were found in vowel duration and vocal intensity among the six Cantonese tones for NL and SE groups. The results suggested that acoustic cues that contributed most to the perception of meaning were not intensity and vowel duration but F_0 contours (Ng *et al.*, 2001).

As far as we know, however, there is no information concerning Mandarin alaryngeal speech. Mandarin is the predominant language used in the mainland of China. As a tone language, Mandarin is characterized by both larger rates of fundamental frequency change and more fundamental frequency fluctuations as a function of time and as a function of the number of syllables as compared to English (Eady, 1982). Gandour *et al.* (1986) noted that "…study of linguistics aspects of alaryngeal speech in different languages is expected to assist in (a) distinguishing those features of alaryngeal speech that are common across languages from those that are specific to particular languages, and (b) providing a more comprehensive assessment of the communicative potential of alaryngeal speakers." Because of the differences in the F_0 patterns between tone and nontone languages, it may be hypothesized that it would be more difficult for Mandarin alaryngeal speakers to produce tonal contrasts.

Although considerable research has been conducted on NL and SE speech in tone languages, several limitations existed in those studies. One is the limited number of subjects concerning SE speech. Only three SE and one NL speaker of Cantonese participated in the study reported by Ching *et al.* (1994). In the study reported by Gandour *et al.* (1986, 1987b, 1988), only two SE speakers of Thai were included. Due to the small number of speakers in these studies, the representative nature is questionable. Another limitation is the limited acoustic parameters in the study of tone languages. Acoustic parameters related to Cantonese SE and NL speakers reported by Ng *et al.* (2001) only included F_0 reading, vocal intensity and vowel duration, and more acoustic parameters such as jitter, mean shimmer, signal-to-ratio (SNR) should be investigated in order to fully understand SE speech of tone languages. The third limitation is the lack of interpretations of acoustic characteristics resulting from the tonal variations. Although acoustic measurements were obtained from different tones, the authors failed to discuss how and why the acoustic characteristics changed as a function of tone. Acoustic characteristics as a function of tone rather than that of mean values of different tones should play a more important role in the perception and production of tone in alaryngeal speech.

Therefore, the purpose of the present investigation was to study the acoustic characteristics of SE and NL speech of Mandarin with considerably larger number of subjects. Considering the intentional F_0 change of Mandarin tones, onset, offset, and range values of acoustic parameters including F_0, intensity, jitter, shimmer, and SNR were calculated and analyzed as well as the mean F_0, mean intensity, word per minute, and number of pauses during the reading.

TABLE I. Citation words used in the present study.

	Tone 1	Tone 2	Tone 3	Tone 4
/ma/	Mother	Hemp	Horse	Scold
/ba/	Eight	Draw	Target	Dam

II. METHODS

A. Subjects

Two groups of adult male speakers of native Mandarin participated in the experiment: eight NL speakers and seven SE speakers. The speakers were age-matched (47–71) with no reported history of speech-language problems except those associated with the laryngectomy. Seven laryngectomized speakers had been using SE speech for the average 6.46 years. Rated as good speakers by the ENT doctors, all the laryngectomees were literate with no problem reading the speech materials. As there is currently neither speech-language pathologist nor formal speech rehabilitation program available in the mainland of China, post-laryngectomy speech rehabilitation and evaluation are conducted by the ENT doctors.

B. Stimuli

In a tone language, it is the accompanying tone that differentiates word meaning. Words with identical phonemic structures but different tones carry different meanings (Ng *et al.*, 1998). Mandarin has four contrastive tone levels, traditionally labeled by using high-level (Tone 1), mid-rising (Tone 2), falling-rising (Tone 3), and high-falling (Tone 4). In the present study, the four tones associated with two sets of isolated Mandarin monosyllables (/ma/ and /ba/) were produced, yielding a total of 8 words (4 tones $\times$ 2 syllables). These specific monosyllables are minimally distinguished by tones and have different meanings (Table I). Each word was embedded in a neutral carrier phrase 'I read___to you' (/wɔ tu___keɪ ni thɪŋ/), so possible contextual effect was eliminated. In order to obtain acoustic information in reading, the speakers read aloud a 136-word passage from a third-grade reading book. The third sentence of this passage was chosen for acoustic analysis. The passage and the selected sentence were the same as that used in the study by Ng *et al.* (2001).

C. Recording

Before the actual reading, each speaker was given a brief practice period to become familiar with the speech materials, recording format, and instrumentation. This was to ensure that the speech sample represented the participant's best production. Each speaker was instructed to read the speech stimuli as though they were conversing with a person at a distance of approximately 1 m. All the recordings took place in a soundproof room. The speech signals collected by a microphone mounted at a distance of 15 cm from the mouth at an angle of 30° above horizontal were amplified by using a multichannel conditioning amplifier (Br el and Kjær, Denmark). This distance was chosen to minimize the potential recording of stoma noise. The sampling frequency is 22 kHz, 16 bits/samples.

During the recording, the subjects were provided with cards on which Chinese characters representing the citation words were printed. Instructions were given to the speakers before the recording of the words. They read the speech materials three times at their normal intensity and speaking rate.

D. Acoustic measurements

Mean F_0 reading values were measured from the third sentence of the reading passage. During the measurements, each sentence was extracted from the passage, re-sampled at 10 kHz, and stored into computer for later analyses. To identify periods from the wave form, markers were placed between each consecutive period of the vocalic portion of the third sentence. The duration of the sample was determined by subtracting the beginning value from the end value of this sentence. Mean F_0 reading values were computed with the use of TF32 (Demo version) which is a time-frequency analysis software program.

A comprehensive signal analysis program, PCQUIRER software (Scicon R&D, Inc.), was used to obtain F_0 contours from the eight citation words. The vowel portions of the syllables were marked period by period manually by the investigator. The F_0 contours associated with different tone levels were then displayed using a frequency-time plot.

The vowel /a/ was extracted based on the difference in time between the onset of the first identifiable period and the offset of the last identifiable period in the vowel. The wave form was visually inspected and edited from the signal. In order to aid the identification of each period, the wave form was low-pass filtered at 1 kHz. The beginning of the vowel /a/ of syllables /ma/ and /ba/ was identified by both the presence of initial vertical striations in the broad-band spectrogram using a resolving filter of 300 Hz bandwidth and a concomitant sharp increase in the amplitude of the signal. The end of vowel was identified by the abrupt attenuation of the signal amplitude. Once the vowel segment was identified, onset, offset, and range values of acoustic variables including F_0, jitter, shimmer, and SNR were competed for /a/ for different tones (only four periods were used for analyses). Details about the algorithm for determining the jitter, shimmer, and SNR can be found in Milenkovic (1987).

Intensity calibration was conducted according to the procedures reported by Robbins *et al.* (1984a). The calibration equation was generated by using a linear regression method after 20 data points were randomly extracted from each of the three digitized calibration signals (at 60, 70, and 80 dB SPL). This calibration equation was used to calculate the actual intensity levels in dB SPL. Then the mean intensity levels of the citation words were calculated by using PCQUIRER software.

WPM and number of pauses were calculated from the entire passage. Time cursors were manually placed at the beginning and end of the passage to determine the duration by subtracting the beginning value from the end value. WPM was then calculated by dividing the number of words in the

TABLE II. Minimal, maximal, and range values of F_0, intensity, jitter, shimmer, and SNR of the whole vowel /a/ produced by normal laryngeal speakers.

		NL speakers							
		/ba/				/ma/			
		Tone 1	Tone 2	Tone 3	Tone 4	Tone 1	Tone 2	Tone 3	Tone 4
F_0 (Hz)	min	122.43	80.10	55.20	87.3	129.23	88.17	68.43	99.6
	max	140.77	143.3	142.63	157.57	137.66	148.07	137.67	167.83
	range	18.34	63.20	87.43	70.27	8.43	59.9	69.23	68.23
Intensity (dB)	min	72.23	69.43	55.90	57.97	66.40	62.97	55.77	65.63
	max	83.57	80.90	81.33	80.47	81.50	80.80	80.63	82.53
	range	11.34	11.47	25.43	22.50	15.10	17.83	24.86	16.90
Jitter (%)	min	0.21	1.46	1.74	1.91	0.14	1.53	1.87	3.43
	max	1.54	4.49	4.65	4.25	0.62	7.87	5.52	4.39
	range	1.33	3.03	2.91	2.34	0.48	6.34	3.65	0.96
Shimmer (%)	min	1.07	7.82	11.35	7.34	1.18	2.94	9.42	5.08
	max	7.28	20.42	22.67	20.63	5.54	16.55	22.63	20.00
	range	6.21	12.60	11.32	13.29	4.36	13.61	13.21	14.92
SNR (dB)	min	22.93	7.83	7.40	5.83	20.13	7.23	7.40	6.33
	max	26.07	15.70	14.57	15.03	22.47	14.50	16.90	14.33
	range	3.13	7.87	7.17	9.20	2.34	7.27	9.50	8.00

passage by the duration of the sample. If the wave form returned to zero, that section of wave form was checked to determine if a pause had occurred. A minimum length of 150 ms was used to be considered as a pause. Then the number of pauses was calculated.

A long-time LPC spectrum was computed from the entire passage using 256 consecutive samples without the low-pass filtering. It took 54.35 s for NL speakers and 72.68 s for SE speakers to finish the reading of this passage. The signal was then multiplied by a 45 ms Hamming window, and LPC coefficient were computed using an autocorrelation method with the order of 28. This procedure was completed with the use of MATLAB.

III. RESULTS

Tables II–IV show the minimal, maximal, and range values of acoustic data of the whole vowel /a/ produced by NL and SE speakers, including F_0 /a/, intensity, jitter, shimmer and SNR as well as mean F_0 reading, mean intensity reading, WPM, and number of pauses. With regard to NL speech, the minimal and maximal values of F_0 vowel /a/ and SNR were higher but jitter and shimmer values are lower than those for SE speech regardless of syllable /ma/ or /ba/. Jitter and shimmer ranges were smaller for NL speech than those for SE speech. F_0 and SNR ranges of Tone 1 were smaller but those of the other three tones were greater than those for SE

TABLE III. Minimal, maximal, and range values of F_0, intensity, jitter, shimmer, and SNR of the whole vowel /a/ produced by standard esophageal speakers.

		SE speakers							
		/ba/				/ma/			
		Tone 1	Tone 2	Tone 3	Tone 4	Tone 1	Tone 2	Tone 3	Tone 4
F_0 (Hz)	min	71.97	59.56	51.16	54.98	62.88	58.18	50.93	54.4
	max	100.50	92.68	88.68	97.90	87.40	87.20	89.45	92.84
	range	28.53	33.12	37.52	42.92	24.52	29.02	38.52	38.44
Intensity (dB)	min	66.22	61.44	57.88	61.91	63.38	64.53	61.34	59.48
	max	79.73	79.56	78.63	81.08	77.38	78.78	75.86	79.09
	range	13.52	18.12	20.75	19.17	14.00	14.25	14.52	19.61
Jitter (%)	min	3.46	6.04	8.19	4.61	3.73	5.79	7.33	7.17
	max	9.17	15.15	20.55	12.92	12.97	17.09	21.51	21.72
	range	5.71	9.11	12.36	8.31	9.24	11.30	14.18	14.55
Shimmer (%)	min	21.69	38.01	39.29	32.20	39.36	38.96	38.75	38.17
	max	45.65	63.45	67.62	66.74	64.54	69.54	66.26	64.59
	range	23.94	25.44	28.33	34.54	25.18	30.58	27.51	26.42
SNR (dB)	min	4.63	4.77	4.04	3.60	4.15	3.93	3.96	4.42
	max	7.84	8.27	7.62	7.22	7.55	7.42	7.33	8.35
	range	3.21	3.50	3.58	3.62	3.40	3.49	3.37	3.93

TABLE IV. Mean F_0, mean intensity, number of pauses, and word per minute of the passage produced by normal laryngeal speakers and standard esophageal speakers.

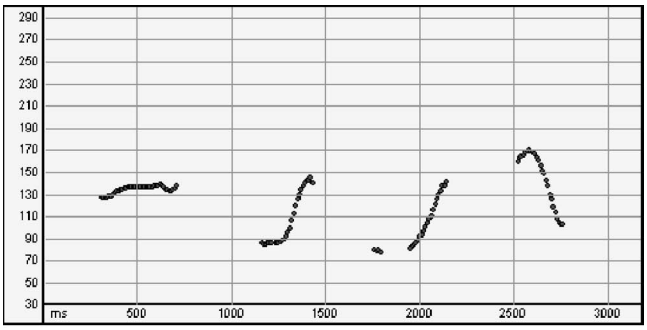
	NL speakers	SE speakers
F_0 reading (Hz)	111.51	84.35
Intensity reading (dB)	72.56	64.81
Number of pauses	16.6	35.8
Word per minute	175.3	118.8

speech, and the maximum intensity values produced by NL speakers were higher than those by SE speakers. In addition, jitter and shimmer values for Tone 1 were lower than those for the other tones. Kruskal-Wallis one-way ANOVA analyses indicated differences between NL and SE speakers among these acoustic parameters ($p<0.05$). With regard to NL speech, significant differences were found in the range values of these variables among four tones ($p<0.05$) and *Man-Whitney U* test indicated that Tone 1 was significantly different from the other tones in the range values ($p<0.05$). However, no significant differences were found among four tones for SE speech. Meanwhile, mean F_0, intensity, and WPM in reading were higher but number of pauses was significantly lower for NL phonation than those for SE phonation (see Table IV).

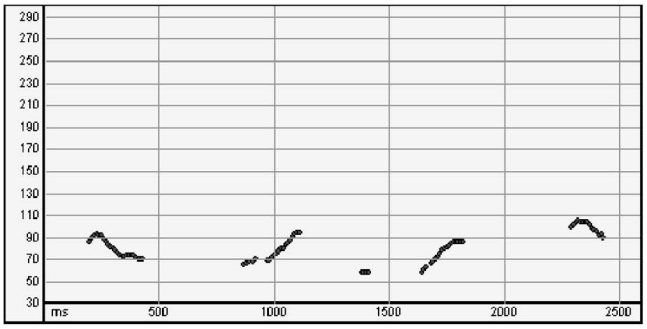
A. Fundamental frequency

Figure 1 shows the typical F_0 contours of syllables /ma/ and /ba/ at the four tones produced by a NL speaker and a SE speaker. The F_0 contours corresponding to Tone 1, Tone 2, Tone 3, and Tone 4 are shown in this figure from the left to the right. In Figs. 1(a)–1(d), the F_0 contours for syllables /ma/ and /ba/ produced by NL and SE speakers are similar. In Fig. 1(a), Tone 1 starts with a high F_0 value (near 130 Hz) and stays around that level throughout the syllable with few changes. Tone 2 starts with a low F_0 (near 90 Hz), then falls slightly before rising throughout the remainder of the syllable. Tone 3 starts with an F_0 value (near 80 Hz) slightly lower than the onset of Tone 2, fails to the lowest F_0 of all the four tones right at the vowel midpoint, then rises sharply to the end of the syllable. Tone 4 starts with the highest F_0 value of the four tones (near 170 Hz), keeps rising before reaching the maximum, then falls abruptly to the end of the syllable. Comparing Figs. 1(a) and 1(b) with Figs. 1(c) and 1(d), the patterns of F_0 contours for SE speakers are similar to that of the NL speakers, but the frequency locations of the contours for SE speakers are obviously lower than those for the NL speakers.

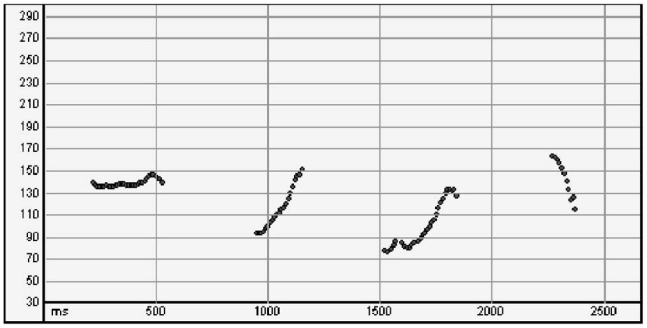
Figure 2 shows the onset and offset F_0 values of vowel /a/ as a function of tone produced by NL and SE speakers. With respect to NL speech, onset F_0 values of Tone 1 and Tone 4 are higher than those of Tone 2 and Tone 3. But opposite results were found for the offset F_0 values, in which Tone 2 and Tone 3 exhibited higher values than Tone 1 and Tone 4. The differences between the onset and the offset values varied as a function of tone. The greatest difference is associated with Tone 4, the smallest difference with Tone 1, intermediated by Tone 2 and Tone 3. With respect to SE speech, similar results were found for F_0 values as a function of tone. The differences between the onset and the offset F_0 values as a function of tone were different between /ma/ and /ba/. The differences of Tone 1 and Tone 2 for /ma/ were lower than those for /ba/.



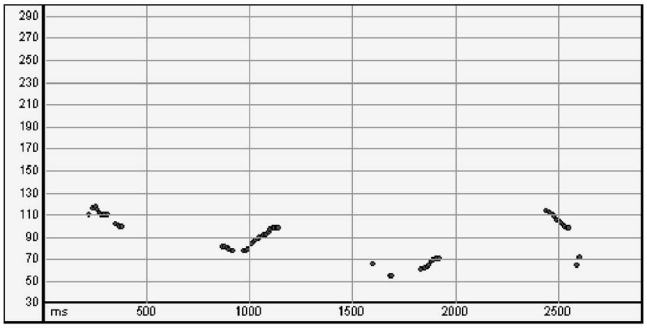
(a). /ma/ for NL



(b). /ma/ for SE



(c). /ba/ for NL



(d). /ba/ for SE

FIG. 1. Typical F_0 contours of (a) syllable /ma/ produced by a NL speaker; (b) syllable /ma/ produced by a SE speaker; (c) syllable /ba/ produced by a NL speaker; (d) syllable /ba/ produced by a SE speaker.

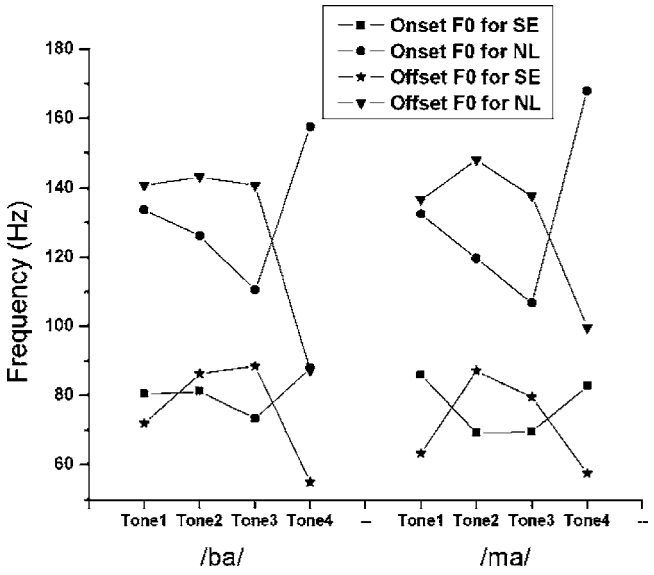


FIG. 2. Onset and offset F_0 values of /a/ as a function of tone for syllables /ma/ and /ba/ produced by NL and SE speakers.

With regard to NL and SE speech, Kruskal-Wallis one-way ANOVA revealed significant differences in onset and offset F_0 values for syllable /ma/ and /ba/ among the four tones ($p<0.02$). *Man-Whitney U* test indicated that onset and offset F_0 values of Tone 2 and Tone 3 were significantly different from those of Tone 1 and Tone 4 ($p<0.05$). For the syllable /ma/produced by SE speakers, *Man-Whitney U* test indicated that mean onset F_0 values for Tone 4 was significantly different from the other three tones ($p<0.05$), but no significant differences were found among these three tones. Similar results were found between Tone 4 and the other tones ($p<0.02$) for the offset F_0 values of syllable /ma/ produced by NL speakers. For both syllables /ma/ and /ba/, onset and offset F_0 values of vowel as a function of tone for NL speech were significantly higher than those for SE speech ($p<0.02$).

B. Intensity

Figure 3 shows the onset and offset intensity values of vowel /a/ at different tone levels produced by NL and SE speakers. With respect to NL speech, Tone 1 and Tone 4 were associated with higher onset values but lower offset values than Tone 2 and Tone 3, which was similar to the pattern of F_0 as a function of tone. The greatest intensity difference between the onset and the offset values was observed in Tone 4, the smallest difference in Tone 2, intermediated by Tone 1 and Tone 3. Similar results were also found for SE phonation, but lower intensity values as a function of tone were found as compared to NL speech.

With regard to NL and SE speakers, Kruskal-Wallis one-way ANOVA revealed significant differences in onset and offset intensity for syllables /ma/ and /ba/ ($p<0.01$) among the four tones. *Man-Whitney U* test indicated no significant differences in the onset and offset intensity values between Tone 2 and Tone 3. For syllables /ma/ and /ba/, onset and offset intensity values as a function of tone for NL speech

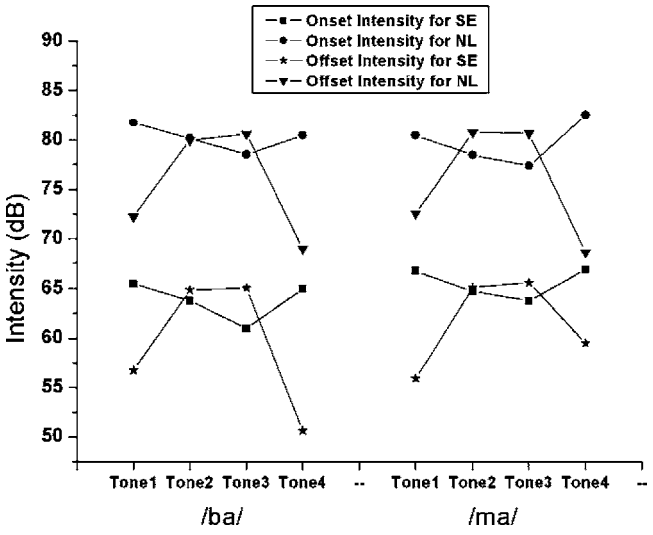


FIG. 3. Onset and offset intensity values of /a/ as a function of tone for syllables /ma/ and /ba/ produced by NL and SE speakers.

were significantly different from those for SE speech ($p<0.01$).

C. Vowel duration

Figure 4 shows the mean vowel duration values of /a/ as a function of tone produced by NL and SE speakers. With regard to NL speech, Tone 3 exhibited the greatest duration values, followed by Tone 2, Tone 1, and Tone 4. With regard to SE speech, Tone 3 also exhibited the greatest duration values, Tone 4 the smallest, intermediated by Tone 2 and Tone 1. Figure 4 shows similar patterns of vowel duration changes as a function of tone between NL and SE speakers. Tone 3 was found to show the greatest duration and Tone 4 the smallest, intermediated by Tone 2 and Tone 1.

With regard to NL speech, Kruskal-Wallis one-way ANOVA revealed significant differences in vowel duration for syllable /ma/ ($\chi^2=48.576, df=3, p<0.01$) and syllable /ba/ ($\chi^2=54.798, df=3, p<0.01$) among the four tones. *Man-Whitney U* test indicated that vowel duration for each

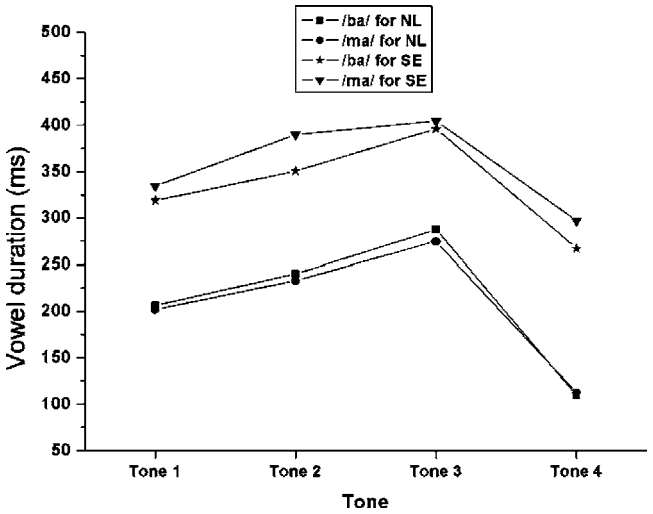


FIG. 4. Mean vowel duration values of /a/ as a function of tone for syllables /ma/ and /ba/ produced by NL and SE speakers.

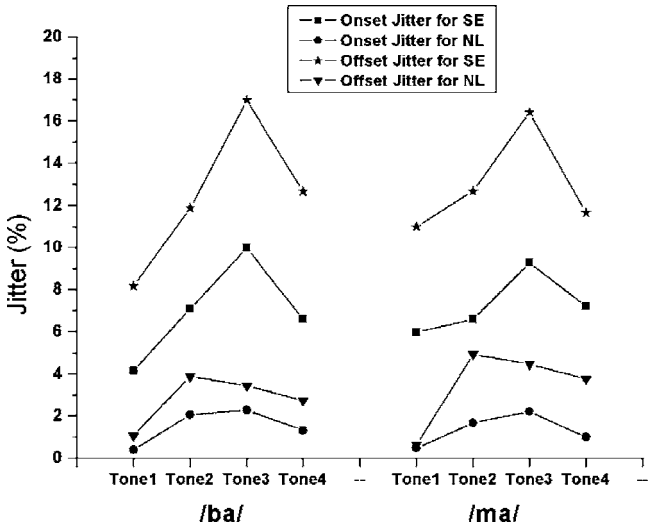


FIG. 5. Onset and offset jitter values of /a/ as a function of tone for syllables /ma/ and /ba/ produced by NL and SE speakers.

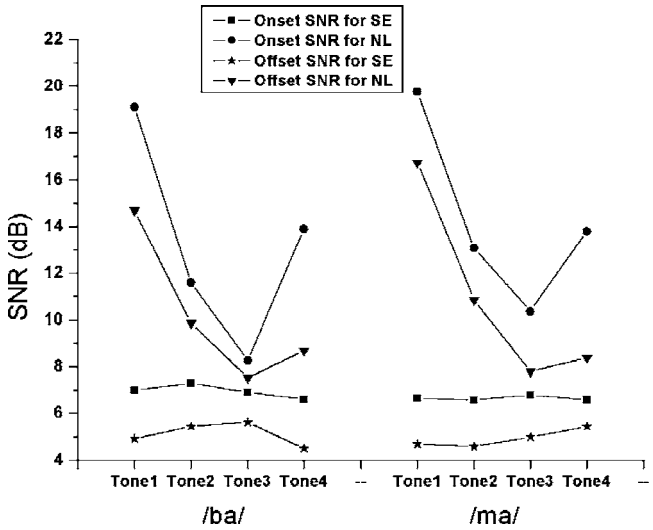


FIG. 7. Onset and offset SNR values of /a/ as a function of tone for syllables /ma/ and /ba/ produced by NL and SE speakers.

was significantly different from those for other tones ($p < 0.05$). With regard to SE speech, significant differences were also found in vowel duration for syllable /ma/ ($\chi^2 = 44.686, df = 3, p < 0.01$) and syllable /ba/ ($\chi^2 = 78.256, df = 3, p < 0.01$) among the four tones. For syllable /ma/, *Man-Whitney U* test indicated no significant difference in vowel duration between Tone 2 and Tone 3. For syllable /ba/, vowel duration for each tone was significantly different from those for the other tones ($p < 0.05$). For syllables /ma/ and /ba/, vowel duration values for SE speech were significantly higher than those for NL speech among the four tones ($p < 0.01$).

D. Jitter and shimmer

Figures 5 and 6 show the onset and offset jitter and shimmer values of vowel /a/ as a function of tone produced by NL and SE speakers. With regard to SE speech, smallest jitter and shimmer values were associated with Tone 1, and the greatest jitter values with Tone 3. The offset jitter and shimmer values as a function of tone were obviously higher than the onset values for syllables /ma/ and /ba/. With regard to NL speech, similar results were found for the onset and offset shimmer values. Tone 1 was associated with the lowest jitter and shimmer values, and Tone 3 exhibited the greatest onset jitter and shimmer values. As compared to SE speech, NL speech was associated with the lower jitter and shimmer values as a function of tone.

With regard to NL and SE speech, Kruskal-Wallis one-way ANOVA revealed significant differences in onset and offset jitter and shimmer values for syllables /ma/ and /ba/ ($p < 0.01$) among the four tones. *Man-Whitney U* test indicated that onset and offset jitter and shimmer values for Tone 1 were significantly lower than those for Tones 2–4 ($p < 0.05$). With regard to NL speech, no significant differences were found between Tones 2 and 4 in onset and offset jitter values for syllables /ma/ and /ba/. For both syllables /ma/ and /ba/, onset and offset jitter and shimmer values as a function of tone for SE speech were significantly higher than those for NL speech ($p < 0.04$).

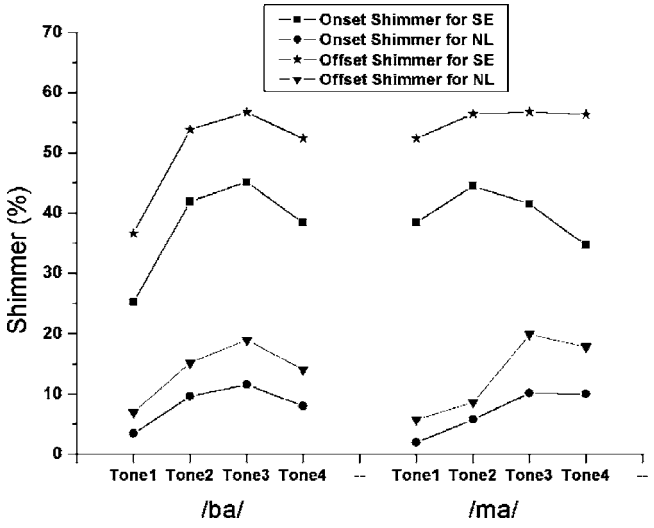


FIG. 6. Onset and offset shimmer values of /a/ as a function of tone for syllables /ma/and /ba/ produced by NL and SE speakers.

E. SNR

Figure 7 shows the onset and offset SNR of vowel /a/ as a function of tone produced by NL and SE speakers. With regard to NL speech, Tone 1 was associated with the greatest onset and offset SNR values, Tone 3 with the smallest, intermediated by Tone 2 and Tone 4. And the onset SNR values as a function of tone were higher than the offset values for syllables /ma/ and /ba/. SE speech exhibited similar onset and offset SNR values among four tones. And onset SNR values as a function of tone were also higher than the offset values.

With regard to NL speech, Kruskal-Wallis one-way ANOVA revealed significant differences in onset and offset SNR values for syllables /ma/ and /ba/ among the four tones ($p < 0.01$). *Man-Whitney U* test indicated that SNR for Tone 1 was significantly higher than those for Tones 2–4 ($p < 0.05$). With regard to SE speech, no significant differences

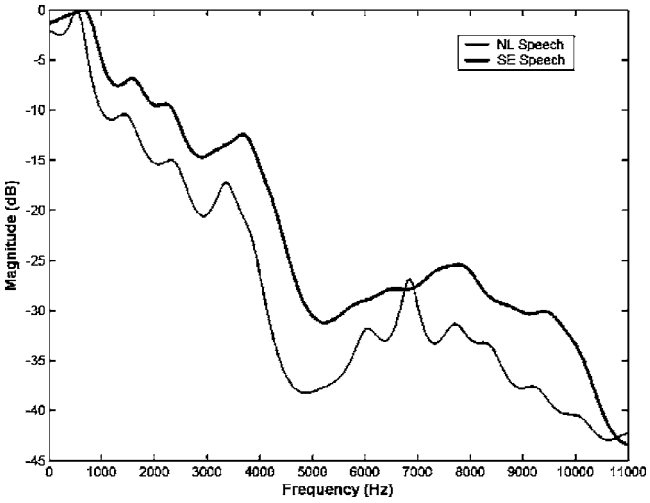


FIG. 8. Long-time LPC average spectra of the reading passage produced by NL speakers and SE speakers.

were found in SNR for syllables /ma/ and /ba/ among the four tones. For syllables /ma/ and /ba/, onset and offset SNR values as a function of tone for NL speech were significantly higher than those for SE speech ($p<0.03$).

F. Long time spectra

Figure 8 shows the long-time speech spectra of NL speakers and SE speakers producing the entire passage. Two general observations were apparent: (1) the formant frequencies were very clear below 4000 Hz, in which higher formant frequencies were obtained in SE speech as compared to NL speech except the third formant peaks; and (2) the SE spectrum was characterized by a more flattened spectral envelope in the high-frequency components (6k–9k Hz) compared with NL speech. This finding coincides with earlier observations made by Sisty and Weinberg (1972) and Weinberg *et al.* (1980). It was noted that the spectrum of SE speech was more flattened than NL speech in the high-frequency components.

IV. DISCUSSION

A. F_0, vocal intensity, and vowel duration

The average F_0 reading value associated with NL phonation was found to be 111.51 Hz (see Table II). This is comparable to the findings reported for NL speakers of English. Robbins *et al.* (1984a) reported a mean F_0 reading value of 102.8 Hz, Blood (1984) reported 120.8 Hz and Bellandese (1998) 115.6 Hz for NL phonation. With the same reading passage, Cantonese NL speakers (Ng *et al.*, 2001) exhibited an average F_0 reading value of 120.5 Hz. Despite the language difference and the use of different reading materials, normal Mandarin and normal English speakers exhibited similar reading F_0 values.

The mean F_0 reading values for Mandarin SE speakers found in the present study were 84.35 Hz, higher than those reported in the literature for English SE speakers (Shipp, 1967; Hoops and Noll, 1969; Filter and Hyman, 1975; Robbins *et al.*, 1984; Blood, 1984; Bellandese, 1998). This value, however, was much lower than that for Cantonese SE speakers (155.2 Hz) as reported by Ng *et al.* (2001).

For all tone levels, the onset and offset F_0 values of NL speech were higher than those of SE speech. Greater F_0 ranges for Tones 2–4 and with a smaller range for Tone 1 were seen in NL speech, when compared with SE speech. This may be due to the characteristics of Mandarin tones and the difference of voice production between NL and SE speakers. While F_0 contours of Tone 1 were similar in both SE and NL phonation, greater rates of F_0 change was found in the other tones (see Fig. 1). In both NL and SE phonation, the range of F_0 fluctuation in Tone 1 was the smallest when compared with other tones. However, the SE speakers use the PE segment for voice production rather that vocal folds as NL speakers. The mass of the PE segment has been found to be greater than the vocal folds and not capable of precise movement as seen in normal larynges (Van den Berg and Moolenaar-Bijl, 1959). The greater mass of the PE segment implies slower rate of vibration, resulting in the lower F_0 values for SE speech. In addition, NL speakers tend to have a better control over the vibration of vocal folds, which means they are able to yield a greater F_0 change for Tones 2–4 and at the same time maintain a steady F_0 value in the production of Tone 1. Meanwhile, due to the poorer control over the vibrating PE segment, SE speakers tend to have a smaller F_0 range for Tones 2–4, and inability to maintain steady F_0 value of Tone 1.

Higher mean F_0 reading values were found for Mandarin SE speakers as compared to English SE speakers. This difference may be related with the mechanism of F_0 regulation in SE speech. Diedrich (1968) noted that F_0 could be increased by tensing the PE segment only to a small extent because of the simple structure and limited nerve supply of the PE segment. Although an increase in the rate of PE segment vibration could be the result of an increase in airflow across the PE segment, SE speaker was limited in the ability to increase airflow because of the limited air supply available. Bellandese (1998) suggested that an increase in airflow to increase F_0 would result in an abrupt depletion of the air supply and less fluent speech. Significantly lower mean F_0 values for TE speakers than those for NL speakers despite the use of the pulmonary air supply (Bellandese, 1998), indicated that a dominant role in F_0 regulation for SE speakers is played by the PE segment rather than the air supply. Ng *et al.* (2001) reported that the mean F_0 for SE speakers were significantly higher than NL speakers in Cantonese. It is believed that Cantonese SE speakers may be able to contract the entire neck area and tense the PE segment, so that a tense and pointed SE segment edge is achieved, then the thin and pointed SE segment edge vibrates at a faster rate and generates higher F_0 values. This explanation agrees with the viewpoint of Bellandese's (1998). Based on this explanation, the results of the present study indicated that Mandarin SE speakers may have a better control of the PE segment to achieve higher F_0 values than English SE speakers.

In the present study, NL speakers exhibited significantly higher mean intensity levels than SE speakers in reading. This is consistent with previous studies (Baggs and Pine, 1983; Blood, 1984; Robbins *et al.*, 1984b). As noted by Rob-

bins *et al.* (1984a), the limited size of air reservoir in SE speech results in low air pressure and flow during SE phonation. Therefore, the diminished air reserve in SE speakers and the reduced airflow during phonation, contributes to the lower vocal intensity in SE speech as compared to NL speech.

The results indicated that SE speakers exhibited longer vowel duration values compared to the NL speakers. This finding is consistent with the results previously reported (Christensen and Weinberg, 1976; Gandour and Weinberg, 1980; Ng *et al.*, 2001). Durational differences between SE and NL speech may be due to the differences in speaking rate. SE speakers achieved a mean value of 139.8 WPM as compared to 175.3 WPM for NL speaker. Gandour and Weinberg (1980) noted that SE speech production systems were not biologically adapted for the production of speech so that SE speakers apparently could not execute articulatory commands as rapidly as NL speakers could. SE speakers are in less efficient use of available air supply for phonation and/or articulation, reducing the efficiency at the vibrating source as compared to NL speakers (Bellandese, 1998). Changes in articulatory aerodynamics and loss of some of the supporting structures for tone production during laryngectomy results in slower articulation rate in SE speakers when compared to NL speakers. As hypothesized by Christensen and Weinberg (1976), the longer vowel duration found in alaryngeal phonation could be attributed to the slower decay in the PE segment vibrations as compared with the laryngeal vibrations in the normal speakers. However, as noted by Sisty and Weinberg (1972), the myoelastic properties of esophageal sphincter were still unclear, care should be taken when making such hypothesis.

B. Jitter, shimmer, and SNR

Higher jitter and shimmer values as a function of tone for SE speakers were found, when compared with those for NL speakers (see Figs. 5 and 6 and Table III). NL speakers demonstrated significantly lower jitter values than SE speaker. This is consistent with the findings reported by Robbins (1984) and Bellandese (1998). SE speakers employ the PE segment as the primary regulator of F_0 (Bellandese, 1998), and higher jitter values in SE speech indicated that SE speakers had a poorer control over the PE segment as compared to NL speakers in controlling their normal larynges, resulting in a reduced vocal stability. The notion of reduced stability over the new vibratory device in SE speakers is supported by the significantly higher shimmer values associated with SE phonation. It is believed that the better the control of the vibratory mechanism is, the more periodic and regular is the vibration, and the lower are the jitter and shimmer values.

It was also found that jitter and shimmer of Tone 1 were significantly lower than those of the other tones for NL speakers. Tone 1 starts with a high pitch value and stays around that level with fewer changes, similar in F_0 contours to the sustained vowels in English. However, more frequency fluctuations were observed in F_0 contours for Tones 2–4, resulting in higher jitter and shimmer values for other three tones. Higher shimmer values indicated that Tones 2–4 exhibited larger amplitude perturbations during phonation than Tone 1.

NL speakers exhibited significantly higher SNR values among the four tone levels than SE speaker. But no significant differences were found among the four tones level for SE speakers. Positive SNR values indicated more periodic signal than aperiodic signal in SE speech. These results do not agree with the previous studies (Bellandese, 1998; Bellandese *et al.*, 2001), which reported negative SNR values of SE English speakers. Our results suggested that SE Mandarin speakers may have a better control in regulating the vibratory source than SE English speakers.

C. Long-time LPC spectra

Examination of individual long-time LPC spectra revealed that the first spectral peak was located around 300–450 Hz for NL speakers and between 450 and 700 Hz for SE speakers (see Fig. 8). These peaks appeared to represent the first formant frequencies of the vocalic portions of the syllables. Higher first formant frequencies for SE speakers were found than that for NL speakers. Similar results were also found in the second format frequencies. The increased values in F1 and F2 are consistent with the previous findings reported by Sisty and Weinberg (1972) that higher formant frequencies in English vowel of SE speech, and Cervera *et al.* (2001) reported similar results in the study of acoustic analysis of Spanish vowels produced by SE speakers. As Sisty and Weinberg (1972) noted, increased average vowel formant frequencies in SE speech could be attributed to the shortened overall vocal tract length caused by the position of neoglottis in SE speakers. However, F3 value appeared to be lower in SE speech than in NL speech. This may be related to the difference in tongue configuration between esophageal speaker and laryngeal speakers. Further data are needed to better understand the possible changes in tongue configuration during speech production after total laryngectomy.

V. CONCLUSION

The present study serves as a preliminary investigation of various acoustic characteristics of Mandarin alaryngeal phonation. Acoustic parameters including mean F_0 (in sustained vowel and passage reading), intensity, vowel duration, jitter, shimmer, SNR, WPM, number of pauses, and long-time LPC spectra were obtained from speech samples produced by NL and SE speakers of Mandarin. Results indicate that mean F_0, intensity, WPM, and SNR values were significantly higher, yet number of pauses, vowel duration, jitter, and shimmer were significantly lower in NL speech than in SE speech. Long-time spectra analyses indicated that higher formant frequencies were found in low-frequency components and more flattened envelopes in high-frequency components for SE speech as compared to NL speech. Despite the language difference, some findings in the present study were consistent with the studies reported previously in English.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (Grant Nos.30070212 and 69925101). The authors express special appreciation to Dr. Yi Xu of Haskins Laboratories, the editor, and two anonymous reviewers for their comments and suggestions on the manuscript.

Baggs, T. W., and Pine, S. J. (**1983**). "Acoustic characteristics: Tracheoesophageal speech," J. Commun. Disord. **16**, 299–307.

Barney, H. L., Haworth, F. E., and Dunn, H. K. (**1959**). "An experimental transistorized artificial larynx," Bell Syst. Tech. J. **38**, 1337–1356.

Bellandese, M. H. (1998). "The relationship between acoustic and perceptual characteristics of laryngeal, excellent tracheoesophageal and excellent esophageal speakers," The Doctoral dissertation, The University of Connecticut.

Bellandese, M. H., Lerman, J., and Gilbert, H. (**2001**). "An acoustic analysis of excellent female esophageal, tracheoesophageal, and laryngeal speakers," J. Speech Lang. Hear. Res. **44**, 1315–1320.

Blood, G. W. (**1984**). "Fundamental frequency and intensity measurement in laryngeal and alaryngeal speakers," J. Appl. Photogr. Eng. **17**, 319–324.

Christensen, J. M., and Weinberg, B. (**1976**). "Vowel duration characteristics of esophageal speech," J. Speech Hear. Res. **19**, 678–689.

Cervera, T., Miralles, J. L., and Gonzalez-Alvarez, J. (**2001**). "Acoustical analysis of Spanish vowels produced by laryngectomized subjects," J. Speech Lang. Hear. Res. **44**, 988–996.

Ching, T. Y., Williams, R., and Van Hasset, A. (**1994**). "Communication of lexical tones in Cantonese alaryngeal speech," J. Speech Hear. Res. **37**, 557–571.

Diedrich, W. M. (**1968**). "The mechanism of esophageal speech," Ann. N.Y. Acad. Sci. **155**, 303–317.

Eady, S. (**1982**). "Differences in the F_0 patterns speech: Tone language versus stress language," Lang Speech **25**, 29–42.

Filter, M. D., and Hyman, M. (**1975**). "Relationship of acoustic parameters and perceptual ratings of esophageal speech," Percept. Mot. Skills **40**, 60–68.

Gandour, J., and Weinberg, B. (**1980**). "Influence of postvocalic consonants on vowel duration in esophageal speech," Lang Speech **23**, 149–158.

Gandour, J., Weinberg, B., and Petty, S. H. (**1987b**). "Voice onset time in Thai alaryngeal speech," J. Speech Hear Disord. **52**, 288–294.

Gandour, J., Weinberg, B., Petty, S. H., and Dardarananda, R. (**1986**). "Rhythm in Thai esophageal speech," J. Speech Hear. Res. **29**, 563–568.

Gandour, J., Weinberg, B., Petty, S. H., and Dardarananda, R. (**1987a**). "Vowel length in Thai alaryngeal speech," Folia Phoniatr Logop **39**, 117–121.

Gandour, J., Weinberg, B., Petty, S. H., and Dardarananda, R. (**1988**). "Tone in Thai alaryngeal speech," J. Speech Hear Disord. **53**, 23–29.

Hoops, H. R., and Noll, J. D. (**1969**). "Relationship of selected acoustic variables to judgments of esophageal speech," J. Commun. Disord. **2**, 1–13.

Karen, C., and Joel, M. (**2000**). "Utilization of microprocessors in voice quality improvement: The electrolarynx," Curr. Opin. Otolaryng. Head Neck Surg., **8**, 138–142.

Milenkovic, P. (**1987**). "Least mean square measures of voice perturbation," J. Speech Hear. Res. **30**, 529–538.

Ng, M., Gilbert, H., and Lerman, J. (**2001**). "Fundamental frequency, intensity, and vowel duration characteristics related to perception of Cantonese alaryngeal speech," Folia Phoniatr Logop **53**, 36–47.

Ng, M., Kwok, C., and Chow, S. (**1997**). "Speech performance of adult Cantonese-speaking laryngectomees using different types of alaryngeal phonation," J. Voice **11**, 338–344.

Ng, M.,Lerman, J., and Gilbert, H. (**1998**). "Perceptions of tonal changes in normal laryngeal, esophageal, and artificial laryngeal male Cantonese speakers," Folia Phoniatr Logop **50**, 64–70.

Robbins, J. (**1984**). "Acoustic differentiation of laryngeal, esophageal, and tracheoesophageal speech," J. Speech Hear. Res. **27**, 577–585.

Robbins, J., Fisher, H. B., Blom, E. C., and Singer, M. I. (**1984a**). "A comparative acoustic study of normal, esophageal and tracheoesophageal speech production," J. Speech Hear Disord. **49**, 202–210.

Robbins, J., Fisher, H. B., Blom, E. C., and Singer, M. I. (**1984b**). "Selected acoustic features of tracheoesophageal, esophageal and normal speech," Arch. Otolaryngol. **110**, 670–672.

Sedory, S. E., Hamlet, S. L., and Connor, N. P. (**1989**). "Comparisons of perceptual and acoustic characteristics of tracheoesophageal and excellent esophageal speech," J. Speech Hear Disord. **54**, 209–214.

Shipp, T. (**1967**). "Frequency, duration, and perceptual measures in relation to judgments of alaryngeal speech acceptability," J. Speech Hear. Res. **10**, 417–427.

Sisty, N., and Weinberg, B. (**1972**). "Vowel format frequency characteristics measured on a wave-by-wave and averaging basis," J. Speech Hear. Res. **15**, 352–355.

Snidecor, J. C., and Curry, E. T. (**1959**). "Temporal and pitch aspects of superior esophageal speech," Ann. Otol. Rhinol. Laryngol. **68**, 623–636.

Van den Berg, J., and Moolenaar-Bijl, A. J. (**1959**). "Crico-pharyngeal sphincter, pitch, intensity and fluency in oesophageal speech," Pract. Otorhinolaryngol. (Basel) **21**, 298–315.

Weinberg, B., and Bennett, S. (**1972**). "Selected acoustic characteristics of esophageal speech produced by female laryngectomees," J. Speech Hear. Res. **15**, 211–216.

Weinberg, B., Horii, Y., and Smith, B. E. (**1980**). "Long-time spectral and intensity characteristics of esophageal speech," J. Acoust. Soc. Am. **67**, 1781–1784.

Yiu, E. M., van Hasselt, C. A., Williams, S. R., and Woo, J. K. S. (**1994**). "Speech intelligibility in tone language (Chinese) laryngectomy speakers," Eur. J. Disord. Commun. **29**, 339–347.

Acoustic-phonetics of coronal stops: A cross-language study of Canadian English and Canadian French

Megha Sundara[a)]
School of Communication Sciences & Disorders, McGill University 1266 Pine Avenue West, Montreal, QC H3G 1A8 Canada

(Received 1 November 2004; revised 24 May 2005; accepted 25 May 2005)

The study was conducted to provide an acoustic description of coronal stops in Canadian English (CE) and Canadian French (CF). CE and CF stops differ in VOT and place of articulation. CE has a two-way voicing distinction (in syllable initial position) between simultaneous and aspirated release; coronal stops are articulated at alveolar place. CF, on the other hand, has a two-way voicing distinction between prevoiced and simultaneous release; coronal stops are articulated at dental place. Acoustic analyses of stop consonants produced by monolingual speakers of CE and of CF, for both VOT and alveolar/dental place of articulation, are reported. Results from the analysis of VOT replicate and confirm differences in phonetic implementation of VOT across the two languages. Analysis of coronal stops with respect to place differences indicates systematic differences across the two languages in relative burst intensity and measures of burst spectral shape, specifically mean frequency, standard deviation, and kurtosis. The majority of CE and CF talkers reliably and consistently produced tokens differing in the SD of burst frequency, a measure of the diffuseness of the burst. Results from the study are interpreted in the context of acoustic and articulatory data on coronal stops from several other languages. [DOI: 10.1121/1.1953270]

PACS number(s): 43.70.Fq, 43.70.Kv, 43.70.−h [AL] Pages: 1026–1037

I. INTRODUCTION

Across the world's language inventories, coronal place is the most favored place for stops (Henton *et al.*, 1992). Both Canadian English (CE) and Canadian French (CF) phonemic inventories include /d/ and /t/. In both CE and CF, /d/ and /t/ are identified by "movement of the tongue from its neutral position," as defined by the feature [+ coronal], and by a constriction in front of the palato-alveolar region, as defined by the feature [+ anterior] (Chomsky and Halle, 1968). However, phonetic descriptions of /d/ and /t/ in CE and CF are different. Like American English (AE), CE coronal stops in initial position are phonetically transcribed as having an alveolar place of articulation; CF coronal stops are transcribed as having a dental place of articulation (Picard, 1987, 2001).

Although place differences across CE and CF have been described phonetically, their acoustic consequences have not been previously investigated for several reasons. Few languages use place of articulation differences within coronal stops to contrast meaning. Consequently, it is difficult to obtain reliable acoustic measures differentiating coronal stops across languages given differences in vowels and the implementation of voicing in the two languages of interest. Furthermore, several researchers (Jongman *et al.*, 1985; Stevens *et al.*, 1985) have suggested that a greater variability in production by talkers is likely to be a direct consequence of having one or the other (but not both) subgroup of coronal stops in the phonetic inventory. Dental allophones of /d/ and /t/ occur in English, specifically preceding interdental consonants; some researchers have also claimed that coronal consonants are dentalized in several dialects of English (Francis, 1958) or even that some English speakers do not distinguish between dental and alveolar stops, often interchanging them (Dixon, 1980). Thus, variability, in addition to that routinely expected across talkers of the same language, is likely to make generalizations regarding acoustic characteristics of coronal stops difficult.

Finally, at present there are no articulatory data for coronal stops in CE and CF. Besides place of articulation differences, researchers have also suggested that subgroups of coronal stops differ in the length of constriction (Chomsky and Halle, 1968) or the active articulator (Stevens *et al.*, 1985). Articulatory recordings of multi-syllabic utterances with coronal consonants in intervocalic position from 20 speakers of American English (AE) and European French (EF) presented by Dart (1991, 1998) illustrate the problem of identifying the articulatory differences between coronal stops in these two languages. Using data from palatograms and linguagrams, Dart investigated whether differences in place of articulation, constriction length, or active articulator underlie the differences between AE and EF coronal stops. She reports that whether coronal stops in AE and EF differ in the active articulator used to produce it, the place of articulation, or the constriction length, varies considerably across individuals. Thus, Dart's results attest to the variability in the articulation of coronal stops in languages that do not have both kinds of coronal stops.

For the reasons stated above, predicting the acoustic characteristics of coronal stops in CE and CF is not straight-

[a)]Present address: Institute of Learning & Brain Sciences, Box 357988, University of Washington, Seattle, WA 98195-7988. Electronic mail: msundara@u.washington.edu

forward. The present study was designed to address how (if at all) the acoustic characteristics of coronal stops, both voiced and voiceless, differ across CE and CF. Apart from providing an acoustic description of language-specific characteristics of coronal stops in CE and CF, results from this study will provide an essential baseline for investigations of monolingual and bilingual acquisition of coronal stops. Finally, the acoustic characteristics of coronal stops in each of the two languages will help predict the articulatory movements underlying coronal stop production in CE and CF.

To determine whether the acoustic characteristics of coronal stops in CE and CF are different, in this study, burst intensity and burst spectral measures were used. Burst intensity and burst spectral measures have been previously applied to identification of coronal stops thought to differ in place of articulation (Jongman *et al.*, 1985; Stoel-Gammon *et al.*, 1994). Jongman *et al.* (1985) first introduced a measure of the intensity of the burst with respect to the following vowel in order to distinguish place differences in voiceless coronal stops produced by three adult male talkers of Malayalam. Malayalam is one of the few languages thought to include both dental and alveolar stops in its phonetic inventory; specifically, in intervocalic position the dental-alveolar place difference for voiceless stops contrasts meaning. Jongman *et al.* predicted that differences in place of articulation alter the nature of turbulent noise generated around the constriction, as well as the direction of airflow as it hits the teeth; therefore, alveolar and dental stops should differ in burst amplitude. Because burst amplitude is likely to be modulated by overall loudness of productions, they measured root mean square (rms) amplitude of the burst relative to the amplitude of the following vowel ($\mathrm{Amp}_{\mathrm{vowel}}/\mathrm{Amp}_{\mathrm{burst}}$; a ratio without units).

Jongman *et al.* (1985) reported that alveolar stops are characterized by a louder burst and consequently relative burst amplitude ratios below 5 (a rms amplitude ratio of 5 corresponds to an intensity difference of about 14–15 dB). Dental stops are characterized by a softer burst and a relative burst amplitude above 5. Subsequently, using a ratio of 5 between vowel and burst rms amplitude as a metric, they successfully classified 95.8% of voiceless coronal stops produced by three new Malayalam speakers. However, when applied to distinguish /d/ and /t/ produced by three male native speakers of AE and Dutch, they had limited success. The dental-alveolar distinction does not contrast meaning in either of the two languages; in initial position AE coronal stops are described as alveolar whereas Dutch coronal stops have been described as dental. Although AE coronal stops were characterized by a louder burst, only about 68.2% of stops produced by AE speakers had relative burst amplitude below 5. Similarly, although Dutch coronal stops were characterized by a softer burst, only 63.2% of the tokens produced by Dutch speakers had relative burst amplitude above 5. Thus, Jongman *et al.* (1985) demonstrated that within as well as cross-language differences in place of articulation for coronal stops can be captured with a relative amplitude measure. However, they reported greater speaker-to-speaker variability in the number of tokens that can be correctly identified using the relative amplitude measure in AE and Dutch—languages with only one of the two coronal stops in their inventories—when compared to Malayalam, where both types of coronal stops are encountered.

More recently, Stoel-Gammon *et al.* (1994) have successfully applied an analogous relative intensity measure ($I_{\mathrm{vowel}}-I_{\mathrm{burst}}$; measured in dB) to distinguish between AE and Swedish coronal stops. Like CF and Dutch coronal stops, Swedish coronal stops are described as dental. Stoel-Gammon *et al.* contrasted /t/ productions in five vowel contexts (/i/, /ɪ/, /e/, /a/, and /u/) in real and nonsense /t/-initial words embedded in carrier phrases by ten female native speakers of AE and ten female native speakers of Swedish. AE alveolar stops had louder bursts and consequently lower relative burst intensity when compared to Swedish dental stops. They reported that relative intensity was significantly different for AE and Swedish stops, successfully demonstrating that this measure can be used to reliably distinguish between alveolar and dental stops even in a cross-language comparison where this distinction is not contrastive.

Stoel-Gammon *et al.* (1994) also measured burst spectra to distinguish alveolar and dental stops. Researchers have previously demonstrated consequences of place differences on the shape of burst spectra (Blumstein and Stevens, 1979). Forrest *et al.* (1988) describe numerical indices using spectral moments analysis to describe spectral shape differences. In this approach, the spectrum is treated like a probability distribution of energy over frequencies, which can then be used to calculate four spectral moments. The four spectral moments index four independent features of the energy distribution over frequency to derive average energy concentration (mean frequency), spectral shape as indexed by spread of frequency around the mean (standard deviation), the symmetry or tilt of the distribution (skewness), and the degree of its peakedness (kurtosis).

Stoel-Gammon *et al.* (1994) used the indices described by Forrest *et al.* (1988) to characterize differences between AE and Swedish bursts. They reported that among the spectral measures, AE and Swedish /t/ differed significantly on standard deviation and kurtosis of burst frequency. AE stops had more compact and more peaked burst spectra as indicated by a smaller standard deviation and higher kurtosis when compared to Swedish stops.

However, as the AE and Swedish corpora were recorded in different physical locations with different equipment, the differences in spectral shape reported by Stoel-Gammon *et al.* need to be interpreted with caution. In a subsequent investigation, Buder *et al.* (1995) documented the effects of recording condition differences on the burst spectra of a calibration signal. They reported small but systematic differences in spectral mean and standard deviation and large differences in the skewness and kurtosis measures in a calibration signal played in the two conditions. Buder *et al.* (1995) then reanalyzed just the spectral mean and standard deviation data from Stoel-Gammon *et al.* (1994) with corrections made for differences in recording condition. Although spectral standard deviation remained significantly different, mean frequency was now also found to be significantly different across the two languages. When compared to Swedish

stops, AE stops had a higher spectral mean frequency. Buder *et al.* (1995) do not report results for skewness or kurtosis measures.

In the present study, the relative intensity and spectral moments measures used by Stoel-Gammon *et al.* (1994) were used to determine the acoustic characteristics of coronal stops, /d/ and /t/, in CE and CF. Given that the vowels in CE and CF are likely to differ in their formant (F1 and F2) structure, and as these differences are also likely to influence burst characteristics, specifically, the mean burst frequency, only vowels that are similar in CE and CF were selected.

However, not only do CE and CF differ in their vowel inventories but they also differ in how voicing is realized. Caramazza *et al.* (1973) have previously demonstrated that CE and CF differ in the voice onset time (VOT) patterns underlying the two-way voicing contrast in each of the two languages. Caramazza *et al.* report that CE talkers, like AE talkers, produce nonoverlapping VOT distributions for voiced and voiceless stops at each place of articulation. Voiced stops in CE are produced typically with short-lag VOT and voiceless stops are produced with long-lag VOT (mean VOT=70 ms). Caramazza *et al.* do not report mean values of VOT for voiced stops. In contrast, CF talkers produce overlapping VOT distributions for voiced and voiceless stops at each place of articulation. Voiced stops in CF are produced with either lead VOT or short-lag VOT and voiceless stops are produced with short-lag VOT (mean VOT =23 ms). Thus, unlike in CE, in CF voiced-voiceless distinctions cannot be uniquely identified by VOT values alone.

Crucial to the present study, VOT values can be expected to influence the burst intensity (Pickett, 1999). Burst intensity differences relating to VOT may be related to the aerodynamic consequences of duration of oral closure. Typically, stops with greater VOT values can be expected to have longer closure durations (Chen, 1970) and, consequently, a greater build-up of oral pressure resulting in louder bursts. Thus, VOT values for /d/ and /t/ in CE and CF are also reported. As VOT alone is not sufficient to signal voicing in CF, in the present study the spectral moments were also analyzed for voicing effects.

Unlike Stoel-Gammon *et al.* and Jongman *et al.*'s investigations, male and female subjects were recorded in this study to provide a comprehensive description of the coronal stops in CE and CF. Furthermore, in view of predictions of greater variability for acoustic measures for noncontrastive segments, in addition to analyzing group differences, individual talker data are also reported. Neither Stoel-Gammon *et al.* (1994) nor Buder *et al.* (1995) report how well (if at all) data from individual subjects conform to group patterns. Finally, burst intensity and spectral measures of coronal stops in CE and CF from this study are related to possible underlying articulatory movements.

II. METHOD

A. Subjects

Six adult monolingual (3 M and 3 F) speakers of CE and six speakers of CF were recorded for analyses (mean age =24; range=22 to 35). Subjects had no history of speech, language, or hearing impairment. Their language background was assessed using a detailed language questionnaire including a self-rating of language ability in both CE and CF on a scale from 1 to 7, where 7 represents nativelike ability whereas 1 represents no ability. Subject-selection criteria were kept stringent because most people educated in Canada receive formal instruction in both languages at school. However, this instruction is mainly in reading and writing with minimal emphasis on speaking or listening skills. Thus, steps were taken to ensure that subject's competence in the non-native language was minimal. For this purpose, a proficient bilingual research assistant interviewed each subject in both languages. Subsequently, a 3-min speech sample describing a picture story [*Frog, where are you?* by Mayer (1969)] was collected from each subject in his or her native language. These samples were presented to three native listeners of CE (or CF). They were asked to rate the sample on a scale from 1 to 7, where 7 represents nativelike ability and 1 represents no ability. Strict criteria were also necessary to make the present study comparable to a parallel investigation of production by bilingual adults.

To be included in the native CE (or CF) group, subjects had to meet the following five criteria. First, subject's parents were monolingual speakers of CE (or CF). Second, subjects were schooled in CE (or CF). Third, they rated their ability in their native language with a minimum of 6 on a scale of 1 to 7. If they had any knowledge of the non-native language, they rated it below 3 on the same scale. The bilingual interviewer confirmed their lack of proficiency in the non-native language. Fourth, they had spent no time in a country where a language other than their native language was spoken. Fifth, native CE (or CF) listeners rated their speech sample describing the picture story with a minimum of 6 on a scale from 1 to 7. Six additional monolingual subjects (two CE male and one CE female; two CF female and one CF male) were recorded but excluded from the analyses because native listeners rated their speech sample lower than 6.

TABLE I. Canadian English and Canadian French stimuli are listed in the columns. Only initial voiced and voiceless coronal stops (/d/ and /t/) were analyzed.

C English		C French	
docile	toffee	docile	toffee
doctor	topic	docteur	topique
dopey	total	doper	total
dodo	topaz	dodo	topaz
deadly	textile	detter	textile
despot	texture	despote	texture
dagger	tablet	dadais	tablette
dapper	taxi	datcha	taxi

B. Stimuli

Subjects were recorded producing bisyllabic real words with coronal stops in word-initial position in a soundproof booth using an AKG C1000S microphone and a Tascam DA-30 digital audio recorder. Subjects read target words (Table I), twice embedded in sentences, followed by twice in

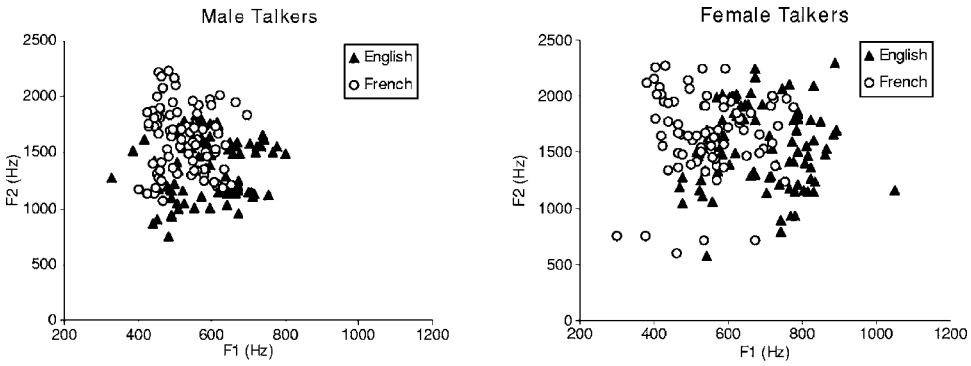


FIG. 1. F1-F2 data from male and female CE and CF talkers.

isolation. On a trial the subject produced the following utterance "Now I say doctor again. Now I say doctor again. Doctor. Doctor." The French words were embedded in the carrier phrase "Maintenant je dis__encore." To avoid list effects, each subject read the sentences in a different order and the sentences of interest were interspersed with 30 other sentences that were not analyzed. Subjects were asked to read at a comfortable rate of speech monitored by the experimenter. In the present study, analyses of isolated tokens—that is, tokens not embedded in sentences—are presented.

Thirty-two target words (16 English and 16 French) were selected to meet three criteria. First, articulatory descriptions and phonetic symbols of the vowels following coronal stops in the target words were identical across the two languages (Picard, 1987, 2001). As a result, target words with the four mid-vowels (/ɛ/, /æ/, /o/ and /ɔ/) in the first syllable were selected. These vowels include vowel height contrasts and front-back distinctions. Vowel formants were measured to confirm overlap in acoustic space for CE and CF tokens. Second, the consonant following the target syllable was a fricative, affricate, or stop. This helped to ensure that syllable boundary could be easily identified on the spectrographic and waveform display. Third, because initial syllables of words are likely to manifest coarticulatory influences of successive segments, segments (consonants and vowels) that are unique to either language were excluded to ensure that differences between contrasting syllables were restricted to those based on place or VOT only. Thus, whenever possible, cognates, defined as words with both identical orthographies and largely overlapping semantics, were selected to minimize differences in the target coronal stop due to differences in the phonetic context in which it was produced in the two languages.

Because few monosyllabic words met all the abovementioned criteria, bisyllabic words were used despite differences in stress allocation in CE and CF. Although there is little research on the effect of stress on burst intensity and spectral measures, a recent study (Cole *et al.* 2003) indicates that at least within English, burst amplitude in stressed and unstressed syllables was not significantly different. The tokens were digitized at 22 050 Hz and 16-bit quantization. Subsequently, the first syllable was excised from target words. Acoustic analyses are reported for these syllables.

C. Acoustic analyses

Analyses of VOT, burst intensity, and burst spectral properties were conducted excluding tokens without clear bursts. A visual inspection of the waveform and spectrograph revealed that all CF talkers, except one, produced some prevoiced /d/ tokens without clearly delineated bursts. These included 8% of /d/ tokens (3 tokens) produced by male talkers and 41.6% of tokens (15 tokens) produced by female talkers. It is possible that with a long prevoicing duration, clear bursts may not be produced due to insufficient build-up of intraoral pressure. To see if the lack of a clear burst was related to the duration of glottal vibration preceding the vowel, correlations were calculated between the relative intensity of the burst and prevoicing duration. There was no significant correlation between the two. Further, these bursts had spectral mean frequency values less than mean +2 SD for the rest of the distribution. Tokens without clear bursts were removed from analysis, as were unclear tokens.

A total of 318 tokens, 179 CE and 139 CF tokens, were analyzed. All analyses were carried out in PRAAT (Boersma and Weenink, 1992). The focus of investigation in the present study was initial /d/ and /t/. Five cursor positions were identified using a waveform display supplemented by a wideband spectrographic display; first periodic pattern before the burst (if any), onset of the burst, offset of burst, first periodic pattern after the burst signaling vowel onset, and vowel offset.

Vowel formants were measured at mid-point between vowel onset and offset to confirm that the vowels in CE and CF overlapped acoustically. Formant frequencies were derived from LPC analysis with a 15-ms hamming window centered at vowel steady state. None of these speakers produced the vowel in the first syllable as a diphthong. Figure 1 plots F_1 vs. F_2 for /ɛ/, /æ/, /o/, and /ɔ/ produced in the context of the syllables analyzed for this manuscript. Although far from identical, there is considerable overlap in the vowel space of CE and CF. The vowel space for CF is shifted upward in F_2 and downward in F_1 for both male and female talkers. The formant data suggest a tongue position that is more posterior and lower for CE and more forward and higher for CF; this difference may be related to articulatory set differences, coronal place differences, or some combination of the two (see Sec. IV for details).

VOT was measured as the time between the onset of the first clearly periodic pattern and the onset of the burst (Lieberman and Blumstein, 1988). Burst intensity and shape of the burst spectrum were calculated over the entire burst duration beginning at consonantal release. The size of the analysis window thus varied from token to token; it was determined by the duration of burst. When calculating burst

intensity measures for voiceless aspirated stops in CE, aspiration was not included in the analysis window. Visual inspection of the spectrograph and waveform was used to distinguish the burst duration from subsequent aspiration. Aspiration was characterized by a sudden drop in intensity and reduced energy at lower frequencies.

Relative burst intensity was calculated relative to the intensity of the following vowel to factor out the effect of differences in overall intensity across speakers. Intensity of the burst (in dB) was subtracted from the maximum intensity of the vowel (in dB) to obtain this measure of relative burst intensity (Stoel-Gammon *et al.*, 1994). On this measure, a softer burst is expected to have a greater intensity difference from the subsequent vowel.

The shape of the burst spectrum as characterized by the four spectral moments—mean, standard deviation, skewness, and kurtosis—was measured (Forrest *et al.*, 1988; Stoel-Gammon *et al.*, 1994). Spectral moments were derived from the power spectra over the entire burst duration for frequencies up to 11 025 Hz. To make the procedure for calculating spectral moments consistent with that used by Forest *et al.* (1998), bursts were preemphasized prior to making spectral measurements; above 1000 Hz the slope was increased by 6 dB/oct. Voiced tokens in CF, and sometimes in CE, are produced with prevoicing. Prevoicing is characterized by regular low-frequency glottal vibration during stop closure and sometimes through the burst. To compare intensity and spectral measures for voiced and voiceless stop consonants, all stops with lead VOT were filtered using a 200-Hz high-pass filter to remove the effects of voicing [a similar technique was used by Jongman *et al.*, (1985)].

III. RESULTS

A. VOT

As the results on the VOT measure are merely a replication of Caramazza *et al.*'s (1973) study, they are reported first. There was no reason to expect gender differences in production of VOT, thus data were pooled across gender for analyses. Group data for VOT are summarized in box plots in Fig. 2. The box stretches from the 25th to the 75th percentile and thus contains the middle half of the distribution; the bar in the middle of the box represents the median or the middle of the distribution—half the tokens have values greater than the median whereas the other half have values less than the median. The lower and upper brackets in the box plots denote the 10th and 90th percentile points, thus 80% of tokens lie within the limits defined by the brackets. Outliers are denoted by a circle (○) and have values between 1.5 and 3 times the box length whereas extremes are denoted by asterisk (*) and have values that are greater than 3 times the box length.

VOT values for tokens produced in isolation in this study were similar, in distribution and range of values, to those reported for CE and CF by Caramazza *et al.* (1973). VOT ranges observed in the present study for CE were also similar to VOT ranges previously reported for AE (Lisker and Abramson, 1964). For descriptive analysis, VOT values were separated into three bins: lead VOT, with values less than 0, short-lag VOT with values between 0 and 30 ms, and long-lag VOT with values greater than 30 ms. Mean, minimum, and maximum values are included in parentheses (mean, min: max) after the percentage of tokens produced with that VOT value.

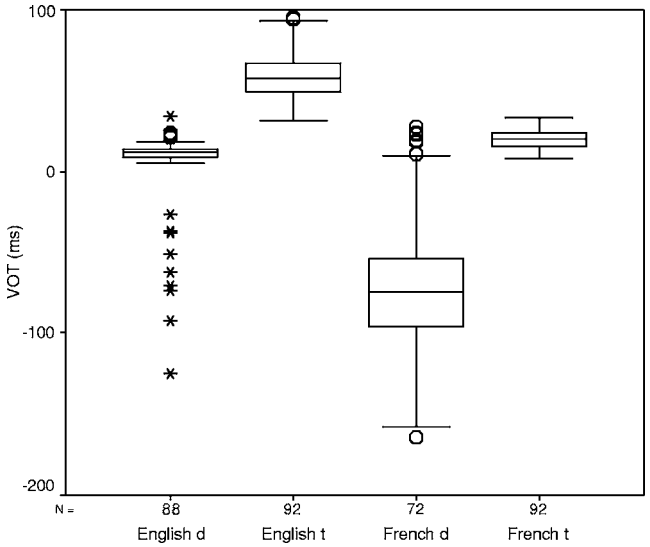


FIG. 2. Box plots of VOT distributions in CE and CF talkers. The middle half of the distributions lies in the box, the line in the middle of the box represents the median or centermost value in the distributions, and 80% of the distribution is within the brackets. Circle (○) denotes an outlier—a value between 1.5 and 3 times the box length, and asterisk (*) denotes extremes—values greater than 3 times the box length.

In CE, 87.5% of /d/ tokens were produced with short-lag VOT (16, 5:29) whereas 12.5% of the tokens were produced with lead VOT (−56, −125:−26); 100% of /t/ tokens were produced with long-lag VOT (60, 31:95). In CF, 90.8% of the /d/ tokens were produced with lead VOT (−82, −164:−17) whereas 9.2% of /d/ tokens were produced with short-lag VOT (19, 10:28); 100% of the /t/ tokens were produced with short-lag VOT (20, 8:30). Also CE talkers produced nonoverlapping distributions of VOT for voiced and voiceless tokens whereas CF talkers produced overlapping distributions of VOT because they produced some /d/ tokens with short-lag VOT. These results replicate Caramazza *et al.*'s findings; both report analyses for words produced in isolation.

For completeness and to make the comparison of VOT values consistent with the comparisons made on burst measures, VOT values for /d/ and /t/ were compared using a general linear model (GLM) repeated measures analysis of variance (ANOVA) with language (CE and CF) as the between-subjects variable and voicing (voiced and voiceless) as the within-subjects variable. A GLM analysis is more powerful for comparing unequal cell sizes. Significant interactions of language and voicing were explored using Bonferonni's *posthoc* analyses to confirm that language effects were significant for /d/ as well as /t/. Group patterns are reported followed by individual performance.

Differences in VOT distribution across CE and CF were confirmed by the analysis of variance. The main effects of language $[F(1,152)=370, p<0.01]$ and voicing $[F(1,152)$

TABLE II. VOT values [Mean (SD, number of tokens)] for each talker in the CE & CF group. Data from a single talker is summarized in each row. Each of the subjects is identified by their language group (CE/CF), gender (M/F), and a number. SD values are not reported when only one token was produced with that value.

	CE talkers					CF talkers			
Subjects	/d/ lead	/d/ lag	/t/ short lag	/t/ aspirated	Subjects	/d/ lead	/d/ lag	/t/ short lag	/t/ aspirated
CEM1	−72.4 (1.8, 2)	35 (6.3,12)		51 (9, 16)	CFM1	−93 (32, 11)	10 (1)	13 (4, 16)	
CEM2	−77 (33, 5)	9 (2, 7)		58 (12, 16)	CFM2	−81 (34, 12)	⋯	22 (5, 16)	
CEM3	−53 (24, 2)	12 (3, 14)		58 (10, 14)	CFM3	−85 (38, 11)	27 (1)	24 (4, 16)	
CEF1	⋯	12 (4, 16)		76 (15, 16)	CFF1	−53 (23, 11)	16 (4, 3)	23 (5, 14)	
CEF2	−26 (1)	14 (4, 15)		61 (10, 14)	CFF2	−100 (35, 12)	23 (1, 2)	18 (5, 14)	
CEF3	−50 (1)	12 (3, 13)		58 (12, 16)	CFF3	−74 (27, 12)	⋯	22 (6, 16)	

$=640, p<0.01$] and the interaction of language and voicing [$F(1,152)=40, p<0.01$] were significant. Language effects were significant for /d/ and /t/ as measured by Bonferroni's *posthoc* tests ($p<0.01$). Thus, as expected, VOT for /d/ as well as /t/ tokens is longer in CE than in CF.

Talker-specific differences in VOT production (Kessinger and Blumstein, 1998; Volaitis and Miller, 1992) as well as perception (Summerfield, 1981) have been previously documented. VOT values for each subject (Table II) also revealed individual variability in this corpus. Talkers in both language groups varied in their production of /d/ tokens. One male CE talker (CEM2) produced about 40% of voiced tokens with lead VOT. In AE, several researchers (Flege and Eefting, 1987; Mack, 1989) have reported that voiced tokens may be produced with lead VOT. Although four out of six CF talkers produced /d/ tokens with short-lag VOT, contributing to the overlap in the distribution of voiced and voiceless tokens in CF, only one female talker (CFF1) was responsible for most of the overlap. She produced 20% of voiced tokens with short-lag VOT. One male (CFM2) and one female (CFF3) CF talker did not produce any /d/ tokens with short-lag VOT. Talker-specific differences in VOT production have been directly attributed to individual differences in rate of speech (Allen *et al.*, 2003) or to social, dialectal, or idiolectal differences.

B. Burst measures

1. Group patterns

Results are reported for each burst measure (relative intensity, mean frequency, SD, skewness, and kurtosis of burst spectra) separately. A GLM repeated measures ANOVA with language (CE and CF) as the between-subjects variable and voicing (voiced and voiceless) as the within-subjects variable was conducted for each burst measure separately. Voicing was included as a variable in the ANOVAs on burst measures as VOT differences are known to influence burst intensity, at least in English (Pickett, 1999), which in turn may influence burst spectral measures. Because spectral measures reflect vocal tract size and shape, which are likely to differ across gender, results are reported separately for each gender. Significant interactions of language and voicing were explored with Bonferroni's *posthoc* analyses ($p<0.01$ are reported) for voicing as well as language effects. Finally, in order to determine the relative contribution of each burst measure in differentiating CE and CF tokens, results from discriminant function analysis with all burst measures and VOT included as predictors are presented.

(a) Relative burst intensity: Relative intensity data from male and female talkers across voicing conditions and across language are summarized in box plots in Fig. 3. Overall as expected, for talkers of both genders, relative intensity of /d/ and /t/ tokens was lower in CE than in CF, confirming that CE bursts are louder than CF bursts. The mean relative intensity levels for CE /t/ and CF /t/ tokens by female talkers in our study are comparable to those reported by Stoel-Gammon *et al.* (1994) for AE and Swedish female talkers respectively.

For male talkers, the main effect of language [$F(1,70)=5.5, p<0.05$] and the interaction of language and voicing [$F(1,70)=14.5, p<0.01$] were significant. As expected, relative intensity for CE was lower than for CF for both /d/ and /t/ tokens, but *posthoc* tests revealed that this difference was significant only for /d/. Voicing differences were only significant in CF, with relative intensity of /d/ tokens greater than that of /t/ tokens. For female talkers, the main effects of language [$F(1,62)=41, p<0.01$], voicing [$F(1,62)=38, p<0.01$], and the interaction of voicing and language [$F(1,62)=12.6, p<0.01$] were significant. *Posthoc* tests

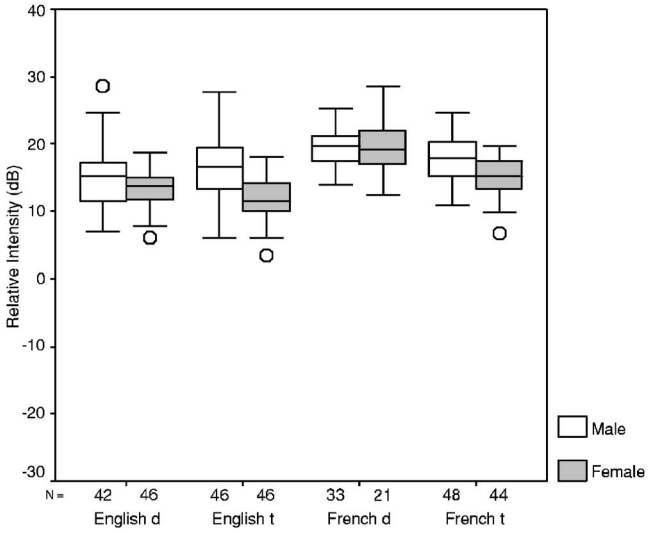


FIG. 3. Box plots of relative intensity for male and female talkers of CE and CF.

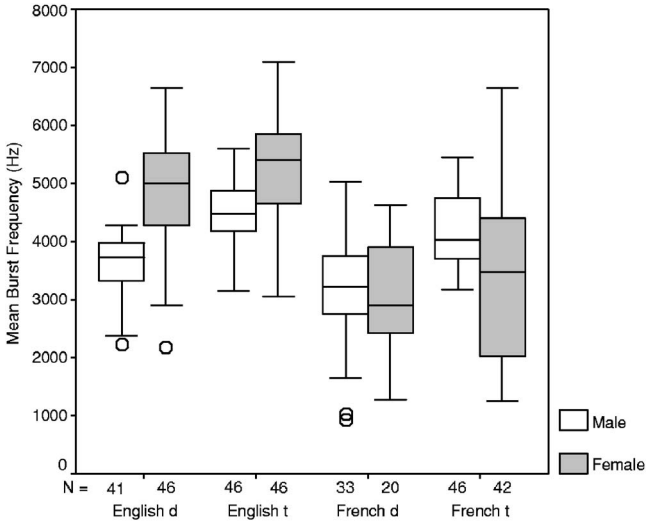


FIG. 4. Box plots of mean burst frequency for male and female talkers of CE and CF.

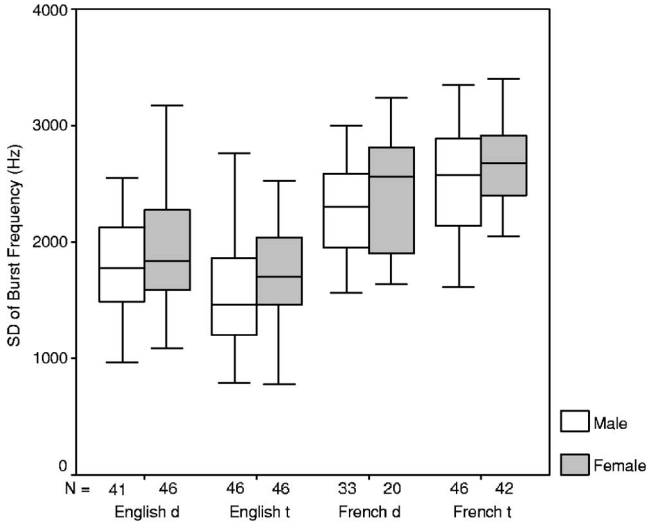


FIG. 5. Box plots of SD of burst frequency for male and female talkers of CE and CF.

confirmed that relative intensity for CE was lower than for CF for both /d/ and /t/. The relative intensity for /d/ was greater than that of /t/ in both CE and CF, but the voicing difference was only significant in CF.

As evidenced by the significant interaction of language and voicing for data from male and female speakers, voicing differences modulated burst intensity. In CE, although the direction of relative intensity difference was consistent with Pickett's (1999) prediction, the voicing difference was not significant for either male or female talkers. In CF, voiceless tokens were significantly louder than voiced tokens. Clearly, burst intensity provides a cue to voicing in CF.

Relative intensity ranges reported in the present study for CE and CF voiced and voiceless tokens are similar to those reported by Jongman *et al.* (1985) for AE and Dutch isolated stops, respectively. Note that the inclusion of the /d/ tokens in CF produced without clear bursts would effectively guarantee a large relative intensity difference between the vowel and the burst. Tokens with large relative intensity difference are consistent with the pattern seen above for CF talkers; in fact, exclusion of tokens without clear bursts, as has been done, underestimates the difference between relative intensity in CE and CF.

(b) Mean burst frequency: Mean burst frequency data from male and female talkers across voicing conditions and across language are summarized in box plots in Fig. 4. Overall as expected, for talkers of both genders, mean burst frequency was higher for CE tokens than CF tokens. Mean burst frequencies reported in our study for CE /t/ produced by female talkers are comparable to those reported by Stoel-Gammon *et al.* for AE /t/ tokens. The values are also consistent with those reported for AE /t/ by Forrest *et al.* (1988). However, compared to mean burst frequency reported by Stoel-Gammon *et al.* for Swedish female talkers, mean burst frequency for CF /t/ tokens produced by female talkers is lower (over 1000 Hz). When compared to the same Swedish corpus corrected for differences in recording conditions (Buder *et al.*, 1995), mean burst frequency of CF /t/ tokens is still lower but the difference is reduced to about 1000 Hz.

For male talkers, only the main effects of language $[F(1,70)=13, p<0.01]$ and voicing $[F(1,70)=56, p<0.01]$ were significant; CE bursts had a higher mean frequency than CF bursts, and voiceless stops had a higher mean frequency than voiced stops. For female talkers, only the main effect of language $[F(1,62)=109, p<0.01]$ was significant. Again, CE bursts had a higher mean frequency than CF bursts.

We know little about the effects of voicing on spectral mean burst frequency of coronal stops because all previous studies measuring burst spectral cues to consonant place have analyzed voiceless stops. In both CE and CF, voiceless stops had a higher mean burst frequency when compared to voiced stops for male and female talkers, but reached significance only for male talkers. Thus, mean frequency of burst may serve as a supplemental cue to stop voicing in both CE and CF.

(c) SD of burst frequency: SD of burst frequency from male and female talkers across voicing conditions and across language is summarized in box plots in Fig. 5. Overall as expected, for talkers of both genders, SD of burst frequency was lower for CE tokens than for CF tokens, confirming that CE bursts are compact whereas CF bursts are diffuse. In other words, energy in CE bursts is spread over a smaller range of frequencies than CF bursts. SDs of burst frequency reported in this study for CE and CF /t/ tokens produced by female talkers are systematically higher (approximately 500 Hz) than those reported by Stoel-Gammon *et al.* for AE and Swedish female talkers, respectively; they are also higher than the corrected values reported by Buder *et al.* (1995).

For male talkers, the main effect of language $[F(1,70)=78, p<0.01]$ and the interaction of language and voicing $[F(1,70)=20, p<0.01]$ were significant. SD for CF was significantly greater than for CE for both /d/ and /t/. Voicing effects were significant in both CE and CF, however they were in opposite directions. For female talkers as well, the main effect of language $[F(1,62)=87, p<0.01]$ and the interaction of language and voicing $[F(1,62)=6.1, p<0.05]$

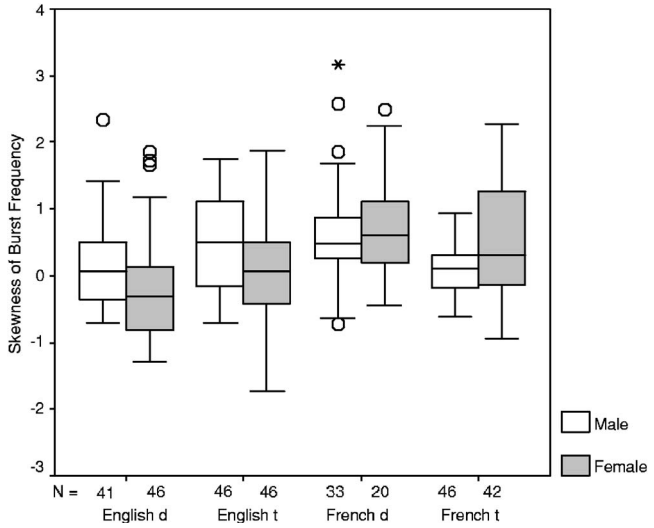


FIG. 6. Box plots of skewness of burst frequency for male and female talkers of CE and CF.

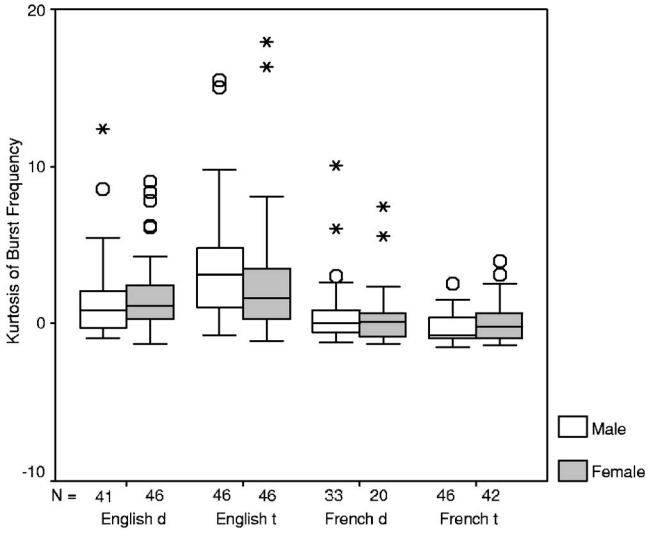


FIG. 7. Box plots of kurtosis of burst frequency for male and female talkers of CE and CF.

were significant. Language effects were significant for both /d/ and /t/, but voicing effects were not significant in either CE or CF. Because voicing effects were not consistent across gender for either language, it is unlikely that differences in SD of burst frequency cue voicing.

(d) Skewness: Skewness of burst frequency from male and female talkers across voicing conditions and across languages is summarized in box plots in Fig. 6. Overall, results were as expected only for female talkers. For female talkers, skewness of burst frequency was lower for CE tokens than for CF tokens. CE bursts have a negative (or 0) spectral tilt, implying that they have a concentration of energy in the frequencies above mean frequency (or are symmetric with respect to distribution of energy). CF bursts have a positive spectral tilt, implying that they have a concentration of energy in the frequencies below mean frequency.

For male talkers, the interaction of language and voicing $[F(1,70)=19.4, p<0.01]$ was significant. Language effects were significant for /d/ as well as /t/. However, the direction of the language effect was not consistent; the skewness of /d/ tokens was greater in CF whereas the skewness of /t/ tokens was greater for CE tokens. Voicing differences, although significant in both CE and CF, were also not consistent. For female talkers, only the main effect of language $[F(1,62)=27, p<0.01]$ was significant, supporting the predicted pattern. Skewness is not a consistent cue for differentiating between CE and CF talkers across gender or for signaling voicing differences within either language.

(e) Kurtosis: Kurtosis of burst frequency from male and female talkers across voicing conditions and across languages is summarized in box plots in Fig. 7. Overall as expected, for talkers of both genders, kurtosis of burst frequency was higher for CE tokens than for CF tokens. CE bursts have positive kurtosis values, implying that they have peaked energy distributions and thus spectra with clearly defined, well-resolved peaks. CF bursts have kurtosis values that are negative or around 0, implying that they have relatively flat spectra with no clear peaks. There was a greater difference between the kurtosis values for CE and CF than those reported for AE and Swedish (Stoel-Gammon *et al.*, 1994).

For male talkers, the main effect of language $[F(1,70)=24.8, p<0.01]$ and the interaction of language and voicing $[F(1,70)=13.8, p<0.01]$ were significant. CE bursts had a greater kurtosis than CF bursts for both /d/ and /t/ but the difference was significant only for /t/. Voicing effects were significant only in CE, with kurtosis of /d/ tokens significantly lower than that of /t/ tokens. For female talkers, only the main effect of language $[F(1,62)=12.1, p<0.01]$ was significant. As in the case of burst SD and skewness, kurtosis differences across voiced and voiceless stops were not consistent and thus are unlikely to provide a cue for stop voicing.

(f) Discriminant function analysis: To ascertain the relative efficiency of burst intensity and spectral measures, each variable was entered into a stepwise discriminant function analysis to predict whether tokens belonged to the CE or the CF group. Because CE and CF tokens were also significantly different on VOT, VOT was included in the discriminant function analysis as a predictor. When the variable to be predicted, in this case CE / CF, is categorical and binary, stepwise discriminant analysis is more appropriate than regression. As in Forrest *et al.* (1988), tokens were averaged to yield one entry per subject so as not to violate the assumption of independence of cases. Analysis was conducted separately for /d/ and /t/. Only variables for which Wilk's lambda was significant ($p<0.05$) are reported.

For /d/ tokens, VOT alone accounted for 58.4% of the variance $[(1,10)=59, p<0.01]$; the only other variable that was significant after VOT was included was SD $[(2,9)=141, p<0.01]$. Together, VOT and SD accounted for 83.1% of the variance. For /t/ tokens as well, VOT accounted for 83.9% of the variance $[(1,10)=106, p<0.01]$; again the only other variable to significantly add to the prediction was SD $[(1,9)=93, p<0.01]$. VOT and SD accounted for 89.9%

TABLE III. Composite CE/CF-ness scores for each talker on each measure. Variables are listed in rows, and subjects in columns.

	Canadian-English						Canadian-French					
Measure	CEM1	CEM2	CEM3	CEF1	CEF2	CEF3	CFM1	CFM2	CFM3	CFF1	CFF2	CFF3
Relative intensity	0.37	0.68	0.93	0.75	0.77	0.87	0.63[a]	0.42[a]	0.81	0.57[a]	0.82[a]	0.82[a]
Mean frequency	0.67	0.64	0.66	0.97	0.73	0.87	0.63	0.54	0.73	0.91	0.73	0.53
SD	0.50	0.96	0.83	0.69	0.87	0.80	0.67	0.77	0.88	0.74	0.86	0.88
Kurtosis	0.33	0.79	0.52	0.53	0.60	0.50	0.96	0.88	0.85	0.83	0.86	0.94

[a]When tokens without bursts were included the proportions increased slightly to 0.64 for CFM1, 0.46 for CFM2, 0.62 for CFF1, 0.85 for CFF2, and 0.88 for CFF3.

of the variance. Recall that even when AE and Swedish talkers were recorded under different conditions, Stoel-Gammon *et al.* (1994) reported a significant difference in SD of burst frequency. Moreover, this difference remained significant even after corrections were made for differences in recording conditions. Thus, apart from VOT, SD appears to be the most robust acoustic cue distinguishing coronal stops in CE and CF.

For both /d/ and /t/, the variable accounting for the highest degree of variance after SD was mean burst frequency. However, because 12 subjects in the analyses provide power for only up to two variables in the discriminant analyses, mean frequency was never significant.

2. Individual patterns

Individuals within each group varied on the various burst measures and the range of their tokens (data available upon request). Recall that Jongman *et al.* (1985) and Stevens *et al.* (1985) have suggested that a direct consequence of having one or the other (but not both) subgroup of coronal stops in the phonetic inventory is greater inter-speaker variability in production by talkers. Because greater variability may potentially lead to overlap in distributions, production by individual talkers needs to be evaluated to obtain some index of overlap on each measure.

Jongman *et al.* (1985) applied a metric generated from voiceless tokens produced in Malayalam to differentiate between AE and Dutch voiced and voiceless tokens. Recall that they used an amplitude ratio of 5, corresponding to an intensity difference of about 15, to differentiate between alveolar and dental stops. To compare the data from this study to Jongman *et al.*'s investigation, the percentage of CE tokens produced with relative intensity values less than 15 dB and the percentage of CF tokens produced with relative intensity values greater than 15 dB were calculated. Sixty percent of tokens in CE had relative intensity lower than 15 dB and 74% of tokens in CF had relative intensity greater than 15 dB. Using this metric, over 80% of tokens produced by three talkers (one CE and two CF), and between 50% and 80% of tokens by six other talkers (three CE and three CF) were correctly classified. Fewer than 50% of tokens were classified correctly for two male CE talkers (13% and 43%) and one female CF talker (43%). Recall that fewer than 36% of stops from one AE and one Dutch speaker were correctly classified using the metric derived from stops in Malayalam; over 85% of tokens produced by Malayalam talkers were correctly classified. Thus, the variability of the relative intensity measure and subsequently the overlap in distribution of relative amplitude is much greater when the comparison is cross-language than when it is within a language.

Jongman *et al.*'s approach relies on a comparison of burst intensity in contrastive and noncontrastive languages. However, because burst spectral measures are not available from Malayalam or other languages where this distinction is contrastive, there is a need for an alternative way to index overlap in distributions with respect to productions by each subject. One such way is to characterize how well tokens produced by each talker conform to group distributions. Consider two distributions with means M1 and M2. For well-separated, nonoverlapping distributions, all tokens in distribution 1 are closer to M1 whereas all tokens in distribution 2 are closer to M2. However, with overlap in distributions, some tokens—specifically ones in the tails of the distribution—will be produced with values closer to the mean of the other distribution.

A CE-like (or CF-like) score was obtained for each individual by calculating the percent CE (or CF) tokens produced by a talker that are closer to the mean of the CE (or CF) group than the mean of the CF (or CE) group. Because distributions are different across voicing conditions, tokens were always compared to the mean of the specific condition. They were finally combined to give a composite CE-like/CF-like score. Skewness was not included because there was no significant group difference between CE and CF tokens on this measure. Besides indexing the overlap in the distribution, this composite score also allows a comparison of the efficiency of each measure in distinguishing between CE and CF tokens by providing a measure of overlap between the CE and CF distribution. The measures on which higher percentage of tokens produced by most CF talkers are CF-like, and tokens produced by most talkers of CE are CE-like, have less overlap, and hence are more efficient in categorizing CE and CF tokens. Composite CE-like/CF-like scores are summarized in Table III.

The composite scores in Table III illustrate the variability and ambiguity of /d/ and /t/ produced in CE and CF. Only half the tokens produced by talkers CEM1, CEM2, CFM1, and CFM2 are likely to be similar in relative intensity and spectral measures to the distribution of their language group. Of the variables measured, talkers across the two groups produced the most distinct tokens on the SD measure. At least half the tokens produced by every talker were classified cor-

rectly using the SD measure; of 12 talkers, 7 (2 male and 2 female CE talkers and 1 male and 2 female CF talkers) produced over 80% of tokens with well-separated SD values. Of 12 talkers, only 5 (1 male and 1 female CE talker and 1 male and 2 female CF talkers) produced over 80% tokens with well-separated relative intensity values. This was the case even when the tokens that had been excluded from analyses due to unclear bursts were included to get an estimate of tokens correctly classified using the relative intensity measure (values in Table III, footnote a). Thus, of all measures, SD most consistently distinguished between individual talkers of CE and CF.

IV. GENERAL DISCUSSION

What emerges from the present study is an acoustic description of both voiced and voiceless coronal stops in CE as well as in CF. Data confirm that for coronal bursts produced in syllable-initial position, CE stops, like AE stops, contrast short-lag VOT with long-lag or aspiration; CE bursts are loud, have a higher mean burst frequency, and are more compact as measured by standard deviation, with a more peaked spectral shape as measured by the kurtosis of burst frequency. CF stops, on the other hand, contrast lead VOT and short-lag VOT values with some overlap; CF bursts are lower in intensity, have a lower mean burst frequency, and are more diffuse as measured by standard deviation, with a less peaked spectral shape as measured by the kurtosis of burst frequency. Thus, acoustic data confirm that coronal stops differ in their phonetic implementation across the two languages.

Analyses of voicing differences in CE and CF as measured by VOT for coronal stop-initial words replicate those reported in Caramazza *et al.*'s (1973) study. CF talkers produce overlapping distributions of VOT for voiced and voiceless stops. The distribution of VOT in CF has been reported to be different from French from France (European French or EF); Caramazza and Yeni-Komshian (1974) report that unlike in CF, VOT for voiced and voiceless stops in EF do not overlap. They attribute the differences between EF and CF to the extensive contact of CF with CE. An overlap in the distribution of VOT for /d/ and /t/ precludes VOT from being a sufficient cue for voicing in CF.

In the present study, in addition to differences in VOT, voiced and voiceless tokens in CF differed systematically on relative burst intensity and mean burst frequency. Thus, in CF, burst intensity and mean burst frequency may supplement VOT differences to cue voicing differences. Although there have been suggestions that burst intensity differences may signal voicing (Pickett, 1999), previously there has been little discussion of mean burst frequency as a cue to voicing. Despite low-pass filtering of voiced tokens, a lower mean frequency for voiced stops may have resulted from greater low-frequency energy accounting for the pattern of results obtained here.

Besides providing information about voicing differences in CF, burst intensity and mean burst frequency were also significantly different across CE and CF. CE and CF stops also differed in SD and kurtosis of burst frequency. Perhaps not surprisingly, the burst intensity and spectral measures were correlated. A strong negative correlation was observed between SD and kurtosis (−0.83 for male talkers and −0.70 for female talkers). Other significant correlations for male talkers include correlations between SD and relative intensity (0.37), SD and skewness (−0.32), skewness and mean frequency (−0.32), skewness and kurtosis (0.56), and relative intensity and kurtosis (−0.38). Significant correlations for female talkers include SD and relative intensity (0.40), SD and mean frequency (−0.40), and mean frequency and skewness (0.70).

A strong correlation between SD and one (or more) of the other three moments would explain why the other three moments did not account for any significant variance in the discriminant analyses. Stoel-Gammon *et al.* (1994) do not report correlations between the various spectral measures. Of the correlations reported above, the strong negative correlation between SD and kurtosis is most remarkable. Specifically, it is possible that SD and kurtosis of burst frequency are consequences of the same underlying articulatory gesture.

Although differences in the intensity of burst are thought to be consequences of place of articulation differences (Jongman *et al.*, 1985), we know little about how the spectral moments map on to articulation. Note that the mean burst frequency or the first spectral moment measured in this experiment is not to be confused with the peak spectral location. The former is the average frequency of the burst power spectra, while the latter is the highest amplitude peak of the FFT spectrum. While the peak spectral location is correlated with the length of the front cavity, there is no evidence that the mean burst frequency is determined by the location of the constriction [Forrest *et al.*, 1988; see also Jongman *et al.* (2000) for evidence of this distinction in the analysis of fricatives].

Instead of being consequences of differences in place of articulation, differences in spectral shape of coronal stops between CE and CF stops may relate to variations in the degree of damping of the active articulator. Tokens produced with longer constriction length are likely to be more damped. Due to greater damping, these tokens are likely to have a greater bandwidth and lesser energy in the higher frequencies. In this study, CF stops have higher SD and lower kurtosis and lesser energy in the higher frequencies than CE stops.

Currently, there are no articulatory data on CE and CF coronal stops to directly test this hypothesis. However, neither the acoustic data presented in this study nor Dart's articulatory data support Stevens *et al.*'s (1985) claim that the active articulator determines the place of articulation. Dart's data indicate that while place and active articulator used are often correlated, individuals may use one or the other or both. Similarly, in the acoustic data presented here, although the relative intensity measure was significantly correlated with all other measures of spectral shape, there was individual variability in the measures used by each talker. Some of the correlations in Dart's articulatory data as well as the acoustic data presented here no doubt arise from anatomical and biomechanical constraints on the movement of articula-

tors. As Dart succinctly points out, "it is very difficult for someone with normal dentition to put the tip of the tongue on the teeth without the blade also touching the base of the teeth" (1998, p. 73), confounding place of articulation with active articulator used to produce coronal stops.

In this study, only burst intensity and spectral measures, one set of cues relating to stop place of articulation, were measured for CE and CF coronal stops. Given that between 8% and 40% of voiced stops produced by CF talkers were produced without clear bursts, place differences must be signaled by acoustic cues relating to more than just the burst. Besides acoustic information in the burst, formant frequency changes or transitions have been previously used to identify place of articulation for stop consonants (Delattre *et al.*, 1955; Kewley-Port, 1983; Klatt, 1979, 1987). However, comparison of formant frequency transitions across languages to identify place distinctions is confounded by systematic differences in the formant values of vowel targets themselves.

A comparison of F_1-F_2 space in CE and CF (Fig. 1) indicates that CE vowels are produced with a tongue position that is more posterior and lower, whereas CF vowels are produced with a tongue position that is more forward and higher. Although this difference in the vowel systems could be attributed to the CE vowels having been produced in alveolar context and CF stops in dental context, it is unlikely for two reasons. First, F_1 and F_2 measurements were made at the mid-point between vowel onsets and offset and the effects of consonant context are less likely to extend to such vowel targets. Second, other researchers (Dart, 1991) have previously reported systematic differences unrelated to phonetic context between the vowel systems of AE and EF. Articulatory and vowel formant data for stops produced in several additional phonetic contexts (i.e., bilabial or velar) are required to disambiguate between these two accounts.

Not only is it problematic to compare transition information across languages in view of systematic differences in vowel targets across the languages unrelated to phonetic context, but there is also evidence to suggest that even in languages that contrast more than one coronal place, transition data are inadequate to specify place information. In a recent investigation of Australian aboriginal languages, Yanyuwa and Yindjibarndi, Tabain and Butcher (1999) report that F_2 transition information (incorporated into locus equations) does not provide sufficient information to uniquely identify place differences within coronal consonants. These aboriginal languages share an extensive set of place contrasts, including four coronal place distinctions, but not voicing or manner contrasts for stops (Busby, 1980; Dixon, 1980).

To summarize, the VOT results presented in this study replicate Caramazza *et al.*'s (1973) findings; of the burst cues measured in this study, CE and CF bursts differ consistently across gender and voicing in mean frequency, SD, and kurtosis of burst spectra. Relative intensity and skewness of burst spectra are less consistent and help to differentiate tokens produced only by female talkers. Analyses of differences in CE and CF coronal stops as measured by burst intensity and spectral cues support and extend investigations by Jongman *et al.* (1985) and Stoel-Gammon *et al.* (1994). Results from the present study provide a quantitative analysis of acoustic correlates of subgroups of coronal stops across gender, voicing, and languages. Moreover, individual patterns are presented and discussed in addition to group data. Given the potential for variability in production of coronal stops cross-language, the discussion of individual patterns is particularly useful.

Data from this study provide the first step in establishing measurable and reliable differences in the language-specific characteristics of bursts associated with CE and CF coronal stops. Apart from providing an acoustic description of language-specific characteristics of coronal stops in CE and CF, results from this study will provide an essential baseline for investigations of monolingual and bilingual acquisition of coronal stops. Empirical data are needed from further investigations to clearly delineate the articulatory characteristics of coronal stops in CE and CF. There is also a need for systematic articulatory-acoustic investigation to ascertain which acoustic properties of the burst map on to observed articulatory differences between CE and CF stops.

ACKNOWLEDGMENTS

This study is part of M. Sundara's doctoral thesis. I would like to thank Linda Polka, Shari Baum, Anders Löfqvist, and two anonymous reviewers for comments on previous versions of this paper, and Georgina Hernandez for her help with recording French speakers. This study was supported by an Internal SSHRC grant (202686) from McGill University to M. Sundara.

Allen, J. S., Miller, J. L., and DeSteno, D. (**2003**). "Individual talker differences in voice-onset-time," J. Acoust. Soc. Am. **113**, 544–551.

Blumstein, S. E., and Stevens, K. N. (**1979**). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," J. Acoust. Soc. Am. **66**, 1001–1016.

Boersma, P., and Weenink, D. (1992). The Praat program, University of Amsterdam.

Buder, E., Williams, K., and Stoel-Gammon, C. (**1995**). "Characterizing the adult target: Acoustic studies of Swedish and American-English /t/ and /p/," in *Proceedings of the 13th International Conference of Phonetic Sciences*, edited by K. Elenius and P. Brandevud, Stockholm, Vol. **3**.

Busby, P. (**1980**). "The distribution of phonemes in Australian aboriginal languages," Papers Austral. Linguist. **4**, 73–139.

Caramazza, A., and Yeni-Komshian, G. H. (**1974**). "Voice onset time in two French dialects," J. Phonetics **2**, 239–245.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., and Carbone, E. (**1973**). "The acquistion of a new phonological contrast: The case of stop consonants in French-English bilinguals," J. Acoust. Soc. Am. **54**, 421–428.

Chen, M. (**1970**). "Vowel length variation as a function of the voicing of the consonant environment," Phonetica **22**, 129–159.

Chomsky, N., and Halle, M. (**1968**). *The Sound Pattern of English* (Harper and Row, New York).

Cole, J., Choi, H., Kim, H., and Hasegawa-Johnson, M. (**2003**). "The effect of accent on the acoustic cues to stop voicing in Radio News Speech," in *Proceedings of the 15th International Congress of Phonetic Sciences*, edited by M. J. Solé, D. Recasens, and J. Romero, 2665–2668.

Dart, S. N. (**1991**). "Articulatory and acoustic properties of apical and laminal articulations," UCLA Work. Pap. Phonetics **79**, 1–155.

Dart, S. N. (**1998**). "Comparing French and English coronal consonant articulation," J. Phonetics **26**, 71–94.

Delattre, P., Liberman, A. M., and Cooper, F. S. (**1955**). "Acoustic loci and transitional cues for consonants," J. Acoust. Soc. Am. **27**, 769–774.

Dixon, R. M. W. (**1980**). *The Languages of Australia* (Cambridge U. P., Cambridge).

Flege, J. E., and Eefting, W. (**1987**). "Cross-language switching in stop consonant perception and production by Dutch speakers of English,"

Speech Commun. **6**, 185–202.
Forrest, K., Weismer, G., Milenkovic, P., and Dougall, R. N. (**1988**). "Statistical analysis of word-initial voiceless obstruents: Preliminary data," J. Acoust. Soc. Am. **84**, 115–123.
Francis, W. N. (**1958**). *The Structure of American English* (Ronald, New York).
Henton, C., Ladefoged, P., and Maddieson, I. (**1992**). "Stops in the world's languages," Phonetica **49**, 65–101.
Jongman, A., Blumstein, S. E., and Lahiri, A. (**1985**). "Acoustic properties for dental and alveolar stop consonants: a cross-language study," J. Phonetics **13**, 235–251.
Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives," J. Acoust. Soc. Am. **108**, 1252–1263.
Kessinger, R. H., and Blumstein, S. E. (**1998**). "Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies," J. Phonetics **26**, 117–128.
Kewley-Port, D. (**1983**). "Measurement of formant transitions in naturally produced stop consonant-vowel syllables," J. Acoust. Soc. Am. **72**, 379–389.
Klatt, D. H. (**1979**). "Synthesis by rule of consonant-vowel syllables," Speech Commun. Group Working Pap. **3**, 93–105.
Klatt, D. H. (**1987**). "Review of text-to-speech conversion for English," J. Acoust. Soc. Am. **82**, 737–793.
Lieberman, P., and Blumstein, S. E. (**1988**). *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge U. P., Cambridge).
Lisker, L., and Abramson, A. S. (**1964**). "A cross-language study of voicing in initial stops: Acoustical measurements," Word **20**, 384–422.
Mack, M. (**1989**). "Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals," Percept. Psychophys. **46**(2), 187–200.
Mayer, M. (**1969**). *Frog, where are you?* (Puffin Pied Piper, New York).
Picard, M. (**1987**). *An Introduction to the Comparative Phonetics of English and French in North America* (Benjamin, Amsteram).
Picard, M. (**2001**). *Phonetics and Phonology for ESL and TESL Teachers: Comparing Canadian English and French* (Concordia Univ., Montreal).
Pickett, J. M. (**1999**). *The Acoustics of Speech Communication* (Allyn and Bacon, Boston).
Stevens, K. N., Keyser, S. J., and Kawasaki, H. (**1985**). "Toward a phonetic and phonological theory of redundant features," in *Invariance and Variability of Speech Processes*, edited by J. S. Perkell and D. H. Klatt (Erlbaum, Hillsdale, NJ), pp. 426–463.
Stoel-Gammon, C., Williams, K., and Buder, E. (**1994**). "Cross-Language Differences in Phonological Acquisition: Swedish and American /t/," Phonetica **51**, 146–158.
Summerfield, Q. (**1981**). "Articulatory rate and perceptual constancy in phonetic perception," J. Exp. Psychol. Hum. Percept. Perform. **7**, 1074–1095.
Tabain, M., and Butcher, A. (**1999**). "Stop consonants in Yanyuwa and Yindjibarndi: locus equation data," J. Phonetics **27**, 333–357.
Volaitis, L. E., and Miller, J. L. (**1992**). "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories," J. Acoust. Soc. Am. **92**, 723–735.

Loudness predicts prominence: Fundamental frequency lends little

G. Kochanski,[a] E. Grabe, J. Coleman, and B. Rosner
The University of Oxford Phonetics Laboratory, 41 Wellington Square, Oxford OX1 2JF, United Kingdom

(Received 4 November 2004; revised 10 March 2005; accepted 6 April 2005)

We explored a database covering seven dialects of British and Irish English and three different styles of speech to find acoustic correlates of prominence. We built classifiers, trained the classifiers on human prominence/nonprominence judgments, and then evaluated how well they behaved. The classifiers operate on 452 ms windows centered on syllables, using different acoustic measures. By comparing the performance of classifiers based on different measures, we can learn how prominence is expressed in speech. Contrary to textbooks and common assumption, fundamental frequency (f_0) played a minor role in distinguishing prominent syllables from the rest of the utterance. Instead, speakers primarily marked prominence with patterns of loudness and duration. Two other acoustic measures that we examined also played a minor role, comparable to f_0. All dialects and speaking styles studied here share a common definition of prominence. The result is robust to differences in labeling practice and the dialect of the labeler.
[DOI: 10.1121/1.1923349]

PACS number(s): 43.70.Fq, 43.66.−x, 43.71.−k [DOS] Pages: 1038–1054

I. INTRODUCTION AND BACKGROUND

In English, some syllables are special and more important, and others less so. The important ones are described, variously, as bearing "stress accents" (Beckman, 1986), as "prominent," or by other terms, but a definition strictly in terms of their acoustic properties has been lacking.

Our central question is the following: Using acoustic data, what property allows the best machine replication of the prominence judgments of human listeners? The experiments here focus on acoustic cues in a window that includes the syllable under consideration and the neighboring syllables. We explore seven dialects and three different styles of speech: lists of sentences, story paragraphs, and a retelling of a story.

Many people have looked at cues to prominence, and they have reached a variety of answers. Passy (1891, pp. 41-42; 1906, p. 27), Sweet (1906, p. 47), Trager and Smith (1951), and others impressionistically described English prosody in terms of "force" or "accent" (equated to loudness), and "intonation" (equated to pitch). Fry (1955, 1958) did early perceptual studies on minimal pairs of synthesized English words that are distinguished by a difference of stress placement (e.g., súbject versus subjéct). He found that the more prominent syllable was marked, in decreasing order of importance, by duration, f_0, and amplitude. His results have achieved wide currency in the linguistic community, despite the study's limitation to single, isolated words.

Other experiments with careful, "laboratory" speech have yielded a variety of results. Lieberman (1960), for instance, described a very early system for deducing lexical stress from acoustics. His work indicates that f_0, amplitude, and duration, are similarly important and that each individually is a good predictor of prominence. However his exceptionally good classification probabilities are due to the explicit selection of clearly enunciated and unambiguous speech: utterances were used as stimuli only when four human judgments all agreed on the stress placement.

Likewise, synthesis studies (discussed in Sec. III E) showed that f_0 bumps can induce the perception of prominence (Gussenhoven *et al.*, 1997; Rietveld and Gussenhoven, 1985; Terken, 1991).

Other laboratory work is often taken to support the importance of f_0. For instance, Cooper *et al.* (1985) and Eady and Cooper (1986) found significantly different f_0 patterns in a sentence as a function of the focus position (roughly, the pattern of prominences). However, this result needs to be interpreted carefully. These papers reported statistically significant changes to the average f_0 of a group of utterances. While this is useful from a descriptive point of view, the usual listener only hears only one utterance at a time and does not have the luxury of averaging several repetitions before responding. Consequently, while averages of two classes may be significantly different, the distributions of individual measurements may overlap enough so that a listener could not usefully decide what has been said, based on a single utterance.

On the other hand, Beckman (1986) saw substantial correlations of prominence with a combination of amplitude and duration. Turk and Sawusch (1996) have conducted synthesis experiments, comparing isolated instances of (e.g., máma versus mamá) to tease apart the relative importance of loudness and duration to perception judgments. They come to two main conclusions. The first is that these two acoustic measures were perceived together as a single percept; the second is that loudness made a negligible contribution to the results of their rating scale experiment.

Tamburini (2003) has had success with a prominence detection system for more natural speech that assumes an

[a]Electronic mail: gpk@kochanski.org

important role for amplitude contrasts between neighboring syllables. However, he did not measure what the differences were between prominent and nonprominent syllables; he simply reported that a particular algorithm achieved 80% correct classification on a corpus of American English.

Another system for automated prominence transcription, built by Silipo and Greenberg (1999, 2000) was tested on an American English corpus with several plausible acoustic correlates of prominence. This study was a first attempt to understand prominence of natural speech, as opposed to careful laboratory speech, although there is not a complete published description of the experiment. In their work, f_0 was shown to have relatively little importance. Comparisons to this work are difficult in that their system had strong assumptions wired in (which we test instead of assuming). For instance, they assumed that f_0 induced prominence only through a single f_0 contour: a symmetrical bump. However, they achieved good performance ($\approx$80% correct classification) by operating their system on the product of syllable-averaged amplitude and vowel duration, which suggests that amplitude and duration are good indicators of prominence. The strong assumptions built into the classifier mean that little can be said about their other, less successful combinations of acoustic features.

In summary, the literature is not completely clear on what acoustic properties of speech communicate prominence, but f_0 is not the complete story. Nevertheless, much work on intonation and prosody, especially in the field of intonational phonology, implicitly assumes that prominence is primarily a function of f_0 [see Terken and Hermes (2000) and Beckman (1986) for reviews].

Prominence of a syllable is sometimes explicitly equated with special f_0 motions in its vicinity. For instance, Ladd (1996) states:

> A pitch accent may be defined as a local feature of a pitch contour—usually, but not invariably a *pitch change*, and often involving a local minimum or maximum—which signals that the syllable with which it is associated is *prominent* in the utterance. … If a word is prominent in a sentence, this prominence is realized as a pitch accent on the 'stressed' syllable of the word.

Similarly equating pitch motions with prominence, Welby (2003) writes: "The two versions [of an utterance] differ in that (1) has a pitch accent, a prominence-lending pitch movement… ." A standard textbook by Roca and Johnson (1999, p. 390) claims that pitch patterns can be used to prove the reality of abstract lexical stress: they state that one can test syllables for stress by looking at pitch in their vicinity. Another textbook, Clark and Yallop (1995, p. 349), gives a less extreme view but still espouses the primary importance of f_0 when discussing the acoustic implementation of lexical stress: "Our perception is in fact likely to be more responsive to the pitch pattern than other factors." Similar views were put forth by Bolinger (1958), 't Hart *et al.* (1990), and others. Since the assumption that pitch implies prominence underlies much work, it needs to be thoroughly tested.

To do this, we studied seven dialects of British English. We looked for patterns in f_0 and other acoustic properties that could separate prominent from nonprominent syllables.

II. DATA AND METHODS

A. Overview

This experiment is conducted on a large corpus of natural speech. Listeners judge the prominence of syllables, and the speech is analyzed to find the acoustic basis of their judgements.

We measure a selection of acoustic properties in a window that centers on a syllable. Listeners mark the syllables as either prominent or not. Five time series are then computed from the speech signal: measures of loudness, aperiodicity, spectral slope, f_0, and a running measure of duration. These measures are transformed into coefficients for Legendre polynomials and fed into a classifier that is trained to reproduce the human prominent/nonprominent decision. Finally, the classifier performance is measured on a test set, and the result reveals how consistently the speakers used each of the measured properties to mark prominent syllables.

The first step in the analysis is the extraction of prominence marks (Sec. II C) from a labeled corpus (Sec. II B). Second, the five time series ("acoustic measures") are computed from the speech; details are in Sec. II D. Third, each property is normalized (Sec. II E), then, fourth, the data are represented as a best-fit sum of Legendre polynomials (Sec. II G 1). The coefficients of the polynomials that result from the fit are a compact representation of the shape of the time series in the window. (Some of these coefficients are easy to interpret: the first coefficient is the average over the window; the second coefficient captures the overall rate of change.)

These coefficients form a feature vector, which is the input for the fifth stage of the analysis. The feature vector specifies a point in a space; hence the Legendre polynomial analysis maps an acoustic time-series into a single point into a, e.g., six-dimensional feature space. Each point in that space (each syllable) is labeled as prominent or nonprominent by a human. Fifth, we build a classifier (Sec. II H) on those vectors to reproduce the human prominence marks as well as possible. We use a quadratic discriminant forest classifer, which should be reasonably efficient for our features, which are roughly multivariate Gaussian and have no obvious complex structure.

We chose this classifier partially because it is a variant of a quadratic classifier, and can capture classes that are linguistically interesting. For instance, if f_0 indicated prominence by being either high or low at the syllable center (and nonprominent by being intermediate), we could capture that behavior. Likewise, if f_0 indicated prominence by slopes or extra variance, a quadratic classifier could capture such classes.

Since each class is defined by a full covariance matrix among all the orthogonal polynomial (OP) coefficients, it can represent complex patterns of low and high pitch combined with large and small standard deviations. Specifically, using this design of classifier will let us test models of prominence where $f_0(t)$ on a syllable is measured relative to any linear

combination of the surrounding f_0 measurements. This includes many plausible normalizations of $f_0(t)$ relative to preceding and/or following syllables, such as a consistent declination slope. We put quantitative limits on the classifier performance in Sec. III D.

Sixth, after the classifiers are built and tested, we compare the error rates for classifiers based on different acoustic measurements (Sec. III) to deduce how much information is carried by each acoustic property.

B. Corpus

We use the IViE (Intonational Variation in English) corpus [Grabe *et al.* (2001)], which is freely available on the Web at http://www.phon.ox.ac.uk/ivyweb. The IViE corpus contains equivalent sets of recordings from seven British English urban dialects: Belfast, Bradford (speakers of Punjabi heritage), Cambridge, Dublin, Leeds, London (speakers of Jamaican heritage), and Newcastle. Speech of six of the twelve speakers per dialect have been intonationally labeled. The speakers were students in secondary schools, with a mean age of 16 years ($\sigma < 1$ year). We use data from three styles of speech: the "sentences," "read story," and "retold story" sections of IViE (abbreviated in the following as "sentences," "read," and "retold").

In "sentences," speakers read lists of sentences like "We were in yellow." or "May I lean on the railings?" The "read" section involved reading a version of the Cinderella story, containing narration and dialog. In the "retold" section, the subjects retold the story in their own words, from memory. The IViE corpus has about 240 min of annotated data, which includes about 7200 intonational phrases and 14 400 accents. For this analysis, we use all the annotated IViE single-speaker data.

C. Prominence marks

The IViE corpus contains files marking prominent syllables. We adopted these as the primary data source. In IViE, all accented syllables are prominent and *vice versa*. The marks were made by two phoneticians (one of whom, EG, is an author), who are experienced in the analysis of English intonation. The phoneticians were native speakers of Dutch and German who acquired RP English before adolescence. They consulted with a third phonetician who was a native speaker of British English. Accented syllables were marked according to the British tradition defined by O'Connor and Arnold (1973) and Cruttenden (1997), using the prosodic prominence hierarchy of Beckman and Edwards (1994). During labeling, the speech was heard and the speech wave form and f_0 trace were displayed on a screen.

Nonprominent syllables were not marked in IViE, but word boundaries were. Using the boundaries, one can deduce the locations of most nonprominent syllables and automatically mark them. We built a dictionary containing the number of syllables in a typical conversational version of each word or word fragment. An analysis program then scanned through the labeled part of the corpus. As each word was encountered, the program placed the correct number of syllable marks, evenly spaced throughout the word. Any syllables that IViE shows as prominent then replaced the nearest automatically generated mark. The remaining nonprominent syllables are needed as a comparison to the prominent syllables because the classifiers are trained and tested on their ability to separate two classes.

The primary set of speech data includes 2173 prominence marks out of an estimated 5962 syllables in the "sentence" style; 1919/5134 in the "read" style; and 805/2341 in the "retold" style. Most are on syllables that have primary or secondary lexical stress. In the "read" style in the primary set, the Belfast and Cambridge dialects had considerably more data labeled than other dialects: 34 and 32 audio files, respectively, versus a total of 18 for the other five dialects. Otherwise, the data were almost evenly balanced between dialects.

Two other sets of prominence marks were produced independently, to ensure that the primary data source reflected widely perceived properties of the language, rather than something specific to the primary labelers. These two secondary sets were smaller, but (unlike the primary data set) they also contained marks for the centers of nonprominent syllables. Data files were chosen randomly from "read" data obtained in Cambridge, Belfast, and Newcastle, from audio files that had transcriptions. The secondary sets were created by two people with significantly different training and dialects from the primary labelers.

In the labeling for the secondary sets, the labelers attempted to mark syllables that perceptually "stand out," giving minimal attention to meaning or syntax. No attempt was made to discriminate between lexical stress, focus, and other causes of prominence. No attempt was made to decide what type of accents were present or to define intonational phrases. One secondary labeler (GK, an author) is a native speaker of American English (suburban Connecticut), trained as a physicist. The GK set has 454 prominence marks out of 1385 syllables. The other secondary labeler (EL) is a native speaker of Scottish English (Glasgow), trained as a Medieval English dialectologist. The EL set has 775 prominence marks among 2336 syllables.

During the secondary labeling, only the speech wave form and word boundaries were displayed; IViE labels were not displayed; and the primary labelers were not consulted. Marks were placed without regard to a detailed phonetic segmentation; syllables were marked somewhere between the center of the voiced region and the temporal center of the syllable. The secondary labelers had the option of not labeling a word if the number of syllables was unclear or if it was a fragment. Otherwise, they marked each syllable as prominent or nonprominent.

The secondary sets include some data that are not in the primary set: 3/12 audio files in the GK set are not in the primary set, 8/24 in the EL set are not in the primary set, and only two audio files are common between the GK and EL sets. The secondary sets thus bring in new data and are almost independent of the primary set, but they have enough overlap to allow some limited comparison of the consistency of label placement.

Overall, the median spacing between neighboring syllable centers in the secondary sets is 180 ms (which is also

the median syllable duration). The median distance between prominence marks is 440 ms in the primary set and about 600 ms in the secondary sets.

D. Acoustic measures

We based the paper on five acoustic measures that are plausibly important in describing prosody. All are time series, and they describe the local acoustic properties. We used approximations to perceptual loudness, and phone duration, a measure of the voicing (aperiodicity), the spectral slope, and the fundamental frequency. In addition to the three classic contenders, we added a spectral slope measure because of the success of Sluijter's spectral slope measurement (Sluijter and van Heuven, 1996). Aperiodicity was added simply as a relatively unexplored candidate: it is sensitive to some prosodic changes (e.g., pressed versus modal versus breathy speech) and so might plausibly be correlated with prominence. Additionally, it is sensitive to the relative durations of vowels and consonants in syllables, and therefore might capture some duration changes associated with prominence. Loudness and duration, together, capture at least some of the acoustic features of vowel quality; reduced vowels tend to be quiet and short; more open vowels tend to be louder.

1. Loudness

The loudness measure is an approximation to steady-state perceptual loudness (Fletcher and Munson, 1933). The analysis implements a slightly modified version of Stevens' Mark VII computation (Stevens, 1971), which is an improved version of the ISO-R532 Method A standard noise measurement. We modified it to use 0.7 octave frequency bins rather than the full- or third-octave bands for which it was originally defined. It operates on the spectral power density derived from an $L=50$ ms wide, $1+\cos(2\pi(t-t_c)/L)$ window, and supposes that the rms speech level in an utterance is 68 dB relative to 20 $\mu\mathrm{N/m}^2$ sound pressure.

The IViE recordings were obtained in whatever spaces were available, so background noise is sometimes audible. The noise could affect our analysis because the weight we assign to acoustic measures depends on the loudness (Sec. II F), and changes in the weight will affect the orthogonal polynomial coefficients (Sec. II G 1). To minimize this problem, we subtracted an estimate of the background noise from the loudness.

The correction was

$$L^3(t) = \max(0, L_r^3(t) - \hat{L}_r^3), \tag{1}$$

where $L_r(t)$ is the raw (Stevens) loudness measure, $L(t)$ is a corrected loudness, excluding the background noise, and $\hat{L}_r$ is an estimate of the background noise loudness. Equation (1) is approximate and assumes that the speech and noise spectrum have the same shape; The ratio of peak speech power to the background noise is typically about 30 dB, however, so the correction only affects the quietest parts of most utterances. $\hat{L}_r$ was conservatively set equal to the fifth percentile of L_r, as all the utterances contained at least 5% silence. In other words, the analysis assumed that the quietest 5% of the data contained no speech and could be used to estimate the background noise level.

2. Running duration measure

The running duration measure, $D(t)$, is a time series whose value at each moment approximately equals the duration of the current phone. It is derived by finding regions with relatively stable acoustic properties and measuring their length. Longer phones, especially sonorants, will tend to have long regions that have nearly constant spectra and will give large values for $D(t)$. Shorter phones will give small values for $D(t)$. Stops are treated as the edge of a sonorant plus a silence, and bursts are effectively treated as separate entities. Short silences have the expected duration, but $D(t)$ is ill-defined for long silences.

To compute the running duration measure do the following.

> For every 10 ms interval in the utterance, compute a perceptual spectrum, $\psi(t_c, j)$, where t_c is the time of the window center and j is the frequency index, in bark. The frequency interval from 300 to 5500 Hz is used. The Fourier transform of the signal is taken over an $L=30$-ms-wide, $1+\cos(2\pi(t-t_c)/L)$ window.[1] Then the power spectrum is normalized by the total power within the window. The spectral power density is then collected into 1-bark-wide bins on 0.5 bark centers, and a cube-root is taken of the power in each bin. (The summed power across all bins is of the order of unity.)

Then, in a second pass, compute the $D(t)$ at each 10 ms interval as follows.

(1) Starting at $t=t_c$ with $\eta=0$, and moving t forward from t_c in 10 ms steps, accumulate $\eta=\eta+\Sigma_j(\psi(t,j)-\psi(t_c,j))^2$. This is a measure of how much the spectrum has changed over the interval between t_c and t.
(2) In the same sweep, accumulate $\Delta_{\mathrm{fwd}}=\Delta_{\mathrm{fwd}}+e^{-\eta/C}$, with $C=600$. As long as the accumulated difference is smaller than C, Δ_{fwd} will approximately equal the time difference, $t-t_c$, but when the spectrum changes and η becomes bigger than C, the accumulation will slow down and stop. The final value of Δ_{fwd} will be approximately equal to how far one can go in the forward-time direction before the spectrum changes substantially.
(3) Do the same in the reverse direction, to compute Δ_{rev}.
(4) The $D(t_c)$ is then $(10\ \mathrm{ms})\cdot(\Delta_{\mathrm{rev}}+\Delta_{\mathrm{fwd}}-1)$, where the final "−1" corrects for double counting of the sample at t_c.

Figure 1 shows a section of acoustic data and the resulting time series of $D(t)$ for a phrase "…go to the ball…," along with the input wave form. The values of $D(t)$ near each sonorant center approximately match the phone duration.

3. Aperiodicity

The aperiodicity measure, $A(t)$, ranges from 0 to the vicinity of 1. It assigns zero to regions of locally perfect periodicity, and numbers near one where the wave form of the signal cannot be predicted. (For stationary signals, the

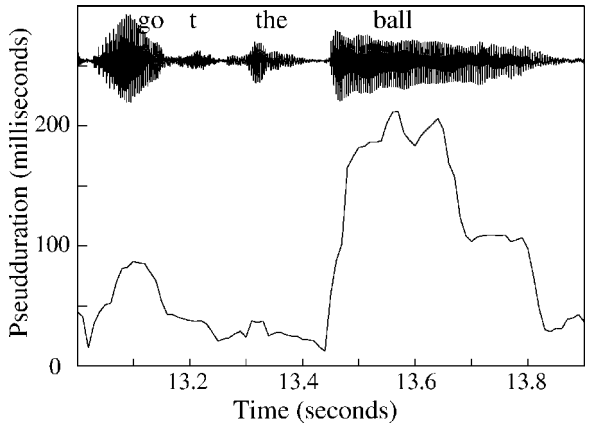


FIG. 1. Running duration measure, $D(t)$ (below, smooth curve) and acoustic waveform (top) for "…go to the ball… ." The sharp downwards step in $D(t)$ near 13.65 s corresponds to the transition between the vowel and liquid in "ball;" the two adjacent sounds have different durations.

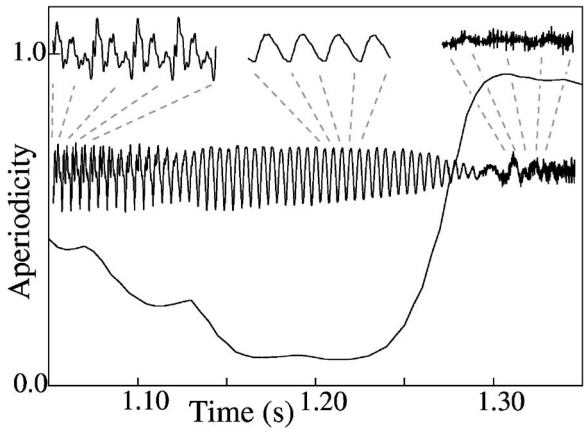


FIG. 2. Aperiodicity measure (below, smooth curve) and acoustic wave form (middle) for the end of "railings." Enlarged sections of the high-pass filtered wave form are shown above. The data show a vowel, nasal, and the final unvoiced fricative. Modest changes from period to period can be seen in the leftmost section of the wave form, leading to an intermediate value of $A(t)$; the middle section is more periodic, so $A(t)$ is close to zero; and the fricative produces a large value of $A(t)$.

maximum value is unity, but amplitude changes, especially on 20 ms or shorter time scales, will locally change the maximum.) It is related to Boersma's harmonics-to-noise ratio (Boersma, 1993) (HNR) and can be approximated by $A(t) \approx (1+10^{\mathrm{HNR}/10})^{-1/2}$. $A(t)$ can also be considered a measure of voicing, as voiced speech is often nearly periodic and unvoiced speech is typically aperiodic.

To compute $A(t)$, the audio signal first had low-frequency noise and DC offsets removed with a 50 Hz fourth-order time-symmetric Butterworth high-pass filter, and then was passed through a 500 Hz single-pole high-pass filter for pre-emphasis. The aperiodicity measure was derived by taking a section of the filtered signal defined by a Gaussian window with a 20 ms standard deviation and comparing it to other sections shifted by 2–20 ms. If the acoustic signal were exactly periodic with f_0 between 50 and 500 Hz, then one of the shifted windows would exactly match the starting window, and the difference would be zero. The value of $A(t)$ is proportional to the minimum rms mismatch between the windows.

To compute the aperiodicity measure:

(1) For each possible shift, δ, between 2 and 20 ms, compute $p_\delta(t) = (\tilde{s}(t+\delta/2) - \tilde{s}(t-\delta/2))^2$, where $\tilde{s}(t)$ is the filtered acoustic wave form at time t.
(2) Compute $P(t) = \tilde{s}^2(t)$.
(3) Convolve $p_\delta(t)$ and $P(t)$ with 20 ms standard deviation Gaussians to yield $\bar{p}_\delta(t)$ and $\bar{P}(t)$, respectively.
(4) Compute $\hat{p}(t) = \min_\delta\{p_\delta(t)\}$, i.e., find the minimum error at each time, minimizing over all the shifts, δ.
(5) The aperiodicity measure is then $A(t) = \hat{p}^{1/2}(t)/(2\bar{P}(t))^{1/2}$.

Figure 2 shows a small section of acoustic data and the resulting time series of $A(t)$ near the end of the word "railings," along with the input wave form and enlarged sections of the preprocessed (high-pass filtered) wave form.

4. Spectral slope

The spectral slope estimator is intended to approximate the average slope of the power spectrum near the glottis, i.e., to be relatively insensitive to the formant structure of the speech. It takes a local spectrum of the speech wave form, computed in a 30 ms window, and collects the power in 1 bark bins (the bins are overlapping, on 0.5 bark centers). Next, a cube-root is taken to yield an approximation to the perceptual response in each frequency band. Finally, the spectral slope estimate, $S(t)$ is the slope of the best fit to the Bark-binned spectrum between 500 and 3000 Hz.

Related measures are described in Heldner (2001), Sluijter and van Heuven (1996), and references therein. Our measure is not identical to prior measures, but should have a substantial correlation with them. We chose it because it could be computed easily and reliably on a large corpus, in a strictly automated manner.

5. Fundamental frequency

We compute an estimate of the fundamental frequency, $f_0(t)$, with the get_f0 program from the ESPS package (Entropic Corp.). The program also produced a voicing estimate, $V(t)$, which was zero or one at each 10 ms interval. Before further analysis, the f_0 tracks were inspected for gross errors. An automated procedure that (a) searched for substantial jumps, and (b) looked for f_0 values close to the subject's minimum and maximum f_0 was used to identify likely problem areas. A roughly equal number of problems were identified during manual inspections driven by various checks not directly associated with f_0. About half of the utterances were manually inspected, and we checked f_0 on every utterance that we inspected. Finally, another set of utterances was inspected because the mean-squared error of the Fourier fit was unusually large.

Once an utterance was identified as having possible problems with its pitch tracking, a labeler inspected each area and marked a change to $f_0(t)$ or $V(t)$ if get_f0 results did not match the perceived sound. The labeler had the option of shifting f_0 up or down by a factor of 1.5, 2, or 3, and/or marking a region as irregularly voiced or unvoiced. In all, 498 regions in 254 utterances were marked, of which 75

regions included upwards octave shifts of f_0, while 15 were f_0 shifts by other factors. The remaining majority were either marked as irregular phonation or no phonation. The median length of the marked regions is 56 ms.

E. Normalization

To compute the orthogonal polynomial coefficients, we take data from a window of width w, centered on the relevant syllable. We then normalized the time axis so that the data ranged from -1 to 1, in preparation for fitting OPs to the data. This converted the t axis to an x axis via $x=2(t-t_c)/w$, where t_c is the time of the center of the syllable.

Additionally, we normalized each acoustic property relative to a weighted average of the corresponding speaker's data of that property over the corpus. For $f_0(t)$, we divided by the 10%-trimmed weighted average[2] of $f_0(t)$. For $A(t)$, we divided by the 35%-trimmed weighted average.[3] For $S(t)$, we subtracted the 10%-trimmed weighted average. Finally, because the microphone placement was not controlled in the recordings in the IViE corpus, we normalized $L(t)$ locally, so motions of the speaker would not have much effect on the normalized amplitude. We normalized $D(t)$ and $L(t)$ by dividing by the 5%-trimmed weighted average over the window. This local normalization reduces the sensitivity of the analysis to changes in the speaking rate or microphone position between one utterance and another.

F. Weighting the data

Not every part of the acoustic measures are equally valuable. For instance, f_0 information is meaningless in unvoiced regions, as is S, D, and A. It is necessary, then, to give a weight to each point in the data when we later compute the orthogonal polynomial fits in Eq. (4). The weight function is written $W_\alpha(t)$, where α indicates one or another of the acoustic measures.

The detailed form of the weight functions are somewhat arbitrary, but we made plausible choices, then tested that they are close to optimal (see Appendix A). All the weight functions are computed from the acoustic measures before normalization.

The weights are different for each acoustic measure, but they share some common features. Specifically, using weights that increase with loudness will emphasize regions that may be more perceptually important. Under real-world conditions, speech more than 15 dB below the peaks is often buried in ambient noise, and thus has less importance.[4]

For f_0, in addition to perceptual importance, we were motivated by considerations of the accuracy and reliability of the pitch tracker. We took $W_{f_0}(t) = L^2(t)\cdot\max(1-A^2(t),0)^2\cdot V(t)\cdot I^2(t)$, where $V(t)$ is the voicing estimate from Sec. II D 5. The component $I(t)$ is a semiautomatic indicator of irregular voicing. It is a product of factors:

- A factor that de-weights the edges of a voiced region to reduce the impact of segmental effects. It is unity everywhere except in the first and last 10 ms of each voiced region where it is 0.5.
- A factor that de-emphasizes unstable f_0 readings: $(1+(\delta/10\text{ Hz})^2)^{-1}$, where δ is the pitch change over the 10 ms interval between samples.
- A factor that is 1, except 0.5 in regions hand-marked as irregularly voiced (see Sec. II D 5).

This weight function forces the orthogonal polynomial fit to $f_0(t)$ to be most precise in loud regions that are periodic, such as syllable centers.

For the spectral slope, we suppressed the unvoiced regions to avoid the large jumps in $S(t)$ that occur across voiced–unvoiced transitions. Thus, we used $W_S(t) = L^2(t)\cdot V(t)$. For aperiodicity and the running duration measure we used $W_A(t)=L^2(t)$ and $W_D(t)=L^2(t)$. Finally, for $L(t)$, we used a uniform weight: $W_L(t)=1$. The net result of our weighting choices is to focus on the peak of the syllable, paying less attention to the margins, especially consonant clusters.

Weighting the data with a power of the loudness gives us some sensitivity to the relative timing of f_0 excursions with respect to syllable centers. For instance, f_0 peaks that appear earlier than syllable centers will have the largest weight applied to their falling edge. The resulting OP coefficients will be biased toward those those of a falling accent. A delayed f_0 peak will have more weight placed on its rising edge and will push the coefficients towards those of a rising accent.

G. Orthogonal polynomials

We use orthogonal polynomials because the intentionally controlled aspects of intonation are, by and large, smooth and continuous. This is especially true for $f_0(t)$ [Kochanski *et al.* (2003, Sec. 1.2), Kochanski and Shih (2000)], because f_0 is controlled by muscle tensions that are smooth functions of time. We chose Legendre polynomials (Hochstrasser, 1972) which have the property of orthogonality:

$$\sum_x P_i(x)\cdot P_j(x)\cdot\omega(x)\cdot dx = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Here, $P_i(x)$ is the ith Legendre polynomial, $\omega(x)$ is the weight function that specifies the family of orthogonal polynomials [$\omega(x)=1$ for Legendre polynomials]. The sum is computed on the 10 ms grid where the acoustic measures are computed. Note that $\omega(x)$ and $W(x)$ are not the same: $\omega(x)$ is a global property of the entire analysis; $W(x)$ is the weight function used to fit the sum of polynomials to a particular utterance.

This orthogonal polynomial analysis is similar to a Fourier transform in that the low-ranking polynomials pick out slowly varying properties and the higher-ranking polynomials pick out successively more rapidly varying properties. The nth Legendre polynomial has $(n-1)/2$ peaks and the

same number of troughs, if we count a high (low) point at an edge of the utterance as half a peak (trough).

1. Deriving coefficients: Fitting the acoustic data

One can derive coefficients that represent an acoustic measurement by fitting it with the sum of Legendre polynomials,

$$y(x;c) = \sum_x c_i \cdot P_i(x), \tag{3}$$

using a regularized, weighted linear regression. In Eq. (3), c_i are the coefficients that multiply each Legendre polynomial, and $y(x;c)$ is a model for the data. The model (y) is x (e.g., time) dependent, and also depends on the coefficients, c. To compute the coefficients that best represent some data $\alpha(x)$, we minimize

$$\mathbb{E}_\alpha = \sum_x W(x) \cdot (y(x;c) - \alpha(x))^2 + \gamma \cdot c_i^2. \tag{4}$$

The first term is the normal sum-squared-error term; the second term is a regularization term. In Eq. (4), $\alpha(x)$ stands for each of the five acoustic time series, and γ is the strength of the regularization. The regularization causes $c_i \to 0$ when $\gamma \to \infty$, and is equivalent to assuming a Gaussian prior probability distribution with a width proportional to γ^{-1} in a maximum a posteriori probability estimator. Descriptions of the method can be found in Press *et al.* (1992, pp. 808–813) and Gelman *et al.* (1995).

We use linear regularization because some of the syllables have $W(x) \approx 0$ over 50% or more of the window; an example might be a syllable with a long fricative when one is fitting $f_0(t)$. In such a case, Eq. (4) becomes nearly degenerate when $\gamma = 0$ and yields large, canceling values of the coefficients c_i. The resulting c_i are far outside the distribution obtained for most syllables and degrade the classifier performance by violating its assumption of Gaussian classes.

Regularization can limit these spurious values of c_i. We chose $\gamma = 10^{-4} \Sigma_x W(x)$, which has the effect of reducing most c_i by only about 1%, but yields fairly good behavior for the hard cases that have large regions in which $W(x) \approx 0$.

By experimentation, we found that good fits to the time series of acoustic data can be obtained by using $1 + w/2\tau_\alpha$ orthogonal polynomials, where w is the length of the analysis window and $\tau_L = 60$ ms, $\tau_D = 70$ ms, $\tau_{f_0} = 90$ ms, $\tau_A = 80$ ms, and $\tau_S = 90$ ms. We make τ_L small because the loudness contours have sharp features which require a higher density of orthogonal polynomials in order to get a good fit; τ_{f_0} is adequate to represent the relatively slow f_0 variations. Others are in between.

The fits are generally quite accurate. The weighted rms error between the normalized time series and the fit is 0.008 (about 1 Hz) for f_0, 0.14 (i.e., 15%) for loudness, 0.09 (i.e., about 8 ms) for $D(t)$, 0.13 (i.e., about 13% of the median) for aperiodicity, and 0.003 (i.e., less than a 1 dB shift in the spectral power density at 3000 Hz relative to the power at 500 Hz). This is probably good enough to be indistinguishable by human perception, so we presumably capture most of the relevant information. Appendix B shows that our results are relatively insensitive to the values of τ_α or (equivalently) to the accuracy of the fit.

The weighted orthogonal polynomial fit to $f_0(t)$ is not strongly affected by small changes in which regions are voiced. Indeed, if $f_0(t)$ were fit exactly, de-voicing small regions would have no effect at all. Much of the $f_0(t)$ time series is indeed smooth and well-fitted by the polynomials, so changes to voicing are primarily captured by $L(t)$ and $A(t)$.

2. Transforming coefficients to make them more Gaussian

Next, we transform the coefficients to remove any obvious nonlinear correlations. We saw that for f_0 and especially loudness, the scatter plot of c_0 vs c_1 was crescent-shaped. However, it could be made much closer to a Gaussian by the following adjustments: $c_0 \Leftarrow c_0 - \kappa c_1^2$. Since the histograms for c_2 and other coefficients also had visible curvature, all the coefficients except c_1 were adjusted via

$$c_i \Leftarrow c_i - \kappa_{\alpha,i} c_1^2. \tag{5}$$

A linear least-squares procedure was used to determine $\kappa_{\alpha,i}$, from the union of the prominent and nonprominent data. For each coefficient (except c_1) and for each acoustic measure, α, the scatter-plot of c_i vs c_1 was fitted to $\hat{c}_i = \eta_{\alpha,i} + \nu_{\alpha,i} c_1 + \kappa_{\alpha,i} c_1^2$, and c_i was then corrected via Eq. (5). We did not need to consider η further, as it is picked up by μ in the classifier, and ν becomes part of the classifier's covariance matrix. The transformed c_i is the feature vector that will be used by the classifier.

H. Classifier

We developed a Bayesian quadratic forest classifier, inspired by the forest approach of Ho (1998). The classifier is a straightforward application of Bayes' theorem. To build the classifier, assume that there are M classes, each defined by a multivariate Gaussian probability distribution

$$\begin{aligned} &P(\boldsymbol{z}|\text{class } i) \\ &\quad = P(\boldsymbol{z}|\boldsymbol{\mu}_i, H_i) \\ &\quad = (2\pi)^{-N/2} \cdot \det(H_i) \cdot \exp(-(\boldsymbol{z} - \boldsymbol{\mu}_i)^T \cdot H_i \cdot (\boldsymbol{z} - \boldsymbol{\mu}_i)) \end{aligned} \tag{6}$$

on the input coordinates, $\boldsymbol{z}$, where N is the dimension of $\boldsymbol{z}$, $\boldsymbol{\mu}_i$ is a vector that defines the center of the ith class, H_i is the inverse of the ith class's covariance matrix, and $\det(H_i)$ is its determinant. There are then M hypotheses: the input coordinates belong to one or another of the M classes ($M = 2$ here, i.e., prominent or nonprominent).

One can then use Bayes' theorem to compute $P(\text{class } i|\boldsymbol{z})$ from the set of $P(\boldsymbol{z}|\text{class } i)$ and the relative frequency with which one observes the various classes. The classifier output is then the class that has the largest probability, given $\boldsymbol{z}$. If the classes are observed equally often, this boils down to picking the class with the largest $P(\boldsymbol{z}|\text{class } i)$. The classifier is defined by a choice of M triplets of (μ_i, H_i, ϕ_i), where ϕ_i is the prior probability of observing of each

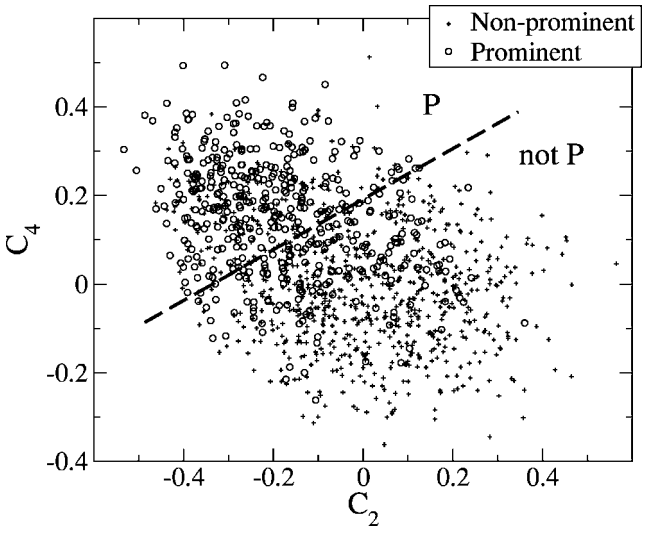


FIG. 3. Scatter plot showing two of the eight components of the feature vector for a loudness classifier for Cambridge data. Each point corresponds to a syllable. Prominent syllables are marked with circles, nonprominent by a plus sign. The dashed line is an approximation to the classifier boundary, derived from the machine classifications of the syllables.

class. The algorithm operates like a linear discriminant analysis in that it chooses ϕ_i, μ_i, and H_i so as to maximize the product of the probabilities that the feature vectors are classified correctly.

Figure 3 shows a sample set of feature vectors that are sent to the classifier. The figure shows loudness data for Cambridge (all styles). Only two of the eight components of the feature vector are shown, so the separation in this two-dimensional projection is not as good as is possible in the full eight-dimensional space. The dashed line is an approximate class boundary derived from the machine classifications of the data.[5]

One limitation of a standard discriminant classifier is that when the number of feature vectors becomes small, the border between classes becomes poorly defined; many algorithms fail entirely when the number of training points is smaller than the number of parameters necessary to define the classifier.

To avoid this, we computed an ensemble of good estimates by way of a Markov Chain Monte Carlo process, rather than limiting ourselves to a single "best-estimate" of the classifier parameters. The Markov Chain Monte Carlo process generates samples of μ, H, and ϕ from the distribution $P(\mu_i, H_i, \phi_i|z)$. This distribution is sharp as long as the number of feature vectors is much larger than the number of parameters that defines the classifier [which is $(M-1)\cdot(N+N\cdot(N+1)/2+1)$]. For small numbers of feature vectors, however, the probability distribution of the covariance matrix will become broad and heavy-tailed, and the prior distribution of $P(\mu_i, H_i, \phi_i)$ becomes important. We chose a prior that is constant, independent of μ, H, and ϕ.

In practice, we found it useful to select the best few from among the classifiers generated by the Markov Chain Monte Carlo algorithm. This makes the algorithm far less sensitive to the termination conditions of the Monte Carlo process and also makes the definition of classification probabilities more comparable to those reported for other classifiers. For this study, we test $Q=10N$ (about 50) candidate classifiers against the training set. Of the Q candidates, we keep the best $Q^{1/2}$, where "best" is defined by the fraction of the training set that is correctly classified.

TABLE I. P[chance] for classifiers based on the different acoustic features. These are the probabilities of correctly classifying the acoustic data, after shuffling so that there is no correlation between prominent/nonprominent labels and acoustic properties.

Loudness	$D(t)$	f_0	Irregularity	Spectral slope
59.1%	59.7%	61.1%	60.0%	61.4%

We split the data, randomly assigning 75% to the training set and 25% to the test set. We built classifiers for 8 different splits into training and test set, to allow us to estimate errors for the classification accuracy. Consequently, there were 8 selected ensembles, $80N$ candidate classifiers, and a total of $8\cdot(10N)^{1/2}\approx 55$ selected classifiers. This approach is a variant on a cross-validation procedure (Webb, 1999, p. 323).

Given this selected ensemble of classifiers, we computed the overall classification accuracy on a test set, averaging the accuracy across the selected classifiers. The accuracy we report is the averaged percent of correct classification, or 100% minus the sum of false-negative and false-positive errors.

One advantage of this Monte Carlo procedure is that it correctly reproduces the longer tails of Student's t-statistic in the one-dimensional case, whereas any quadratic classifier with a single best value of μ and H cannot.

III. RESULTS AND DISCUSSION

We express results in terms of the classifier effectiveness,

$$K = \frac{F[\text{correct}] - P[\text{chance}]}{1 - P[\text{chance}]}, \tag{7}$$

where F[correct] is the fraction of the test set that is correctly classified, and P[chance] is the accuracy of the classification in the absence of acoustic information. Consequently, $K=0$ implies the acoustic data were useless and yielded chance performance, while $K=1$ implies perfect classification.

P[chance] is determined by randomly shuffling the labels to break any association with the acoustic parameters (Table I). We classified such decorrelated data for five different window widths between 446 and 458 ms, with the average w chosen to match the $w=452$ ms that the bulk of the paper discusses (Sec. III A and thereafter). P[chance] depends on the acoustic measure, ranging from 59.0% for the loudness classifier to 61.4% for the spectral slope classifier. The differences are significant [$F(4,524)=7.1$; $P<0.01$], but not large. Presumably they are due to differences in the shape of the distributions. The performance on decorrelated data is slightly worse than the theoretical limit of 63.6% derived by predicting all syllables to be nonprominent.

Unless noted, all results will be for classifiers that are trained for a particular dialect and style of speech (e.g. "read

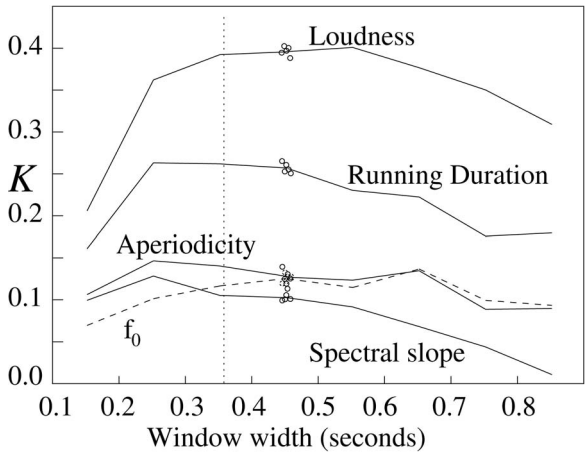


FIG. 4. Classifier performance versus the size of the analysis window, w. Each curve shows performance of classifiers based on a different acoustic feature (f_0 is shown dashed to separate it from its neighbors). The vertical axis is the K value, which shows how well each classifier performs relative to chance (shown as zero) and exact duplication of the human labels (shown as one). Plotted K values are averages over seven dialects and three styles of speech. The vertical dotted line marks where the window includes neighboring syllable centers. The small clusters of points near $w=0.45$ s show the reproducibility of the classifiers, derived from five classifier runs with slightly different window sizes.

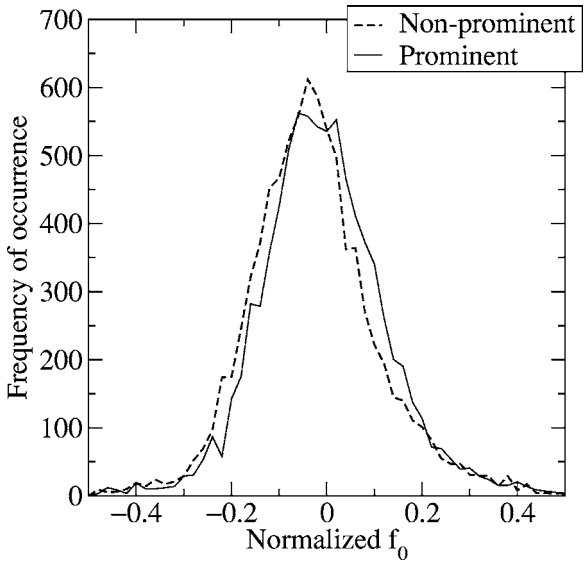


FIG. 5. Histograms of f_0 at the center of a $w=0.152$ s window for prominent (solid) and nonprominent (dashed) syllables. The distributions are nearly identical, showing that neither the central f_0 nor variance can effectively separate the two classes of syllables.

passages in Leeds"). This corresponds to communication within a dialect group. When we present a single value for K, it will refer to the average of all classifiers over the entire corpus. We chose this approach of building many dialect-specific classifiers because of the strong dialect-to-dialect variation that was seen in the IViE corpus in f_0 contours [Grabe *et al.* (to be published); Fletcher *et al.* (2004)] and well-known cross-dialect differences in the question-statement distinction [Cruttenden (1997)].

Figure 4 shows the performance as a function of window size for classifiers built from each of the five acoustic measures. The plotted performance is the average over 21 dialect/style combinations. Each classifier separates prominent from nonprominent syllables based on acoustic time series in a window centered on the syllable.

Three important results appear in Fig. 4. First, the classifiers based on loudness consistently outperform other classifiers by a substantial margin: they are about 50% better than classifiers based on running duration and more than twice as good as classifiers based on f_0 for most window sizes.

Second, the absolute performance of the f_0 classifiers is unimpressive. With a 452 ms window, the average f_0 classifier predicts only 66.3% of the syllable prominences correctly, which is little better than the 61.1% that can be achieved without the data (P[chance]). We found this surprising, as the prominence marks in the primary data set were made by labelers who expected that pitch motions often induced prominence. Further, they worked under labeling rules that encouraged the association of prominences with pitch events. This result contradicts the widespread view that a set of commonly employed f_0 patterns underlie the perception of prominence or accent.

The classifier can separate a wide variety of f_0 patterns. It can separate prominent from nonprominent patterns if they consistently differ over any 100 ms region within the analysis window, either in f_0 or slope of f_0. It can also discriminate if there are large differences in variance between prominent and nonprominent syllables. Additionally, because we have enough feature vectors to compute the full covariance matrix of all the orthogonal polynomial coefficients for each class, the classifier can also separate classes based on a combination of f_0, its slope and variance. These capabilities are sufficient to yield good classification of syllables based on loudness or duration; the poor results for f_0 then suggest that f_0 simply is not strongly correlated with prominence. Quantitative examples are in Sec. III D.

Figure 5 supports this observation that f_0 is not usefully correlated with prominence. It shows that histograms of f_0 at the center of prominent and nonprominent syllables overlap strongly. (Values of f_0 are computed at the window center from the orthogonal polynomial fits to the entire window; this provides interpolation into unvoiced regions.) While the mean f_0 for the entire set of prominent syllables is significantly (in the statistical sense) larger than for nonprominent syllables, that fact is nearly useless to a listener who is attempting to classify a single syllable as prominent or not. For any given f_0, there are roughly equal numbers of prominent and nonprominent syllables, so no measurement of central f_0 for a single syllable provides much evidence as to whether the syllable is prominent or not.

The third result shown in Fig. 4 is that the loudness and running duration classifiers improve *dramatically* until the window encompasses the neighboring syllables. This means that prominence depends not just on the loudness or duration of a syllable, but (as one might expect) on a contrast between a syllable and its neighbors. The decline in classifier performance beyond $w \approx 600$ ms is not understood in detail, but some of the decline is certainly caused by longer windows running off the ends of the utterances. Part of it may also be

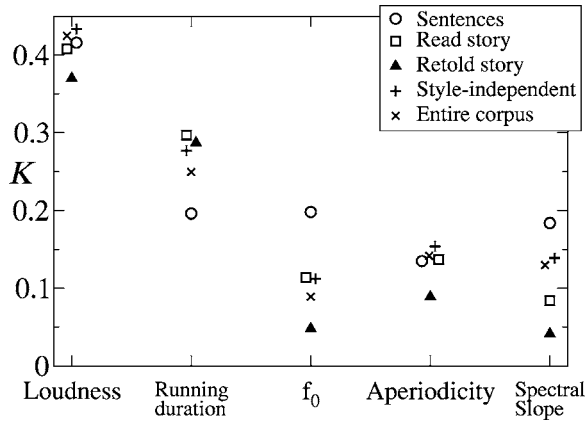


FIG. 6. Classifier performance versus acoustic measure and style of speech. The vertical axis shows performance on a scale where $K=0$ corresponds to chance and $K=1$ corresponds to perfect prediction of prominence marks. Classifiers based on loudness perform substantially better than the others for all three styles. The plus sign shows style-independent classifiers, and the cross marks classifiers that are both dialect- and style-independent.

due to the increasing complexity of the classifiers relative to the constant amount of data.

Given that the loudness classifiers continue improving up to $w \approx 500$ ms, and given that the f_0 classifiers are simpler for the same window size since $\tau_{f_0} > \tau_L$, the f_0 classifiers should make efficient use of the available information up to about 500 ms or beyond. In other words, since the loudness classifiers substantially improved by including neighboring syllables, the classifier complexity is probably not limiting the performance for f_0. Thus, the small change in f_0 classifier performance as the neighboring syllables are included in the analysis window suggests that the f_0 of neighboring syllables carries little information.

In the remainder of the paper, we focus on classifiers with $w=452$ ms, unless noted. We chose this size because it gives nearly peak performance for each acoustic measure.

A. Dependence of *K* values on acoustic measure, dialect, and style of speech

Figure 6 shows the classifier K values separated by acoustic measure and speech style, for classifiers trained on a single style/dialect combination. The results from classifiers built from the loudness measure are substantially and significantly ($P<0.001$) better than classifiers based on f_0, spectral slope, or aperiodicity. This conclusion holds true across all styles of speech. However, there are statistically significant differences between different styles of speech (e.g. "retold" versus "sentences"), so one cannot always rank one acoustic measure as better or worse than another. For instance, classifiers built on running duration outperform classifiers built on f_0 for the "read" and "retold" styles, but are effectively equal for the "sentence" style.[6] The statistical errors on these points were derived from the classifier's cross-validation estimates. They are not uniform, but average to $\sigma_K=0.02$. Most differences larger than 0.06 are significant at the 0.05 level.

Figure 6 also shows the results of classifiers that are trained on all the styles of speech together (e.g., a classifier is built for all of Belfast speech, rather than just Belfast "read" speech). The average performance of these style-independent classifiers is then plotted as a plus sign (+). These classifiers embody the assumption that prominence is marked the same way in all styles of speech. Finally, a style- and dialect-independent classifier (cross "×," trained on the entire corpus) was built for each acoustic measure. This classifier embodies the assumption that prominence is marked the same way in all dialects and all styles of speech that we studied. These more broadly defined classifiers perform about as well as the average of the style-specific classifiers; this suggests that all of the corpus indeed shares the same definition of prominence.

TABLE II. Performance of classifiers trained on increasingly broad portions of the corpus. The top line shows the performance of classifiers trained on a single dialect/style combination; the bottom line is for classifiers of the same complexity, trained on the entire corpus. In the rightmost column, K is averaged over all five acoustic measures.

Number of dialect/style combinations	Scope of the classifer	K
1	One style of speech in one dialect	0.201
3	All styles of speech in one dialect	0.208
7	One style of speech, covering all dialects	0.223
21	All styles of speech in all dialects	0.207

Table II shows the performance of classifiers that make different assumptions about the breadth of application of the definition of prominence. All the classifiers in the table have the same size feature vector, so they are equally capable of representing the classes. If each dialect had a unique definition of prominence, dialect-independent classifiers that attempt to represent seven dialects with one set of classes should give poor performance. Likewise, if prominence were encoded differently in the three styles of speech, the style-independent classifiers that use a common definition of prominence for all three styles should give a low K. Instead, different scopes yield nearly the same performance, differing by only 0.03 in K. The near-equality of K values in Table II implies that there is a useful common definition of prominence across all these dialects and styles of English.

Figure 7 shows the dependence of K values on acoustic

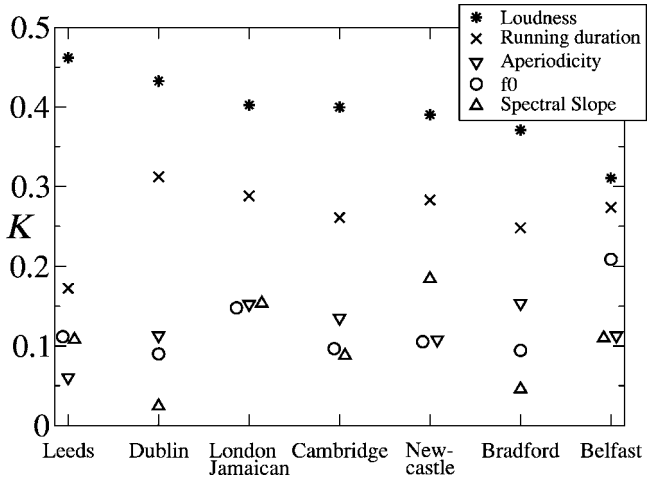


FIG. 7. Classifier performance for the five acoustic measures as a function of dialect. Each classifier is trained on a single dialect/style combination; symbols show the average over the three styles of speech.

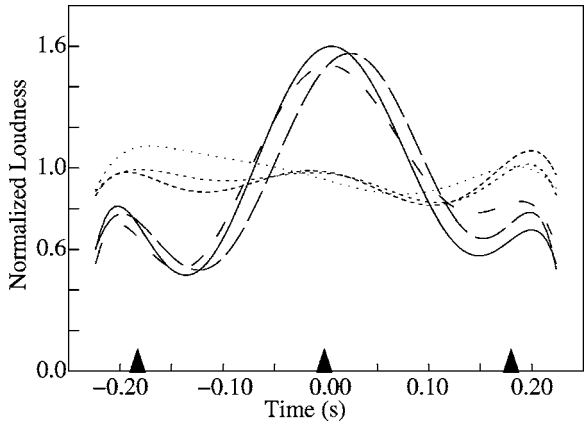


FIG. 8. Reconstructed loudness profiles for prominent (long dashes) and nonprominent (short dashes) syllables, for the primary and two secondary data sets. In each group, the primary data set is plotted with the most ink, followed by secondary sets GK then EL. The closed triangles mark the median position of syllable centers. Zero on the time axis corresponds to the prominence mark.

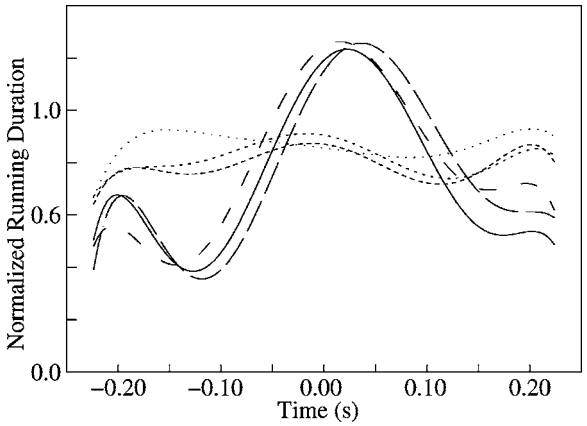


FIG. 9. Reconstructed running duration contours for prominent (long dash) and nonprominent (short dash) syllables. See Fig. 8.

measure and dialect. Again, classifiers based on loudness consistently outperform all others, with running duration in second place. Some dialect-to-dialect variations exist: most notably, $D(t)$ is relatively unimportant for Leeds, and f_0 is relatively important in Belfast. On average, the classifier's cross-validation error estimates for these points are $\sigma_K = 0.03$, so most differences larger than 0.09 are significant.

B. Reconstructing the acoustic properties

The dependence of K_L on window size in Fig. 4 implies that prominence depends on a loudness pattern. Reconstructing a loudness profile within the window reveals the details of this pattern. The reconstruction starts with the style- and dialect-independent classifier that represents the entire corpus. We then take all the syllables in a class (e.g., prominent syllables) that are correctly classified. The correctly classified points are represented by OP coefficients, which one can think of as points in a multidimensional space. (Each appears multiple times, once for each classifier in the forest that classified it correctly.) We then compute the centroid of this cloud of points to get the OP coefficients corresponding to a typical, correctly classified prominent syllable.

Next, these OP coefficients for each class of syllables are converted back into a loudness contour via Eq. (3). The resulting curves are averages but are quite representative of individual contours. As we include only contours where the human and machine classifications agree, these resulting contours emphasize the ones where loudness consistently induces a prominence judgment in the listener.

Figure 8 shows loudness reconstructions, as described earlier. Prominent syllables typically have a loudness peak near the labeled position ($t=0$), which follows an unusually quiet preceding syllable ($t \approx -180$ ms). The prominent syllable is nearly three times as loud as its predecessor. The following syllable is also quieter than average, but the difference is less dramatic. In contrast, nonprominent syllables typically lie in the midst of a fairly flat loudness profile, with the preceding syllable being slightly louder on average. The secondary data sets are discussed further in Sec. IV A.

Figure 9 shows a similar reconstruction of $D(t)$. Prominent syllables have a longer region of stable acoustic properties (presumably a longer vowel), following a relatively short preceding syllable. The prominent syllable is nearly three times as long as the preceding syllable and twice as long as the following syllable.

Figure 10 shows the equivalent $f_0(t)$ reconstruction. As expected, prominent syllables typically show a peak in fundamental frequency, but the peak is not large (about 20% in f_0, or about 30 Hz). This plot represents only those utterances which are correctly machine-classified on the basis of f_0. A similar plot based on all utterances would be diluted by the large number of utterances that cannot be correctly classified on the basis of f_0, and would show much less contrast between prominent and nonprominent syllables.

Similar plots would show that prominent syllables have a lower aperiodicity and a more positive spectral slope than their neighbors (i.e., they have more regular voicing and have more high frequency power in voiced regions). As with the other measures, the contrasts are strongest with the preceding neighbor.

Overall, a variety of differences appear between prominent and nonprominent syllables, perhaps extending beyond the vowel into consonantal regions. Furthermore, the acous-

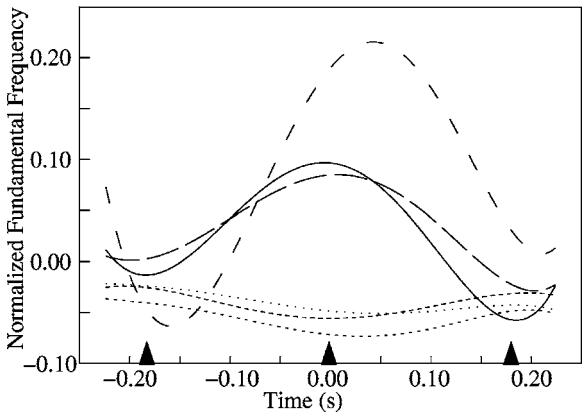


FIG. 10. Reconstructions of the time dependence of $f_0(t)$. See Fig. 8.

tic markers for prominence are not restricted to the prominent syllable; contrasts between a syllable and neighboring ones are important. These reconstructions are an acoustical representation of the alternating metrical pattern of English.

C. Qualitative limits of the analysis

(1) We search for patterns of f_0 and other acoustic measures defined in terms of absolute time offsets from the syllable center. If the patterns stretched as syllable durations changed, so that the the positions of peaks and valleys would move, the features will be blurred in this analysis. The classifier would then be unable to make full use of the information in such patterns. This is another possible explanation for the fall of classifier performance for $w > 600$ ms: duration changes accumulate across the window, so the position of the second- or third-nearest-neighbor syllable is correspondingly less certain than the nearest neighbor. However, this effect will strike classifiers with small τ_α first, so the logic (in Sec. III) that implies good efficiency of f_0 classifiers still holds. We are confident that f_0 and its contrasts with adjacent syllables carry relatively little information about prominence.

(2) The analysis does not take account of position in the utterance or intonational phrase. For instance, final lengthening doubtless dilutes the results based on running duration by introducing a population of long syllables that are only occasionally prominent. Likewise, initial syllables tend to be loud, but are not especially likely to be prominent. This will reduce the K of the loudness-based classifiers.

(3) We analyze f_0, not pitch. Although the correlation between f_0 and pitch is quite tight for pure tones, there has been less work on the psychophysics of speech-like sounds; perhaps the correlation is weaker. Or, perhaps the linguistic usage of the term "pitch" does not agree with the psychophysical definition of the term.

(4) We ignore the dependencies of the acoustic measures on segmental structure. For instance, /m/ and /s/ have intrinsically different values of aperiodicity. This acts as an extra source of noise in our classification, increasing the class variances relative to the difference between the means, thus reducing K.

(5) Loudness and duration are correlated in our corpus, so a decision of which of the two is more important may not be completely reliable.

D. Quantitative limits of the analysis

As the analysis does not detect strong correlations of f_0 with prominence, we should confirm that the weak result for f_0 is not an artifact of our analysis procedure.

We explored the limits of the analysis procedure by adding in an artificial f_0 component to the prominent syllables between normalization and OP fitting. We repeated the analysis, then adjusted the size of the artificial component until $K_{f_0} \approx 0.5$. This reveals how large the motion of f_0 would have to be for detection by the classifier. Since $K_{f_0} \ll 0.5$ with the unmodified data, this allows us to set an upper limit to the size of f_0 motions that might be associated with prominent syllables.

We first explored the possibility that a locally raised f_0 marked prominence. To check this, we added bumps in the shape of a $\sigma = 100$ ms Gaussian, centered on the prominence mark. The classifiers detected these bumps, reaching $K = 0.5$ when the bump size was 2.4 semitones (about 25 Hz for a speaker with mean f_0 of 170 Hz). Since K is much smaller than that for our unmodified data, we can exclude the possibility that prominence is commonly associated with such an f_0 bump or larger, because the analysis would have detected it. This is a conservative upper limit, as we base the limit on $K = 0.5$, whereas the unmodified f_0 data yielded only $K = 0.12$.

However, a standard assumption in the intonation literature is that many different pitch patterns can lend prominence to a syllable [e.g., Ladd (1996); Cruttenden (1997)]. A bump centered on a syllable is only one of many options. Background on this topic can be found in Wichmann *et al.* (1997).

We tested three more patterns to map out more limits of the analysis:

(1) A region of sloping f_0. A bump in the form

$$\frac{(t - t_c)}{\sigma} e^{-(t - t_c)^2/2\sigma^2}$$

was added, with $\sigma = 100$ ms. This function has a broad peak 100 ms after the prominence mark, a valley 100 ms before the mark, and a smooth slope in between. It was detected with $K = 0.5$ when the peak-to-valley difference was 2.8 semitones (about 27 Hz), and the slope was 14 semitones/s.

(2) A region of increased variance of f_0. We used a random mixture of the Gaussian bumps and the sloping contours, above. Instead of using a single amplitude, the amplitudes were chosen from a zero-mean Gaussian distribution. This corresponds to the possibility that prominence is marked by *either* a bump, a dip, a peak-valley pattern, a valley-peak pattern, or some mixture thereof. Nonprominence would presumably be indicated by relatively flat contours. This choice can generate a very broad range of intonational patterns, covering many of the suggested possibilities. Even with this wide variety of possible f_0 patterns, the classifier reached $K = 0.5$ when the standard deviation of the bump amplitude was 3.1 semitones, along with a 3.8 semitone standard deviation for the peak-to-valley difference for the slope component.

(3) We added a Gaussian bump with $\sigma = 100$ ms, but we let the amplitude and position vary from prominence to prominence. The bump center was chosen from a $\sigma = 100$ ms Gaussian probability distribution, to simulate random choices of peak alignment, and the amplitude was chosen from a zero-mean Gaussian. This corresponds to the possibility that prominence is marked by either an f_0 bump or dip, whose timing is not precisely tied to the syllable center. The analysis was not as effec-

tive at this test, detecting it at $K=0.5$ only when the standard deviation of the normalized amplitude was 0.5, corresponding to 10 st (about 85 Hz).

The limits that this analysis can set on f_0 excursions depend on the complexity of the pattern and the accuracy with which it is anchored to the prominence mark. However, most 1/2 octave motions would be easily detectable, if they existed in the data. The analysis can exclude most f_0 features that have a fixed time-alignment with the syllable center and are larger than 3 semitones.

While we do not categorically rule out f_0 as an indicator of prominence, we do rule out many simple associations of f_0 with prominence. Most of the possibilities that we do not exclude would involve fairly complex patterns and/or rather loose associations between the position of the pattern and the syllable center. It is currently uncertain whether such a tenuous association of f_0 with a syllable is sufficient to communicate the prominence to a human listener.

E. Comparison to synthesis experiments

When comparing these results to other studies in the literature, it is important to maintain the distinction between acoustic properties that can induce the perception of prominence and acoustic measures that are actually used to mark prominent syllables. They need not be identical. Speech is only one of several inputs that the human auditory mechanism processes, and other uses, such as monitoring environmental sounds, might define the way the auditory system functions. Additionally, articulatory constraints may make certain ways of inducing prominence easier than others.

This distinction is crucial for understanding the synthesis-based experiments that show that f_0 can induce the perception of prominence (Gussenhoven *et al.*, 1997; Rietveld and Gussenhoven, 1985; Terken, 1991). Despite appearances, their results are consistent with this study, because they use larger f_0 excursions than are normally found in our corpus. For instance, a typical stimulus in these papers above contains a sharp triangular f_0 excursion of 1/2 octave amplitude and a full-width at half-maximum of 200 ms or less.

We looked for such peaks in our database by computing a peak-height statistic h matched to the shapes used in the above synthesis-based studies. We take f_c as an average of f_0 over a 50-ms-wide region centered on a syllable. The average is weighted with $W_{f_0}(t)$ from Sec. II F. Similarly, f_e is an average of f_0 over a pair of regions between 100 and 150 ms to the left and to the right of the prominence mark. We then compute $h=\log_2(f_c/f_e)$; this statistic is close to zero for linear f_0 contours and nearly equal to the bump height (in octaves) for contours used in the papers cited above.

For prominent syllables in the IViE corpus, h has an approximately Gaussian distribution (Fig. 11) with a standard deviation of $h=0.11$ and a mean of zero. Only 2% had bump heights exceeding a quarter octave. Most of the stimuli studied in these papers have bumps that are larger than that. Consequently, they studied bumps that are larger than those commonly found in British English. Their results are thus completely consistent with our conclusion that f_0 is relatively unimportant for prominence, as English is normally spoken. Their experiments, like ours, indicate that a 10% pitch change induces little prominence

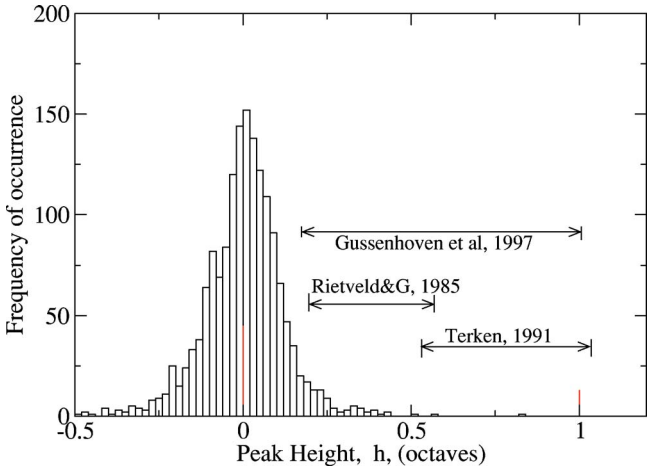


FIG. 11. Peak height statistic, h, for prominent syllables in the IViE corpus (histogram). For comparison, the ranges of the f_0 swings used as experimental stimuli in Gussenhoven *et al.* (1997), Rietveld and Gussenhoven (1985), and Terken (1991) are shown.

F. Importance of the spectral slope

Our result that K_S is small is somewhat unexpected, given work by Heldner (2001) and Sluijter and van Heuven (1996). Heldner found his spectral emphasis measure to be a good predictor of prominence. However, Heldner's measure is different, and is applied to a different language (Swedish). His measure is the difference between the power in the first harmonic and the rest of the spectrum in voiced regions, and is zero in unvoiced regions. So, his measure obtains almost all its information from the low-frequency parts of the spectrum, mostly below $3f_0 \approx 600$ Hz, unlike ours, which extends up to 3000 Hz. A further difference is that his measure responds differently to voiced/unvoiced distinctions than ours.

Sluijter and van Heuven, consistent with this work, found that syllables with contrastive focus have a flatter spectrum. Their experiment yields a strong effect of spectral slope, but that is expected, as their classification task is far easier. Their sentences were read carefully by speakers instructed to produce contrastive focus on certain words. The authors then selected sentences for a clear contrast between the +FOCUS and −FOCUS versions. Thus, they allowed no ambiguous utterances, Their paired comparison between ±FOCUS renditions of the same word in the same position in an utterance also allows for a more sensitive comparison than is normally available to a human listener to natural speech. They proved that speakers *can* produce contrasts in spectral slope, not that speakers normally *do* produce such contrasts.

IV. FURTHER EXPLORATIONS

A. Comparison with secondary data sets

It might be argued that the similarity of our results between dialects is due to the fact that the same pair of labelers marked each dialect rather than because of an intrinsic simi-

TABLE III. Alignment and agreement of syllable and prominence marks between the various data sets. The "alignment" column counts marks that match within 60 ms. Of the aligned marks, the right-hand column counts what fraction agree in terms of prominence/nonprominence judgments.

Comparison	Agreement on alignment	Agreement of syllables that align
Primary versus GK	84%	73%
Primary versus EL	79%	72%
GK versus EL	75%	84%

larity. To check this, we conducted the same analysis on the two secondary data sets. Our secondary labelers speak different dialects and are trained differently from the primary labelers. If the process of labeling says more about the labeler than about the speech, the secondary data sets should give substantially different results from the primary data set.

Table III compares the primary and two secondary sets. Inspection of a sample of the marks reveals some disagreements about prominence, some disagreements about the number of syllables (primarily nonprominent syllables), a few unlabeled words in the secondary sets, and a few long syllables where the labelers agree but placed marks more than 60 ms apart. It is hard to compare these alignment and agreement numbers with the literature [e.g., Yoon *et al.* (to be published), and references therein], because published studies of intertranscriber reliability typically have trained the transcribers to a specific standard in an attempt to minimize the disagreement. In contrast, we wished to find the natural limits of the idea of "prominence," so we did not train labelers.

As can be seen in Figs. 8–10, the reconstructions of the primary and secondary sets are quite similar. The most obvious discrepancy is that the EL set is shifted about 20 ms later, relative to the marks. This is unimportant, as a review of the marks indicates that EL placed labels slightly earlier in the syllable than the other labelers. Reconstructions for the irregularity and spectral slope measures (not shown) are also similar to reconstructions based on the primary data set. The classifier performance on the primary and secondary sets also match well (Fig. 12). These figures suggest that the dialect and academic background of the labeler makes little difference. This supports our main conclusion and suggests that a perceptual, theory-independent definition of prominence may be possible.

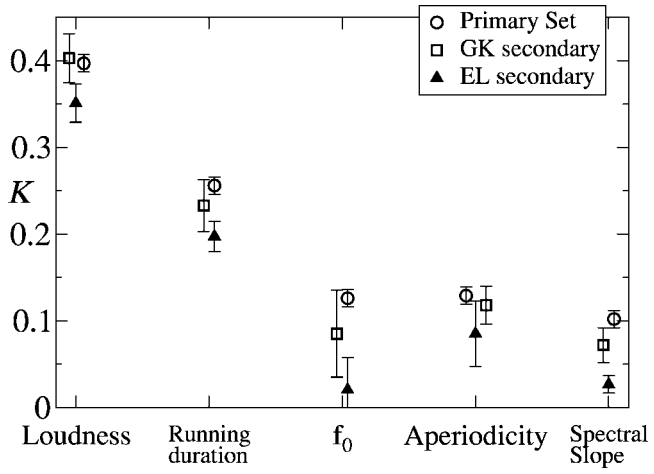


FIG. 12. Classifier performance comparisons between the primary and secondary data sets.

The secondary sets of labels also gave an opportunity to check that the algorithm we used to assign the location of nonprominent syllables in the primary was adequate. The agreement of the secondary and primary sets confirms that the automatic generation of nonprominent syllable positions is good enough.

B. Loudness versus rms versus peak-to-peak

We use a loudness measure rather than the more common rms amplitude or peak amplitude measurements because the former is a better match to the listener's perception. However, to allow comparison with prior work, we also built classifiers based on common amplitude measures. We computed rms amplitude by filtering with a fourth-order, 60 Hz, time-symmetric Butterworth high-pass filter, squaring, and smoothing with a 15 ms standard deviation Gaussian kernel. Peak-to-peak amplitude was computed by high-pass filtering, then finding the positive peak amplitude by taking the maximum over a 20 ms window centered at each point, and then subtracting a similarly defined negative peak amplitude.

Perhaps surprisingly, there is no substantial difference between the loudness, rms, and peak-to-peak classifier performance: K_{rms} is 0.01 lower than K_L, and $K_{\text{peak-to-peak}}$ is 0.03 higher. The similarity between our loudness and rms intensity results means that our results appear to conflict with the findings of Sluijter and van Heuven (1996) and Sluijter *et al.* (1997), who found that intensity is relatively unimportant. The difference may relate to their experimental conditions, which were rather more formal, to the different language, or some other factor.

C. Combining different acoustic properties

Finally, we built one more classifier to see if information from the other acoustic properties could improve the behavior of a classifier based on loudness. To do this, we took the forest of classifiers and constructed a feature vector for each syllable by counting the fraction of classifiers that labeled it as prominent, for each of the five acoustic measures. The feature vector for each syllable is thus $(F_{f_0}, F_L, F_D, F_A, F_S)$, with each of the F_α in the range $[0, 1]$. It is input for a second-stage classifier, operating on the outputs of the first stage Gaussian Forest classifiers. This is a "bagging" or classifier fusion approach (Breiman, 1996; Kittler *et al.*, 1998; Wolpert, 1992; Huang and Suen, 1995). The second stage classifier is a logistic discriminant classifier (Webb, 1999, pp. 124–132).

The resulting distribution of K across dialects and style is fairly narrow distribution, with $\sigma=0.07$, $\sigma/K=0.14$. All dialects seem to be about equally good at marking prominence acoustically. The average K is 0.479 (based on P[chance] for loudness), and P[correct]=0.786.

To see how much information the other acoustic features are contributing, we can compare the $K=0.479$ for the com-

bined classifier to to a similar logistic discriminant classifier that is fed only F_L. Such a loudness-only classifier achieves $K=0.430$. The improvement caused by attaching F_D, F_{f_0}, F_A, and F_S to the feature vector is statistically significant ($t=4.2$, $df=21$, $P<0.001$), but it is not large. The probability of correct classification only increases from 76.6% to 78.6%. This is not completely unexpected: $D(t)$ is correlated with $L(t)$ and the other acoustic measures are generally less effective at classifying syllables than loudness or running duration. We conclude that $D(t)$, $f_0(t)$, $S(t)$, and $A(t)$ do contain some information not present in the loudness, but not very much. This result is not inconsistent with claims that loudness and duration are perceived as a unit (Turk and Sawusch, 1996).

V. CONCLUSION

Prominent syllables are marked by being louder and longer than the previous syllable. Of the two, loudness is the better predictor. However, these two acoustic measures are correlated enough so that distinguishing the effect of the two may not be completely reliable.

Contrary to the common assumption, there is no pattern of f_0 detectable by our analysis that is more than a weak predictor of prominence. Many prominent syllables do indeed have high pitch, but many nonprominent syllables also do. Thus, taking the listeners' point of view, the observation of high pitch does not usually allow the listener to conclude that a syllable is prominent. We found that prominence cannot be usefully distinguished on the basis of local f_0 values, local f_0 changes, or the local variance of f_0. We see no evidence that long f_0 patterns are relevant to the prominence decision.

We do not disagree with the common assumption that dramatic changes in f_0 *can* cause listeners to label syllables as prominent; however, we find that our speakers *do not* normally use this mechanism. They almost never produce the large pitch excursions that are presumably necessary to induce a listener to judge a syllable as prominent. The fact that the labelers were able to consistently mark prominent syllables is clear proof that special f_0 patterns are not necessary near prominent syllables.

All the dialects and styles of speech in our corpus have a similar definition of prominence. We suggest that this definition of prominence could be a feature of most English dialects, as seems consistent with the work of Silipo and Greenberg (2000) and of Beckman (1986). The definition of prominence also seems independent of the labeler's dialect and academic training.

These results have several implications for linguistics. First, prominence and pitch movements should be treated as largely independent and equally important variables. Prominence has a clear acoustic basis, although metrical expectations may also play some role.

Second, these results raise a puzzle. Individual utterances where prominence seems to be due to large pitch excursions are not hard to find in the literature. Are they simply unusual contours that were selected for their tutorial value, or do they represent another style of speech that is not represented in the IViE corpus? Do people produce large f_0 excursions in certain experiments and not in others?

Third, too much attention may have been focused on f_0. Various authors have assigned f_0 the tasks of communicating emotion, contrastive focus, marking the introduction of new topics and new words, separating declaratives from interrogatives, and helping to separate pairs of words. Perhaps it has been assigned too many tasks. At the least, it seems that f_0 does not normally play a role in signaling the prominent words in a sentence.

ACKNOWLEDGMENTS

The authors would like to thank Eleanor Lawson for comments and many hours of labeling, and Chilin Shih for valuable comments. This research was supported by award No. RES000-23-0149 from the UK Economic and Social Research Council; the IViE corpus was funded by UK ESRC Award No. RES000-23-7145.

APPENDIX A: SENSITIVITY ANALYSIS—FORM OF WEIGHT FUNCTION

If W were changed, one might expect K to change, since different weight functions emphasize different parts of the syllables. We examined this possibility by picking three new sets of weight functions and re-analyzing the data: [A] All weights (see Sec. II F) raised to the 0.5 power. This means that the analysis is not focused as strictly on syllable centers: syllable edges contribute more. It also puts more nearly even weights on prominent and nonprominent syllables. [B] All weights raised to the 1.5 power, thus focusing the analysis more tightly toward syllable centers. [C] Changing $W_S(t)$ to $W_S(t)=L^2(t)$, thus including unvoiced regions in the OP fits to the spectral slope data.

None of these results differed much from the default case for any of the acoustic measures: K values changed by no more than 0.03. We conclude that our weight function is adequate and that changes to them would probably not substantially affect our results.

APPENDIX B: SENSITIVITY ANALYSIS—ORDER OF POLYNOMIAL FIT

To check that we used the appropriate number of orthogonal polynomials, we ran the same analysis with 20% more or fewer orthogonal polynomials by altering the τ_α. Most changes to the K values were small, within 0.02 for all acoustic measures, except f_0.

However, the classifiers built from f_0 data showed a trend toward better performance as they were simplified: K_{f_0} increased by 0.055 as τ_{f_0} was increased from 75 to 112 ms. To see whether this increase would continue as the f_0 classifiers became even simpler, we recalculated with $\tau_{f_0}=141$ ms and saw no further increase in K. It seems that the optimal classifier for f_0 therefore involves 4 ($\tau_{f_0}=141$ ms) or 5 ($\tau_{f_0}=112$ ms) orthogonal polynomials in the analysis window.

These tests show that the results do not depend strongly on the number of polynomials used. Fitting the data more

accurately would not substantially improve the classification results for any acoustic measure. Indeed, this check suggests that only the simplest f_0 patterns (those describable by low-order polynomials) carry prominence information.

APPENDIX C: LOUDNESS NORMALIZATION

The loudness normalization (Sec.II E) is arguably too severe: by setting the rms loudness in the window to a constant, it means that the classifier cannot recognize that the analysis window as a whole might be unusually loud. To check whether this is an important limitation, we also computed K values where we normalized the loudness by dividing by the speaker's overall rms loudness. This normalization removes interspeaker differences but preserves all other loudness differences. The average K was little different; it was reduced by 0.03 ± 0.01. The extra information available in this analysis was probably overwhelmed by the increase in loudness variability associated with inter-utterance differences.

[1]Experiments show that the running duration of unusually quiet reasons (e.g., loudness less than 20% of the peak) is sensitive to the choice of L.

[2]To compute an X%-trimmed weighted average, sort the data into order, then trim off data from the bottom until you have removed X% of the total weight; repeat from the top down; then do a weighted average on the remainder. Trimmed weighted averages are insensitive to a few unusually high or low points.

[3]We trim much more weight from the loudness than the other acoustic measures because our goal is to make the normalization insensitive to the amount of silence within the window. All the acoustic measures except loudness have weights (Sec. II F) that go to zero in silent regions. Consequently, one can trim off silent regions by trimming off just a small amount of weight. For loudness, on the other hand, $W_A(t)=1$, both in and out of silent regions. To make the normalization insensitive to modest amounts (e.g., 35%) of silence within the window, one then needs to trim off a large amount (e.g., 35%) of the weight.

[4]Pavlovic (1984) reports that typical speech levels of A-weighted sound pressure level (SPL) in normal speech are about 63 dB A at a normal 1 m conversational distance, rising by 0.46 dB for every decibel of noise over 50 dB. We combine this with an estimate that 50% of the US urban population lives in dwellings with SPL$>$60 dB A (US-EPA). By Pavlovic's model, these people normally communicate at mean signal-to-noise ratios (SNR) of 9 dB or less, implying a peak-to-noise SNR of no more than 15 dB.

[5]The class boundary is actually a segment of a different hyper-ellipse for each classifier in the forest, so there is no unique line that separates syllables classified as prominent from those that are classified nonprominent. In this projection, there is a region whose width is about 0.2 where different classifiers in the forest may yield different answers, or where different values of the six unplotted coefficients may change the classification. The line is drawn by hand down the center of that region.

[6]It is easy to speculate that classifiers built on f_0 perform well for sentences because the sentences are nearly completely voiced, therefore (on average) sentences simply have more f_0 data available per syllable. Likewise, since the sentences are almost completely sonorant, the running duration measure has relatively slow and weak spectral changes to work with, so its results may be less reliable. However, the tasks differ in many important ways, so we have no firm conclusions.

Beckman, M. E. (**1986**). *Stress and Non-Stress Accent*, Netherlands Phonetic Archive Vol. 7 (Dordrecht, Foris).

Beckman, M. E. and Edwards, J. (**1994**). "Articulatory evidence for differentiating stress categories," in *Phonological Structure and Phonetic Form*, edited by P. Keating (Cambridge University Press, Cambridge), Papers in Laboratory Phonology III, pp. 7–33.

Boersma, P. (**1993**). "Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound," Institute of Phonetic Sciences, University of Amsterdam, Proceedings Vol. 17, pp. 97–110, URL http://www.fon.hum.uva.nl/paul/papers/Proceedings_1993.pdf.

Bolinger, D. (**1958**). "A theory of the pitch accent in English," *Word* **7**, 199–210, reprinted in *Forms of English: Accent, Morpheme, Order*, (Harvard University Press, Cambridge, MA, **1965**).

Breiman, L. (**1996**). "Bagging predictors," Mach. Learn. **26**, 123–140.

Clark, J. and Yallop, C. (**1995**). *An Introduction to Phonetics and Phonology*, 2nd ed. (Blackwell, Oxford).

Cooper, W. E., Eady, S. J., and Mueller, P. R. (**1985**). "Acoustical aspects of contrastive stress in question/answer contexts," J. Acoust. Soc. Am. **77**, 2142–2156.

Cruttenden, A. (**1997**). *Intonation*, 2nd ed. (Cambridge University Press, Cambridge).

Eady, S. J. and Cooper, W. E. (**1986**). "Speech intonation and focus location in matched statements and questions," J. Acoust. Soc. Am. **80**, 402–415.

Fletcher, H. and Munson, W. A. (**1933**). "Loudness, its definition, measurement, and calculation," J. Acoust. Soc. Am. **5**, 82–108.

Fletcher, J., Grabe, E., and Warren, P. (**2004**). "Intonational variation in four dialects of English: The high rising tune," in *Prosodic Typology: The Phonology of Intonation and Phrasing*, edited by S.-A. Jun (Oxford University Press, Oxford).

Fry, D. B. (**1955**), "Duration and intensity as physical correlates of linguistic stress," J. Acoust. Soc. Am. **27**, 765–768.

Fry, D. B. (**1958**). "Experiments in the perception of stress," Lang Speech **1**, 126–152.

Gelman, A. B., Carlin, J. S., Stern, H. S., and Rubin, D. B. (**1995**). *Bayesean Data Analysis* 1st ed. (Chapman and Hall/CRC, London).

Grabe, E., Kochanski, G., and Coleman, J. "Quantitative modelling of intonational variation," in Proceedings of SASRTLM 2003 (Speech Analysis and Recognition in Technology, Linguistics and Medicine), URL http://kochanski.org/gpk/papers/2004/2003SASRLTM (to be published).

Grabe, E., Post, B., and Nolan, F. (**2001**). "Modelling intonational variation in English. The IViE system," in Proceedings of Prosody 2000, edited by S. Puppel and G. Demenko, pp. 51–57.

Gussenhoven, C., Repp, B. H., Rietveld, A., Rump, H. H., and Terken, J. (**1997**). "The perceptual prominence of fundamental frequency peaks," J. Acoust. Soc. Am. **102**, 3009–3022.

Heldner, M. (**2001**). "Spectral emphasis as an additional source of information in accent detection," in *Prosody in Speech Recognition and Understanding*, Paper No. 10, 22–24 October, Molly Pitcher Inn, Red Bank, NJ.

Ho, T. K. (**1998**). "The random subspace method for constructing decision forests," IEEE Trans. Pattern Anal. Mach. Intell. **20**, 832–844, URL http://csdl.computer.org/comp/trans/tp/1998/08/i0832abs.htm.

Hochstrasser, U. W. (**1972**). "Orthogonal polynomials," in *Handbook of Mathematical Funtions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York), pp. 771–802.

Huang, Y. S. and Suen, C. Y. (**1995**). "A method of combining multiple experts for the recognition of unconstrained handwritten numerals," IEEE Trans. Pattern Anal. Mach. Intell. **17**, 90–94.

Kittler, J., Hatef, M., Duin, R. P. W., and Matas, J. (**1998**). "On combining classifiers," IEEE Trans. Pattern Anal. Mach. Intell. **20**, 226–239.

Kochanski, G., Shih, C., and Jing, H. (**2003**). "Hierarchical structure and word strength prediction of Mandarin prosody," Int. J. Speech Technology **6**, 33–43, URL http://dx.doi.org/10.1023/A:1021095805490.

Kochanski, G. P. and Shih, C. (**2000**). "Stem-ML: Language independent prosody description," in Proceedings of the Sixth International Conference on Spoken Language Processing, Vol. 3, pp. 239–242, URL http://prosodies.org/papers/2000/stemml_2000.pdf, Beijing, China.

Ladd, D. R. (**1996**). *Intonational Phonology* (Cambridge University Press, Cambridge).

Lieberman, P. (**1960**). "Some acoustic correlates of word stress in American English," J. Acoust. Soc. Am. **32**, 451–454.

O'Connor, J. D. and Arnold, G. F. (**1973**). *Intonation of Colloquial English*, 2nd ed. (Longman Group, London).

Passy, P. (**1891**). *Etude sur les Changements Phonétiques et Leurs Caractères Généraux* (Firmin-Didot, Paris).

Passy, P. (**1906**). *Petite Phonétique Comparée des Principales Langues Européenes* (Teubner, Leipzig).

Pavlovic, C. V. (**1984**). "Speech spectrum considerations and speech intelligibility predictions in hearing aid evaluations," J. Speech Hear Disord. **54**, 3–8.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (**1992**).

Numerical Recipes in C: The Art of Scientific Computing, 2nd ed. (Cambridge University Press, New York).
Rietveld, A. C. M. and Gussenhoven, C. (**1985**). "On the relation between pitch excursions and prominence," J. Phonetics **13**, 299–308.
Roca, I. and Johnson, W. (**1999**). *A Course in Phonology* (Blackwell, Oxford).
Silipo, R. and Greenberg, S. (**1999**). "Automatic transcription of prosodic stress for spontaneous English discourse," in Proceedings of the XIVth International Congress of Phonetic Sciences (ICPhS99), pp. 2351–2354.
Silipo, R. and Greenberg, S. (**2000**). "Prosodic stress revisited: Reassessing the role of fundamental frequency," in Proceedings of the NIST Speech Transcription Workshop.
Sluijter, A. M. C. and van Heuven, V. J. (**1996**), "Spectral balance as an acoustic correlate of linguistic stress," J. Acoust. Soc. Am. **100**, 2471–2485.
Sluijter, A. M. C., van Heuven, V. J., and Pacilly, J. J. A. (**1997**). "Spectral balance as a cue in the perception of linguistic stress," J. Acoust. Soc. Am. **101**, 503–513.
Stevens, S. S. (**1971**). "Perceived level of noise by Mark VII and decibels," J. Acoust. Soc. Am. **51**, 575–602.
Sweet, H. (**1906**). *A Primer of Phonetics* (Clarendon, Oxford).
't Hart, J., Collier, R., and Cohen, A. (**1990**). *A Perceptual Study of Intonation: An Experimental-phonetic Approach to Speech Melody* (Cambridge University Press, Cambridge).
Tamburini, F. (**2003**). "Prosodic prominence detection in speech," in Seventh International Symposium on Signal Processing and its Applications, pp. 385–388.
Terken, J. (**1991**). "Fundamental frequency and percieved prominence of accented syllables," J. Acoust. Soc. Am. **89**, 1768–1776.
Terken, J. and Hermes, D. J. (**2000**). "The perception of prosodic prominence," in *Prosody: Theory and Experiment, Studies Presented to Gösta Bruce* (Kluwer Academic, Dordrecht), pp. 89–127.
Trager, G. L. and Smith, H. L. (**1951**). *An outline of English structure*, Studies in Linguistics: Occasional Papers No. 3 (American Council of Learned Societies, Washington DC).
Turk, A. E. and Sawusch, J. R. (**1996**). "The processing of duration and intensity cues to prominence," J. Acoust. Soc. Am. **99**, 3782–3790, URL doi:.
US-EPA (1974). "Information on levels of environmental noise requisite to protect public health and welfare with an adequate margin of safety," Report 550/9-74-004, U. S. Environmental Protection Agency Office of Noise Abatement and Control, USGPO, Washington, DC 20402, URL http://www.nonoise.org/library/levels74/levels74.htm.
Webb, A. (**1999**). *Statistical Pattern Recognition* (Arnold, London).
Welby, P. (**2003**). "Effects of pitch accent position, type, and status on focus projection," Lang Speech **46**, 53–81.
Wichmann, A., House, J., and Rietveld, T. (**1997**). "Peak displacement and topic structure," in Intonation: Theory, Models and Applications—Proceedings of ESCA Workshop on Intonation.
Wolpert, D. H. (**1992**). "Stacked generalization," Neural Networks **5**, 241–260.
Yoon, T., Chavarria, S., Cole, J., and Hasegawa-Johnson, M. (**2004**). "Intertranscriber reliability of prosodic labeling on telephone conversation using ToBI," in Proceedings of the ISCA International Conference on Spoken Language Processing, Jeju, Korea, pp. 2729–2732, URL http://prosody.beckman.uiuc.edu/pubs/Yoon-etal-ICSLP2004.pdf

Speaker recognition with temporal cues in acoustic and electric hearing[a)]

Michael Vongphoe[b)] and Fan-Gang Zeng[c)]
Hearing and Speech Research Laboratory, Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, and Otolaryngology—Head and Neck Surgery, University of California, Irvine, California 92697-1275

(Received 6 January 2004; revised 6 May 2005; accepted 6 May 2005)

Natural spoken language processing includes not only speech recognition but also identification of the speaker's gender, age, emotional, and social status. Our purpose in this study is to evaluate whether temporal cues are sufficient to support both speech and speaker recognition. Ten cochlear-implant and six normal-hearing subjects were presented with vowel tokens spoken by three men, three women, two boys, and two girls. In one condition, the subject was asked to recognize the vowel. In the other condition, the subject was asked to identify the speaker. Extensive training was provided for the speaker recognition task. Normal-hearing subjects achieved nearly perfect performance in both tasks. Cochlear-implant subjects achieved good performance in vowel recognition but poor performance in speaker recognition. The level of the cochlear implant performance was functionally equivalent to normal performance with eight spectral bands for vowel recognition but only to one band for speaker recognition. These results show a disassociation between speech and speaker recognition with primarily temporal cues, highlighting the limitation of current speech processing strategies in cochlear implants. Several methods, including explicit encoding of fundamental frequency and frequency modulation, are proposed to improve speaker recognition for current cochlear implant users. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1944507]

PACS number(s): 43.71.Bp, 43.71.Es, 43.72.Fx, 43.66.Fe [KWG] Pages: 1055–1061

I. INTRODUCTION

Natural speech utterances carry information not only about what is being said but also about who says it (e.g., gender, age, ethnicity, and emotional state). Acoustic cues encoding "what" and "who" are widely distributed in both gross and fine spectral and temporal domains and can be influenced by physiological, behavioral, and cultural factors (Ladefoged and Broadbent, 1958; Stevens *et al.*, 1968; Atal, 1972; Johnson *et al.*, 1984; Childers and Wu, 1991; Wu and Childers, 1991; Stevens, 2002). For example, spectral peaks or formant frequencies that are critical for speech recognition also carry information regarding a speaker's identity as they reflect the individual speaker's anatomical and physical properties such as vocal tract size, shape and position (Fellowes *et al.*, 1997; Remez *et al.*, 1997). Conversely, temporal waveform periodicity or fundamental frequency (F0) that is typically correlated with a speaker's gender can influence speech recognition (Whalen *et al.*, 1993; Holt *et al.*, 2001) or directly carry lexical information in tonal languages (Liang, 1963).

While traditional research has focused on spectral cues, the temporal waveform envelope has been extensively studied in speech recognition (Van Tasell *et al.*, 1987; Rosen, 1992; Shannon *et al.*, 1995). It has been found that, in both real and simulated cochlear implant implementation, high levels of speech intelligibility can be achieved by encoding relatively slowly varying temporal envelopes that are extracted from one to several numbers of frequency bands (Wilson *et al.*, 1991; Dorman and Loizou, 1998; Zeng *et al.*, 2002). Recently, the utility of the temporal envelope cue has been extended to Mandarin tone recognition (Fu *et al.*, 1998; Xu and Pfingst, 2003; Zeng *et al.*, 2005) as well as other aspects of spoken language processing such as speaker identification (Cleary and Pisoni, 2002; Kong *et al.*, 2003; McDonald *et al.*, 2003; Fu *et al.*, 2004; Gonzalez and Oliver, 2005).

Cleary and Pisoni (2002) tested the effect of linguistic content (fixed sentence versus varied sentence) on talker discrimination between two females in 44 school-aged deaf children who had used the cochlear implant for at least 4 years. They found that these children achieved significantly higher than chance (50%) performance (mean percent correct score =68%) when the sentence was fixed, but produced essentially chance level performance at 57% correct when the sentence was varied. Their results suggest that the cochlear-implant users could not reliably recognize an unfamiliar talker's voice when the linguistic content varied. McDonald *et al.* (2003) replicated this finding using word stimuli in 21 adult cochlear-implant users and 24 normal-hearing listeners who listened to processed stimuli simulating the Nucleus SPEAK strategy (6 of 20 channel peaking). They found a similar linguistic effect on talker discrimination by both groups of subjects.

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

[b)]Current address: USC School of Dentistry, Los Angles, California 90089.
[c)]Corresponding author: 364 Med Surge II, University of California, Irvine, California 92697-1275. Telephone: (949) 824-1539; fax: (949) 824-5907; electronic mail: fzeng@uci.edu

Fu *et al.* (2004) used vowel materials to test gender discrimination in 11 adult cochlear-implant users and found a great deal of variability in performance ranging from 70% to 95% correct. They also performed the same task in a group of normal-hearing listeners and varied systematically the number of spectral bands and the temporal envelope cutoff frequencies. The result showed that the implant users produced performance equivalent to normal performance with four to eight spectral bands. Most interestingly, they found a contrast between speaker and vowel recognition in the four-band condition: Speaker identification was significantly improved as a function of the temporal envelope frequency from 20 to 320 Hz but vowel recognition did not under the same condition.

Gonzalez and Oliver (2005) also examined both gender and speaker identification as a function of the number of spectral bands in normal-hearing listeners. They used Spanish sentence materials and processed them using either sinusoidal and noise carriers. Similar to the findings of Fu *et al.*, Gonzalez and Oliver found that gender and speaker identification is systematically improved with the number of bands. In addition, they found a surprising result that the sinusoidal carrier produced significantly better performance than the noise carrier, particularly when the number of spectral bands was small. Previous studies on speech recognition in quiet found no such carrier effect (Dorman *et al.*, 1997), although recent studies on speech recognition in noise have hinted at a similar carrier effect (Nie *et al.*, 2003). One interpretation of this surprising carrier effect on speaker identification is that the temporal envelope cue, particularly the fundamental frequency, is better encoded by the sinusoidal carrier than the noise carrier (Gonzalez and Oliver, 2005). Another interpretation is that the sinusoidal carrier produces better modulation detection than the noise carrier and possibly resolved sidebands, particularly at low frequencies, to allow the normal-hearing listeners to directly hear out the voice pitch cue (Kohlrausch *et al.*, 2000; Zeng, 2003).

The above-mentioned studies implicated strongly that current cochlear implant users do not receive sufficient acoustic cues to support speaker identification and underscored the importance of extracting and encoding the temporal fine structure in cochlear implants (Oppenheim and Lim, 1981; Smith *et al.*, 2002; Nie *et al.*, 2005). Oppenheim and Lim (1981) independently manipulated Fourier amplitude and phase spectra and sometimes mixed one stimulus's amplitude spectrum with another stimulus's phase spectrum to demonstrate that phase provides critical information for auditory and visual perception. However, the importance of phase information has been largely ignored in the implant field until the Smith *et al.* chimera experiment (Smith *et al.*, 2002). Smith *et al.* mixed up one sound's temporal envelope with another sound's temporal fine structure to demonstrate their independent contributions to speech recognition and pitch perception. To overcome the difficulty of encoding the relatively rapid-varying temporal fine structure, Nie *et al.* (2005) derived slowly varying frequency modulations around the center frequency of a particular subband and found them to be effective in separating one speaker from another to achieve better speech recognition in noise.

Our goal for the present study was twofold. The first goal was to delineate the relative contributions of temporal envelope and fine structure to speech and speaker recognition. The second goal was to identify novel coding strategies to improve speaker identification in cochlear-implant users. To achieve these goals, the present study used the same vowel stimuli to collect systematically both vowel and speaker identification in normal-hearing and cochlear-implant subjects. The normal-hearing subjects listened to vowel syllables (in /hVd/ context) from ten speakers. These syllables included both the original unprocessed stimuli and the processed stimuli to contain either the temporal envelope cue or additionally the slowly varying frequency modulation cue. Performance for the processed stimuli was measured as a function of the number of frequency bands from 1 to 32. The cochlear-implant subjects performed the same task, but with only the original unprocessed stimuli. As a control, vowel recognition was also measured using identical stimuli from the same ten speakers in both normal-hearing and cochlear-implant subjects.

II. METHODS

A. Subjects

Six normal-hearing adults between the ages of 18 and 32 years and ten post-lingually deafened implant users between the ages of 49 and 74 years participated in the experiments. The implant subjects included 1 Ineraid device user (with a Med El CIS speech processor), 6 Nucleus users (with 3 SPEAK and 3 ACE users), and 3 Clarion users (with 1 CIS and 2 PSP users). Each implant subject had used the device for at least one year at the time of test. All participants were native English speakers. Additional demographic information can be found in Table I.

B. Stimuli

Stimuli consisted of 12 vowel tokens in the /hVd/ context and were originally recorded and analyzed by Hillenbrand and his colleagues (Hillenbrand *et al.*, 1995). Instead of the traditionally used sentence materials, the vowel stimuli were chosen because they could be used repetitively for the large number of experimental conditions employed in the present study, and additionally they were generally free of linguistic and speech rate/rhythm cues. In the speaker identification experiment, only two sets of three vowels were selected. One set (/had/, /heed/, and /hawd/) was used for practice and training purpose while the other set (/herd/, /hid/, and /hoed/) was used for the experiment. These tokens were chosen to ensure each set had high/high, high/low, and low/high F1/F2 values. Ten speakers including three men, three women, two boys, and two girls were used to form a total of 60 tokens. In the vowel recognition experiment, all 12 vowels were used. The same ten speakers produced a total of 120 tokens that were used for both practice and experiment purposes.

The original Hillenbrand stimuli were first pre-emphasized by a first-order high-pass Butterworth filter at 1200 Hz. The pre-emphasized stimuli were then bandpassed using fourth-order elliptic bandpass filters to produce 1, 4, 8,

TABLE I. Biographical data on cochlear implant subjects.

Subject #	Gender	Age (years)	Age of loss	Year of implantation	Etiology	Device	Strategy
1	M	71	39	1978	Meniere's	Ineraid	CIS
2	M	62	40	1990	Trauma	Nucleus 22	SPEAK
3	M	62	45	1995	Genetic	Nucleus 22	SPEAK
4	F	70	30	1998	Otosclerosis	Nucleus 22	SPEAK
5	F	49	9	1999	Unknown	Nucleus 24	ACE
6	F	69	44	1997	Virus	Nucleus 24	ACE
7	F	70	30	2000	High fever	Nucleus 24	ACE
8	F	68	34	1998	Autoimmune	Clarion I	CIS
9	F	72	46	2001	Nerve	Clarion II	PSP
10	F	74	57	2000	SNHL	Clarion II	PSP

16, and 32 subbands (Greenwood, 1990). The temporal envelope was extracted from each sub-band by full-wave rectification and low-pass filtering with a 500 Hz cutoff frequency. The slowly varying frequency modulation was extracted from each sub-band using a pair of phase-orthogonal demodulators with a cosine and sine carrier at the center frequency of each sub-band (Nie *et al.*, 2005). The frequency modulation had a bandwidth of 500 Hz and a modulation rate at 400 Hz. The frequency modulation extracted and preserved both the within-band and the cross-band phase information.

To produce stimuli with primarily temporal cues, the band-specific envelope was used to amplitude modulate a fixed carrier whose frequency was equal to the sub-band center frequency. To produce stimuli containing the slowly varying frequency modulation, the phase component was first recovered by integration of the frequency modulations before applying amplitude modulation by the temporal envelope (Nie *et al.*, 2005). Finally, before the summation of the sub-band signals, the same bandpass filter as the analysis bandpass filter was applied to both AM and AM+FM sub-band stimuli. This bandpass filter would effectively remove spectral differences between AM and AM+FM stimuli.

C. Procedure

Computer interfaces using MATLAB were developed for both speaker and vowel recognition experiments. Push buttons displayed on a computer monitor were created to correspond to a closed set of choices. For the speaker identification interface, ten push buttons were displayed in two rows. Row 1 corresponded to Male 1, Boy 1, Male 2, Boy 2, Male 3; Row 2 corresponded to Female 1, Girl 1, Female 2, Girl 2, Female 3. For the vowel recognition experiment, 12 corresponding pushbuttons were displayed on the interface.

Experiments were conducted in a double-walled sound treated booth (IAC). Stimuli were presented at 65 dBA via either a Sennheiser headset (HDA200) monaurally to normal-hearing listeners or a TANNOY Reveal speaker to cochlear-implant listeners. In the speaker identification experiment, all subjects received one to two hour training by systematically and/or randomly selecting push buttons to listen to the corresponding speaker's voice. All subjects underwent five practice rounds before formal data collection for the experimental condition. Each practice round consisted of 60 stimuli, including 2 presentations of 1 set of 3 vowels from 10 speakers. During testing, stimuli were presented randomly and the subject was subsequently asked to choose the correct speaker. Feedback was given after each selection by indicating whether the subject's choice was correct or incorrect via highlighting the push button corresponding to the correct answer.

After five practice rounds, the experimental test was conducted using the other set of three vowels to which the subject had not been exposed. In the vowel recognition experiment, the same procedure as in the speaker identification experiment was used except for a different task (recognizing 12 vowels instead of 10 speakers), and a different user interface. Only one practice session was conducted.

III. RESULTS

A. Original stimuli

Figure 1 shows training data from 5 practice sessions, as well as the test session (#6) for both normal-hearing (tri-

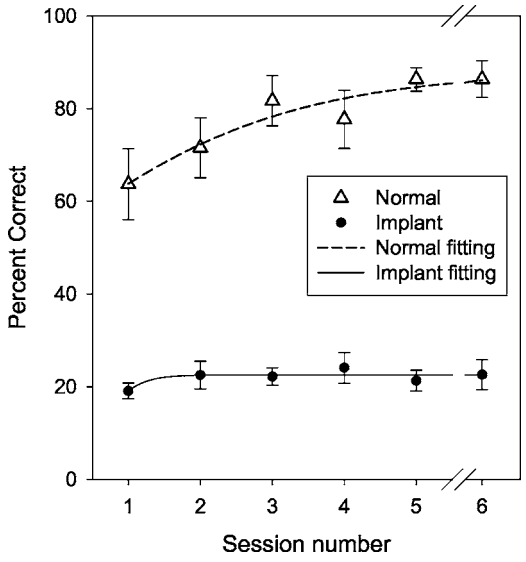


FIG. 1. Learning curve for speaker recognition in normal-hearing (open triangles) and cochlear-implant (filled circles) subjects. Sessions #1–5 represent the training period while session #6 represents the test run. Error bars represent plus and minus one standard error. The lines represent fitting of the learning curve with a sigmoidal function.

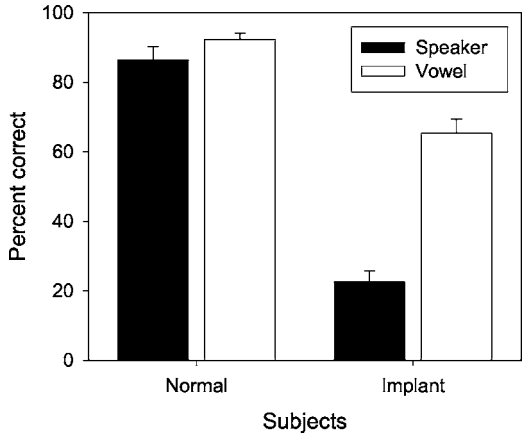


FIG. 2. Average performance for speaker (filled bars) and vowel (open bars) recognition in normal-hearing and cochlear-implant subjects. Error bars represent plus and minus one standard error. The chance performance is 10% for speaker recognition and 8% for vowel recognition.

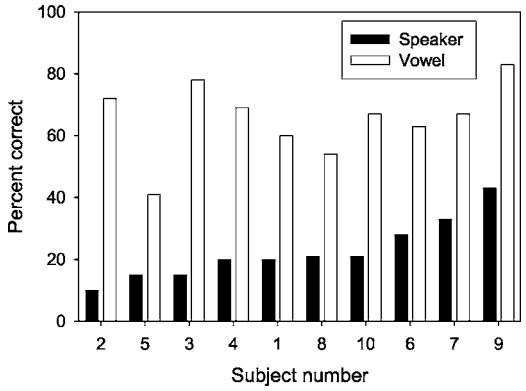


FIG. 3. Individual performance for speaker (filled bars) and vowel (open bars) recognition in cochlear-implant subjects. Individual data are ranked by the speaker recognition performance with the subject number corresponding to that in Table I.

angles) and cochlear-implant (circles) subjects. The normal subjects showed a significant learning effect with average performance increasing from 64% correct in session 1 to a plateau at about 84% in session 3 (paired t test, $p<0.01$). In contrast, the implant subjects performed significantly more poorly than normal-hearing subjects with a plateau at approximately a 20% correct level. In addition, the three-percentage point training effect between sessions one and six was not significant ($p>0.1$). A sigmoid function was used to fit the training data, showing an asymptotic performance of 88% and 23% correct for normal and implant subjects, respectively. Finally, there was no significant difference ($p>0.1$) between the last practice run (session five) and the test run (session six) for both normal and implant subjects.

Figure 2 contrasts the overall performance between speaker (filled bars) and vowel (open bars) recognition in both normal and implant users. ANOVA with a between-subjects, faxed-factor design revealed a highly significant main effect for both the subjects [$F(1,28)=165.1, p<0.01$] and the tasks [$F(1,28)=47.8, p<0.01$]. The normal subjects performed significantly better than the implant subjects on both tasks, with 86% correct for speaker recognition and 92% correct for vowel recognition, as opposed to 23% correct for speaker recognition and 65% correct for vowel recognition in implant subjects. The difference between speaker and vowel recognition was insignificant in normal subjects ($p>0.05$) but was significant in cochlear-implant subjects ($p<0.05$).

Figure 3 shows individual data from the ten implant subjects to further highlight the difference between speaker and vowel recognition. The individual score increases from 10% (chance performance) to 43% correct for speaker recognition. Had there been a strong correlation between the two tasks, a similar increasing trend would be observed for the individual performance for vowel recognition. Instead, an insignificant correlation was found ($r=0.37, p>0.05$), accounting for only 14% variability in the data.

B. Processed stimuli

Figure 4 compares the performance in both speaker (left panel) and vowel (right panel) recognition as a function of the number of spectral bands. The performance with the temporal envelope cue is represented by filled triangles (AM) while the performance with the additional frequency modulation cue is represented by open circles (AM+FM). For comparison, the cochlear implant performance is also included as the hatched bar, with its height corresponding to the mean score and its position on the x axis indicating the equivalent number of the AM bands.

In the speaker recognition task, the overall performance for the AM only condition was increased from 28% correct with one band to 76% correct with 32 bands. The corresponding performance for the AM+FM condition was from 46% to 82% correct. A repeated ANOVA shows a significant effect for both the processing [AM vs AM+FM: $F(1,18)=564.7$, $p<0.01$] and the number of bands [$F(4,18)=504.2, p<0.01$]. With four bands, the AM+FM condition produced the greatest improvement of 36 percentage points over the AM condition. Even with 32 bands, the AM+FM condition still resulted in significantly better performance than the AM condition by 6 percentage points ($p<0.05$).

The vowel recognition performance was similar to the

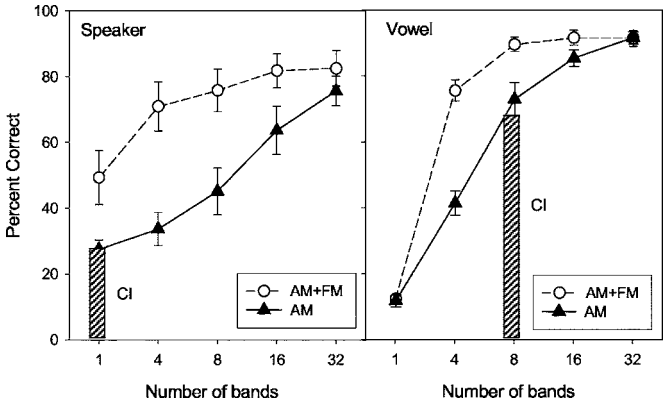


FIG. 4. Average performance for speaker (left panel) and vowel (right panel) recognition as a function of spectral bands in normal-hearing subjects. Filled triangles represent data obtained with the amplitude modulation cue (AM) while open circles represent data with both the amplitude and frequency modulation cues (AM+FM). Cochlear-implant performance is represented by the hatched bar, with its height corresponding to the mean score and its position on the x axis, indicating the equivalent number of the AM bands.

TABLE II. Stimulus-response or confusion matrix for speaker identification in cochlear-implant subjects.

	Man 1	Boy 1	Man 2	Boy 2	Man 3	Woman 1	Girl 1	Woman 2	Girl 2	Woman 3
Man 1	**9**	7	4	6	1	4	2	9	2	6
Boy 1	4	**8**	2	4	0	6	9	5	6	4
Man 2	4	2	**20**	1	16	0	1	3	0	1
Boy 2	7	1	4	**6**	1	5	5	8	6	5
Man 3	7	2	6	0	**32**	0	0	1	0	0
Woman 1	7	4	3	5	0	**7**	5	10	4	3
Girl 2	2	4	2	9	0	4	**11**	10	4	3
Woman 2	3	9	2	7	0	4	8	**10**	7	2
Girl 2	4	6	1	7	0	4	6	7	**7**	6
Woman 3	5	4	3	7	0	10	8	3	6	**3**

speaker recognition performance. However, several apparent differences were present, including a significant interaction between the processing and the task $[F(1,18)=35.6, p<0.01]$, and between the number of bands and the task $[F(1,18)=107.8, p<0.01]$. To demonstrate this interaction, first note the one-band results showing significantly better performance for speaker recognition than for vowel recognition for both the AM and the AM+FM conditions ($p<0.05$). Second, note the insignificant difference in performance between the speaker and vowel recognition with four bands ($p>0.05$). Third, note the reversed pattern showing better vowel recognition than speaker recognition with eight and more bands.

Notice, finally, that the equivalent number of bands for the cochlear-implant subjects is highly dependent on the task. In the speaker recognition task, the implant subjects performed at a level that was equivalent to the performance achieved by normal subjects with only one band. In contrast, the implant subjects were able to achieve a high level of performance on the vowel recognition task that was equivalent to eight bands for the normal subjects.

IV. DISCUSSION

A. Speaker versus vowel recognition

The most striking finding in the present study is the apparent disassociation of the use of temporal envelope cues between speaker and vowel recognition. This disassociation can best be observed by poor performance for speaker recognition but good performance for vowel recognition in the cochlear-implant subjects (Fig. 2). This disassociation is further enhanced by a lack of correlation between speaker and vowel recognition in the implant subjects (Fig. 3). In cochlear implant simulation, this disassociation is best illustrated by the interaction between the number of bands and the listening tasks (Fig. 4). With only 1-band envelope, speaker recognition was 16 percentage points significantly better than vowel recognition; but with 8-band envelopes, speaker recognition was 28 percentage points significantly worse. The disassociation results suggest that depending on the availability of acoustic cues, the brain may use different strategies to process information regarding speaker and vowel recognition.

The disassociation results also suggest that speaker and speech (vowel) recognition may place different weights on different acoustic cues. Speaker recognition relies more on low-frequency cues that can be derived from temporal envelopes, while vowel recognition relies more on high-frequency spectral cues that require a large number of bands. This suggestion is consistent with the traditional view that acoustic cues carrying speaker information are highly related to fundamental frequency and that acoustic cues carrying speech information are highly related to formant frequencies, particularly the second formant frequency (French and Steinberg, 1947).

B. Analysis of error patterns

The speaker pool used in the present study included both gender and age factors. Although both actual and simulated implant performance was low for speaker recognition, it was still possible that the implant subjects could identify gender and age, but were only confused within categories. To analyze the error patterns in speaker recognition, classic sequential information transfer analysis (SINFA) was performed (Wang and Bilger, 1973).

Table II shows the confusion matrix for speaker recognition in eight of the ten cochlear-implant subjects. Subject #1 and #2 were not available as their data were collected before the information regarding the speaker confusion pattern was recorded in the program. The stimuli were represented as rows while the responses were represented as columns. A total of 488 tokens were pooled from 8 subjects with each contributing to 61 responses (10 speakers $\times 3$ tokens $\times 2$ presentation $+1$ randomly selected token). The overall score was 23.2% correct, indistinguishable from the 22.6% score obtained from the 10 implant subjects.

SINFA (Wang and Bilger, 1973) shows that the cochlear-implant subjects were only able to receive 4.7% gender information and 2.5% age information, corresponding to 62.7% and 60.9% overall percent correct, respectively. The percent information transmitted improved slightly to 9.3% for gender discrimination when age had been accounted for and to 4.3% for age discrimination when the gender had been accounted for. SINFA was also used to analyze the error patterns with one- and four-band conditions in normal-hearing subjects and found generally similar results to the implant pattern. Together, the present analysis of error

patterns demonstrated that temporal envelope cues do not provide reliable information for speakers either across or within both gender and age categories.

C. Implications for cochlear implants

The present results highlight the limitation of current speech processing strategies in cochlear implants. Except for the infrequently used analog strategies, only temporal envelope information from several bands is extracted while the temporal fine structure information is discarded in the process. Given the limited number of functional channels available in current cochlear implants, it is clear that the temporal envelope cue is not sufficient to support reliable speaker recognition.

There are at least three ways to redesign current speech processing strategies to improve speaker recognition performance in cochlear-implant users. One way is to explicitly encode the fundamental frequency information. In an earlier but now abandoned speech processing strategy (Skinner *et al.*, 1991), information regarding the fundamental frequency along with the first and/or second formant frequencies was extracted and delivered to the cochlear implant via pulsatile stimulation patterns following the changes in fundamental frequency. To our knowledge, no study had been performed to directly evaluate this earlier strategy's performance in speaker recognition. It is possible that the fundamental frequency information can be reintroduced as a carrier frequency in the modern speech processing strategies. A simulation of such a processing strategy has been shown to improve Mandarin tonal recognition (Lan *et al.*, 2004).

Another way to improve upon the current cochlear implants is to extract the slowly varying frequency modulation and deliver it to cochlear implants (Nie *et al.*, 2005). The slowly varying frequency modulation does not explicitly extract fundamental frequency but does contain information regarding the direction and rate of both fundamental and formant movements. The present simulation result shows that this slowly varying frequency modulation could produce significantly better performance in speaker recognition, even with one or four bands.

A third way to improve upon the current cochlear implants is to introduce a high-frequency or noise conditioner to improve frequency representation in the temporal envelope domain at the auditory nerve level (Rubinstein, 1995; Morse and Evans, 1996; Litvak *et al.*, 2003). The hope is that the conditioner would improve frequency discrimination (Zeng *et al.*, 2000), which would in turn improve speaker identification based on relatively low frequencies in the envelope domain. While it is not clear which exact strategy or a combination of strategies might work, it is clear from the present data that current speech processing strategies need to be changed to improve speaker recognition performance in cochlear implant users.

V. CONCLUSIONS

Speaker and vowel recognition performance was measured in ten cochlear-implant and six normal-hearing subjects. The speakers included three men, three women, two boys, and two girls. The stimuli were 12 vowels in /hVd/ context from the Hillenbrand study (Hillenbrand *et al.*, 1995). The main findings are as follows.

(1) Current cochlear-implant users are able to achieve good performance in vowel recognition (65% correct) but poor performance in speaker recognition (23% correct).
(2) Implant performance is functionally equivalent to normal performance of eight spectral bands with temporal envelopes for vowel recognition but only one band for speaker recognition.
(3) A slowly varying form of frequency modulation can improve significantly the speaker recognition performance and should be encoded in future cochlear implants.
(4) The present result supports the hypothesis that speaker and speech recognition with primarily temporal cues involves two independent processes.

ACKNOWLEDGMENTS

We thank Dr. Kaibao Nie for providing the MATLAB program to process the stimuli, University College London for the SINFA program, and Abby Copeland, Molly Pitassi, Ken Grant and two anonymous reviewers for their helpful comments on the manuscript. The work was supported by the National Institutes of Health (2RO1DC02267) and UCI Undergraduate Research Opportunity Program (UROP to Vongphoe who was also selected as Researcher of the Month in September, 2002).

Atal, B. S. (**1972**). "Automatic speaker recognition based on pitch contours," J. Acoust. Soc. Am. **52**, 1687–1697.
Childers, D. G., and Wu, K. (**1991**). "Gender recognition from speech. Part II: Fine analysis," J. Acoust. Soc. Am. **90**, 1841–1856.
Cleary, M., and Pisoni, D. B. (**2002**). "Talker discrimination by prelingually deaf children with cochlear implants: Preliminary results," Ann. Otol. Rhinol. Laryngol. Suppl. **189**, 113–118.
Dorman, M. F., and Loizou, P. C. (**1998**). "The identification of consonants and vowels by cochlear implant patients using a 6-channel continuous interleaved sampling processor and by normal-hearing subjects using simulations of processors with two to nine channels," Ear Hear. **19**, 162–166.
Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2411.
Fellowes, J. M., Remez, R. E., and Rubin, P. E. (**1997**). "Perceiving the sex and identity of a talker without natural vocal timbre," Percept. Psychophys. **59**, 839–849.
French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.
Fu, Q. J., Chinchilla, S., and Galvin, J. (**2004**). "The role of spectral and temporal cues in voice gender discrimination by normal-hearing listeners and cochlear implant users," J. Assoc. Res. Otolaryngol. **5**, 253–260.
Fu, Q. J., Zeng, F. G., Shannon, R. V., and Soli, S. D. (**1998**). "Importance of tonal envelope cues in Chinese speech recognition," J. Acoust. Soc. Am. **104**, 505–510.
Gonzalez, J., and Oliver, J. C. (**2005**). "Gender and speaker identification as a function of the number of channels in spectrally reduced speech," J. Acoust. Soc. Am. (in press).
Greenwood, D. D. (**1990**). "A cochlear frequency-position function for several species—29 years later," J. Acoust. Soc. Am. **87**, 2592–2605.
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.
Holt, L. L., Lotto, A. J., and Kluender, K. R. (**2001**). "Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement?" J. Acoust. Soc. Am. **109**, 764–774.

Johnson, C. C., Hollien, H., and Hicks, J. W. (**1984**). "Speaker identification utilizing selected temporal speech features," J. Phonetics **36**, 93–100.
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.
Kong, Y. Y., Vongphoe, M., and Zeng, F. G. (**2003**). "Independent contributions of amplitude modulation and frequency modulation to auditory perception: II. Melody, tone and speaker identification," Abstract of the 26th Annual Midwinter Research Meeting, Vol. 26, pp. 213–214.
Ladefoged, P., and Broadbent, D. E. (**1958**). "Information conveyed by vowels," J. Acoust. Soc. Am. **29**, 98–104.
Lan, N., Nie, K. B., Gao, S. K., and Zeng, F. G. (**2004**). "A novel speech-processing strategy incorporating tonal information for cochlear implants," IEEE Trans. Biomed. Eng. **51**, 752–760.
Liang, Z. A. (**1963**). "Auditory perceptual cues in Mandarin tones," Acta Phys. Sin. **26**, 85–91.
Litvak, L. M., Delgutte, B., and Eddington, D. K. (**2003**). "Improved temporal coding of sinusoids in electric stimulation of the auditory nerve using desynchronizing pulse trains," J. Acoust. Soc. Am. **114**, 2079–2098.
McDonald, C. J., Kirk, K. I., Krueger, T., Houston, D., and Sprunger, A., (**2003**). "Talker discrimination by adults with cochlear implants," *Abstracts of the 26th Annual Midwinter Research Meeting of the Association for Research in Otolaryngology*, Daytona Beach, Florida.
Morse, R. P., and Evans, E. F. (**1996**). "Enhancement of vowel coding for cochlear implants by addition of noise," Nat. Med. **2**, 928–932.
Nie, K., Stickney, G., and Zeng, F. G. (**2003**). "Independent contributions of amplitude modulation and frequency modulation to auditory perception: I. Consonant, vowel and sentence recognition," *Abstracts of the 26th Annual Midwinter Research Meeting of the Association for Research in Otolaryngology*, Daytona Beach, Florida.
Nie, K., Stickney, G., and Zeng, F. G. (**2005**). "Encoding fine structure to improve cochlear implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.
Oppenheim, A. V., and Lim, J. S. (**1981**). "The importance of phase in signals," Proc. IEEE **69**, 529–541.
Remez, R. E., Fellowes, J. M., and Rubin, P. E. (**1997**). "Talker identification based on phonetic information," J. Exp. Psychol. Hum. Percept. Perform. **23**, 651–666.
Rosen, S. (**1992**). "Temporal information in speech: Acoustic, auditory and linguistic aspects," Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Rubinstein, J. T. (**1995**). "Threshold fluctuations in an *N* sodium channel model of the node of Ranvier," Biophys. J. **68**, 779–785.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Skinner, M. W., Holden, L. K., Holden, T. A., Dowell, R. C., Seligman, P. M., Brimacombe, J. A., and Beiter, A. L. (**1991**). "Performance of post-linguistically deaf adults with the Wearable Speech Processor (WSP III) and Mini Speech Processor (MSP) of the Nucleus Multi-Electrode Cochlear Implant," Ear Hear. **12**, 3–22.
Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). "Chimaeric sounds reveal dichotomies in auditory perception," Nature (London) **416**, 87–90.
Stevens, K. N. (**2002**). "Toward a model for lexical access based on acoustic landmarks and distinctive features," J. Acoust. Soc. Am. **111**, 1872–1891.
Stevens, K. N., Williams, C. E., Carbonell, J. R., and Woods, B. (**1968**). "Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material," J. Acoust. Soc. Am. **44**, 1596–1607.
Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). "Speech waveform envelope cues for consonant recognition," J. Acoust. Soc. Am. **82**, 1152–1161.
Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.
Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (**1993**). "FO gives voicing information even with unambiguous voice onset times," J. Acoust. Soc. Am. **93**, 2152–2159.
Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M., (**1991**). "Better speech recognition with cochlear implants," Nature (London) **352**, 236–238.
Wu, K., and Childers, D. G. (**1991**). "Gender recognition from speech. Part I: Coarse analysis," J. Acoust. Soc. Am. **90**, 1828–1840.
Xu, L., and Pfingst, B. E. (**2003**). "Relative importance of temporal envelope and fine structure in lexical-tone perception," J. Acoust. Soc. Am. **114**, 3024–3027.
Zeng, F. G. (**2003**). "Compression and cochlear implants," *Compression: From Cochlea to Cochlear Implants*, edited by S. P. Bacon, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), Vol. 17, pp. 184–220.
Zeng, F. G., Fu, Q. J., and Morse, R. (**2000**). "Human hearing enhanced by noise," Brain Res. **869**, 251–255.
Zeng, F. G., Grant, G., Niparko, J., Galvin, J., Shannon, R., Opie, J., and Segel, P. (**2002**). "Speech dynamic range and its effect on cochlear implant performance," J. Acoust. Soc. Am. **111**, 377–386.
Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C. G., and Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Nat. Acad. Sci. USA **102**, 2293–2298.

Evaluating models of vowel perception[a)]

Michelle R. Molis[b)]
Department of Psychology, University of Texas at Austin, Austin, Texas 78712

(Received 30 June 2004; revised 5 May 2005; accepted 9 May 2005)

There is a long-standing debate concerning the efficacy of formant-based versus whole spectrum models of vowel perception. Categorization data for a set of synthetic steady-state vowels were used to evaluate both types of models. The models tested included various combinations of formant frequencies and amplitudes, principal components derived from excitation patterns, and perceptually scaled LPC cepstral coefficients. The stimuli were 54 five-formant synthesized vowels that had a common F1 frequency and varied orthogonally in F2 and F3 frequency. Twelve speakers of American English categorized the stimuli as the vowels /ɪ/, /ʊ/, or /ɜ˞/. Results indicate that formant frequencies provided the best account of the data only if nonlinear terms, in the form of squares and cross products of the formant values, were also included in the analysis. The excitation pattern principal components also produced reasonably accurate fits to the data. Although a wish to use the lowest-dimensional representation would dictate that formant frequencies are the most appropriate vowel description, the relative success of richer, more flexible, and more neurophysiologically plausible whole spectrum representations suggests that they may be preferred for understanding human vowel perception. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1943907]

PACS number(s): 43.71.Es, 43.71.An [PFA] Pages: 1062–1071

I. INTRODUCTION

Vowel perception commonly is characterized in terms of formant frequencies, often the lowest two formant frequencies alone (Peterson and Barney, 1952). Although formant frequencies provide a concise account of vowel perception, formant-based perceptual models are problematic due to practical difficulties in reliable real-time formant extraction by artificial (and, presumably, biological) systems. In order to address the shortcomings of formant peak models, explicit formant extraction is sometimes abandoned altogether in favor of representation of the vowel spectrum as a whole. Whole-spectrum vowel representations can take many forms such as critical-band spectra (Bladon and Lindblom, 1981; Plomp *et al.*, 1967) or cepstral coefficients (Hermansky, 1990; Zahorian and Jagharghi, 1993); however, all have in common a preservation of a richer description of the frequency spectrum than is provided by the formant frequency values alone.

Formant-based and whole-spectrum representations are often pitted against one another and presented as mutually exclusive alternatives (Hillenbrand and Houde, 1995; Ito *et al.*, 2001; Kiefte and Kluender, 2001; Klatt, 1982a, 1982b; Nearey and Kiefte, 2003; Zahorian and Jagharghi, 1993). However, these investigations have found it difficult to select a decisive favorite between the two. The difficulty arises because information about formant frequencies is present implicitly in nearly all whole-spectrum representations (Broad and Clermont, 1989; Nearey and Kiefte, 2003; Plomp *et al.*, 1967; Pols *et al.*, 1969; Zahorian and Jagharghi, 1993) thereby making it difficult to accept one representation to the exclusion of the other.

A. Formant-based representations

Formant frequencies have much to recommend them as the preferred vowel representation. The formant resonances are a consequence of the vocal tract configurations of the speaker and formant frequencies can provide a relatively simple, low-dimensional description of vowel spectra (Fant, 1960). Formant locations can account for the results of many vowel identification studies (for a review of this topic, see Rosner and Pickering, 1994). This is highlighted by the finding that judgments of phonetic distance are affected primarily by formant frequency changes, whereas judgments of psychophysical distance are affected by many spectral shape details other than formants (e.g., changes in spectral tilt or filtering) (Carlson *et al.*, 1970; Klatt, 1982a).

Although formant frequencies have been useful in the study of vowel production and perception, formant-based vowel representations have also received criticism. Bladon (1982) described three objections to such representations based on reduction, determinacy, and perceptual adequacy. According to the reduction objection, to describe vowels with reference only to formant center frequencies results in a nontrivial reduction of spectral information. The determinacy objection stems from observation that automatic formant frequency extraction tends to be unreliable. This is especially true for vowels in noise, for nasalized vowels, and for high fundamental frequencies. Moreover, errors produced by formant tracking algorithms are not always consistent with those made by human listeners (Klatt, 1982b); formant trackers are prone to introduce false peaks, to merge closely spaced peaks, or to miss peaks entirely, resulting in extreme identification errors. In contrast, human listeners tend not to

[a)]Portions of this research were presented at the 143rd meeting of the Acoustical Society of America (J. Acoust. Soc. Am. **111**, 2433–2434). Aspects of these data were also analyzed in Maddox *et al.* [Percept. Psych. **64**, 584–597 (2002)].

[b)]Present address: Army Audiology and Speech Center, Walter Reed Army Medical Center, 6900 Georgia Ave., Washington, DC 20307; electronic-mail: michelle.molis@na.amedd.army.mil

make these extreme misidentifications; instead, aspects of overall spectral shape better explain the pattern of errors. Finally, according to Bladon's perceptual adequacy objection, formants alone cannot account for the entire set of empirical results. Specifically, nonlinearities are observed in judgments of perceptual distance among vowels; equal shifts of formant frequency do not correspond to equal shifts in vowel quality (Bladon, 1983).

The most typical vowel description is the lowest two or three formant frequencies (Peterson and Barney, 1952). These values may be estimated from spectrograms, LPC analysis, or may be the product of various formant-tracking algorithms. This basic description may be elaborated; for instance, including relative formant amplitudes along with formant frequencies provides crude spectral shape information. When simple synthetic stimuli are used, identification studies frequently find vowel category boundaries that are linear with respect to formants expressed in log-like units (e.g., mel or Bark) and parallel to the first and second formant axes or that can be characterized as simple linear combinations of formant frequencies (Carlson *et al.*, 1970; Hose *et al.*, 1983; Karnickaya *et al.*, 1973).

However, linear boundaries may not be sufficient to characterize the boundaries between more complex synthetic stimuli or naturally produced vowels. Through introduction of the squares and cross products of formant frequencies, nonlinear boundaries can be represented. In previous investigations, Nearey and Kiefte (2003) found that adding nonlinear (quadratic) formant terms improved the fit of the multinomial logistic regression analysis of a large F1-F2-F3 vowel space relative to the linear terms alone. Maddox *et al.* (2002) investigated decision processes in the categorization of vowels using both linear and nonlinear formant models and also found improved fits when nonlinear boundaries were allowed. Hillenbrand and Gayvert (1993) evaluated a nonlinear classification method for spoken vowels, and compared their results with earlier studies that used linear analyses. They reported no benefit to classification based on linear versus nonlinear frequency scales.

B. Whole-spectrum representations

Each of the vowel representations introduced so far is predicated on the accurate extraction of formant frequencies. Another class of models calls into question the need for explicit formant extraction at all. These models seek to describe vowels in terms of properties of the spectrum as a whole rather than as sets of individual formant frequency values. Further, formant-based models typically rely on spectral analysis independent of auditory processing (e.g., LPC analysis), whereas, whole-spectrum models often incorporate aspects of auditory processing and, as such, provide formant information in a neurophysiologically plausible form (however, for an exception, see Assmann and Summerfield, 1989).

By incorporating characteristics of the peripheral auditory system, such as critical-band filtering, they can perhaps account for some of the perceptual nonlinearities observed in perceptual distance judgments (Bladon, 1983) and vowel quality matching (Carlson *et al.*, 1970). For example, auditory frequency resolution has been implicated in the appearance of formant mergers in response to vowel stimuli (Chistovich and Lublinskaya, 1979; Chistovich *et al.*, 1979).

However, a critical-band spectrum difference metric, such as developed by Carlson and Granström (1979) or Bladon and Lindblom (1981), is highly sensitive to differences in relative formant amplitude. Because neither variations in formant amplitudes nor spectral tilt have as much effect on phonetic distance judgments as changes in formant frequencies, Klatt (1982b) rejected vowel representations based on level or loudness across frequency. He proposed that this difficulty may be overcome through comparison of differences in spectral slope rather than spectral level. In theory, this would serve to enhance the importance of differences between spectral peaks relative to other spectral regions and would remove overall spectral tilt.

Pols and colleagues sought a more general representation of the differences among vowels than could be provided by formant frequencies alone (Klein *et al.*, 1970; Plomp *et al.*, 1967; Pols, 1970; Pols, 1977; Pols *et al.*, 1969). Their alternative approach contrasted a detailed frequency analysis, as exemplified by formant frequency extraction, with a less detailed frequency analysis, more in line with the frequency resolution of the ear (Pols, 1977). The frequency analysis capabilities of the ear were represented for vowel stimuli in terms of the sound pressure levels (in dB) in a series of successive, nonoverlapping frequency bands, the width of which was on the order of the frequency resolution of the cochlear filters. Because the sound pressure levels in these frequency bands are not independent, the number of stimulus dimensions can be reduced to a small set of independent factors through multivariate analysis. Such an analysis allows efficient reduction of spectral information, without the computationally complex process of formant extraction. In an evaluation with naturally produced Dutch vowels, low-dimension solutions were found that accounted for much of the variance among vowel tokens (Plomp *et al.*, 1967; Klein *et al.*, 1970). The researchers concluded that the same information contained in the F1/F2 plane also could be found within a multivariate analysis of the critical-band representation (Plomp *et al.*, 1967; Pols *et al.*, 1969).

Characteristics of the whole spectrum can also be represented through cepstral coefficients (Hermansky, 1990; Zahorian and Jagharghi, 1993)—the number of coefficients determines the degree of spectral detail that can be represented. Hermansky's (1990) perceptual linear predictive (PLP) analysis recovers spectral shape properties in the form of nonlinearly scaled cepstral coefficients. The PLP processing is similar to LPC analysis but involves a nonlinear scaling on both the frequency and amplitude axes. Hermansky demonstrated that a fifth-order PLP model was more consistent with the output of peripheral auditory processing than traditional linear predictive analysis. In addition, the cepstral coefficients are correlated with formant frequencies (Broad and Clermont, 1989). In a similar vein, Zahorian and Jagharghi (1993) investigated the contribution of spectral shape cues versus formant frequencies for the automatic classification of naturally produced vowels. Their description of

global spectral shape was provided by the discrete cosine transform coefficients (DCTCs) of nonlinearly scaled spectra. The DCTCs are similar to cepstral coefficients. A comparison of automatic classification accuracy between the coefficients and the formant frequencies indicated that the coefficients provided slightly better category discrimination. Moreover, automatic classification performance based on the nonlinear cepstral coefficients was more highly correlated with the perceptual confusions made by human listeners than were classifications based on formant frequencies.

C. Evaluation of vowel representations

This study aims to evaluate a variety of formant-based and whole-spectrum vowel representations for their relative ability to account for vowel identification. Because of the inherent overlap of the information provided by both approaches, testing identification of only a few good category exemplars will have less chance to differentiate among representations. Therefore, a relatively dense, continuously sampled stimulus space is investigated in order to reveal subtle differences in model performance. Listeners were asked to identify synthetic vowel stimuli as belonging to one of three vowel categories, and parameter sets generated from competing vowel representations were compared via separate logistic regression analyses for each listener. Because individual listeners could employ response strategies that favor one approach over another, subject responses are not pooled. Instead, each listener's performance is considered separately and rank orders of success of the various models are compared across listeners.

II. METHOD

A. Stimuli

Fifty-four, five-formant vowel stimuli were synthesized using a KLATT88-type cascade resonance synthesizer (Klatt and Klatt, 1990) implemented on a PC at a sampling rate of 10 kHz. The stimuli varied orthogonally in F2 and F3 frequency: F2 ranged from 9.0 to 13.4 Bark (1081 to 2120 Hz) and F3 ranged from 10.0 to 15.2 Bark (1268 to 2783 Hz). Both formants were sampled in 0.4 Bark intervals; however, the sampling was staggered so that for any given F2 frequency, F3 was sampled at 0.8 Bark intervals and vice versa. This stimulus region included three vowel categories of American English: /ɪ/ as in "*hid*," /ʊ/ as is "*hood*," and /ɝ/ as in "*heard*." Previous investigations (Molis *et al.*, 1998) indicated that this region of F2/F3 variation contained category boundaries between the three possible vowel contrasts. The implications of restricting the stimulus set in this way will be addressed in Sec. IV.

The frequency values of F1, F4, and F5 were held constant at 4.5 Bark (455 Hz), 16.2 Bark (3250 Hz), and 17.0 Bark (3700 Hz), respectively. Formant amplitudes were allowed to vary with changes in formant frequency according to the cascade synthesis algorithm and were not independently manipulated.

A constant fundamental frequency of 132 Hz was maintained for the initial 150 ms and then fell linearly to 127 Hz over the final 75 ms. All stimuli were 225 ms in length to approximate the average measured intrinsic duration of these vowels produced in citation form (Hillenbrand *et al.*, 1995). Stimuli were ramped on and off with a 10 ms half-cosine function and were normalized for rms amplitude.

B. Data acquisition and stimulus representations

In order to compare competing vowel representations, category identification data were obtained from twelve adult males ranging in age from 18 to 37 years with a mean age of 27.2 years. They reported no history of hearing loss or neurological disorders. All listeners were raised in or around the metropolitan areas of Austin, Houston, or Dallas, TX; therefore, this was a relatively homogeneous dialect group (Central Texas). Listeners were paid for their participation. Each listener completed participation in this task in one 2 h session.

Listeners were tested individually, seated in a sound-attenuating chamber. Stimuli were presented at a level of 70 dB SPL over Beyer DT-100 headphones. Fourteen randomized blocks of the 54 stimuli were presented (756 trials/listener). Listeners were asked to identify each stimulus by pressing one of three response buttons labeled with the key words "*hid*," "*hood*," and "*heard*." Listeners were given up to 2 s to make a response. After either a response or the 2 s response interval, another 1 s elapsed before the next token was presented.

For each listener, the response frequencies for the three vowel categories were used as the input to logistic regression analyses in order to evaluate models of vowel representation. This approach is similar to evaluations of model performance described by Maddox *et al.* (2002) and Nearey and Kiefte (2003) and some of the data modeled here also served as the basis for a comparison of visual and auditory models of categorization by Maddox and colleagues (2002). Through logistic regression analysis, any number of categorical or continuous independent variables can be used to model the categorical outcome of a dependent variable (Hosmer and Lemeshow, 1989). A parameter set was developed for each tested model and then used as prediction variables in the logistic regression. The following sections describe in greater detail the parameter sets evaluated. A variety of stimulus representations, ranging from formants to whole spectra, were considered.

1. Representations based on formant frequencies and amplitudes

The parameters used to describe these models were the frequencies and amplitudes measured from the synthesized vowel stimuli.[1] The formant frequencies (in Bark) and relative amplitudes (in dB) of the first three formants were estimated from the output of a 14-pole LPC analysis on 25.6 ms (256 point) Hamming-windowed segments of the synthetic vowel stimuli.

Models were generated using linear and nonlinear combinations of the formant frequencies (F) and amplitudes (A). Table I lists the five parameter sets composed of these com-

TABLE I. The five formant frequency (F) and amplitude (A) based models evaluated. Model b also includes the squares and cross product of the second and third formants.

Sets of input parameters
(a) F2 and F3
(b) F2, F3, $F2^2$, $F3^2$, F2×F3
(c) F1, F2, F3
(d) F2, F3, A2, A3
(e) F1, F2, F3, A1, A2, A3

binations. A parameter set that includes the squares and cross products of the second and third formant frequencies was also evaluated.

2. Whole-spectrum representations

a. Excitation patterns. This representation is derived from auditory excitation patterns. The excitation pattern simulates the effects of peripheral auditory frequency analysis and can be considered to be the combined outputs of the auditory (critical-band) filters across frequency. Filter bandwidths increase with increasing center frequency so that individual harmonics are resolved at low frequencies while only gross spectral detail, such as formant locations, is preserved at higher frequencies. This representation is similar to those described previously by Klatt (1982a, 1982b), Bladon and Lindblom (1981) and Plomp and Pols (Klein *et al.*, 1970; Plomp *et al.*, 1967; Pols *et al.*, 1969).

The excitation patterns were calculated by applying the algorithm described by Moore and Glasberg (1983). Briefly, for each stimulus, the magnitude spectrum of a 1024-point FFT was processed through a bank of overlapping critical-band filters spaced at roughly 20 Hz. The summed output of this filterbank is meant to represent the excitation that would occur in the cochlea at every sampled frequency.

As calculated, the excitation pattern comprises more data points than reasonably can be used as predictors in the logistic regression. Therefore, similar to the approach taken by Plomp and Pols (Klein *et al.*, 1970; Plomp *et al.*, 1967; Pols *et al.*, 1969), the number of predictor variables was reduced through principal components factor analysis. Through factor analysis it is possible to extract a small set of factors that capture the variability among the entire set of excitation patterns.

A principal components analysis of the first 225 points of the 54 excitation patterns (frequencies up to about 4.5 kHz) with varimax rotation produced six factors with eigenvalues greater than one, accounting for a total of 99.6% of the variance (the first two factors accounted for 59.0% and 23.6% of the total variance, respectively). These six factors were used as the predictor variables in the logistic regression analysis. A regression of the varimax rotated principal components on the Bark-scaled formant frequencies produced R^2 values of 0.97 for predictions of both F2 and F3, indicating that formant information is well represented in the factor solution.

b. Spectral slope based on the excitation pattern. The continuous slope representation was derived from the 225-point excitation patterns described in the previous section. The linear slope was computed at each point along the frequency axis relative to the preceding point producing a 224-point representation [i.e., $(x_i - x_{i-1})$]. This representation was also reduced through principal components analysis with varimax rotation. The analysis produced ten factors with eigenvalues greater than one, accounting for 98.73% of the variance.

c. Perceptual linear prediction (PLP) cepstral coefficients. Representations based on cepstral coefficients have been investigated by Hermansky (1990), Zahorian and Jagharghi (1993), and Nearey and Kiefte (2003). The specific implementation employed here followed Hermansky (1990). It simulates auditory processing through the use of critical-band filtering, but also includes estimates of the intensity-to-loudness conversion performed by the auditory system. PLP coding is similar to LPC analysis, but the spectra are nonlinearly scaled on both frequency and amplitude axes before the coefficients are calculated. Unlike LPC, the PLP analysis exhibits frequency resolution characteristics consistent with human hearing (Hermansky, 1990).

The calculation of PLP coefficients for each stimulus began with a 512-point FFT carried out on a Hamming-windowed stimulus segment. The spectrum was transformed along the frequency axis into Bark and convolved with an array of overlapping critical-band filters with center frequencies equally spaced along the Bark scale. The output of this step was further modified with equal-loudness preemphasis and cubic-root amplitude compression. Finally, cepstral coefficients were determined for a fifth-order all-pole PLP model. A complete description of this process can be found in Hermansky (1990). Hermansky found that a fifth-order model was effective in suppressing speaker-dependent characteristics of the auditory spectrum (i.e., F0) while preserving the gross spectral shape. The predictor variables used in the logistic regression were the fifth-order PLP cepstral coefficients.

III. RESULTS

Figure 1 displays pooled response frequencies for the twelve listeners. Each vowel category is displayed on a separate panel: (a) /ɪ/; (b) /ʊ/; and (c) /ɝ/. Bubble sizes correspond to the response count for each category: the larger the bubble size, the more times a stimulus was chosen as a member of the category indicated on the panel. For the most part, the different vowel categories are contained in different regions of the stimulus space and the response regions correspond well with the production averages reported for these vowel categories by Hillenbrand *et al.* (1995). This indicates that the listeners were fairly consistent in their responses. In the following analyses, the vowel identification performance of individual subjects will be considered.

A. Comparison of vowel representations

Vowel representations were compared for individual listeners via logistic regression analyses using the predictor variables generated for each model. Whenever possible, contrasting regression models were expressed in hierarchical or nested form providing, at least in the context of the logistic representation, a direct likelihood ratio test of the relative

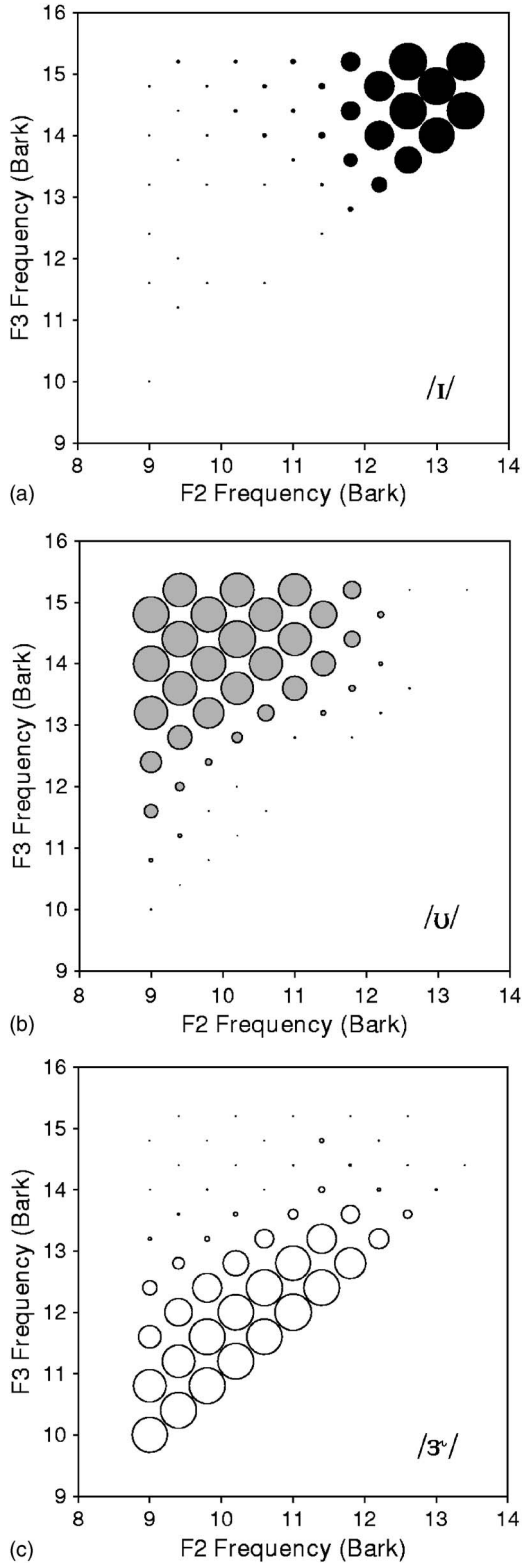


FIG. 1. Bubble plots of pooled vowel identification for all 12 listeners. Each category is displayed on a separate panel: (a) /ɪ/; (b) /ʊ/; (c) /ɜ˞/. Bubble size corresponds to the total number of times a stimulus was identified as a member of that category.

merit of sequential fits. The improvement for nested models using maximum-likelihood methods can be assessed through a likelihood ratio test. The ratio of the −2 log-likelihood (−2LL) of the model of interest to the saturated model is the residual deviance (D) and is distributed asymptotically as chi-square. The difference in residual deviance for nested models can be compared by: $G=D$ [for the model without the variable(s)]$-D$ [for the model with the variable(s)] with degrees of freedom equal to the difference in the number of parameters in the two models (Hosmer and Lemeshow, 1989).

Comparisons of nonhierarchical models are more problematic. Decisions regarding the degree to which one model accounts for the data relative to another must be made with some caution and usually involve considerations beyond a simple statistical test. In order to compare the performance of non-nested models, Akaike's information criterion (AIC) was used to assess goodness-of-fit (Akaike, 1974). For a given model, the AIC is calculated using

$$\mathrm{AIC} = -2 \ln L + 2n, \tag{1}$$

where L is a maximum likelihood estimate and n is the number of free parameters. The first term indicates the fit of the model to the data set while the second term penalizes the use of additional free parameters. When comparing competing non-nested models, the model associated with the smallest value of AIC is chosen as the best fitting model. The AIC also allows for more than two models to be compared simultaneously.

It should be noted that caution must be used in the interpretation of any count data, such as those used in all the analyses reported here, due to the possibility of underestimating the sampling variances and covariances, or "overdispersion." According to McCullagh and Nelder (1989), unless the data meet the expected variance assumptions, the presence of overdispersion should be assumed.

Table II shows the number of model comparisons across listeners for which the models listed by row were determined to provide a superior fit over models listed by column according to the AIC (ties were eliminated). The comparisons for which one model provided a significantly better fit than a competing model according to a sign test ($p<0.05$) are shown in bold; those for which one model provided a significantly worse fit than a competing model ($p<0.05$) are shown in bold italic. The two rightmost columns of Table II summarize the number of times each model in the row performed better or worse than all the other models considered. Although nested formant models are included in this table, specific comparisons among these models are more appropriately made with the likelihood ratio tests described in the next section.

1. Models based on formant frequencies and amplitudes (models a–e)

A number of nested models based on the measured formant frequencies and amplitudes were evaluated. The simplest of these models used the measured frequencies of F2 and F3 alone. Additional models included the measured frequency of F1, relative formant amplitudes, and nonlinear (quadratic) terms.

Table III lists the G statistics and probabilities associated with the likelihood ratio tests comparing models that also include information about F1 and formant amplitudes in addition to F2 and F3 frequencies. Significant differences at

TABLE II. Number of cases (out of 12) for which the model on the row provided a better fit to listeners' data than did the column model according to the AIC. Tied cases are omitted. The bold entries indicate the comparisons for which the difference was significant according to a sign test ($p<0.05$). The right-hand side summarizes the number of times each model in each row performed better or worse than all the other models considered. Models: (a) measured F2 and F3; (b) measured nonlinear; (c) F1, F2, F3; (d) F2, F3, A2, A3; (e) F1, F2, F3, A1, A2, A3; (f) excitation patterns; (g) spectral slope; (h) PLP.

	Models								Summary comparison to other models	
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	Better	Worse
(a)	⋯	***2***	8	5	4	3	8	6	**0**	***1***
(b)	**10**	⋯	**10**	**10**	**10**	8	**12**	**10**	**6**	***0***
(c)	4	***2***	⋯	3	4	***2***	9	5	**0**	***2***
(d)	7	***2***	9	⋯	7	5	9	6	**0**	***1***
(e)	8	***2***	8	5	⋯	3	9	8	**0**	***1***
(f)	9	3	**10**	7	9	⋯	**11**	9	**2**	***0***
(g)	4	***0***	3	3	3	***1***	⋯	***2***	**0**	***3***
(h)	6	***2***	7	6	4	3	**10**	⋯	**1**	***1***

$p<0.05$ (with Bonferroni correction) are indicated in bold. Including the relative amplitudes of F2 and F3, along with their frequencies, produced a significant improvement in fit for seven of the twelve listeners over measured frequencies of F2 and F3 alone (model a vs d). Using a criterion for the underlying probability of a significant difference due to chance for a single comparison of 0.05, the probability of a significant difference due to chance in seven out of twelve comparisons is exceedingly small ($p<0.001$).

Although the nominal F1 frequency used in the synthesis did not vary among the stimuli, the influence of F2 and F3 frequency could be observed in slight variations in measured F1 frequency and amplitude. Including F1 frequency significantly improved the fit for only 2 of the 12 listeners (model a vs c) ($p=0.118$). Likewise, adding both the F1 frequency and amplitude information improved the fit over F2 and F3 frequencies and amplitudes (model d vs e) for only two listeners. These comparisons indicate that there was no additional advantage from the inclusion of F1 information.

Finally, models containing both linear and nonlinear terms for the measured formant frequencies were also tested. Table IV lists the G statistics and probabilities for these comparisons. Once again, significant differences at $p<0.05$ (with Bonferroni correction) are indicated in bold. The addition of the nonlinear (quadratic) terms of F2 and F3 resulted in a

TABLE III. Likelihood ratio tests for formant frequency and amplitude. A brief description of the models tested is indicated within the brackets.

	Model comparison					
	a vs d (F2 and F3 vs F2, F3, A2, and A3)		a vs c (F2 and F3 vs F1, F2, and F3)		d vs e (F2, F3, A2 and A3 vs F1, F2, F3, A1, A2, and A3)	
	$df=8$		$df=4$		$df=4$	
Subject	G	p	G	p	G	p
S1	3.4	0.493	3.6	0.165	10.1	0.039
S2	16.1	**0.003**	0.1	0.951	2.1	0.717
S3	18.4	**0.001**	2.2	0.333	2.2	0.699
S4	52.8	**0.000**	2.6	0.273	25.0	**0.000**
S5	22.2	**0.000**	2.3	0.317	7.6	0.107
S6	2.5	0.645	4.6	0.100	3.0	0.558
S7	38.9	**0.000**	8.1	0.017	15.1	0.004
S8	20.6	**0.000**	0.8	0.670	13.2	0.010
S9	22.5	**0.000**	11.5	**0.003**	5.7	0.223
S10	12.8	0.012	11.5	**0.003**	22.0	**0.000**
S11	8.6	0.072	1.7	0.427	2.4	0.663
S12	9.2	0.056	0.2	0.905	2.4	0.663

TABLE IV. Likelihood ratio tests for linear vs nonlinear formant components. A brief description of the models tested is indicated within the brackets.

	Model comparison	
	a vs b (Measured linear vs nonlinear)	
	$df=6$	
Subject	G	p
S1	17.3	0.008
S2	28.7	**0.000**
S3	16.7	0.010
S4	57.6	**0.000**
S5	34.9	**0.000**
S6	8.9	0.179
S7	38.7	**0.000**
S8	5.0	0.544
S9	47.7	**0.000**
S10	53.1	**0.000**
S11	18.7	0.005
S12	40.1	**0.000**

significant improvement in fit for seven out of twelve listeners for the measured formant frequencies (model a vs b) ($p<0.001$).

2. Models based on whole-spectrum representations (models f–h)

According to AIC (Table II), the six factors extracted from the principal components analysis of the excitation patterns (model f) provided an improved fit to the data for nine listeners when compared to the measured formant frequencies (model f vs a) ($p=0.073$). The excitation pattern model also performed significantly ($p<0.05$) better than the first three measured formant frequencies (model c) and the factors derived from the spectral slope representation (model g).

In general, the ten factors obtained from the spectral slope representation (model g) did not perform well as predictors for the logistic regression. Sign tests indicated that this model performed significantly worse than the nonlinear formant model (b), the excitation pattern model (f), and the PLP model (h) ($p<0.05$). The PLP (model h) did about as well as the measured formant frequencies (model h vs a or c), even when formant amplitude was added (model h vs d or e); however, it produced significantly poorer fits when compared to the nonlinear model ($p<0.05$) (model h vs b) and was surpassed by the excitation pattern model for nine of the twelve listeners ($p=0.073$).

B. Evaluating competing models

The relative success of the competing models in the logistic regression analyses was assessed by how often a model performed significantly better (or worse) than the other models considered and by a cumulative ranking of a model's success.

TABLE V. Models ordered by cumulative ranking of success according to the AIC. The ranking score represents the weighted sums of the models' rank order across all listeners.

Model	Ranking score
(b) Measured Nonlinear	83
(f) Excitation Patterns	71
(d) F2, F3, A2, A3	58
(e) F1, F2, F3, A1, A2, A3	56
(h) PLP	51
(a) F2 and F3	49
(c) F1, F2, F3	42
(g) Spectral slope	29

As shown in Table II, the nonlinear formant model (b) provided significantly better fits ($p<0.05$) to listeners' data than most other models. The nonlinear measured formant model (b) outperformed six of the other seven competing models, excepting the excitation pattern model (f). The excitation pattern model itself outperformed two other models. The nonlinear formant model and the excitation pattern model were the only models evaluated that did not perform significantly worse than any other model. In contrast, several models fared significantly worse than competing models. For instance, the spectral slope model (g) provided a worse fit to listeners' data for a significant number of cases in comparison to three other models.

The cumulative rankings of AIC values allowed assessment of how often a model was ranked among the best- or worst-fitting models across listeners. As a measure of cumulative ranking, representations were rank ordered according to AIC value for each listener. Representations were then assigned scores based on these rankings (e.g., representations ranked first received scores of 8, representations ranked second received scores of 7, etc.). Scores were summed for each representation and are listed in Table V. By this metric, the nonlinear formant model (b) once again performed the best, followed closely by the excitation pattern model (f). The performance of these two models was ranked within the top five positions for eleven of the twelve listeners and was within the top two positions for seven listeners.

IV. DISCUSSION

Competing models of vowel representation can be evaluated based on various criteria such as simplicity (i.e., parameter count minimization), goodness-of-fit to a particular set of data, and biological plausibility. Initially, the formant models appear to be clear winners: they use the fewest number of parameters[2] to provide superior fits to listeners' data. However, they are perhaps too reductionist in their attempts to simplify the description to the detriment of some other considerations, including those noted above. When implementability and generalizability are considered, the preference for these models may have to be tempered.

Models of vowel perception can also be assessed according to their flexibility and robustness in the presence of speaker variability and adverse listening conditions produced by noise, reverberation, and/or hearing impairment. Listeners encounter vowels produced by many different speakers with a range of vocal tract sizes and average F0s; nevertheless, vowel-identification accuracy does not differ substantially across speakers as a result of these variations (Hillenbrand *et al.*, 1995; Klatt, 1982a). Likewise, circumstances that present difficulties for accurate formant extraction do not necessarily lead to poorer performance in human listeners. Assmann and Summerfield (1989), for example, reported that listeners can accurately identify concurrent vowels, even when the individual formants have considerable overlap in the spectrum. For normally hearing listeners, vowel identification accuracy is fairly robust in noise (Nábělek *et al.*, 1992) and with reduced spectral resolution (Dubno and Dorman, 1987; Fu *et al.*, 1998). Additionally, although cochlear hearing impairment results in broadening of auditory filters and loss of spectral contrast, in the absence of background noise, vowel identification performance can remain high for listeners with mild to moderate losses (Nábělek *et al.*, 1992; Owens *et al.*, 1968; Van Tasell *et al.*, 1987). A successful model of vowel perception will need to account for the levels of performance observed in each of these less-than-ideal situations.

Some of the models of vowel representation evaluated here were more successful than others in reflecting the listeners' classifications across the three vowel categories. Logistic regression analyses used to model response probabilities indicated that spectral peak frequencies alone (models c and e) did not provide the best account of listeners' vowel categorization data. As a group, the representations that included only formant frequency information performed poorly. Improvement in the fit to listeners' identification performance was obtained with the addition of either relative formant amplitudes (models d and e) or the quadratic formant terms (model b).

Among all models, the one that included the quadratic formant terms provided the best fit to the response probability data for the majority of listeners (7 out of 12). Previous researchers have emphasized the apparent linearity of vowel category boundaries in F1/F2 space (Carlson *et al.*, 1970; Hose *et al.*, 1983; Karnickaya *et al.*, 1973). In the current study and others (Maddox *et al.*, 2002; Nearey and Kiefte, 2003), although linear boundaries provided an acceptable account of the category identification data, nonlinear boundaries often yielded a significant improvement. This was true even though the boundary locations appear very similar.

There is little argument that formants can be a useful construct for understanding vowel perception (Klatt, 1982a, 1982b). However, the question remains of how crucial explicit formant extraction is to the mechanism of vowel perception in human listeners. For example, formant-based representations have considerable difficulty accounting for the effects of the spectral shape manipulations described by Ito *et al.* (2001). In that study, formant frequencies alone could not predict the categorization of vowels where either first or second formant peaks were locally suppressed but the balance of the amplitude spectrum was unaltered.

The whole-spectrum model based on the factors extracted from excitation patterns (model f) also performed reasonably well according to the logistic regression analyses. This was true despite the gross level of spectral representation provided by the model. Nevertheless, the extracted factors supplied information about formant locations, as well as allowing for nonlinear boundaries between categories. The nonlinear model did not perform significantly better than this model.

The relative success of the excitation pattern model used here and of similar models (Klein *et al.*, 1970; Plomp *et al.*, 1967; Pols *et al.*, 1969) demonstrates that whole-spectrum representations can be distilled into relatively low-dimensional descriptions and effectively account for vowel identification performance. These models rely on implicit formant information rather than explicit formant locations. While models developed using this approach lack the apparent simplicity of formant models, they offer other potential benefits in terms of ease of implementation, information content, and generalizability (Pols, 1977).

The stimuli used in this study were based on a limited set of vowels with variation restricted to F2 and F3. They were developed based on earlier investigations that tested hypotheses concerning the critical limit of formant integration (Molis *et al.*, 1998). That research indicated that some of the boundaries among these three vowel categories are described in terms of interactions between the second and third formants. Therefore, by restricting the stimuli to variation in F2 and F3, the stimulus set provided continuous and relatively dense sampling across the three vowel categories examined here and allowed further exploration of these interactions.

However, any practicable model of vowel perception will need to address a much wider vowel space that also incorporates variation in F1. In the context of the entire American English vowel inventory, the pattern of results could change dramatically if models were assessed according to variation in the F1/F2 plane. For example, variation in frequency selectivity across the frequency range usually results in resolution of individual harmonics within formants in the F1 region, whereas the unresolved harmonics in the F2–F3 frequency regions tend to produce peaks in the excitation patterns that correspond more closely to the formant resonances only. The frequency of the first formant must usually be inferred from a number of adjacent harmonics, leading to some uncertainty as to where the formant peak is actually located. Moreover, if F1 is not a multiple of F0, the resolved harmonic with maximum amplitude in the excitation pattern may not correspond with the formant frequency. Holding both F0 and F1 constant in this set of vowel stimuli bypassed the ambiguity potentially introduced by formant-harmonic interactions.

Selecting a different set of vowel stimuli, with variation in F1 as well as F2 and F3 would certainly have produced different factors extracted from the factor analysis of the excitation patterns. Plomp and Pols (Plomp *et al.*, 1967; Pols *et al.*, 1969) included a normal range of F1 variation in their stimuli, demonstrating that meaningful factors can be extracted from a broad, naturally produced vowel inventory.

However, by restricting the stimuli here to variation only on F2 and F3, the factors emerging from the principal-components analysis would be largely immune to extraneous variability associated with the formant-harmonic interactions.

The models tested here do not exhaust the possibilities. Other promising models, such as the missing data model of de Cheveigné and Kawahara (1999) and the narrow band pattern matching model of Hillenbrand and Houde (2003), propose that vowels may be recognized directly from an unsmoothed harmonic spectrum. These models do not fall easily into the categories of "formant-based" or "whole-spectrum" representations investigated here, but instead may include valuable characteristics associated with each type. Models developed along these lines have the potential to avoid some of the pitfalls of the more restricted representations, while exploiting other important cues to vowel perception. Although this study focused on purely spectral representations and excluded other properties important to vowel identity such as intrinsic duration and fundamental frequency differences, and formant dynamics, these additional attributes undoubtedly come into play in the perception of natural speech.

V. CONCLUSION

In this study, normal-hearing listeners identified synthetic vowel stimuli that varied in F2 and F3 in equal steps along the Bark scale. Through logistic regression analysis, a number of formant peak and whole-spectrum vowel representations were evaluated for their ability to predict listeners' identification response patterns. Based on the results of these analyses, it is possible to eliminate some representations from future consideration while revealing the relative benefits of the remaining representations.

Formant frequencies were the most successful model of the listeners' responses provided that nonlinear category boundaries could be represented. However, considering that a simplified excitation pattern model could perform almost as well as the most successful formant models, the whole-spectrum approach should not be abandoned in favor of the descriptive advantage provided by formant frequency models; rather, the use of each approach should be matched to the relative strengths that each offers.

ACKNOWLEDGMENTS

This research was conducted as part of a doctoral dissertation at the University of Texas at Austin. The work was supported by NIH (R01 DC00427-13,-14). The author wishes to thank Randy L. Diehl, Marjorie R. Leek, James M. Hillenbrand, and one unnamed reviewer for comments on an earlier draft. The opinions or assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

[1]Evaluations of the measured formant frequencies and the nominal formant values used in the synthesis indicated a considerable similarity between the two and no predictive advantage for either. Therefore, the measured formant frequencies were used in all the formant-based model comparisons reported here.

[2]In the case of the nonlinear formant models, there are only two measured parameters—the additional three parameters are derived from the first two.

Akaike, H. (**1974**). "A new look at the statistical model identification," IEEE Trans. Autom. Control **19**, 716–723.

Assmann, P. F., and Summerfield, Q. (**1989**). "Modeling the perception of concurrent vowels: Vowels with the same fundamental frequency," J. Acoust. Soc. Am. **85**, 327–338.

Bladon, R. A. W. (**1982**). "Arguments against formants in the auditory representation of speech," in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granström (Elsevier Biomedical Press, Amsterdam), pp. 95–102.

Bladon, R. A. W. (**1983**). "Two-formant models of vowel perception: Shortcomings and enhancements," Speech Commun. **2**, 305–313.

Bladon, R. A. W., and Lindblom, B. (**1981**). "Modeling the judgment of vowel quality differences," J. Acoust. Soc. Am. **69**, 1414–1422.

Broad, D. J., and Clermont, F. (**1989**). "Formant estimation by linear transformation of the LPC cepstrum," J. Acoust. Soc. Am. **86**, 2013–2017.

Carlson, R., and Granström, B. (**1979**). "Model predictions of vowel dissimilarity," *Speech Transmission Laboratory-Quarterly Progress and Status Report (STL-QPSR)* **3–4**, 84–104.

Carlson, R., Granström, B., and Fant, G. (1970). "Some studies concerning perception of isolated vowels," *STL-QPSR* **2–3**, 19–35.

Chistovich, L. A., and Lublinskaya, V. V. (**1979**). "The 'center of gravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli," Hear. Res. **1**, 185–195.

Chistovich, L. A., Sheikin, R. L., and Lublinskaya, V. V. (**1979**). " 'Centers of gravity' and spectral peaks as the determinants of vowel quality," in *Frontiers of Speech Communication Research*, edited by B. Lindblom and S. Ohman (Academic, London), pp. 143–157.

de Cheveigné, A., and Kawahara, H. (**1999**). "A missing data model of vowel identification," J. Acoust. Soc. Am. **105**, 3497–3508.

Dubno, J. R., and Dorman, M. F. (**1987**). "Effects of spectral flattening on vowel identification," J. Acoust. Soc. Am. **82**, 1503–1511.

Fant, G. (**1960**). *The Acoustic Theory of Speech Perception* (Mouton, The Hague).

Fu, Q.-J., Shannon, R. V., and Wang, X. (**1998**). "Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing," J. Acoust. Soc. Am. **104**, 3586–3596.

Hermansky, H. (**1990**). "Perceptual linear predictive (PLP) analysis of speech," J. Acoust. Soc. Am. **87**, 1738–1752.

Hillenbrand, J., and Gayvert, R. T. (**1993**). "Vowel classification based on fundamental frequency and formant frequencies," J. Speech Hear. Res. **36**, 694–700.

Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). "Acoustic characteristics of American English vowels," J. Acoust. Soc. Am. **97**, 3099–3111.

Hillenbrand, J., and Houde, R. A. (**1995**). "Vowel recognition: Formants, spectral peaks and spectral shape representations," J. Acoust. Soc. Am. **98**, 2949.

Hillenbrand, J. M., and Houde, R. A. (**2003**). "A narrow band pattern-matching model of vowel perception," J. Acoust. Soc. Am. **113**, 1044–1055.

Hose, B., Langer, G., and Scheich, H. (**1983**). "Linear phoneme boundaries for German synthetic two-formant vowels," Hear. Res. **9**, 13–25.

Hosmer, D. W., and Lemeshow, S. (**1989**). *Applied Logistic Regression* (Wiley, New York).

Ito, M., Tsuchida, J., and Yano, M. (**2001**). "On the effectiveness of whole spectral shape for vowel perception," J. Acoust. Soc. Am. **110**, 1141–1149.

Karnickaya, E. G., Mushnikov, V. N., Slepokurova, N. A., and Zhukov, S. J. (**1973**). "Auditory processing of steady-state vowels," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, New York), pp. 37–53.

Kiefte, M., and Kluender, K. R. (**2001**). "Spectral tilt versus formant frequency in static and dynamic vowels," J. Acoust. Soc. Am. **109**, 2294–2295.

Klatt, D. H. (**1982a**). "Speech processing strategies based on auditory models," in *The Representation of Speech in the Peripheral Auditory System*, edited by R. Carlson and B. Granström (Elsevier Biomedical, Amster-

dam), pp. 181–196.
Klatt, D. H. (**1982b**). "Prediction of perceived phonetic distance from critical-band spectra: A first step," *IEEE, ICASSP*, 1278–1281.
Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.
Klein, W., Plomp, R., and Pols, L. C. W. (**1970**). "Vowel spectra, vowel spaces and vowel identification," J. Acoust. Soc. Am. **48**, 999–1009.
Maddox, W. T., Molis, M. R., and Diehl, R. L. (**2002**). "Generalizing a neuropsychological model of visual categorization to auditory categorization of vowels," Percept. Psychophys. **64**, 584–597.
McCullagh, P., and Nelder, J. A. (**1989**). *Generalized Linear Models*, 2nd ed. (Chapman & Hall, London).
Molis, M., Diehl, R. L., and Jacks, A. (**1998**). "Phonological boundaries and the spectral center of gravity," J. Acoust. Soc. Am. **103**, 2981.
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.
Nábělek, A. K., Zbigniew, C., and Krishnan, L. A. (**1992**). "The influence of talker differences on vowel identification by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **92**, 1228–1246.
Nearey, T. M., and Kiefte, M. (**2003**). "Comparison of several proposed perceptual representations of vowel spectra," in Proceedings of the XVth International Congress of Phonetic Sciences, pp. 1005–1008.
Owens, E., Talbott, C., and Schubert, E. (**1968**). "Vowel discrimination of hearing impaired listeners," J. Speech Hear. Res. **11**, 648–655.
Peterson, G. E., and Barney, H. L. (**1952**). "Control methods used in a study of vowels," J. Acoust. Soc. Am. **24**, 175–184.
Plomp, R., Pols, L. C. W., and Van der Geer, J. P. (**1967**). "Dimensional analysis of vowel spectra," J. Acoust. Soc. Am. **41**, 707–712.
Pols, L. C. W. (**1970**). "Perceptual space of vowel-like sounds and its correlation with frequency spectrum," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 463–470.
Pols, L. C. W. (**1977**). *Spectral Analysis and Identification of Dutch Vowels in Monosyllabic Words* (Institute for Perception TNO, Soeterberg, The Netherlands).
Pols, L. C. W., van der Kamp, L. J. T., and Plomp, R. (**1969**). "Perceptual and physical space of vowels sounds," J. Acoust. Soc. Am. **46**, 457–467.
Rosner, B. S., and Pickering, J. B. (**1994**). *Vowel Perception and Production* (Oxford University Press, New York).
Van Tasell, D. J., Fabry, D. A., and Thibodeau, L. M. (**1987**). "Vowel identification and vowel masking patterns of hearing-impaired subjects," J. Acoust. Soc. Am. **81**, 1586–1597.
Zahorian, S. A., and Jagharghi, A. J. (**1993**). "Spectral-shape features versus formants as acoustic correlates for vowels," J. Acoust. Soc. Am. **94**, 1966–1982.

Age-related differences in weighting and masking of two cues to word-final stop voicing in noise[a)]

Susan Nittrouer[b)]
Utah State University, UMC 6840, Logan, Utah 84322-6840

(Received 30 August 2004; revised 2 May 2005; accepted 4 May 2005)

Because laboratory studies are conducted in optimal listening conditions, often with highly stylized stimuli that attenuate or eliminate some naturally occurring cues, results may have constrained applicability to the "real world." Such studies show that English-speaking adults weight vocalic duration greatly and formant offsets slightly in voicing decisions for word-final obstruents. Using more natural stimuli, Nittrouer [J. Acoust. Soc. Am. **115**, 1777–1790 (2004)] found different results, raising questions about what would happen if experimental conditions were even more like the real world. In this study noise was used to simulate the real world. Edited natural words with voiced and voiceless final stops were presented in quiet and noise to adults and children (4 to 8 years) for labeling. Hypotheses tested were (1) Adults (and perhaps older children) would weight vocalic duration more in noise than in quiet; (2) Previously reported age-related differences in cue weighting might not be found in this real-world simulation; and (3) Children would experience greater masking than adults. Results showed: (1) no increase for any age listeners in the weighting of vocalic duration in noise; (2) age-related differences in the weighting of cues in both quiet and noise; and (3) masking effects for all listeners, but more so for children than adults. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1940508]

PACS number(s): 43.71.Ft, 43.71.An [ALF] Pages: 1072–1088

I. INTRODUCTION

The way speech is perceived depends upon the native language of the listener. For example, native Japanese speakers fail to use the third formant ($F3$) transition when deciding whether a word initial liquid is /ɹ/ or /l/ (MacKain, Best, and Strange, 1981; Miyawaki *et al.*, 1975). Native English speakers, on the other hand, show /ɹ/–/l/ labeling responses that reveal a strong dependence on the $F3$ transition. This language-related difference in attention paid (or weight assigned) to the $F3$ transition is observed even though the same sets of English- and Japanese-speaking listeners (who provided the labeling results) have been found to discriminate nonspeech spectral glides in the region of $F3$ equally well (Miyawaki *et al.*, 1975).

Cross-linguistic differences have also been described for decisions regarding the voicing of word-final stops. When a speaker produces words that differ phonetically only in the voicing of a final stop (e.g., cap/cab, wait/wade, and duck/dug), differences in articulation occur throughout the syllable: The jaw lowers faster and farther in syllables with voiceless, rather than voiced, final stops (Gracco, 1994; Summers, 1987). The jaw remains open (Summers, 1987) and the tongue retains its vowel-related posture longer in syllables with voiced final stops (Raphael, 1975). And, the relative timing of the offset of laryngeal vibration and of vocal-tract closure differs across words depending on voicing of final stops. For words with voiceless final stops, laryngeal vibration is halted before vocal-tract closure is achieved. For words with voiced final stops, laryngeal vibration continues into the closure until sub- and supraglottal pressures are equalized. All these articulatory differences create numerous acoustic differences between words with voiced and voiceless final stops: Words with voiced final stops have longer vocalic segments than words with voiceless final stops. Formant transitions at the ends of the vocalic portions differ depending on the voicing of the final stop; in particular, the first formant ($F1$) is generally higher at voicing offset when the final stop is voiceless, rather than voiced. Words with voiced final stops have voicing present during the closure; words with voiceless final stops do not. The frequency of ($F1$) at syllable center tends to be higher for words with voiceless final stops than for words with voiced final stops. But, of all these acoustic differences between words with voiced and voiceless final stops, the perceptual influence on adults' voicing decisions of the duration of the vocalic syllable portion and of syllable-final formant transitions (particularly $F1$) have been most studied (e.g., Crowther and Mann, 1992; 1994; Denes, 1955; Fischer and Ohde, 1990; Flege and Wang, 1989; Hillenbrand *et al.*, 1984; Raphael, 1972; Raphael, Dorman, and Liberman, 1980; Wardrip-Fruin, 1982). Collectively these studies have shown that adult speakers of all languages examined weight $F1$ transitions at voicing offset similarly. However, listeners differ in the extent to which they weight the duration of the vocalic portion depending on native language background. Speakers of languages that permit obstruents in syllable-final position and that demonstrate a vocalic-length distinction based on the voicing of those final obstruents, such as English, weight vocalic length more strongly than speakers of languages that either do not permit syllable-final obstruents, such as Mandarin (Flege and Wang, 1989), or that do not demonstrate a vocalic-length distinction based on the voicing

[a)]Portions of this work presented at the 145th meeting of the Acoustical Society of America, Nashville, April–May, 2003.
[b)]Electronic mail: nittrouer@cpd2.usu.edu

 0001-4966/2005/118(2)/1072/17/$22.50

of those final obstruents, such as Arabic (Crowther and Mann, 1992; 1994; Flege and Port, 1981). Such cross-linguistic results suggest that some learning must be involved for speakers of specific languages to know which properties of the signal demand our perceptual attention, and which may largely be ignored.

Studies comparing labeling results for children and adults support that suggestion. Young children do not weight properties of the speech signal as adults do who share their native language. A characterization of these age-related differences is that children tend to prefer the dynamic resonances arising from the continuously changing cavities of the vocal tract over other acoustic properties, such as static, aperiodic noises, and durational differences.[1] Adults, on the other hand, seem to know when a nondynamic property can come in handy in making a phonetic decision in their native language. Several lines of investigation bolster this characterization of developmental changes in perceptual weighting strategies for speech. Studies investigating the labeling of fricatives have shown that children rely more on the voiced formant transitions in the vicinity of those fricatives than adults do, but rely less on the fricative noises themselves (Mayo *et al.*, 2003; Nittrouer, 1992; Nittrouer *et al.* 2000; Siren and Wilcox, 1995). Studies investigating voicing decisions for syllable-final obstruents have shown that children (3 to 6 years of age) rely more on voiced formant transitions preceding vocal-tract closure and less on the length of the vocalic portion than adults (Greenlee, 1980; Krause, 1982; Nittrouer, 2004; Wardrip-Fruin and Peach, 1984), although these strategies may begin to take on the characteristics of adults in the native language community by 5 years of age (Jones, 2003).

The idea that children initially focus on dynamic signal components in perception parallels suggestions that children first master global vocal-tract movements in their productions. Dynamic signal components arise from global movements of upper vocal-tract articulators. There is evidence from investigators such as MacNeilage and Davis (e.g., 1991) to suggest that articulators operate as a common coordinative structure in children's early speech production: that is, articulators work in synchrony, largely following jaw action. This suggestion is perfectly consistent with more general ideas concerning the development of movement control. For example, the work of Thelen and colleagues on the development of leg movements shows that initially these movements are "global and inflexible," but gradually "…limb segments become both disassociated from these global synergies and reintegrated into more complex coalitions." (from abstract of Thelen, 1985.) Similarly, children gradually acquire the ability to organize movements of isolated regions of the vocal tract in order to make refined constriction shapes, with precise timing patterns. However, in the case of speech, these precise patterns are language specific. Consequently, their acquisition is likely shaped by the child's emerging attention to details of the speech signals produced by others. It appears that burgeoning abilities both to produce more precise articulatory gestures and to attend to details of the speech signal develop in lock step, with each facilitating the other.

But, not all experiments examining developmental changes in perceptual strategies for speech have found differences between children and adults in their perceptual weighting of dynamic and nondynamic signal components. In some cases, this is as expected. For example, adults and children alike weight formant transitions greatly and fricative noises hardly at all in place decisions for /f/ versus /θ/ (Harris, 1958; Nittrouer, 2002). This result is not surprising because /f/ and /θ/ noises are spectrally indistinguishable (Nittrouer, 2002). In other experiments, results are difficult to interpret. For example, Mayo and Turk (2004) reported that children weighted formant transitions less and acoustic voice onset time more than adults in voicing decisions for syllable-initial stops. However, it is not clear why these investigators would ever have expected formant transitions to influence voicing decisions for syllable-initial stops much, if at all, given that voiced and voiceless initial stops that share the same place of constriction share the same formant trajectories. The only difference is that larger portions of those trajectories are excited by aspiration noise rather than by a voiced source in words with voiceless initial stops. In Mayo and Turk, the stimuli did not replicate natural tokens in that they contained no portion of aspirated formant transitions. Instead, vocalic segments of identical length were constructed with different onset frequencies for the formants and placed at different temporal distances from a preceding burst noise. This design meant that there were significant silent gaps separating the two syllable portions. Results from Murphy, Shea, and Aslin (1989) showed that children are incapable of integrating acoustic segments in speech stimuli that are separated by silent gaps of several tens of milliseconds. Consequently, the children in Mayo and Turk's study were likely unable to integrate the initial release burst with the following vocalic segment for stimuli with silent gaps longer than their integration thresholds, and so may have been basing decisions on whether they heard one segment or two. Mayo and Turk's stimulus design also meant that syllable duration was perfectly confounded with voice onset time, making it impossible to know whether responses of listeners of any age were due to changes in voice onset time or to changes in overall syllable duration.

In another study, Sussman (2001) investigated vowel perception by adults and 4-to-5-year-olds developing language normally. Stimuli were synthetic /bib/ and /bæb/, with 40-ms transitions on either side of 280-ms steady-state formants. In one condition, 220-ms sections of the steady-state vocalic portions were inserted between transitions for the incongruent vowel. When asked to label the vowel in this condition, all listeners responded with the label associated with the 220-ms steady-state section. From this result, Sussman concluded that listeners of all ages weight steady-state formants most strongly in vowel recognition. However, there is another obvious explanation: The steady-state stimulus sections so overwhelmed the dynamic regions of the syllables that it is little wonder that listeners based their decisions on those steady-state sections. In sum, there remains no strong evidence contradicting the suggestion that as children gain experience with a native language they modify the relative amounts of perceptual attention paid to various signal

properties. Besides, it must be the case *a fortiori* that children's perceptual strategies for speech change through childhood because adults have different perceptual weighting strategies depending on their native language.

In particular, young listeners seem to prefer dynamic resonances of the speech signal, and then learn what additional properties of their native language can help in phonetic decisions. This suggestion makes sense in light of experiments showing that dynamic components play a central role in speech perception, even for adults. In laboratory experiments, investigators have traditionally crafted single-syllable stimuli with great attention to nondynamic signal components, such as aperiodic noises and length distinctions. These stimuli tend to have long steady-state vocalic segments. Rarely do experimental stimuli have more than one region of spectral change (i.e., formant transitions). But, natural speech is intrinsically dynamic. Vocal-tract resonances are constantly changing, rarely, if ever, exhibiting regions of stable formants as long as 220 ms. Relatively recently in the history of speech research, stimulus generation techniques, such as sine wave speech, have been developed to capture the continuously changing nature of these resonances while eliminating other signal attributes. Results of experiments using these stimuli demonstrate that the dynamic resonant patterns by themselves can support accurate speech recognition for adults listening to their native language (e.g., Remez *et al.*, 1981). In turn, this kind of finding suggests that dynamic resonances may be viewed as the "backbone" of the speech signal, so to speak, providing the listener with necessary and almost sufficient information for speech perception. From this perspective it makes sense that dynamic signal components would be what children focus on first.

Of course, there are challenges to the suggestion that speech perception can be accomplished with only dynamic resonances. For one, some experience is generally required for listeners to be able to interpret time-varying sinusoids as phonetically relevant. Even with experience, the perceptual organization required to hear these signals as phonetically coherent forms remains remarkably fragile. Listeners can easily be provoked into abandoning the perceptual posture needed to hear the signals as indivisible structures, instead segregating an individual component from the spectral whole (Remez *et al.*, 2001). In addition, the fact is that speakers do not produce signals that provide information only about global changes in vocal-tract shapes. Instead, speakers go to the trouble of fashioning precise constrictions and carefully timed syllables. Presumably these behaviors serve a purpose, or else they would not have been selected through evolution. In fact, more than 50 years of traditional research into speech perception has shown us that nondynamic components of the speech signal, such as release bursts, fricative noises, and length differences, can affect phonetic perception for adult listeners. Finally, there is no evidence that listeners can understand impoverished signals such as time-varying sinusoids in the noisy listening conditions that generally exist in the real world. Thus, it may be that properties of the signal other than dynamic resonances provide phonetic information more immune to degradation by natural listening environments. This notion is generally referred to under the general heading of "speech redundancy," and has been the prevailing account of why there are several different "cues" to any one phonetic decision. For example, Edwards (1981) wrote of redundancy, "By integrating information from many acoustic cues… the perceptual mechanism is able to accommodate a large variety of source and channel variations." (p. 535). Assman and Summerfield (2004) wrote, "Speech is a highly efficient and robust medium for conveying information because it combines strategic forms of redundancy to minimize loss of information." (p. 231). In particular, it has been suggested that these other properties might help the listener when listening to speech in noisy backgrounds (e.g., Coker and Umeda, 1976).

In this study the roles of vocalic duration and formant offsets in voicing decisions for syllable-final stops were examined when words were presented in noise. Noise was used as a contextual variable that might influence the relative amounts of attention given to various acoustic properties (or cues) precisely because it is the most commonly offered natural condition in which speech redundancy is thought to provide an advantage: if one property is masked, listeners can use a different one. Stimuli varying in the voicing of syllable-final stops provided a particularly appropriate way of exploring the possibility that children might gradually increase the weight given to signal properties other than dynamic resonances because other properties might be more immune to degradation in natural conditions, such as in noisy backgrounds. Several studies have reported that children (3 to 6 years of age) from American English backgrounds fail to weight vocalic duration as strongly as adults from the same language background when making voicing decisions about final stops (Greenlee, 1980; Krause, 1982; Lehman and Sharf, 1989; Wardrip-Fruin and Peach, 1984). Three of these studies also reported that children weight syllable-final formant transitions more than adults (Greenlee, 1980; Krause, 1982; Wardrip-Fruin and Peach, 1984). An earlier study from this laboratory extended those results, showing that children (ages 6 and 8 years) weighted formant transitions more and vocalic duration less than adults when synthetic stimuli were used (Nittrouer, 2004). However, that study also found that adults weighted syllable-final formant transitions more than reported by earlier investigations and more similarly to children when stimuli were created by editing natural tokens; that is, by reiterating or deleting pitch periods in the steady-state vocalic portion of words ending in voiceless and voiced stops, respectively. Figure 1 illustrates findings for adults and 6-year-olds for synthetic and edited-natural *buck/bug* stimuli. In this figure, vocalic duration changes in a continuous fashion. Steps along this continuum are represented on the *x* axis from shortest to longest. Separate functions are plotted for stimuli with offset transitions appropriate for either voiced (filled symbols) or voiceless (open symbols) final stops. In this case the separation between functions is an index of the weight assigned to offset transitions.[2] Slope (i.e., change in units on the *y* axis per unit of change on the *x* axis) is an index of the weight assigned to vocalic duration. The functions on the left, for synthetic stimuli, are fairly close together and steep, although they are more separated and shallower for 6-year-olds than for adults.

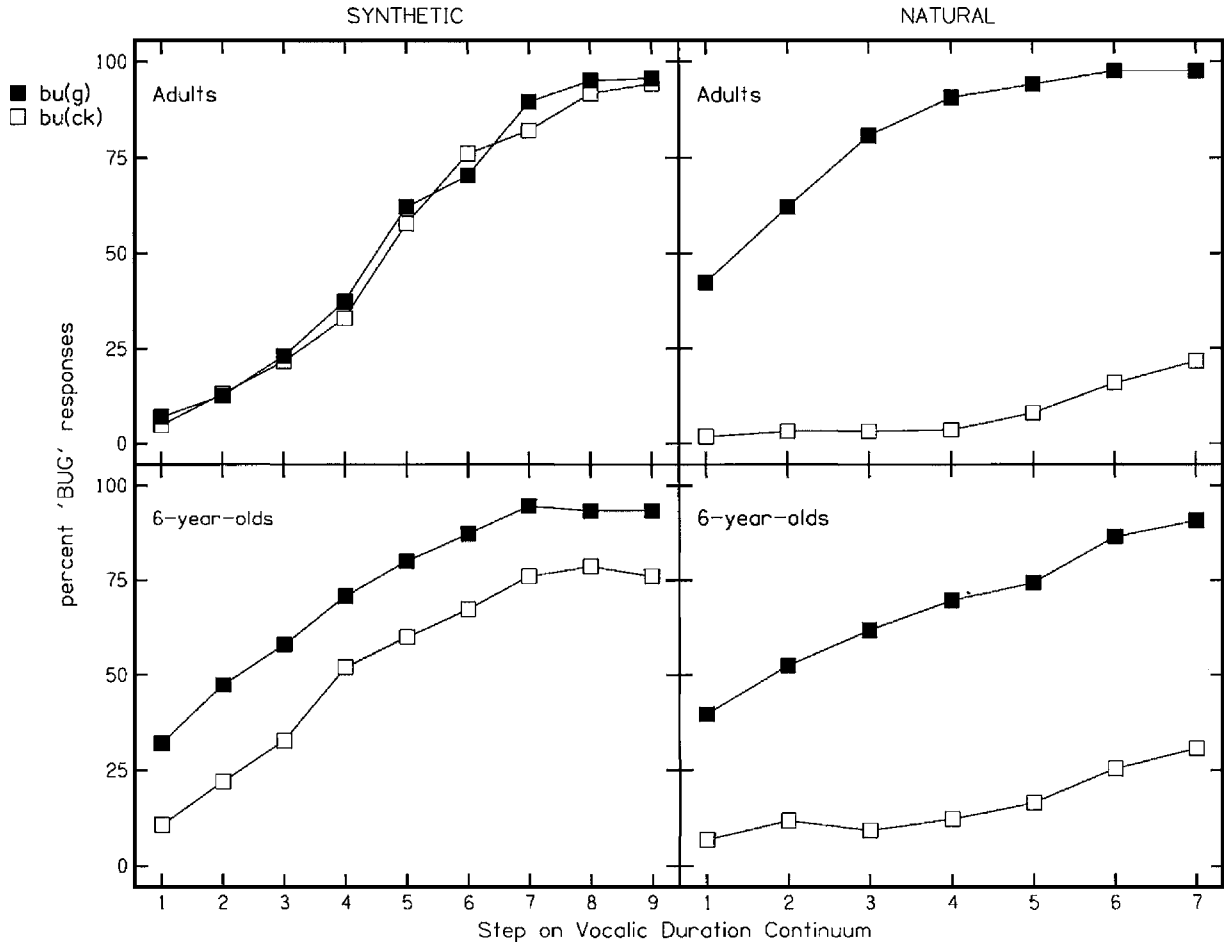


FIG. 1. Labeling functions for *buck/bug* stimuli, synthetic and edited-natural, for adults and 6-year-olds, adapted from Nittrouer (2004).

This pattern indicates that listeners weighted vocalic duration greatly and offset transitions much less so. The functions on the right, for edited-natural stimuli, are more widely separated and shallower, particularly for functions obtained from originally voiceless final stops. The greater separation between functions indicates that both adults and children increased the weight assigned to formant offset transitions when natural stimuli were used instead of synthetic stimuli: Mean separation between functions changed (from synthetic to natural stimuli) from 0.24 to 7.89 steps on the vocalic duration continuum for adults and from 2.80 to 7.27 for 6-year-olds. The age-related difference was statistically significant for synthetic stimuli (0.24 versus 2.80), but not for natural stimuli (7.89 versus 7.27). The shallower functions for natural stimuli indicate that listeners decreased the weight they assigned to vocalic duration when listening to these stimuli instead of the synthetic ones: Mean slopes (across both functions) changed from 0.53 to 0.48 for adults (from synthetic to natural stimuli) and from 0.37 to 0.28 for 6-year-olds. In summary, listeners of all ages (but more so adults than children) increased the weight assigned to offset transitions and decreased the weight they assigned to vocalic duration when edited-natural stimuli were used instead of synthetic stimuli. Working backwards, these results suggest that traditional stimulus synthesis might have caused adults in earlier studies to decrease the weight they normally assign to offset transitions from what they normally do, likely because only $F1$ at syllable offset differed across the ostensibly voiced and voiceless stimuli. Although this effect may have led to erroneous conclusions about what listeners do in the real world, it might also be a demonstration of just how well adults are able to use signal redundancy.

A review of the literature on the perception of voicing for syllable-final stops shows that most studies commonly cited to support the notion that adults base voicing judgments largely on vocalic length have used either synthetic stimuli (e.g., Crowther and Mann, 1992; 1994; Denes, 1955; Fischer and Ohde, 1990; Raphael, 1972) or edited-natural stimuli created only from originally voiced final stops (e.g., Hillenbrand *et al.*, 1984; Hogan and Rozsypal, 1980). In fact, Hogan and Rozsypal explicitly stated that pilot work failed to evoke voiced judgments from adult listeners when stimuli were created by lengthening the steady-state portion of syllables with originally voiceless final stops, so they did not use those syllables in stimulus generation. Findings from Nittrouer (2004) reflect their observation: Figure 1 shows that adults' labeling function for stimuli created from natural, originally voiceless final stops (open symbols in the right-hand plot) is almost completely flat with few "voiced" judgments, even at longer vocalic durations. This pattern contrasts with what is seen for responses to synthetic stimuli with offset transitions appropriate for voiceless final stops (open symbols in the left-hand plot): that function is much steeper, with voiced judgments at longer vocalic durations.

The change in adults' functions across stimulus sets for stimuli with voiced offset transitions (filled symbols) was not as remarkable. In particular, functions were similarly steep for the edited-natural and the synthetic stimuli. Although mean slopes of 0.53 and 0.48 were obtained for synthetic and natural stimuli, respectively, by averaging adults' responses across stimuli with voiced and voiceless offset transitions, a different picture is obtained by looking at functions for voiced and voiceless offsets separately. Mean slopes for adults were the following: synthetic, voiced offset transitions=0.55; natural, voiced offset transitions=0.58; synthetic, voiceless offset transitions=0.51; and natural, voiceless offset transitions=0.37. Thus, all of the change in mean slopes from the synthetic to the natural stimulus conditions was due to functions for stimuli with voiceless offset transitions being much shallower when edited-natural, rather than synthetic, stimuli were used. Nittrouer concluded that this finding (along with the dramatic increase in weight assigned to offset transitions in the natural, compared to the synthetic, condition) mandates modification in our commonly and collectively held assumptions about the roles of vocalic duration and offset transitions in voicing decisions for final stops: Even adult native speakers of languages with a vocalic-length distinction weight formant transitions strongly, at least when words are naturally produced and heard in quiet.

At the same time, the finding gleaned by looking across studies, that adults greatly attenuate the weight they assign to offset transitions when synthetic stimuli are used, served as the impetus for the current study. Perhaps, the thinking went, there are natural conditions in which formant-offset cues are attenuated, as they have been in the synthetic stimuli of earlier experiments. Under these circumstances it would benefit listeners to shift their perceptual attention to vocalic duration. Accordingly, the current study was designed to examine whether adults (and perhaps older children) would show increased weighting of vocalic duration in voicing decisions for word-final stops when listening in a naturalistic condition that might mask formant offsets.

This study also permitted the examination of other possibilities. For one, it is possible that the whole idea that age-related differences in weighting strategies for speech exist might be an artifact, so to speak, of laboratory methods. Perhaps it is only because adults are so skilled at adjusting their listening strategies to fit the conditions that differences between children and adults have ever been found in laboratory studies. When some cues are artificially constrained, perhaps adults are able to turn their attention to other cues, but children cannot. In the real world, where the availability of cues might differ from that of the laboratory, perhaps adults and children use the same strategies. By replicating one condition in the real world this possibility could be tested. And so, another purpose of the current study was to see if age-related differences in weighting strategies exist in natural listening conditions. Although the hypothesis that adults and children would weight acoustic properties similarly in natural conditions (which in this case meant noisy conditions) conflicts with the hypothesis that adults, but not children, would use cue redundancy to make voicing judgments in noise, both hypotheses could be tested by the experimental design.

There is a solid basis for suggesting that noise might mask offset transitions more than it masks vocalic duration. Transitions tend to be lower in amplitude than syllable nuclei because they occur at syllable margins: in this case, when the vocal tract is closing, and so amplitude is falling. Differences in the durations of syllable nuclei primarily account for differences in vocalic duration between voicing conditions. Consequently, voicing-related differences in vocalic duration should remain salient even in noisy conditions. Mature listeners might benefit from paying particular attention to this cue in noise. Accordingly, children would need to learn to attend perceptually to this cue, a skill that native listeners of languages that either do not have syllable-final obstruents or that do not differentiate vocalic duration based on the voicing of those obstruents apparently never acquire.

To test this idea, stimuli differing in vocalic duration created by editing natural words with voiced or voiceless final stops were presented in quiet and in noise to adults and children for labeling. For the quiet condition, listeners of all ages were expected to show labeling functions similar to those for edited-natural stimuli in Fig. 1: that is, functions were expected to be widely separated and shallow (at least those for stimuli created from words with voiceless final stops). If adults (and older children) showed a perceptual weighting shift when words were presented in noise, labeling functions would be expected to resemble those for synthetic stimuli in Fig. 1. That is, they would be less separated and both would be steep. This pattern would indicate that perceptual attention shifted away from offset transitions, and towards vocalic duration.

Two word pairs were used that differed in the frequency of *F*1 at syllable center: *boot/booed* and *cop/cob*. The reason for this was that the extent of the *F*1 transition near voicing offset might affect how robust the *F*1-transition cue is to masking. The frequency of *F*1 at voicing offset is lower for voiced than for voiceless final stops, but this voicing-related difference is greater for words with high medial *F*1 frequencies. Words with low medial *F*1 frequencies fail to show much of a difference in final *F*1 frequency because *F*1 is low throughout the syllable. Consequently, it may be that words with lower medial *F*1 frequencies (such as *boot/booed*) might be more affected by noise masking than words with higher medial *F*1 frequencies (such as *cop/cob*). For these words with low medial *F*1 frequencies listeners, especially adults, might show more of a weighting shift for vocalic duration.

Care was given to the decision of what kind of noise to use as a masker. In general, studies of speech perception in noise have used either speech babble or speech-shaped noise. The reason is that often the speech of others masks the speech signal of interest in natural environments. However, environmental noises (e.g., air-handling equipment, computers, fax machines, traffic, wind, etc.) can mask speech, as well, and these environmental noises have flatter spectra. Therefore, the decision was made to use flat-spectrum noise, which should replicate the combined effects of speech and other environmental maskers.

The decision regarding which signal-to-noise ratio(s) (SNRs) to use when presenting words in noise was also carefully made. Children generally recognize speech less accurately than adults when speech is presented at the same SNRs to both groups (Nittrouer and Boothroyd, 1990). However, this result was found for speech-shaped noise maskers. There was no way of knowing how children would perform with a flat-spectrum masker before this experiment was undertaken, but one report suggested that adults could be expected to perform more poorly with a flat-spectrum than with a speech-shaped masker (Kuzniarz, 1968). Ideally, the SNR selected for each listener would provide the same amount of masking across listeners, indicated by similar overall speech recognition scores across age groups. Initially, the belief was that SNR would likely need to be adjusted among age groups to provide the same amount of masking. Pilot testing, however, showed that recognition scores for consonant–vowel–consonant words were similar for listeners of different ages at a variety of SNRs. Consequently, the decision was made to present words ending with voiced and voiceless stops at one SNR (in addition to quiet) to all listeners: 0 dB, which resulted in roughly 45%–50% correct recognition for listeners of all ages. In addition, adults heard the stimuli at one poorer SNR: −3 dB. General speech recognition scores for CVC words were obtained from all listeners participating in the labeling experiment, even though pilot testing showed similar results for listeners of all ages, just to document that these specific listeners showed similar recognition scores at each SNR. Finally, recognition scores in quiet were also obtained to ensure that recognition scores in noise actually reflected masking effects, rather than merely indexing how well listeners can recognize the particular words used.

In summary, the hypothesis was tested that listeners (especially adults) would decrease the weight they assigned to offset transitions and increase the weight they assigned to vocalic duration when conditions changed from quiet to noise. This hypothesis would be supported by steeper functions that were closer together. At the same time, the hypothesis was tested that adults and children might perform similarly when real-world conditions were simulated. The hypothesis was also tested that children might experience more masking of formant transitions than adults experienced. Nittrouer and Boothroyd (1990) found that children's recognition scores were generally poorer at every SNR than those of adults, although the effects of linguistic context were similar, leading to the conclusion of greater masking for children. Unfortunately, Nittrouer and Boothroyd had no way of determining which part(s) of the speech signal was particularly masked for children. In the current experiment, the signal properties that could be used for phonetic decisions were restricted to formant offsets and vocalic duration. As already proposed, there was good reason to suspect that formant offsets would be vulnerable to masking, but vocalic duration would not be.

II. METHOD

A. Listeners

Adults and children of the ages 8, 6, and 4 years participated in this experiment. To participate, listeners needed to be native speakers of American English. They had to pass a hearing screening of the pure tones 0.5, 1.0, 2.0, 4.0, and 6.0 kHz presented at 25 dB HL. Children needed to be within −1 and +5 months of their birthdays: for example, all 4-year-olds were between 3 years, 11 months and 4 years, 5 months. Children needed to score at or above the 30th percentile on the Goldman-Fristoe 2 Test of Articulation, Sounds-in-Words subtest (Goldman and Fristoe, 2000). Children had to be free from significant, early histories of otitis media with effusion, defined as six or more episodes during the first 2 years of life. Adults needed to be between 18 and 40 years of age. Adults needed to demonstrate at least an 11th-grade reading level on the reading subtest of the Wide Range Achievement Test-Revised. (Jastak and Wilkinson, 1984). Meeting these criteria were 22 adults (mean age =26 years), 20 8-year-olds, 22 6-year-olds, and 24 4-year-olds. However, four of the 4-year-olds were unable to reach the minimum criteria for participation in two of the three tasks they were asked to do (word recognition in quiet and noise, *boot/booed* labeling in quiet and noise, and *cop/cob* labeling in quiet and noise), and so their data were not included for any of the tasks.

B. Equipment and materials

Testing took place in a sound-proof booth with the computer that controlled the various tasks in an adjacent control room. The hearing screening was done with a Welch Allen TM262 audiometer and TDH-39 earphones. All stimuli were stored on a computer and presented through a Creative Labs Soundblaster card, a Samson headphone amplifier, and AKG-K141 headphones at a 22.05-kHz sampling rate. The experimenter recorded responses using a keyboard.

For the labeling tasks, two hand-drawn pictures (8 ×8 in) were used to represent each response label: for example, a picture of a police officer was used for *cop* and a picture of a corncob was used for *cob*. Game boards with ten steps were also used with children. They moved a marker to the next number on the board after each block of stimuli. Cartoon pictures were used as reinforcement and were presented on a color monitor after completion of each block of stimuli. A bell sounded while the pictures were being shown and served as additional reinforcement.

C. Stimuli

1. General speech recognition in noise

For evaluating speech recognition at various SNRs, 20 lists were used, each with ten phonetically balanced consonant–vowel–consonant (CVC) words. These word lists were taken from Mackersie, Boothroyd, and Minniear (2001), and were similar to ones used by Boothroyd and Nittrouer (1988) and Nittrouer and Boothroyd (1990). Each word was recorded three times by a male adult speaker, and the token of each word with the flattest intonation but without any vocal glitches was selected for use in this study. Level was equalized for all words, and then words were mixed with randomly generated white noise (i.e., flat spectrum) low-pass filtered with a cutoff frequency of 11.03 kHz (the upper cutoff of the speech stimuli). The level of the

noise relative to the speech stimuli varied in five equal steps between −6 and +6 dB, a range that results of Boothroyd and Nittrouer (1988) and Nittrouer and Boothroyd (1990) suggested should provide recognition scores between 25% and 75% correct for all listeners in this experiment. Four word lists were presented at each of the five SNRs used. Speech stimuli were mixed with the noise for each listener separately such that different lists were presented at each of the five SNRs across listeners. Furthermore, order of presentation of the lists varied across listeners so that the order of presentation of SNR was randomized. Word level was held constant at 68 dB SPL during testing.

2. Labeling words with voiced and voiceless final stops in noise

Stimuli used for the labeling tasks were taken from the second experiment of Nittrouer (2004). These were natural tokens of a male adult speaker saying *cop*, *cob*, *boot*, and *booed*. Three tokens of each word were used so that there was natural variation in properties such as fundamental frequency and intonation. Although efforts are always made to select tokens with similar fundamental frequencies and flat intonation contours, some variation inevitably exists. When listeners are hearing tokens from only two word categories, these slight variations could influence phonetic decisions if only one token of each word is used. Having several tokens of each word, with all the natural variability that entails, controls for this possible confound.

For each word, the release burst and any voicing during closure was deleted. Vocalic length was manipulated either by reiterating a single pitch period from the most stable region of the vocalic portion or by deleting pitch periods from that stable region. Thus, formant offset transitions were left intact. Care was taken to ensure that the points in the waveform where pitch periods were either reiterated or deleted subsequently lined up at zero crossings to avoid any clicks in the signal. Seven stimuli were created for each token in this way, varying in length from the mean length of the three tokens of the word ending in a voiceless stop to the mean length of the three tokens ending in a voiced stop. For *cop/cob*, the continua varied from 82 to 265 ms. For *boot/booed*, the continua varied from 97 to 258 ms. Steps were kept as equal in size as possible across the continua. Mean *F*1 frequency at voicing offset was 300 Hz across the three tokens of *boot*, 268 Hz across the three tokens of *booed*, 801 Hz across the three tokens of *cop*, and 625 Hz across the three tokens of *cob*. Clearly there was a greater difference in *F*1-offset frequency between *cop* and *cob* than between *boot* and *booed*. In summary, four continua were generated: one each with *boot*, *booed*, *cop*, and *cob* formant offsets. Each continuum had seven stimuli of different lengths, and three tokens of each of those stimuli.

For listening conditions that required that these stimuli be presented in noise, noise was generated in the same way as for the speech recognition task. As with those stimuli, all labeling stimuli were equalized in amplitude before being combined with the noise. And again, the level of the words was held constant at 68 dB SPL. All listeners heard the labeling stimuli presented in noise at 0-dB SNR; adults also heard the stimuli presented at −3-dB SNR.

D. Procedures

Testing took place over two test sessions on different days at least 3 days apart, but not more than 2 weeks apart. This separation between sessions was used to diminish the possibility that there would be learning effects for the word lists, which were presented at both sessions.

The screening tasks were the first things done on the first day, followed by the 20 word lists, either in quiet or in noise. These word lists were the first thing presented on the second day. Half the listeners heard them in quiet on the first day and in noise on the second day, and half heard them in the opposite order. After hearing the 20 word lists, listeners were presented with the labeling stimuli. Children were presented with four sets of stimuli for labeling: *cop/cob* and *boot/booed*, both in quiet and at 0-dB SNR. Adults were presented with six sets of stimuli for labeling: *cop/cob* and *boot/booed*, in quiet and at both 0- and −3-dB SNR. The order of presentation of these stimulus sets was randomized across listeners, with certain restrictions. Children had to hear one set of each word pair (*cop/cob* or *boot/booed*) at each session, and one of these sets had to be presented in quiet and the other at 0-dB SNR. Adults were restricted from hearing the same word pair consecutively, and they had to hear stimuli in each listening condition at each of the two sessions. So, for example, at the first session an adult listener might hear *cop/cob*, *boot/booed*, and then *cop/cob* again in the conditions of 0-dB SNR, quiet, and −3-dB SNR, respectively. At the next session this listener would hear *boot/booed*, *cop/cob*, and *boot/booed*, in the listening conditions of −3-dB SNR, quiet, and 0-dB SNR, respectively.

The task when listening to the 20 word lists at varying SNRs was to repeat the word. The experimenter recorded onto the computer whether the response was correct or not. The word had to be completely correct to be counted as such. Listeners had to recognize correctly at least 180 words on the 20 word lists (90%) when presented in quiet to have their data included in this analysis. This requirement served as a check that all listeners could understand the words used and perform the repetition task.

During the labeling tasks, listeners responded by saying the label and pointing to the picture that represented their selection. Listeners had to pass preliminary tasks with two sets of stimuli in order to proceed to testing. First, unedited versions of the words (i.e., with the release bursts and voicing during closures intact) were presented. Each of the six words (e.g., three tokens of *boot* and three tokens of *booed*) was presented twice. The listener had to respond correctly to at least 11 of the 12 (92%) without feedback to proceed to the next preliminary task. This requirement ensured that all listeners could perform the task and that they all recognized the voicing distinction for word-final stops presented in quiet. This first preliminary task was administered only prior to the first presentation of either set of words (*boot/booed* or *cop/cob*). The second preliminary task was administered

prior to the presentation of each set of words, in each condition. In this task the best exemplars of the six stimuli in the listening condition about to be tested were presented twice each. The term "best exemplar" is used here to refer to the stimulus in which formant transitions and vocalic duration most clearly signaled a specific voicing decision. These stimuli had the release bursts and any voicing during closure removed. So, for example, the best exemplars of *cop* were the three tokens taken from the speaker saying *cop* (so that formant transitions were appropriate for the voiceless stop), with the shortest vocalic portions. The listener needed to respond correctly to at least 11 of the 12 presentations of best exemplars (92%) to proceed to testing. This requirement ensured that all listeners were able to make voicing judgments based on one or the other of the available cues, or a combination of those cues. If listeners do not base the phonetic decision they are being asked to make on the cues available to them, it is pointless to ask questions about the relative weighting of those cues. This preliminary task also serves as a general check on the quality of the stimuli created: If a large number of listeners, particularly adults, cannot hear the presumed best exemplars of each category with near-perfect accuracy, it suggests that the stimuli do not validly replicate natural tokens.

During testing, ten blocks of the 14 stimuli were presented (i.e., stimuli with formant transitions appropriate for a voiced or voiceless final stop, at each of the seven vocalic durations). Because there were actually three tokens with each kind of offset transitions (voiced or voiceless), the program was designed to select randomly one of the three to present during the first block, and then repeat this random selection during the next block without replacement. After three blocks the process was repeated until ten blocks had been presented. For children, cartoon pictures were displayed on the monitor and a bell sounded at the end of each block. They moved a marker to the next space on a game board after each block as a way of keeping track of how much more time they had left in the test. Listeners had to respond correctly to at least 80% of the best exemplars during testing to have their data included in the final analysis. This requirement is commonly viewed as providing a check that the listener paid attention to the task. The reasoning is that, if listeners could respond with better than 90% accuracy to these best exemplars during the preliminary task, then they should be able to respond with at least 80% accuracy during testing, if general attention is maintained during testing. It might also be argued that a listener who labels the best exemplars of each phonetic category with better than 90% accuracy during the preliminary task and then fails to label accurately those same tokens during testing was operating on the perceptual edge, so to speak, during the preliminary task. That is, the listener may have just barely been able to label the best exemplars using the available cues during the preliminary task when no additional demands were present. The increased demands of having to listen to many ambiguous tokens might be enough to disrupt his/her abilities to integrate those cues into a phonetic percept. Regardless, however, of whether the cause of uncertain responding to best exemplars is a general lack of attention or disrupted perceptual processing, there is little to be learned from labeling functions that hover around the 50% line for the length of the function. That sort of responding only means that the listener could not perform the labeling task with the available cues.

Each listener's labeling responses were used to construct cumulative distributions of the proportion of one response (the voiced response in this experiment) across levels of the acoustic property manipulated in a continuous fashion (vocalic duration in this experiment) for each level of the acoustic property manipulated in a dichotomous fashion (formant offsets in this experiment). Best-fit lines were then obtained using probit analysis (Finney, 1964). From these probit functions slopes and distribution means (i.e., phoneme boundaries) were computed. Generally, phoneme boundaries are given in physical units for the property manipulated in a continuous fashion, such as Hz or ms. However, in this experiment step size differed slightly for the two sets of stimuli. Consequently, phoneme boundaries are given using steps as the units of description. Similarly, slope is generally given as the change in probit units per unit change on the physical continuum. In this experiment, slope is given as change in probit units per step. Probit analysis can extrapolate so that phoneme boundaries outside of the range tested can be obtained. For this work, the values that extrapolated phoneme boundaries could take were limited to 3.5 steps beyond the lowest and highest values tested.[3] Mean slope of the function is taken as an indication of the weight assigned to the continuously varied property: the steeper the function, the more weight that was assigned to that property. The separation between functions at the phoneme boundaries (for each level of the dichotomously set property) is taken as an indication of the weight assigned to that dichotomously set property, as long as settings of the property clearly signal each phonetic category involved: the greater the separation, the greater the weight that was assigned. Because these stimuli were created from natural tokens, offset transitions clearly signaled either voiced or voiceless final stops.

Some investigators (e.g., Turner *et al.*, 1998) have computed partial correlation coefficients between each acoustic property and the proportion of one response as a way to describe the weighting of acoustic properties. Nittrouer (2004) compared results for partial correlation coefficients and slopes/phoneme boundaries and found that conclusions reached by the two kinds of metrics were largely the same. However, slopes and separations in phoneme boundaries were found to provide slightly more sensitive estimates of weighting strategies. Furthermore, slopes and separations in phoneme boundaries correspond more directly to visual impressions gleaned from graphed labeling functions. For both these reasons the decision was made to analyze slopes and phoneme boundaries in this study.

III. RESULTS

A. SNR

One 6-year-old and one 4-year-old failed to meet the requirement that they recognize correctly 90% of the words presented in quiet, and so their data were not included. Figure 2 shows mean percent-correct recognition scores for each

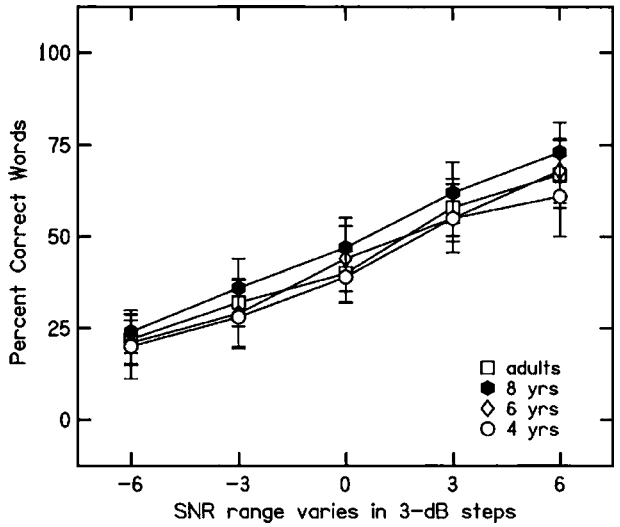


FIG. 2. Percent-correct word recognition for CVC words heard at five SNRs.

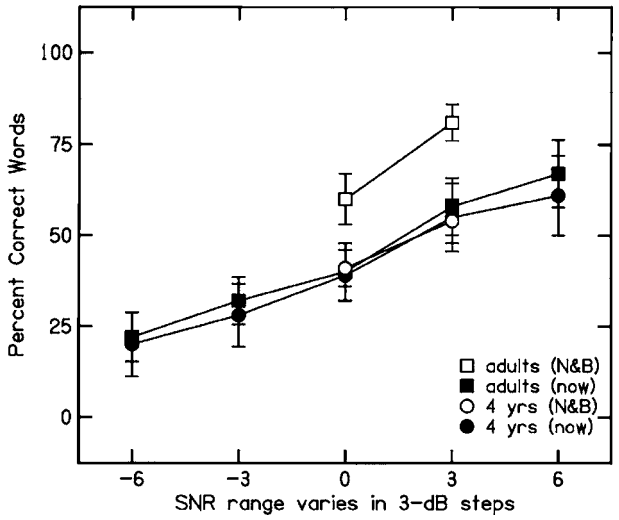


FIG. 3. Percent-correct word recognition for CVC words for adults and 4-year-olds from this experiment (now) and from Nittrouer and Boothroyd (1990) (N&B).

age group at each SNR. Recognition scores are similar across age groups at all SNRs, with the minor exception that 8-year-olds' scores are roughly 5 percentage points better than the other groups at all SNRs. Across SNRs, mean recognition scores (and standard deviations) were: 43.7 (1.7) for adults; 48.6 (1.9) for 8-year-olds; 43.6 (1.9) for 6-year-olds; and 40.6 (1.8) for 4-year-olds. A two-way analysis of variance (ANOVA) was performed on these recognition scores, with age as the between-subjects factor and SNR as the within-subjects factor. The main effect of age was significant, $F(3,78)=10.80$, $p<0.001$, as was the main effect of SNR, $F(4,312)=498.64$, $p<0.001$.[4] Most likely, the significant age effect was largely due to the better recognition scores exhibited by 8-year-olds, rather than to a linear developmental trend. The age $\times$ SNR interaction was not significant.

Although not the focus of this study, it is interesting to compare these results for speech recognition in flat-spectrum noise with those from Nittrouer and Boothroyd (1990) for speech recognition in speech-shaped noise. Figure 3 shows results for adults and 4-year-olds from the current study, and for adults and 4-year-olds from Nittrouer and Boothroyd. Results from Nittrouer and Boothroyd are shown for only 0- and 3-dB SNR because these are the only SNRs that study used. Four-year-olds from the two studies performed identically, but adults from Nittrouer and Boothroyd showed roughly a 20% advantage over adults from this study. It seems that adults benefit from the high-frequency signal portions that are readily available when speech is embedded in speech-shaped noise (as in Nittrouer and Boothroyd) rather than in flat-spectrum noise (as in this experiment). Children, on the other hand, apparently do not achieve this benefit.

B. Labeling results for all age groups, quiet and 0-dB SNR

1. Boot/booed

One 6-year-old and one 4-year-old failed to meet the requirement that they recognize correctly 80% of the best exemplars during testing, and so their data were not included. These were different children from those who failed to meet the criterion for participation with the word lists presented in varying SNRs.

a. Adults versus children. Figure 4 shows labeling functions for all age groups for *boot* and *booed* presented in quiet and at 0-dB SNR. This figure indicates that functions were similarly placed for children and adults when stimuli were heard in quiet, particularly for stimuli with *boot* offset tran-

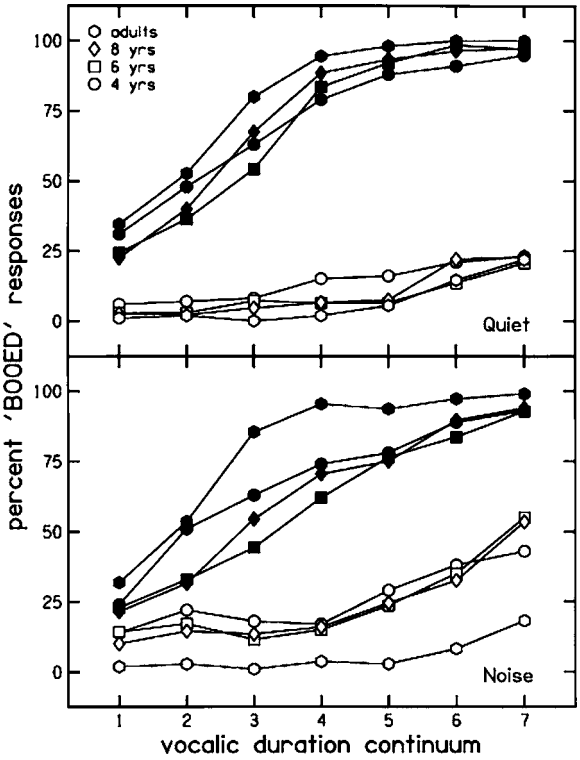


FIG. 4. Labeling functions for *boot/booed* stimuli presented in quiet and at 0-dB SNR, plotted with all age groups together. Filled symbols indicate responses to stimuli with booed formant offsets; open symbols indicate responses to stimuli with *boot* formant offsets.

sitions. Children's labeling functions for stimuli with *booed* offset transitions are slightly more to the right (i.e., towards longer vocalic durations) than those of adults, indicating that children did not weight these offset transitions quite as strongly as adults did. When these stimuli were heard in noise, children's labeling functions for stimuli with *booed* and *boot* offset transitions show less separation than those of adults. Graphically, children's functions are closer to the center of the plot than are those of adults. That is, functions for stimuli with *booed* offset transitions are more to the right (towards longer vocalic durations) and stimuli with *boot* offset transitions are more to the left (towards shorter vocalic durations). This pattern indicates that children did not weight offset transitions as strongly as adults. It is difficult to tell how slopes of the functions may have changed, if at all, across noise conditions, but it does appear as if adults' function for the originally voiced stimuli might be steeper than those of children in the noise condition.

To investigate the apparent age-related effects observed in Fig. 4, simple effects analyses were done on phoneme boundaries for each listening condition separately, with age as a between-subjects' factor and formant offsets as a within-subjects' factor. Simple effects analysis is often a reasonable selection of statistical test for experiments with several independent factors because it permits the examination of effects for one or more of those factors at each level of another factor, while using the overall estimate of error variance.

For stimuli presented in quiet, only the main effect of formant offsets was significant for phoneme boundaries, $F(1,78)=808.30$, $p<0.001$. The age × formant offsets interaction was close to significant, $F(3,78)=2.47$, $p=0.068$, probably reflecting the slight difference in placement of adults' and children's functions for stimuli with *booed* offset transitions. For stimuli presented in noise, the main effect of formant offsets was again significant, $F(1,78)=731.70$, $p<0.001$, and this time the age × formant offsets interaction was significant, $F(3,78)=12.40$, $p<0.001$. Therefore, it seems fair to conclude that listeners of all ages placed labeling functions in roughly the same locations when stimuli were heard in quiet, indicating that formant offset transitions were weighted similarly. However, when stimuli were heard in noise, children's labeling functions were actually less separated, indicating that they decreased the weight they assigned to those formant offset transitions from the quiet condition.

Simple effects analysis was done on slopes for each of the four functions separately with age as the between-subjects factor. Only the function for stimuli with *booed* offsets presented in noise showed a significant age effect, $F(3,78)=10.35$, $p<0.001$, although it was close to significant for stimuli with *booed* offsets presented in quiet, $F(3,78)=2.37$, $p=0.077$. Clearly adults weighted vocalic duration more than children for stimuli presented in noise (at least for those stimuli with *booed* offsets): Mean slopes for stimuli with *booed* offsets presented in noise for individual age groups were 0.79 (0.39) for adults; 0.46 (0.21) for 8-year-olds; 0.40 (0.11) for 6-year-olds; and 0.43 (0.25) for 4-year-olds. For stimuli presented in quiet, mean slopes for stimuli with *booed* offsets presented in quiet were 0.74 (0.36) for adults; 0.70 (0.31) for 8-year-olds; 0.55 (0.31) for 6-year-olds; and 0.52 (0.32) for 4-year-olds. Thus, it appears that children actually assigned slightly more weight to vocalic duration for stimuli presented in quiet than they did for stimuli presented in noise.

b. Effects for each age group. The information above compared results for children and adults, which was necessary to do in order to address two of the three hypotheses posed. However, the third hypothesis to be tested, that adults (and perhaps older children) would weight vocalic duration more in noise than in quiet, could only be addressed by comparing results in quiet and noise for each age group separately.

Figure 5 shows labeling functions for each age group separately for *boot/booed* when stimuli were presented in quiet and at 0-dB SNR. Regarding the weight assigned to offset transitions, evidence can be gathered from the separation between labeling functions. Adults' labeling functions appear similar for stimuli presented in quiet and at 0-dB SNR, but children's labeling functions appear less separated for stimuli presented in noise, rather than in quiet. To see whether these noise-related changes were significant, simple effects analyses were performed on phoneme boundaries for each age group separately. The term of most interest was the noise × formant offset interaction because a significant interaction would indicate that indeed the direction of change in phoneme boundaries across listening conditions was different for stimuli with *boot* and *booed* offset transitions (i.e., functions were "moving towards the center"). Results are shown in Table I and show that all three children's groups had significant noise × formant offset interactions. This pattern of the functions moving closer to one another when the stimuli were presented in noise, rather than in quiet, indicates that children weighted those offset transitions less when stimuli were heard in noise than when they were heard in quiet. Because adults' labeling functions did not differ in location for stimuli heard in noise and in quiet, it can be concluded that they weighted offset transitions equally in both conditions.

Of course, the focus of this particular experiment was on the possibility that listeners (in particular, adults) might increase the weight assigned to vocalic duration in decisions of word-final voicing when speech is heard in noisy backgrounds. If the perceptual weight for vocalic duration increased when listening in noise, then the slopes of labeling functions for stimuli heard in noise would be steeper than those of stimuli heard in quiet. To evaluate this possibility, a simple effects analysis was performed on slopes for each age group separately. The effect of noise was examined separately for the *booed* and *boot* functions. Only results from 8-year-olds showed a significant noise effect, and only for stimuli with *booed* offset transitions, $F(1,78)=8.48$, $p=0.004$, although 6-year-olds' results for stimuli with *booed* offset transitions were close to significant, $F(1,78)=3.42$, $p=0.068$. However, instead of functions being steeper when stimuli were heard in noise, rather than in quiet, they were shallower: for 8-year-olds, mean slope=0.46 probit units in noise versus 0.70 in quiet; for 6-year-olds, mean slope =0.40 probit units in noise versus 0.55 in quiet. These results

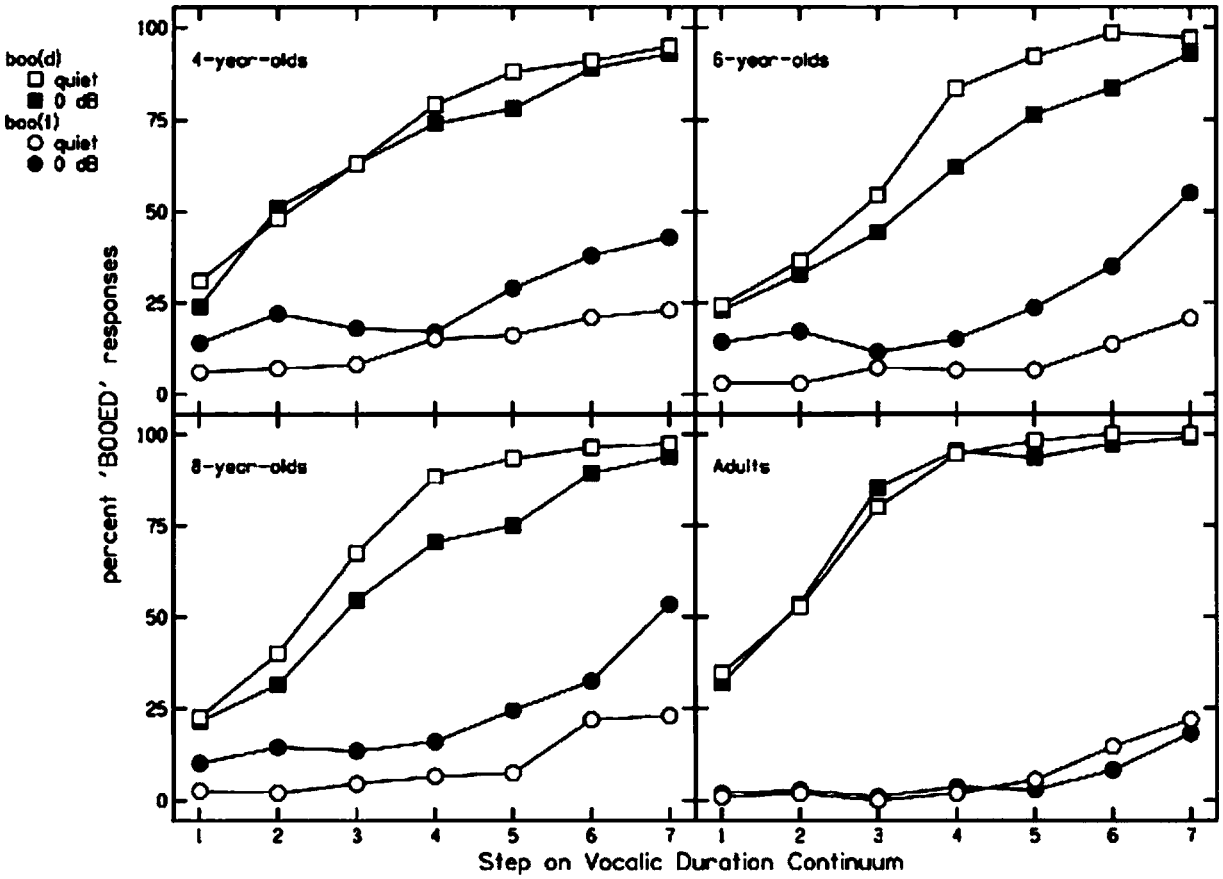


FIG. 5. Labeling functions for *boot/booed* stimuli presented in quiet and at 0-dB SNR, plotted with each age group separately.

indicate that 8-year-olds (and probably 6-year-olds) weighted vocalic duration less when stimuli were in noise, rather than in quiet. These changes are opposite to predictions.

2. Cop/cob

All listeners were able to complete this task.

a. Adults versus children. Figure 6 shows labeling functions for all age groups for *cop* and *cob* presented in quiet and at 0-dB SNR. As with *boot/booed* stimuli presented in quiet, it appears that listeners of all ages had similarly placed labeling functions for these *cop/cob* stimuli when they were presented in quiet. The results of the simple effects analysis done on phoneme boundaries for each listening condition separately confirmed this impression: for the quiet condition, only the main effect of formant offsets was significant, $F(1,80)=680.22$, $p<0.001$. In particular, the age $\times$ formant offsets interaction was not significant, nor close to significant. Although it appears from the lower half of Fig. 6 that children's labeling functions for *cop/cob* stimuli presented in noise may be less separated than those of adults, the simple effects analysis done on phoneme boundaries does not confirm this impression: As with stimuli presented in quiet, only the main effect of formant offsets was significant, $F(1,80)=692.13$, $p<0.001$. Consequently, the conclusion may be drawn that adults and children weighted formant offsets similarly for these stimuli, in both noise and quiet.

As with results for *boot/booed* stimuli, simple effects

TABLE I. Results of simple effects analysis (for each age group separately) for phoneme boundaries, *boot/booed* stimuli presented in quiet and in noise at 0-dB SNR. The main effect of noise refers to whether stimuli were heard in quiet or in noise. The main effect of formant offsets refers to whether formant offset transitions were consistent with a final voiced or voiceless stop. Degrees of freedom were 1, 78 for all effects.

	Noise		Formant offsets		Noise $\times$ formant offsets	
	F	*p*	*F*	*p*	*F*	*p*
Adults	0.12	NS	437.35	<0.001	1.03	NS
8-year-olds	0.46	NS	190.34	<0.001	11.26	=0.001
6-year-olds	0.04	NS	226.86	<0.001	26.29	<0.001
4-year-olds	2.24	NS	237.36	<0.001	6.86	=0.011

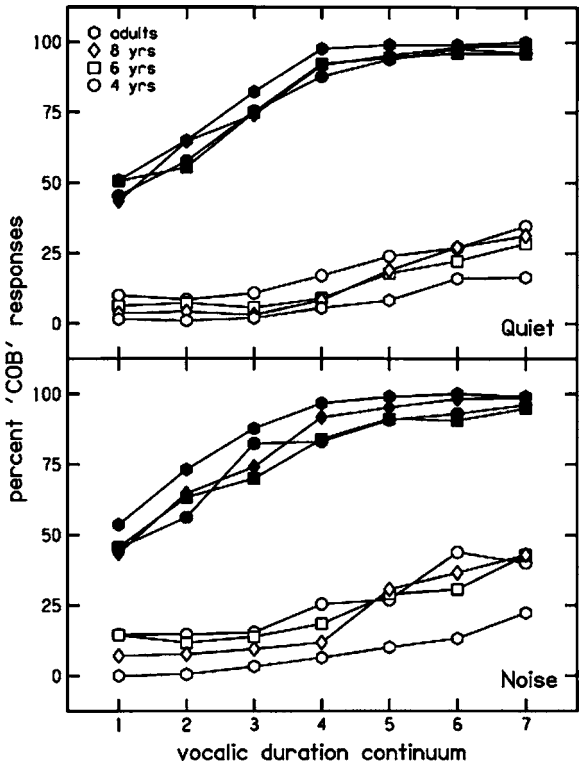


FIG. 6. Labeling functions for *cop/cob* stimuli presented in quiet and at 0-dB SNR, plotted with all age groups together. Filled symbols indicate responses to stimuli with *cob* formant offsets; open symbols indicate responses to stimuli with *cop* formant offsets.

analyses were conducted on slopes for each function separately, with age as the between-subjects factor. The main effect of age was significant for both functions from stimuli presented in noise: for stimuli with *cop* offsets, $F(3,80)=5.00$, $p=0.003$; for stimuli with *cob* offsets, $F(3,80)=7.02$, $p<0.001$. For stimuli with *cop* offsets presented in noise, mean slopes were 0.34 (0.22) for adults; 0.31 (0.22) for 8-year-olds; 0.19 (0.11) for 6-year-olds; and 0.17 (0.11) for 4-year-olds. For stimuli with *cob* offsets presented in noise, mean slopes were 0.66 (0.34) for adults; 0.50 (0.24) for 8-year-olds; 0.33 (0.14) for 6-year-olds; and 0.41 (0.23) for 4-year-olds. Looking at stimuli presented in quiet, the main effect of age was close to significant for stimuli with *cob* offsets only, $F(3,80)=2.30$, $p=0.084$. For these stimuli, mean slopes were 0.74 (0.37) for adults; 0.61 (0.28) for 8-year-olds; 0.46 (0.35) for 6-year-olds; and 0.56 (0.41) for 4-year-olds. Consequently, the conclusion may be drawn that adults weighted vocalic duration more than children for stimuli presented in noise, and possibly for stimuli presented in quiet when formant offsets were appropriate for a voiced final stop.

b. Effects for each age group. Figure 7 shows labeling functions for *cop/cob* for each age group separately. Unlike Fig. 5 showing functions for *boot/booed*, it appears that listeners in all groups performed similarly for stimuli presented in quiet and noise. Statistical analyses support that conclusion. Table II shows results of the simple effects analysis done on phoneme boundaries for *cop/cob*. No age group showed a significant noise × formant offsets interaction. As with results for *boot/booed*, simple effects analyses were

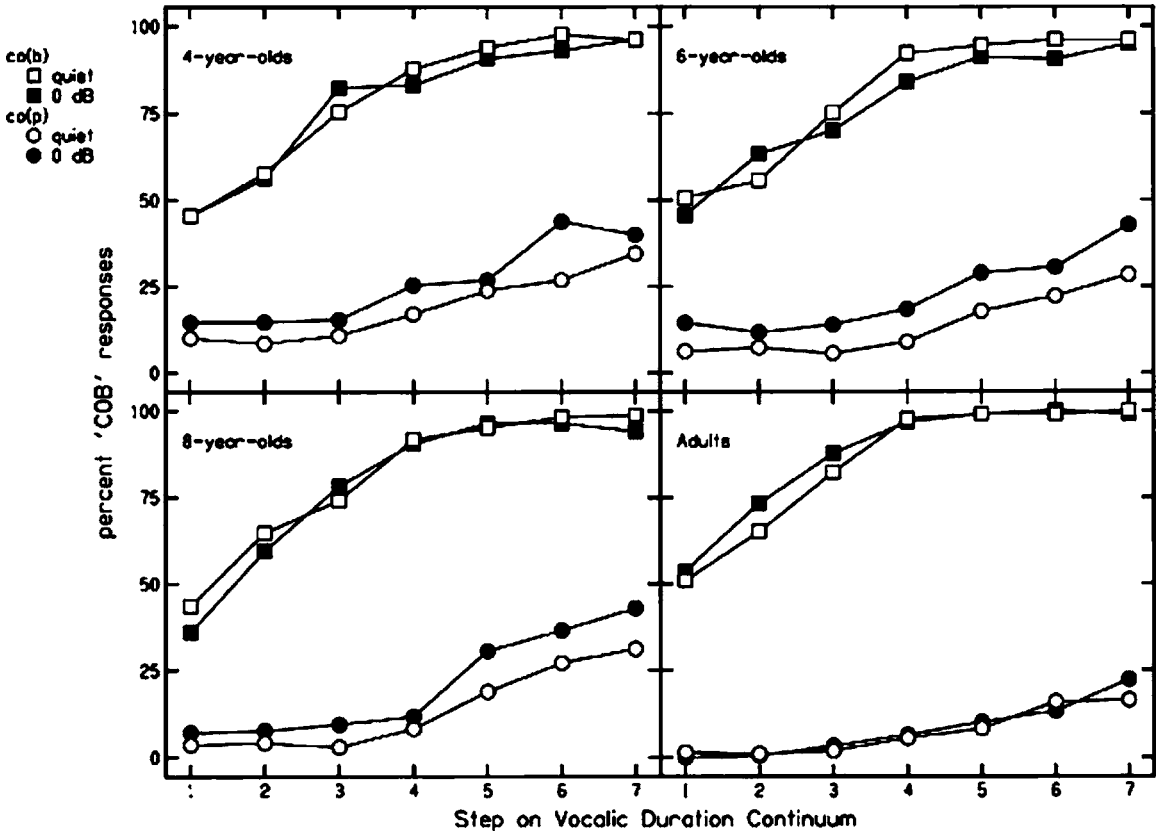


FIG. 7. Labeling functions for *cop/cob* stimuli presented in quiet and at 0-dB signal-to-noise ratio, plotted with each age group separately.

TABLE II. Results of simple effects analysis (for each age group separately) for phoneme boundaries, *cop/cob* stimuli presented in quiet and in noise at 0-dB SNR. Degrees of freedom were 1,80 for all effects.

	Noise		Formant offsets		Noise × formant offsets	
	F	*p*	*F*	*p*	*F*	*p*
Adults	2.41	NS	252.03	<0.001	0.06	NS
8-year-olds	3.27	=0.074	169.81	<0.001	0.76	NS
6-year-olds	1.80	NS	245.16	<0.001	0.31	NS
4-year-olds	1.88	NS	212.30	<0.001	0.04	NS

conducted on slopes for *cop* and *cob* functions separately, for each age group. No significant or marginally significant noise effects were found. Thus, listeners of all ages weighted vocalic duration similarly in noise and quiet.

The finding of significant age effects for slopes when stimuli were presented in noise, but not in quiet (other than the marginally significant age effect for stimuli with *cob* offsets presented in quiet) suggests that listeners modified their weighting of vocalic duration based on whether stimuli were presented in noise or in quiet, and children showed greater shifts than adults. These shifts were enough to create significant age effects for stimuli presented in noise that were not seen for stimuli presented in quiet, but not enough for any one listener group to show a significant noise effect. Of importance to the current study, the slight, nonsignificant shifts in weighting of vocalic duration from the quiet to the noise condition were in the direction of less weight being assigned to vocalic duration in noise than in quiet for all groups. This shift is opposite to the prediction.

C. Labeling results for adults, quiet, 0-dB SNR, and −3-dB SNR

Results of the analyses done on labeling functions for stimuli presented in quiet and at a 0-dB SNR (described above) revealed no significant differences for adults in placement or steepness of functions for *boot/booed* or *cop/cob* presented in these two conditions. However, adults also heard stimuli at an even poorer SNR (−3 dB) to see if this decrement in SNR would affect their performance.

Figure 8 shows labeling functions for adults for stimuli presented in the three listening conditions (quiet, 0-dB SNR, and −3-dB SNR), for *boot/booed* and *cop/cob*. From this figure it appears that functions are similar across listening conditions for the *cop/cob* stimuli. For the *boot/booed* stimuli, functions for the quiet and 0-dB SNR conditions are similar, as demonstrated by the simple effects analysis for adults described in the previous section, but functions for the −3-dB SNR condition appear to be closer to the middle of the figure. This trend is similar to that observed for children at the 0-dB SNR. To evaluate these impressions, two-way ANOVAs were performed on phoneme boundaries for the *boot/booed* and *cop/cob* stimuli separately, with noise and formant offsets as within-subjects' factors. Results of these analyses are shown in Table III. Of particular importance, the noise × formant offsets interaction was significant for *boot/booed* phoneme boundaries, as it had been for children for stimuli presented in quiet and at 0-dB SNR. This result supports the observation that functions were less separated (i.e., closer to the middle of the plot) when stimuli were presented at −3-dB SNR. This trend was not found for phoneme boundaries for *cop/cob* stimuli.

Simple effects analyses were done on slopes for each function separately. Only the *booed* slopes showed a significant effect of noise, $F(2,42)=6.24$, $p=0.004$. Slopes for the *booed* functions were 0.74 (0.36), 0.79 (0.39), and 0.50 (0.28) for the quiet, 0-dB, and −3-dB conditions, respectively. The noise effect was close to significant for *cob* slopes, $F(2,42)=3.08$, $p=0.057$. Slopes for the *cob* functions were 0.74 (0.37), 0.66 (0.34), and 0.50 (0.20) for the quiet, 0-dB, and −3-dB conditions, respectively. Clearly there is evidence that adults decreased the weight assigned to vocalic duration at the poorest SNR. As with the shift in weighting for vocalic duration observed for children at the 0-dB SNR,

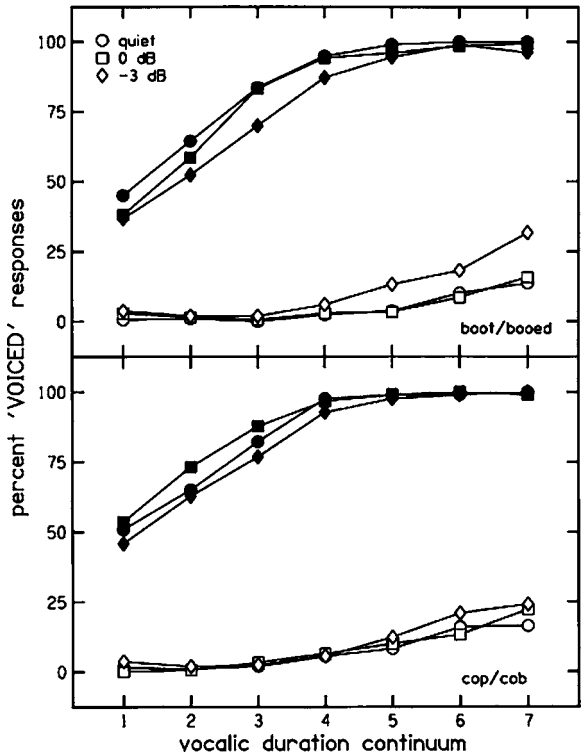


FIG. 8. Labeling functions for *boot/booed* and *cop/cob* stimuli presented in quiet and in noise at 0-dB and −3-dB SNR, adults only. Filled symbols indicate responses to stimuli with voiced formant offsets; open symbols indicate responses to stimuli with voiceless formant offsets.

TABLE III. Results of ANOVAS for phoneme boundaries for *boot/booed* and *cop/cob* stimuli heard in quiet and at 0-dB and −3-dB SNR, adult listeners only. Degrees of freedom were 2, 42 for the main effect of noise, and the noise × formant offsets interaction, and 1,21 for the main effect of formant offsets.

	Noise		Formant offsets		Noise × formant offsets	
	F	*p*	*F*	*p*	*F*	*p*
Boot/booed	2.11	NS	268.20	<0.001	6.36	=0.004
Cop/cob	1.41	NS	363.19	<0.001	0.61	NS

this shift for adults is in the opposite direction to the prediction.

IV. DISCUSSION

The purpose of this experiment was to test the hypothesis that developmental changes in perceptual strategies for speech occur because some acoustic properties in the speech signal are more resistant to noise than others, and mature speech perception takes advantage of this signal redundancy by shifting weight away from the more vulnerable properties and towards the more resistant properties as listening conditions dictate. According to this account, the ability to adjust perceptual weighting strategies depending on listening conditions would be a skill that children would need to develop. However, no evidence was found to support the suggestion that adults make use of redundant cues in this way, at least not for voicing decisions of word-final stops. It does appear, however, that noise masked the signal property weighted most heavily by all listeners in decisions of syllable-final stop voicing: formant offset transitions. That is, the weight assigned to this acoustic property decreased when stimuli were heard in noise, and presumably this decrease was due to those transitions being less available perceptually (i.e., they were masked). This masking could degrade speech recognition in naturally noisy settings. But, even when faced with this degradation, listeners did not shift the focus of their perceptual attention to vocalic duration.

Differences were observed between the two sets of stimuli in how resistant offset transitions were to masking by the noise. *Cop/cob* stimuli had high *F*1 frequencies at syllable center, and so *F*1 fell substantially for cob. As a result, the difference in the frequency of *F*1 at voicing offset between *cop* and *cob* was considerable (176 Hz). *Boot/booed* stimuli, on the other hand, had low *F*1 frequencies at syllable center, and so *F*1 did not fall very much going into closure for the voiced cognate. As a result, there was little difference in the frequency of *F*1 at voicing offset between *boot* and *booed* (32 Hz). These latter stimuli showed greater masking effects for formant offset transitions than did the former stimuli. Whether this is because the offset transitions did not distinguish as strongly between the voiced and voiceless cognates for *boot/booed* as for *cop/cob* or because *F*1 offset transitions are not as extensive for *boot/booed* as for *cop/cob* cannot be determined by this study because the two factors covaried.

Eight-year-olds demonstrated two results that seemed to run counter to developmental trends. First, their recognition scores for words presented in noise were 5 percentage points better than those of adults, at every SNR. No obvious explanation for this result can be offered, but it is probably not important. In spite of having better overall recognition scores than adults, 8-year-olds performed similarly to younger children on the labeling task in noise. Second, 8-year-olds were the only group to show a significant effect of noise on the slope of labeling functions (although the effect was marginally significant for 6-year-olds), and this effect was restricted to stimuli with *booed* offset transitions. For these stimuli, mean slopes in quiet and noise, respectively, were 0.74 vs 0.79 for adults, 0.70 vs 0.46 for 8-year-olds, 0.55 vs 0.40 for 6-year-olds, and 0.52 vs 0.43 for 4-year-olds. Thus, 8-year-olds were able to attend to vocalic duration when stimuli were presented in quiet, but this attention was disrupted when stimuli were presented in noise. A similar trend is observed for 6-year-olds, and to even a lesser extent for 4-year-olds, although effects for these listeners did not reach statistical significance.

A secondary question addressed by this work was whether age-related differences in perceptual weighting strategies would be found when stimuli were presented in more naturalistic conditions than most laboratory studies offer. The most revealing finding in these data is that adults had steeper labeling functions than children (with *p* values for age effects of <0.10) for five of the eight functions obtained. Three of these functions were obtained in the noise condition. In fact, only one labeling function obtained in noise failed to show an age effect on slope, supporting the proposal that there is a genuine age-related difference in the weighting of vocalic duration, even in real-world conditions. This finding fits with the more general notion that children initially attend primarily to the slowly changing, global resonances of the vocal tract, and gradually incorporate information from other sources (such as vocalic duration) as they acquire experience with their native language.

There was also an age-related difference in the amount of masking produced by noise, particularly for *boot/booed* stimuli. This result matches findings of Nittrouer and Boothroyd (1990), who concluded that peripheral masking likely accounts for this difference in masking effect between children and adults. But, greater central masking for children could also explain this difference. Since Nittrouer and Boothroyd was published, two studies have shown that children experience greater masking effects than adults for multitonal maskers (Oh, Wightman, and Lutfi, 2001; Wightman *et al.*, 2003). This sort of masking is thought to be central in nature.

However, these experiments with multitonal maskers used nonspeech stimuli presented simultaneously with the maskers, and so the implications for speech stimuli are not clear. In addition, Wright *et al.* (1997) showed that children with specific language deficits experienced greater backward masking than children developing language normally. Backwards masking is also considered a central effect (Plack, Carlyon, and Viemeister, 1995). Perhaps the difference reported by Wright *et al.* between two groups of children based on the presence or absence of a disorder could reflect a more general, developmental trend. Perhaps children experience more central masking than adults, accounting for the greater reduction in weighting of offset transitions when stimuli are presented in noise. However, Wright *et al.*'s finding was also obtained with nonspeech stimuli, possibly limiting its relevance to speech.

Still one other possible explanation should be considered for the apparent age-related difference in noise masking. Brady, Shankweiler, and Mann (1983) examined recognition in noise of words and environmental sounds by 8-year-olds with reading disorders, and by 8-year-olds learning to read normally. Both groups of listeners were able to recognize words and environmental sounds in quiet with near-perfect accuracy. When the environmental sounds were presented in flat-spectrum noise at a 0-dB SNR, both groups showed similar masking effects. When the words were presented in noise, again at 0-dB SNR, the children who were learning to read normally showed better recognition, compared to their results for environmental sounds. The children with reading disorders showed no improvement in recognition for words in noise over environmental sounds in noise. Brady *et al.*'s conclusion was that the ability to recognize phonetic structure in the acoustic speech signal actually provides some "release from masking," so to speak, for skilled listeners. According to this explanation, the increased masking for speech signals experienced by younger listeners is a consequence of poorer (less-mature) language abilities, rather than a source of those poorer abilities. Unfortunately, this study can shed no light on whether the age-related differences found in the reduction of weighting of formant offset transitions when stimuli were presented in noise can best be explained by central masking effects or by differences in linguistic processing (specifically, in abilities to recover phonetic structure).

In the end, this study leaves open the question of why children gradually modify their perceptual weighting strategies for speech. The reason that has been suggested previously arises from evidence indicating that developmental changes in speech perception strategies and in the abilities to recover and use phonetic structure co-occur. Evidence from different investigators shows that children gradually modify their perceptual weighting strategies for speech (e.g., Greenlee, 1980; Krause, 1982; Nittrouer, 1992; Parnell and Amerman, 1978; Siren and Wilcox, 1995; Wardrip-Fruin and Peach, 1984) and that they gradually acquire skills such as counting word-internal phonetic units (Liberman *et al.* 1974), judging similarity of phonetic structure between different words (Walley, Smith, and Juszcyk, 1986), and using phonetic structure for storing words in working memory (Nittrouer and Miller, 1999). One study showed that these developmental changes in speech perception and abilities to access and use phonetic structure co-occur in the same group of children (Mayo *et al.*, 2003). Finally, several studies have found that when one developmental change is delayed, the other is as well (Nittrouer, 1999; Nittrouer and Burton, 2001; 2005). In light of such evidence, the suggestion has been made that the acoustic properties that gradually, through childhood, come to be weighted more are ones that facilitate the recovering of phonetic structure in the child's native language. Certainly none of the data reported here contradict that suggestion.

In conclusion, this experiment was designed largely to test the hypothesis that children's perceptual weighting strategies for speech change through childhood to allow them to take advantage of signal redundancy in natural listening environments where some acoustic properties may be masked. This hypothesis would have been supported if adults (and perhaps older children, as well) shifted perceptual attention away from formant offset transitions and toward vocalic duration when listening to signals in noise. However, this result was not observed, and so the suggestion that the need to take advantage of redundancy in speech signals motivates developmental shifts in perceptual weighting strategies for speech is not supported. In fact, no evidence was found to support the generally held perspective that skilled perceivers of speech shift the focus of their attention from one acoustic cue to another as listening conditions dictate. The data reported here were consistent, however, with the notion that there is a developmental shift in perceptual weighting strategies for speech in which phonetically relevant signal properties, other than global resonance patterns, come to be weighted more strongly. Finally, a developmental decrease in the masking effects of environmental noise was observed.

ACKNOWLEDGMENTS

The author wishes to thank the following people for their help on this project: Tom Creutz at Boys Town National Research Hospital for writing the software to present stimuli in noise; Melanie Wilhelmsen, Kathi Bodily, and Jennifer Smith for help with data collection; and Carol A. Fowler, John Kingston, Joanna H. Lowenstein, and Donal G. Sinex for their comments on an earlier draft of this manuscript. This work was supported by Research Grant No. R01 DC00633 from the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health.

[1]In keeping with investigators such as Kewley-Port, Pisoni, and Studdert-Kennedy (1983), the term "dynamic" is used here to refer to signal properties involving formant movement. These authors took this term directly from distinctive feature theory. According to this theory, "static" properties contrast with dynamic properties, and refer to broadband spectral patterns that remain stable over at least a few milliseconds, such as steady-state vowel formants, fricative noises, and release bursts. In this work, a temporal property (vocalic length) is considered, and along with static properties is subsumed under the general term of "nondynamic."

[2]In order for the separation between labeling functions to be a valid indicator of the weight assigned to the dichotomously set property (i.e., the property that defines each continuum), the two settings of that property must unambiguously signal each of the two phonetic labels that listeners are

being asked to use. The dichotomously set property in Nittrouer (2004) was formant offsets. Because the stimuli were constructed from natural stimuli, this property clearly signaled voiced and voiceless final stops unambiguously.

[3]Although the value of these limits is somewhat arbitrary, they essentially establish numerical markers for functions that never cross the 50% line. Importantly, the use of these markers only serves to constrain the probability of obtaining statistically significant results, and so cannot bias procedures to show effects where there are none.

[4]F and p values are reported for any results with $p < 0.10$. Results with $p > 0.10$ are described as "not significant" (NS).

Assmann, P., and Summerfield, Q. (**2004**). "The perception of speech under adverse acoustic conditions," in *Speech Processing in the Auditory System, Springer Handbook of Auditory Research*, edited by S. Greenberg and W. Ainsworth (Springer, New York), pp. 231–308.

Boothroyd, A., and Nittrouer, S. (**1988**). "Mathematical treatment of context effects in phoneme and word recognition," J. Acoust. Soc. Am. **84**, 101–114.

Brady, S., Shankweiler, D., and Mann, V. (**1983**). "Speech perception and memory coding in relation to reading ability," J. Exp. Child Psychol. **35**, 345–367.

Coker, C. H., and Umeda, N. (**1976**). "Speech as an error-correcting process," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, New York), pp. 349–364.

Crowther, C. S., and Mann, V. (**1992**). "Native language factors affecting use of vocalic cues to final consonant voicing in English," J. Acoust. Soc. Am. **92**, 711–722.

Crowther, C. S., and Mann, V. (**1994**). "Use of vocalic cues to consonant voicing and native language background: The influence of experimental design," Percept. Psychophys. **55**, 513–525.

Denes, P. (**1955**). "Effect of duration on the perception of voicing," J. Acoust. Soc. Am. **27**, 761–764.

Edwards, T. J. (**1981**). "Multiple features analysis of intervocalic English plosives," J. Acoust. Soc. Am. **69**, 535–547.

Finney, D. J. (**1964**). *Probit Analysis* (Cambridge University, Cambridge, England).

Fischer, R. M., and Ohde, R. N. (**1990**). "Spectral and duration properties of front vowels as cues to final stop-consonant voicing," J. Acoust. Soc. Am. **88**, 1250–1259.

Flege, J. E., and Port, R. (**1981**). "Cross-language phonetic interference: Arabic to English," Lang Speech **24**, 125–146.

Flege, J. E., and Wang, C., (**1989**). "Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/–/d/ contrast," J. Phonetics **17**, 299–315.

Goldman, R., and Fristoe, M. (**2000**). *Goldman Fristoe 2: Test of Articulation* (American Guidance Service, Inc., Circle Pines, MN).

Gracco, V. L. (**1994**). "Some organizational characteristics of speech movement control," J. Speech Hear. Res. **37**, 4–27.

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.

Harris, K. S. (**1958**). "Cues for the discrimination of American English fricatives in spoken syllables," Lang Speech **1**, 1–7.

Hillenbrand, J., Ingrisano, D. R., Smith, B. L., and Flege, J. E. (**1984**). "Perception of the voiced-voiceless contrast in syllable-final stops," J. Acoust. Soc. Am. **76**, 18–26.

Hogan, J. T., and Rozsypal, A. J. (**1980**). "Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant," J. Acoust. Soc. Am. **67**, 1764–1771.

Jastak, S., and Wilkinson, G. S. (**1984**). *The Wide Range Achievement Test-Revised* (Jastak Associates, Wilmington, DE).

Jones, C. (**2003**). "Development of phonological categories in children's perception of final voicing," Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Kewley-Port, D., Pisoni, D. B., and Studdert-Kennedy, M. (**1983**). "Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants," J. Acoust. Soc. Am. **73**, 1779–1793.

Krause, S. E. (**1982**). "Developmental use of vowel duration as a cue to postvocalic stop consonant voicing," J. Speech Hear. Res. **25**, 388–393.

Kuzniarz, J. (**1968**). "Masking of speech by continuous noise," Pol. Med. J. **7**, 1001–1008.

Lehman, M. E., and Sharf, D. J. (**1989**). "Perception/production relationships in the development of the vowel duration cue to final consonant voicing," J. Speech Hear. Res. **32**, 803–815.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., and Carter, B. (**1974**). "Explicit syllable and phoneme segmentation in the young child," J. Exp. Child Psychol. **18**, 201–212.

MacKain, K. S., Best, C. T., and Strange, W. (**1981**). "Categorical perception of English /r/ and /l/ by Japanese bilinguals," Appl. Psycholinguist. **2**, 369–390.

Mackersie, C. L., Boothroyd, A., and Minniear, D. (**2001**). "Evaluation of the Computer-Assisted Speech Perception Assessment Test (CASPA)," J. Am. Acad. Audiol **12**, 390–396.

MacNeilage, P. F., and Davis, B. (**1991**). "Acquisition of speech production: Frames, then content," in *Attention & Performance XIII*, edited by M. Jeanerod (Erlbaum, New York), pp. 453–476.

Mayo, C., and Turk, A. (**2004**). "Adult–child differences in acoustic cue weighting are influenced by segmental context: Children are not always perceptually biased toward transitions," J. Acoust. Soc. Am. **115**, 3184–3194.

Mayo, C., Scobbie, J. M., Hewlett, N., and Waters, D. (**2003**). "The influence of phonemic awareness development on acoustic cue weighting strategies in children″s speech perception," J. Speech Lang. Hear. Res. **46**, 1184–1196.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M., Jenkins, J. J., and Fujimura, O. (**1975**). "An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English," Percept. Psychophys. **18**, 331–340.

Murphy, W. D., Shea, S. L., and Aslin, R. N. (**1989**). "Identification of vowels in 'vowel-less' syllables by 3-year-olds," Percept. Psychophys. **46**, 375–383.

Nittrouer, S. (**1992**). "Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries," J. Phonetics **20**, 351–382.

Nittrouer, S. (**1999**). "Do temporal processing deficits cause phonological processing problems?" J. Speech Lang. Hear. Res. **42**, 925–942.

Nittrouer, S. (**2002**). "Learning to perceive speech: How fricative perception changes, and how it stays the same," J. Acoust. Soc. Am. **112**, 711–719.

Nittrouer, S. (**2004**). "The role of temporal and dynamic signal components in the perception of syllable-final stop voicing by children and adults," J. Acoust. Soc. Am. **115**, 1777–1790.

Nittrouer, S., and Boothroyd, A. (**1990**). "Context effects in phoneme and word recognition by young children and older adults," J. Acoust. Soc. Am. **87**, 2705–2715.

Nittrouer, S, and Burton, L. (**2001**). "The role of early language experience in the development of speech perception and language processing abilities in children with hearing loss," Volta Review **103**, 5–37.

Nittrouer, S., and Burton, L. (**2005**). "The role of early language experience in the development of speech perception and phonological processing abilities: Evidence from 5-year-olds with histories of otitis media with effusion and low socioeconomic status," J. Commun. Disord. **38**, 29–63.

Nittrouer, S., and Miller, M. E. (**1999**). "The development of phonemic coding strategies for serial recall," Appl. Psycholinguist. **20**, 563–588.

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.

Oh, E. L., Wightman, F., and Lutfi, R. A. (**2001**). "Children's detection of pure-tone signals with random multitone maskers," J. Acoust. Soc. Am. **109**, 2888–2895.

Parnell, M. M., and Amerman, J. D. (**1978**). "Maturational influences on perception of coarticulatory effects," J. Speech Hear. Res. **21**, 682–701.

Plack, C. J., Carlyon, R. P., and Viemeister, N. F. (**1995**). "Intensity discrimination under forward and backward masking: Role of referential coding," J. Acoust. Soc. Am. **97**, 1141–1149.

Raphael, L. J. (**1972**). "Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English," J. Acoust. Soc. Am. **51**, 1296–1303.

Raphael, L. J. (**1975**). "The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English," J. Phonetics **3**, 25–33.

Raphael, L. J., Dorman, M. F., and Liberman, A. M. (**1980**). "On defining the vowel duration that cues voicing in final position," Lang Speech **23**, 297–307.

Remez, R. E., Pardo, J. S., Piorkowski, R. L., and Rubin, P. E. (**2001**). "On the bistability of sine wave analogues of speech," Psychol. Sci. **12**, 24–29.

Remez, R. E., Rubin, P. E., Pisoni, D. B., and Carrell, T. D. (**1981**). "Speech perception without traditional speech cues," Science **212**, 947–949.

Siren, K. A., and Wilcox, K. A. (**1995**). "Effects of lexical meaning and practiced productions on coarticulation in children's and adults' speech," J. Speech Hear. Res. **38**, 351–359.
Summers, W. V. (**1987**). "Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses," J. Acoust. Soc. Am. **82**, 847–863.
Sussman, J. E. (**2001**). "Vowel perception by adults and children with normal language and specific language impairment: Based on steady states or transitions?" J. Acoust. Soc. Am. **109**, 1173–1180.
Thelen, E. (**1985**). "Developmental origins of motor coordination: Leg movements in human infants," Dev. Psychobiol. **18**, 1–22.
Turner, C. W., Kwon, B. J., Tanaka, C., Knapp, J., Hubbartt, J. L., and Doherty, K. A. (**1998**). "Frequency-weighting functions for broadband speech as estimated by a correlational method," J. Acoust. Soc. Am. **104**, 1580–1585.
Walley, A. C., Smith, L. B., and Jusczyk, P. W. (**1986**). "The role of phonemes and syllables in the perceived similarity of speech sounds for children," Mem. Cognit. **14**, 220–229.
Wardrip-Fruin, C. (**1982**). "On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants," J. Acoust. Soc. Am. **71**, 187–195.
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.
Wightman, F. L., Callahan, M. R., Lutfi, R. A., Kistler, D. J., and Oh, E. (**2003**). "Children's detection of pure-tone signals: Informational masking with contralateral maskers," J. Acoust. Soc. Am. **113**, 3297–3305.
Wright, B. A., Lombardino, L. J., King, W. M., Puranik, C. S., Leonard, C. M., and Merzenich, M. M. (**1997**). "Deficits in auditory temporal and spectral resolution in language-impaired children," Nature (London) **387**, 176–178.

Decline of speech understanding and auditory thresholds in the elderly[a)]

Pierre L. Divenyi[b)]
Speech and Hearing Research, Veterans Affairs Medical Center and East Bay Institute for Research and Education, Martinez, California 94553

Philip B. Stark
Department of Statistics, University of California, Berkeley, California 94720-3860

Kara M. Haupt
Speech and Hearing Research, Veterans Affairs Medical Center, Martinez, California 94553

(Received 12 October 2004; revised 14 May 2005; accepted 18 May 2005)

A group of 29 elderly subjects between 60.0 and 83.7 years of age at the beginning of the study, and whose hearing loss was not greater than moderate, was tested twice, an average of 5.27 years apart. The tests measured pure-tone thresholds, word recognition in quiet, and understanding of speech with various types of distortion (low-pass filtering, time compression) or interference (single speaker, babble noise, reverberation). Performance declined consistently and significantly between the two testing phases. In addition, the variability of speech understanding measures increased significantly between testing phases, though the variability of audiometric measurements did not. A right-ear superiority was observed but this lateral asymmetry did not increase between testing phases. Comparison of the elderly subjects with a group of young subjects with normal hearing shows that the decline of speech understanding measures accelerated significantly relative to the decline in audiometric measures in the seventh to ninth decades of life. On the assumption that speech understanding depends linearly on age and audiometric variables, there is evidence that this linear relationship changes with age, suggesting that not only the accuracy but also the nature of speech understanding evolves with age. [DOI: 10.1121/1.1953207]

PACS number(s): 43.71.Lz, 43.71.Pc, 43.71.Ky, 43.71.An [KWG] Pages: 1089–1100

I. INTRODUCTION

From early adulthood on, but especially after age 60, auditory sensitivity and other aspects of hearing gradually deteriorate. More troubling for communication, and thus for quality of life, is the progressive degradation of speech understanding, especially when there is noise such as background speech or reverberation. The decline of auditory communication ability is interesting as basic science and for its clinical implications: understanding this loss would shed light on the relationship between hearing and speech comprehension and could lead to the development of devices to alleviate the deficit.

The nature, degree, and rate of hearing deterioration with age have been studied both diachronically, by testing a group of subjects longitudinally as they age, and synchronically, by comparing subjects of different ages. Longitudinal study is more direct and has fewer inherent ambiguities, but it is costly, difficult, and fraught with unpredictable obstacles, such as dropouts or illnesses affecting auditory functions. Cross-sectional study of subjects of different ages is easier and less expensive; however, it is more vulnerable to confounding. Interestingly, differences between findings reported in longitudinal and cross-sectional studies with large sample sizes are minor: the comprehensive pictures of auditory aging are quite consistent (Brant and Fozard 1990). The gradual presbycusic (i.e., sloping, high-frequency) loss of hearing sensitivity with age has been studied extensively. Pure-tone thresholds tend to increase as individuals age: longitudinal and cross-sectional studies put the rate of decay at about 5.5 to 9 dB/decade for the *better* ear, depending on frequency (Davis *et al.*, 1990; Gates *et al.*, 1990; Ostri and Parving, 1991); the worse ear often deteriorates almost 50% faster. Although one longitudinal study of changes in hearing level covering a 23-year span (the Baltimore Longitudinal Study on Aging) estimated the rate of decline to be almost twice that found in other studies (4.5 to 14.7 dB/decade, Pearson *et al.*, 1995; Morrell *et al.*, 1996), the discrepancy may be due in part to the different ways the rate of decline was calculated. Several studies also observed that decline in men is faster in the younger advanced years and slower later, whereas in women, the rate of decline is slow at first and accelerates later—with the result that by the eighth or ninth decade of life, almost no gender differences remain (Gates *et al.*, 1990; Pearson *et al.*, 1995; Morrell *et al.*, 1996).

Auditory thresholds in the two ears are seldom exactly equal. The ear with the higher thresholds is often the left, and its relative disadvantage increases with age (Gates *et al.*, 1990; Gates and Cooper, 1991). There is also typically a left-ear disadvantage in processing some suprathreshold speech presented in interference. In particular, perception of left-ear targets of dichotic sentences deteriorates with age

[a)]Portions of the findings were reported at the 139th meeting of the Acoustical Society of America, Atlanta, GA, on 31 May 2000.
[b)]Electronic mail: pdivenyi@ebire.org

relative to perception of right-ear targets (Jerger *et al.*, 1990; Jerger and Jordan, 1992), even in individuals with little or no change in their hearing sensitivity over a period of several years (Stach *et al.*, 1985). In contrast, the right-ear advantage for brief, dichotically presented nonsense syllables, digits, or spondees does not appear to change (Gelfand *et al.*, 1980; Martini *et al.*, 1988). These reports suggest that a right-ear advantage increases both for peripheral hearing sensitivity and for speech tasks that rely on centrally mediated processes. This increase of asymmetry may be due to a breakdown of the integration of binaural information, possibly caused by gradual demyelination of the callosal pathway necessary for interhemispheric transfer (Jerger *et al.*, 1993; Chmiel and Jerger, 1996; Chmiel *et al.*, 1997).

Decline of speech understanding after age 60 also has been investigated extensively. Generally, understanding of undistorted speech presented in quiet does not change significantly (Blumenfield *et al.*, 1969; Gelfand *et al.*, 1985; van Rooij and Plomp, 1990; Tun, 1998), at least until the eighth decade of life (Bergman *et al.*, 1976; Pedersen *et al.*, 1991). In contrast, understanding of speech presented in interference deteriorates steadily and sometimes dramatically. For example, understanding of speech with added speech-spectrum random noise or multi-talker babble deteriorates with age, especially from the seventh decade on (Gelfand *et al.*, 1986; Dubno *et al.*, 1997); reverberation increasingly affects speech understanding as listeners age (Duquesnoy and Plomp, 1980; Nabelek and Robinson, 1982); thresholds for word, consonant, and even vowel identification increase with age (van Rooij and Plomp, 1990); and age predicts understanding of speech in the presence of reverberation (Humes and Christopherson, 1991; Divenyi and Haupt, 1997a) or an interfering ipsilateral or contralateral sentence (Jerger and Hayes, 1977; Divenyi and Haupt, 1997a). Speech understanding by older listeners is also affected by time compression (Konkle *et al.*, 1977; Tun, 1998), periodic interruption (Bergman *et al.*, 1976), and low-pass filtering (Cheesman *et al.*, 1995)—especially when the filtering is paired with reverberation (Humes and Christopherson, 1991).

Can the decline of speech understanding be attributed to presbycusic threshold shifts? Some studies suggest so for some aspects of speech perception (e.g., Duquesnoy and Plomp, 1980; Humes and Roberts, 1990; Humes, 1991; Dubno *et al.*, 1997), but other studies indicate that deterioration in speech understanding occurs *in addition* to deterioration in hearing sensitivity (Jerger and Hayes, 1977; Bergman, 1980; Humes and Christopherson, 1991; Pedersen *et al.*, 1991; Divenyi and Haupt, 1997a, c; Jerger and Chmiel, 1997). That is, changes in speech understanding in aging, although influenced by hearing sensitivity, also include components separate from the presbycusic process.

Hearing and understanding speech in interference or distortion deteriorate later in life. But how rapidly? Do hearing and speech understanding decline at the same rate? Does the decline in hearing account for the decline in understanding? To answer these questions requires tracking changes in individuals' hearing sensitivity and speech understanding over time. Such changes provide clues to relationships among different auditory functions; in turn, these relationships might hint at the underlying causes.

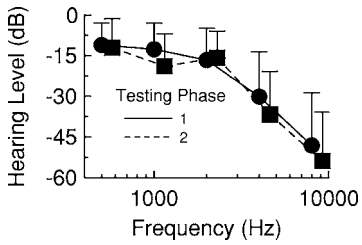


FIG. 1. Average of right- and left-ear audiograms of the 29 subjects measured in the first phase (round symbols/solid lines) and second phase (square symbols/broken lines). Error bars are one standard deviation.

This article presents such a longitudinal study. Although it spans only about 5 years and includes only 29 subjects, according to a recent survey (Divenyi and Simon, 1999) no comparable study has been conducted before. The present investigation focused on how speech understanding under unfavorable conditions declines in elderly individuals whose auditory performance is initially better than average for the general population at their age. Limiting the study to individuals with relatively good initial hearing reduces confounding from presbycusis and makes it easier to separate the changes in the two auditory functions.

II. METHODS

A. Subjects

Twenty-nine subjects, 21 females and eight males with above-average results on a large battery of auditory performance tests (Divenyi and Haupt, 1997a), were selected from an original group of 100 elderly individuals. The subjects, whose ages ranged from 60.0 to 83.7 years (mean: 69.6 years, SD: 6.1 years) at the time of initial testing, were retested using the same battery 4.0 to 8.5 years after their first test. Ages ranged from 65.2 to 88.6 years (mean: 74.9 years, SD: 5.8 years) at retesting. The mean speech recognition threshold (SRT) of the group was 13.6 dB HL (SD: 8.62 dB HL) at initial testing and 15.0 dB HL (SD: 10.85 dB HL) at retesting. Pure-tone thresholds were measured at 0.5, 1, 2, 4, and 8 kHz (Fig. 1). At each frequency, in both testing phases, and for all subjects, the thresholds were symmetrical within 15 dB in 280 of 290 right-left threshold pairs, and always within 25 dB. Figure 1 shows that the hearing loss of our group is presbycusic, typical of persons in their seventh to ninth decades, and does not exceed what is considered mild-to-moderate. Air-conduction thresholds matched bone-conduction thresholds (i.e., any loss was of the sensorineural type); tympanometry was normal; acoustic reflexes were present at 0.5, 1, and 2 kHz; there was no decay at either 0.5 or 1 kHz; and word recognition in quiet was 80% or better (at 45 dB *re:* SRT). At the time of the initial testing, the subjects' self-assessed hearing was "good" or "excellent." Our inquiry, both at the initial screening and at the time of retesting, revealed no history of unusual noise exposure. None of the subjects had ever complained of tinnitus or vertigo or worn a hearing aid. All were in good general health

and were native English speakers—a requirement for the interpretation of speech understanding results.

To estimate the rate of change in the auditory performance of the subjects prior to their first test, we compared their results to those of 11 healthy, normal-hearing individuals with an average age of 22.18 years (SD: 3.15 years) and no prior experience as listeners. Their average auditory performance was used as an estimate of the auditory performance of the elderly subjects 40 to 60 years before the study started.

B. Apparatus and test battery

Audiological status of the subjects was assessed using acoustic immittance, pure-tone, and speech audiometry. All tests were conducted in a sound-attenuated testing booth. Tests were performed using an Interacoustics AC30 clinical audiometer and a set of Telephonic TDH-39 earphones with MX 41/AR type cushions.

Speech understanding was measured on the word and sentence levels for spectrally and temporally distorted speech as well as for speech presented with various types of interference. Some tests were performed under more than one condition, producing several performance measures; some performance measures were calculated by combining results for several test conditions. The seven tests described below gave a total of 14 measures, each obtained separately for either the right and left ears or the right and left hemifields (for the SPIN tests presented in virtual free field). Assessing performance in the right and left ears or hemifields allowed us to estimate changes in lateral asymmetry. The acronyms in bold type are used throughout the rest of the paper.

(1) Pure-tone thresholds at 0.5, 1, 2, 4, and 8 kHz, leading to three measures: From the 0.5-, 1-, 2-, and 4-kHz thresholds, we calculated **PTA4**, the average pure-tone threshold of the four lowest frequencies, as well as **PTSLP**, the difference between the 0.5- and 4-kHz thresholds. We used the 8-kHz threshold by itself (**PT8k**).
(2) Speech Recognition Threshold (**SRT**) for spondee words.
(3) Low-pass filtered speech (see Lynn and Gilroy, 1972, 1975): listeners were asked to attend to and repeat NU-6 words low-pass filtered at 750 Hz (**LP750**), presented at 50 dB *re:* the average pure-tone threshold at 0.5, 1, and 2 kHz.
(4) Time-compressed speech (see Beasley *et al.*, 1972): listeners were asked to repeat words from the W-22 list compressed by 60% (**TC60**), presented at 50 dB *re:* **SRT**.[1]
(5) Speech understanding in reverberation, using the modified rhyme reverberation test (see Nabelek and Robinson, 1982) at reverberation times of 0.45, 0.85, and 1.25 s or without reverberation, presented at 50 dB *re:* **SRT**. Listeners were asked to identify which of six similar-sounding words (differing only with respect to the initial or final consonant) they heard. **RT75** is the estimated reverberation time corresponding to 75% correct identification.[2]
(6) Competing sentence test (Willeford, 1985): pairs of sentences were presented simultaneously, with one sentence in each pair spoken by a female and the other sentence spoken by a male. In the monaural condition, both sentences in each pair were presented in the same ear at 45 dB SL for the test sentences and 50 dB SL for the nontest sentences (**CSTI**), i.e., a −5-dB signal-to-noise ratio (S/N). Listeners were asked to repeat the sentence said by the female speaker. In the binaural condition (**BCST**), one sentence was presented in each ear at 50 dB SL, i.e., 0-dB S/N. Listeners were asked to repeat both sentences.
(7) Speech Perception In Noise test [SPIN, originally developed by Kalikow *et al.* (1977), revised by Bilger *et al.* (1984), the "Illinois SPIN test"], presented either monaurally or binaurally. Listeners were asked to repeat the final word of a sentence spoken by a male speaker. The last word was either easy to predict (a "high-probability" item) or hard to predict (a "low-probability" item) from the sentence context. High- and low-probability items were scored separately. In the monaural condition, the target was presented at 50 dB SL (*re:* babble threshold) and the speech-to-babble ratio (S/B) was 4 dB. The monaural results yielded **SPMon**, the average of the high- and low-probability scores, and **SPMonCEf**, the difference between the scores for high- and low-probability items, a measure of the effect of context. In the binaural conditions, the sentences were presented in virtual free field at 45° right or left, flanked by two independent four-talker (two female, two male) babble sources on either side of the target speaker, with all speakers positioned on the periphery of a 2-m radius circle. There were three spatial SPIN conditions differing in the angular separation between adjacent babble speakers; this separation was 22.5°, 45°, or 72°, resulting in a total azimuthal span of 90°, 180°, or 360°, respectively.[3] The presentation level at the nearer ear was 50 dB SL for both the target sentences and the babble. Three measures were derived from the spatial SPIN data: (1) The overall performance measure,**SPSpat**, is the average of high- and low-probability item scores across the three angular separations. (2) The context effect, **SPSpatCEf**, is the difference between scores for high- and low-probability items averaged across the three conditions. (3) The spatial separation effect, **SPSpatEf**, is the least-squares estimate of the slope of performance as a function of azimuthal separation, weighted by the logarithm of the overall spatial performance. This last measure was included because speech understanding in babble was seen to depend on the azimuthal separation between target and interference for individuals without significant hearing loss, whether young (Shinn-Cunningham *et al.*, 2001) or elderly (Gelfand *et al.*, 1988; Divenyi and Haupt, 1997b). The weighting prevents masking of the effect by consistently poor or consistently excellent performance.

The battery thus produced four audiometric measures (**PTA4**, **PTSLP**, **PT8k**, and **SRT**) and ten measures of

speech understanding: two of temporally or spectrally distorted speech (**TC60** and **LP750**) and eight of speech in interference (**RT75**, **SPMon**, **SPMonCEf**, **SPSpat**, **SPSpatCEf**, **SPSpatEf**, **BCST**, and **CSTI**). Testing procedures and presentation levels were those recommended by the developers of each test except for the SPIN test.[4] For tests presented to both ears or in both hemifields, the first ear tested was counterbalanced across subjects and the order of test presentation was pseudo-random. Procedures for the initial and second tests were identical. In each phase, subjects were tested for a total of 8–10 hours over an average span of 3 weeks. Each testing session lasted approximately 90 min. Breaks were provided as needed. Units of the performance measures differed by task, but larger values correspond to better performance—thus, a positive difference between the first and second phases shows deterioration in performance. Percent correct measures were logit-transformed to emphasize differences at the high and low ends of the range. SPIN test results were logit-transformed before any further data manipulation, such as averaging and/or differencing. The remaining measures represent threshold values, expressed in dB or in seconds (for **RT75**).

C. Statistical methods

The results were analyzed using nonparametric methods. In some cases, parametric tests were also performed for comparison. We eschew common parametric methods, such as ANOVA and the *t*–test, because their assumptions are difficult to justify in the present study and, arguably, in audiological studies in general. Significance levels for common parametric tests are computed assuming that the data are independent samples from normally distributed populations, which is questionable for several reasons: (1) Test and retest data for the same measure for the same individual may be dependent, as may data for a given subject's two ears. (2) There is little reason to believe that test values or measurement errors have normal distributions. (3) Measures can be difficult to compare when they have different scales or are bounded by different limits. (4) Intrasubject variability is large, and differs across individuals (Marshall, 1981). The significance levels of parametric tests on audiological data can be misleading, causing one to reject the null hypothesis not because the alternative is a better explanation of the data, but because the null hypothesis is unrealistic. Conversely, violations of the statistical assumptions can obscure real effects. Nonparametric methods obviate the need for some of these assumptions and provide a robust alternative to parametric tests (Lehmann and D'Abrera, 1988). In the last two decades there has been an explosion in methods based on resampling the data, which also require fewer assumptions about the distribution of the data than parametric methods do (e.g., Efron, 1982; Davison and Hinkley, 1997). The present study uses nonparametric methods and resampling to test hypotheses about the changes in audiometric and speech understanding measures as subjects age.

III. RESULTS

Analysis focused on three points: (1) comparing performance between the two testing phases, (2) determining the rate of change in audiometric and speech understanding measures, and (3) assessing the relationship between speech understanding performance and both age and auditory thresholds. The first point was addressed by analyzing the data using univariate methods; the second, by a comparison between the results for elderly subjects and reference data for young, normal-hearing subjects; and the third, by bootstrap tests of a simple linear model that represents speech understanding as a function of age and hearing sensitivity.

A. Univariate statistics and hypothesis tests

Means and standard deviations of the 14 measures of performance over the two testing phases averaged across the right and left ears of the elderly subjects are illustrated in Fig. 2. In all analyses in this section, each subject's measurements in the second testing phase were replaced with linearly interpolated or extrapolated estimates of the subject's performance 5.27 years (the mean lapse) after the first testing phase, in order to normalize the elapsed time between the two testing phases. (The largest extrapolation interval was 1.27 years; the largest interpolation interval was 3.23 years.) Figure 2 shows that essentially all performance measures deteriorated between phase 1 and phase 2. The average puretone threshold (**PTA4**) increased by 3.33 dB (SD: 2.47 dB), corresponding to 6.32 dB/decade (SD: 4.69 dB), a result very similar to threshold changes reported by other investigators (Davis *et al.*, 1990; Gates *et al.*, 1990; Ostri and Parving, 1991) and comparable to the lower limit of decay estimated by Pearson *et al.* (1995). Figure 1 shows that the audiometric decline is dominated by the threshold increases at 4 kHz (7.67 dB, SD 9.23 dB, or 14.61 dB/decade) and 8 kHz (5.97 dB, SD 9.09 dB, or 11.32 dB/decade), which explains the relatively large average change of the audiogram slope (**PTSLP**, 5.78 dB, SD 6.16 dB, or 10.98 dB/decade). Because increased thresholds at higher frequencies impair the perception of consonantal information in speech (Dubno *et al.*, 1984; Frisina and Frisina, 1997), it is likely that threshold increases at 4 and 8 kHz contributed to the decline of speech understanding in interference. However, the presentation level was threshold dependent (SL or *re:* SRT), so the speech signal was always audible even in the 4-kHz range and, because threshold increases at high frequencies also impede perception of the interfering speech, loss of speech understanding is unlikely to be solely the result of simultaneous masking of the signal.[5] Moreover, although we included the 8-kHz threshold as an indicator of the subject's hearing status, no speech material presented had information beyond 5 kHz—the developers of the tests were careful to minimize any direct effect of high-frequency loss. The relatively large standard deviation of the threshold changes at the high frequencies indicates that presbycusic hearing loss increased substantially between testing phases, at least for some subjects.

Differences in the 14 measures by testing phase and laterality were examined using nonparametric methods based

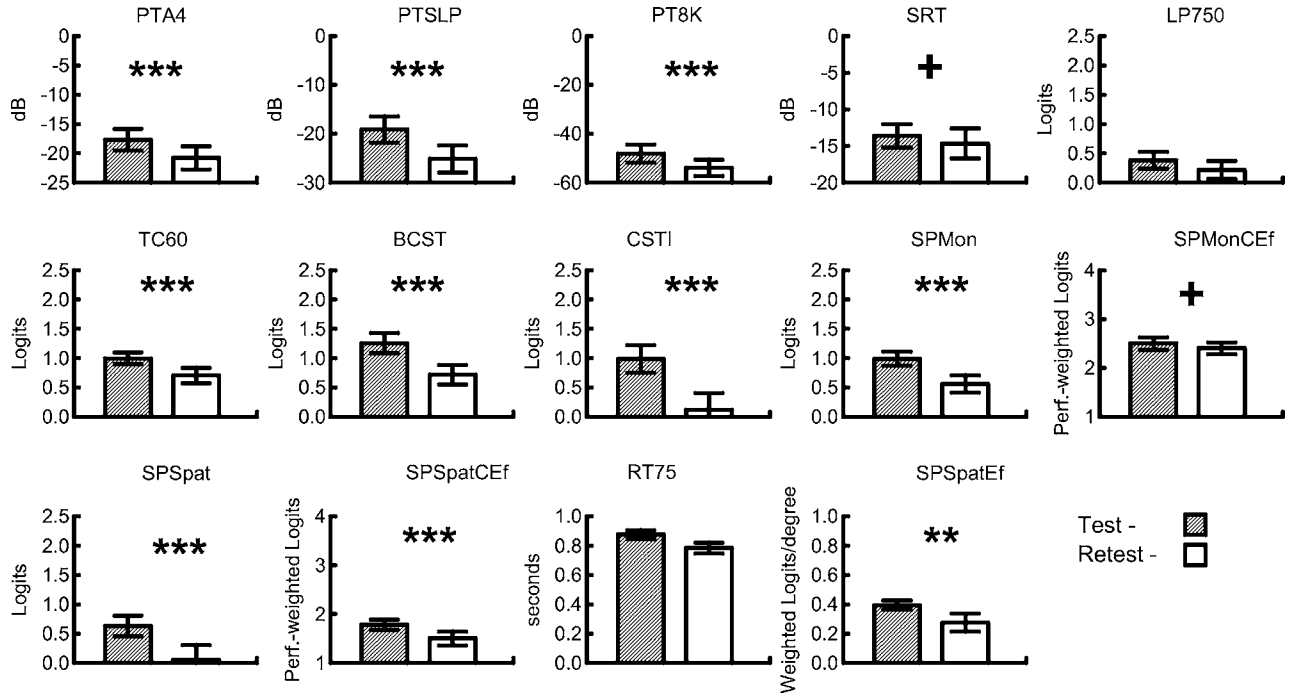


FIG. 2. Performance means and standard deviations for the 29 subjects in the first phase (solid bars) and second phase (open bars) for four audiometric tests and ten tests of speech understanding in interference, averaged across the right and left ears. Data for individual subjects in the second testing phase were interpolated to correspond to a 5.27-year lapse between phases. The *P*–values above the data bars indicate the significance level for a permutation test of the difference between phases for the raw data. The measure acronyms are **PTA4**: average of four pure-tone thresholds; **PTSLP**: slope of the audiogram (4 kHz threshold minus 0.5–kHz threshold); **PT8K**: threshold at 8 kHz; **SRT**: speech recognition threshold; **LP750**: recognition of speech low-pass filtered at 750 Hz; **TC60**: recognition of speech compressed 60%; **CSTI**: sentence intelligibility in the presence of a monaurally mixed sentence by another speaker; **BCST**: sentence intelligibility in the presence of a sentence presented contralaterally by another speaker. **SPMon**: average monaural SPIN test score; **SPMonCEf**: monaural SPIN sentence context effect; **SPSpat**: average scores for spatially distributed speech target and four babble sources; **SPSpatCEf**: sentence context effect for spatially distributed SPIN test; **RT75:** reverberation time estimated to correspond to 75% word recognition. Statistical significance of the difference between results obtained in the two testing phases (see Table I) is indicated for each measure ($^{***}: P<0.001$, $^{**}: P<0.005$, $+: P<0.05$).

on randomization. Observations were paired by subject and by ear across the two phases, which gave 812 differences (14 measures $\times$ 29 subjects $\times$ 2 ears). We sought to determine whether the observed differences between phases were larger than chance would account for if the measurements in each pair had been labeled randomly as arising from phase 1 or from phase 2, as if by tossing a fair coin 812 times. To that end, each difference was coded as −1, 0, or 1, depending on whether the subject's measure for that ear at phase 1 was larger than (−1), equal to (0), or smaller than (+1) at phase 2. We then compared the observed mean of these coded data with the distribution of means that would arise from random assignment by coin tossing.[6] This allowed us to assign a *P*–value to the hypothesis that there was no change in each measure between testing phases. We made the analogous test for all the measures pooled together, and for the measures pooled by laterality. (These tests involving the mean of the coded responses are equivalent to tests involving the medians of the original responses.) The null hypothesis for these tests is that the collection of scores is fixed in advance, and that the measurement process is equivalent to a random assignment of one of each pair of responses to each of the two phases or two ears. In this paper, *P*–values not exceeding 0.05 are considered to be statistically significant.

Table I shows *P*–values[7] for the test of differences in individual measures by testing phase and laterality. For comparison, corresponding *P*–values obtained in one-tailed *t*–tests are also shown; when they differ, we trust the nonparametric test for reasons mentioned in Sec. II.[8] Table I also shows *P*–values for differences pooled across measures. The observed differences are statistically significant at level

TABLE I. *P*–values of differences under sign-of-data permutation test and *t*–test.

	Phase 1-phase 2 difference		Right-left laterality difference	
Measure	Permutation test	*t*–test	Permutation test	*t*–test
PTA4	0.0000	0.0000	0.2051	0.2908
PTSLP	0.0000	0.0000	0.0063	0.0043
P8KT	0.0002	0.0002	0.9703	0.9972
SRT	0.0214	0.0350	0.9186	0.9718
SPMon	0.0000	0.0000	0.0435	0.0073
SPMonCEf	0.0267	0.0001	0.7441	0.5838
SPSpat	0.0001	0.0000	0.3470	0.4055
SPSpatCEf	0.0000	0.0000	0.2559	0.0496
SPSpatEf	0.0027	0.0008	0.0004	0.0000
RT75	0.8209	0.2887	0.9760	0.9334
LP750	0.1144	0.0135	0.2483	0.2402
TC60	0.0001	0.0000	0.0220	0.0142
BCST	0.0006	0.0006	0.0099	0.0017
CSTI	0.0000	0.0000	0.0012	0.0004
Overall	0.0000	0.0000	0.0001	0.4235

0.0001: the evidence that the median of the measurements is different for the two epochs is quite strong. For all but two individual measures (**LP750** and **RT75**), the difference between phase 1 and phase 2 is statistically significant. There is an overall right-ear superiority, but it is statistically significant for fewer than half of the individual measures. Phase-laterality interaction was found to be statistically insignificant (P–value 0.35): laterality did not change significantly between the two testing phases.

The analyses above do not distinguish among audiometric and speech-related measures. Since the median of most of the measures—both audiometric and speech related—seems to have changed between phases, identical processes could drive their deterioration. However, if the variability changed differently for the two groups of measures, that could indicate that more than hearing loss is responsible for the decline in speech understanding. To investigate this possibility, we tested the hypotheses that the variability of the speech understanding measures is the same in both phases and that the variability of the audiometric measures is the same in both phases, against the alternative hypothesis that the variability is higher in the second phase.[9] Results indicate that the variability changed significantly only for the speech-related variables: the test does not reject the null hypothesis for the audiometric variables (P–value approximately 0.46), but rejects it for the speech-related variables (P–value approximately 0.03). This suggests that changes in central factors may play a role in the degradation of speech understanding—as proposed by other investigators (e.g., Frisina and Frisina, 1997; Gordon-Salant and Fitzgibbons, 1997).

B. Rate of decline of hearing sensitivity and of speech understanding in interference

A primary objective of the present research was to explore whether a decline of central function with age—beyond the decline of peripheral function—plays a role in the accelerating decline with age of the ability to understand speech. Comparing the rate of decline before the first testing phase with the rate of decline between the first and second testing phases can help address this question. A young group was used to estimate the elderly subjects' performance at a young age, so the analysis in this section has both cross-sectional and longitudinal aspects. For each elderly subject i and measure j, a "decline difference sign" d_{ij} was computed for each of the four audiometric variables and each of the eight measures of speech understanding in interference, **RT75**, **SPMon**, **SPMonCEf**, **SPSpat**, **SPSpatCEf**, **SPSpatEf**, **BCST**, and **CSTI**:

$$d_{ij} = \mathrm{Sign}\left[\frac{(m_{ij1} - m_{j0})}{(a_{i1} - a_0)} - \frac{(m_{ij2} - m_{ij1})}{(a_{i2} - a_{i1})}\right], \quad (1)$$

where m_{j0} is the reference value of measure j in the young control group, m_{ij1} is measure j for subject i at phase 1, m_{ij2} is the corresponding datum at phase 2, a_0 is the reference (young) age, a_{i1} is the age of subject i at phase 1, and a_{i2} is the subject's age at phase 2.[10] The decline difference sign compares the average slope between the reference age (22.18 years) and age at phase 1 to the average slope between age at phase 1 and age at phase 2. If the deterioration accelerates—i.e., if the measure declines more rapidly on average between phase 1 and phase 2 than between the reference age and phase 1—then d_{ij} is −1; if the two rates are the same, d_{ij} is 0; and if the deterioration decelerates, d_{ij} is +1. Using the signs of the differences of rates puts all the measurements on an equal footing, making it possible to pool across measures to increase the power of the test, and reducing the influence of outliers.

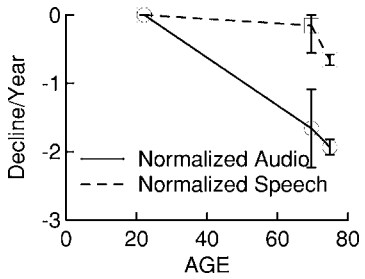


FIG. 3. Average normalized decline of three audiometric measures (solid line) and eight measures of speech understanding in interference (broken line), as a function of the subjects' average age.

A permutation test was used to test the null hypothesis that the change in rate of decline of a measure does not depend on whether the measure is audiometric or speech related against the alternative hypothesis that the change in rate of decline for the eight speech-related variables tends to be larger than that of the four audiometric variables. The test statistic was the sum of the decline difference signs for the speech-related measures. The P–value for the null hypothesis was estimated by comparing the observed value of this test statistic with the distribution of values of the test statistic obtained by randomly labeling $4 \times 29 = 116$ slope ratios as audiometric and $8 \times 29 = 232$ slope ratios as speech related, 10 000 times. The estimated P–value was 0.032: the decline with age of speech-related measurements accelerates faster than the decline with age of audiometric measurements by an amount that is statistically significant. Figure 3 shows the performance decline per year as a function of age at the testing phase separately for the set of four audiometric variables and the set of eight measures of speech understanding in interference. Performance change at the reference age was, by definition, zero. The larger acceleration of decline in the speech-related measures compared to the acceleration of decline in the audiometric measures is clearly visible.

C. Differences in the relationship between audiometric and speech variables across the two testing phases

The analyses above indicate that both audiometric and speech understanding measures declined between the first and second testing phases, and that the variability of speech understanding increased between testing phases. The acceleration analysis in the previous section shows that speech understanding declined more rapidly later in life than audiometric measures did, relative to their decline between about age 20 and about age 60. These findings, however, do not explore the relationships among the variables. We have seen

TABLE II. Varimax rotated loading matrices from principal component analysis.

Audiometric measures			Speech measures					
Principal components	1	2		1	2	3	4	5
Percent variance accounted for	47.55	41.36		29.75	17.21	16.93	15.86	14.34
SRT	0.989	0.075	**SPSpCEf**	−0.943	−0.163	−0.186	−0.067	0.130
PTA4	0.904	0.378	**SPMonCEf**	−0.878	−0.270	−0.095	−0.117	−0.259
			SPSpatEf	0.736	0.180	0.277	0.151	0.479
PTSLP	0.076	0.923	**TC60**	0.610	0.268	0.396	0.504	0.260
PT8K	0.320	0.809						
			SPMon	0.222	0.800	0.136	0.285	0.379
			RT75	0.470	0.753	0.297	0.161	0.196
			BCST	0.241	0.128	0.919	0.201	0.014
			CSTI	0.208	0.385	0.633	0.560	0.018
			LP750	0.106	0.201	0.254	0.872	0.293
			SPSpat	0.138	0.323	−0.025	0.262	0.886

that measures of speech understanding are associated with both peripheral hearing and age, but is the association the same at the two testing phases? In this section, we test the hypothesis that the relationship between audiometric variables, age, and speech understanding is constant with time.

Testing this hypothesis requires much stronger assumptions than those made so far: (1) individual cases are independent and identically distributed, as if the subjects were drawn at random from a population,[11] (2) speech-related performance variables at the two phases have a linear relationship to the explanatory variables—age and audiometric variables, and (3) the speech performance variables have additive measurement errors that have the same distribution, have zero expected value, and are independent for different individuals, but that need not be independent across testing phases nor across measurements for a given individual.

To reduce the dimensionality of the problem, the largest principal components of the four audiometric and the ten speech-related variables were used instead of all 14 raw measures. Enough principal components were retained to account for about 90% of the total variance of the raw measures: the top two audiometric principal components accounted for 88.9% of the variance of the audiometric measures, and the top five speech principal components accounted for 94.0% of the variance of the speech understanding measures. The loading matrices (rotated for easier interpretability) are shown in Table II. The first principal component of speech, $\mathbf{Speech_1}$, is related mainly to SPIN sentence context effect, spatial separation of speech target and babble sources, and temporal distortion; $\mathbf{Speech_2}$, to monaural processing of speech in babble and reverberation; $\mathbf{Speech_3}$, to sentence-based interference; $\mathbf{Speech_4}$, to spectral distortion; and $\mathbf{Speech_5}$, to spatially distributed speech processing in babble. $\mathbf{Audio_1}$ summarizes overall hearing level and $\mathbf{Audio_2}$, high-frequency hearing. The model in the null hypothesis is

$$\mathbf{Speech}_{ij} = a_0 + a_1\mathbf{Age}_{ij} + a_2\mathbf{Audio}_{1ij} + a_3\mathbf{Audio}_{2ij} + error_{ij}, \quad i = 1,2,\ldots,5; \quad j = 1,2,\ldots,29. \tag{2}$$

We wish to test the null hypothesis that, for $i=1,2,\ldots,5$, each of the principal components of the speech measures, the coefficients a_0, a_1, a_2, and a_3, are the same at both phases. The test statistic compares the fit of a model that requires each coefficient a_{ij} to be the same at the two phases (the *restricted* model) with the fit of a model that allows the coefficients to differ (the *unrestricted* or *full* model). The full model will always fit at least as well as the restricted model, because it contains the restricted model as a special case. The measure of fit we use is related to the Chow statistic (Kennedy, 1998, pp. 229–230):

$$\frac{(RSS_R - RSS_U)/k}{RSS_U/(N-k)}, \tag{3}$$

where RSS_R is the sum of squared residuals from the restricted model, RSS_U is the sum of squared residuals from the unrestricted model, k is the number of variables in each regression ($k=4$ here), and N is the number of data in the regression (twice the number of subjects, because there are two testing phases).[12] To test the overall hypothesis that any of the relationships changed, we need to combine the Chow statistics for the five speech variables. Using the maximum of the Chow statistics would be particularly sensitive to a change in only one of the relationships; using the sum of the Chow statistics is more sensitive to small changes in several of the relationships.[13] We report the result of tests based on the sum of the Chow statistics.[14]

To determine whether the observed improvement in this summary measure of fit would occur frequently by chance if the model were correct and the coefficients were constant across phases, a bootstrap approach was used.[15] Starting with the best-fitting model that restricts the coefficients to be equal across phases, we calculate 29 vectors, each consisting

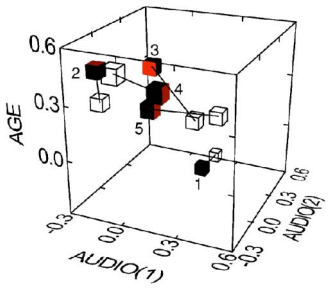


FIG. 4. Estimated coefficients in the linear model of Eq. (2) for the top five principal components of speech understanding measures as a function of age and the top two principal components of audiometric measures, in phase 1 (solid symbols) and phase 2 (open symbols).

of five pairs of residuals for each of the 29 subjects at the two phases (one pair for each of the five speech principal components at the two phases). Then, vectors of residuals are drawn at random with replacement from the population of 29 vectors of residuals. We add these residuals to the predictions of the best-fitting model and fit the restricted and unrestricted models to these synthetic data. By repeating this process 1000 times, the sampling distribution of the sum of the Chow statistics was estimated under the assumption that the null hypothesis is true. That allowed us to assign an approximate P–value to the null hypothesis.

Differences in the coefficients across phases were statistically significant ($P \approx 0.017$): On the assumption that individual cases are independent and identically distributed, we would reject the hypothesis that the first five principal components of the speech-related variables have the same linear relationship to age and the first two principal components of the audiometric variables at the two phases. Figure 4 shows the five pairs of coefficients of the three predictors, age, and the first two audiometric principal components. The coefficient vectors differ between phases, that is, the five principal components of the speech measures are associated with age and hearing sensitivity differently at phase 1 and at phase 2. With increasing age, the monaural perception of speech in babble and in reverberation changes, as does comprehension of sentences in the presence of another simultaneous sentence. As overall hearing threshold level changes, so does the perception of spectrally distorted and spatially distributed speech. As high-frequency loss changes, so does the perception of temporally distorted speech and the utility of sentence context.

IV. DISCUSSION

The most compelling observation of this study is that auditory performance in advanced age declines significantly over a period as short as 5 years in subjects whose initial hearing loss is only mild to moderate. This decline manifests in pure-tone thresholds and in almost all speech understanding performance measures. Although auditory thresholds might not characterize the status of the peripheral hearing system completely, understanding speech in interference clearly depends on auditory thresholds. In every reasonable model of speech perception, the peripheral auditory system extracts spectral-temporal features of the stimulus, features that comprise the input to higher centers. Using one such simple model, in which speech perception measures depend linearly on age and pure-tone thresholds (Fig. 4), we found that the relationships between measures of speech perception on the one hand, and age and auditory thresholds on the other, changes with age. This suggests that the nature of speech perception changes with age. While auditory thresholds gauge the role of the peripheral auditory system in speech perception, the effect of age includes the degradation of central structures responsible for low-rate temporal processing, among other things. These two effects can be regarded as complementary and, as Fig. 3 shows, they progress at different rates.

Audiometric thresholds increased by roughly 6.32 dB/decade, which is consistent with threshold changes reported by others (Davis *et al.*, 1990; Gates *et al.*, 1990; Ostri and Parving, 1991). The decline in speech understanding measures, illustrated in Fig. 2 and Table I, agrees with studies showing an increasing presbycusic loss accompanied by deteriorating speech perception performance (e.g., Bergman *et al.*, 1976; Pedersen *et al.*, 1991; Frisina and Frisina, 1997).

The variability of speech-related measures increased significantly between phases, but the variability of audiometric measures did not. Since speech understanding depends on the combination of auditory and cognitive processes, it is not surprising that the variability of this combination might change with age even if the variability of audibility alone does not. The increase of intersubject variability of speech understanding with age may be a facet of the large individual differences observed across a variety of cognitive functions in aged populations (Salthouse, 1991). Speech perception relies in a complex way on many cues, which in turn depend, most likely in a nonlinear fashion, on the acoustic signal impinging on listeners' ears. Thresholding effects that cause some cues to become inaudible are another form of nonlinearity. These nonlinearities could also account for the increased variability of speech understanding relative to audibility. So could individual differences in how audible cues are used. The increased variability of differences of speech understanding measures compared with measures of audibility could also arise from differences between the processes involved in their declines. This hypothesis is also supported by Fig. 3, which shows an increase in the rate of decline of speech understanding in interference, i.e., an acceleration. Taken together, the comparative increase in variability of speech understanding compared with audiometric measures, the comparative acceleration of the decline in speech understanding compared with audiometric measures, and the apparent change with time in the relationship between speech understanding and age and audiometric variables[16] suggest that the decline in auditory performance during the later decades of life has two major components: (1) presbycusis due to deterioration of the peripheral auditory structures, and (2) loss of speech understanding involving a nonperipheral (i.e., central) component.

In fact, a centrally originating component of auditory decline has been proposed by many other investigators (Bergman, 1980; Jerger *et al.*, 1989; Pedersen *et al.*, 1991;

Jerger *et al.*, 1994; Pichora-Fuller *et al.*, 1995; Gordon-Salant and Fitzgibbons, 1997; Pichora-Fuller, 1997; Schneider *et al.*, 1998; Divenyi and Simon, 1999; Gordon-Salant and Fitzgibbons, 1999). What is the functional significance of this central component? Previous results (Bergman *et al.*, 1976; Gelfand *et al.*, 1988; Gordon-Salant and Fitzgibbons, 1993; Divenyi and Haupt, 1997a) suggest the decline affects most aspects of the temporal processing of simple and complex sounds (Gordon-Salant and Fitzgibbons, 1997)—and thus is quite likely to impede the ability to separate simultaneous speech sounds. Most temporal processing deficits affect the time range most useful for speech processing, i.e., amplitude modulation frequencies of 3 to 6 Hz corresponding to the syllabic rate (e.g., Takahashi and Bacon, 1992; Humes *et al.*, 1994; Fitzgibbons and Gordon-Salant, 1995). In our study, the **CSTI** and **BCST** measures are affected by the speech of even a single interfering talker, which can degrade the envelope information at these low AM frequencies. However, speech also can be degraded at higher modulation rates (20 to 50 Hz), where fluctuations important for segmental processing occur. One example is reverberation, which reduces intelligibility by degrading speech envelopes at modulation rates up to 50 Hz (Avendano and Hermansky, 1996). The decline in understanding of reverberant speech, also reported by others (Duquesnoy and Plomp, 1980; Nabelek and Robinson, 1982; Nabelek and Dagenais, 1986; Helfer and Huntley, 1991; Helfer, 1992; Gordon-Salant and Fitzgibbons, 1995, 1999), indicates that temporal information processing in both the syllabic and segmental ranges is impaired in the elderly. Deterioration of temporal processing in the 3- to 50-Hz envelope frequencies thus could cause the decline of speech understanding in interference observed here and elsewhere (Bergman *et al.*, 1976). Babble, reverberation, and other modulated nontarget sounds have syllabic-rate fluctuations difficult to distinguish from those of the target. If the ability to discern those fluctuations declines, understanding of reverberant speech or speech in babble noise presumably will be affected. Our laboratory has found that the ability of elderly listeners to segregate nonspeech streams using only envelope information is highly correlated with their ability to understand speech in babble and reverberation (Divenyi, 2004a). Therefore, studying the connection between speech understanding in interference and the decline of auditory temporal processing at low modulation rates might give insight into auditory losses beyond presbycusis. Unfortunately, neither the precise characteristics of the age-related decline in temporal processing nor its precise role in speech perception impairments is yet clear, although research in the area is progressing (Fitzgibbons and Gordon-Salant, 1996; Gordon-Salant and Fitzgibbons, 1999; Snell and Frisina, 2000). It seems that the decline in speech understanding found here is likely to reflect a general deficit of auditory temporal processing in the syllabic and subsyllabic range (Fitzgibbons and Gordon-Salant, 2004; Gordon-Salant and Fitzgibbons, 2004), i.e., a time scale significantly longer than changes in the auditory periphery would account for.

Right-ear superiority in elderly subjects has been reported by several investigators (Divenyi and Haupt, 1997a; Jerger and Chmiel, 1997; Greenwald and Jerger, 2001). The present study confirms this (see Table I). Because one of our subject selection criteria was symmetrical audiograms, it is not surprising that we saw no laterality effect for audibility measures. The right-ear advantage we find in the present study therefore strongly suggests a central imbalance. This conclusion is consistent with physiological-anatomical observations that attribute much of the right-ear advantage (or, rather, the left-ear disadvantage) in aging to cortical impairment—a deterioration of the fibers in the corpus callosum (Jerger *et al.*, 1995; Chmiel *et al.*, 1997). Although we might have predicted an increase in the right-ear advantage from testing phase 1 to testing phase 2, no significant increase was observed, suggesting that the development of central imbalance occurs over a longer time than the duration of our study.

We find that the effect of sentence context can be measured with the SPIN test, although the context effect interacts nonmonotonically with the overall SPIN score: both poor and excellent performances reduce the apparent utility of sentence context, but for opposite reasons. Subjects who cannot understand speech in noise at all cannot pick out context to use it, and for subjects who can understand every word, context is redundant. To be a useful indicator of the higher-order cortical function of sentence processing, the context effect needs to be weighted by performance level.

It is perhaps counterintuitive that the effect of spatial separation of speech and babble sources was statistically significant but small, possibly because several subjects had difficulty understanding the targets at any separation. There is, however, another likely explanation: auditory localization deteriorates with age (Herman *et al.*, 1977; Divenyi and Haupt, 1997a), so poor localization could have reduced or obliterated the benefits of increasing the spatial separation between sources. Because auditory localization in the elderly has been measured only with simple signals (clicks or tone pips), the last explanation remains a possibility until tested directly, for example, by replicating on an elderly population the (virtual) free-field speech masking level difference experiments of Shinn-Cunningham and her colleagues (Shinn-Cunningham *et al.*, 2001).

We want to reiterate the importance and utility of nonparametric methods for analyzing auditory performance data. When the assumptions underlying parametric tests are satisfied, nonparametric tests tend to be more conservative, possibly increasing the risk of missing a real effect. But because their assumptions are less restrictive, nonparametric tests, such as the permutation-based tests used in this study, tend to behave well in many situations in which the parametric tests do not, and we believe that the assumptions of usual parametric tests often are not satisfied in studies on hearing. Moreover, many nonparametric tests are in fact simpler and easier to understand than the corresponding parametric tests. We recommend these nonparametric methods, especially when the sample size is small, when subjects are not chosen at random, and when it is not reasonable to assume that the variables of interest have Gaussian distributions.

Can conclusions of the present study be generalized to the entire aging population? Because the criteria for enroll-

ment in this study excluded people with hearing loss worse than moderate or with significant medical problems, our subject sample does not represent the elderly population as a whole. Had the criteria for inclusion been relaxed, intersubject variability due to illness and auditory status might well have interfered with the effects we wanted to examine. We believe this study shows important effects of normal aging on speech understanding and that it can serve as a baseline against which to compare results on subject samples having auditory or other pathologies.

To conclude, our results suggest that the decline of speech understanding in interference in the seventh to ninth decades of life is similar to the presbycusic deterioration of hearing, with important differences: deterioration of speech understanding accelerates and its variability increases compared with the rate of deterioration and variability of absolute hearing thresholds, suggesting an accelerating decline of central auditory processing. This might have been discovered sooner if central auditory decline with age had received as much attention as peripheral. The central component of these deficits may also explain the limited success elderly hearing aid users have understanding conversations in noisy situations, especially in speech interference. Providing robust speech understanding in adverse listening conditions might eventually be possible using intelligent, automatic speech-separation devices still in development.[17]

ACKNOWLEDGMENTS

We are grateful to Brian Gygi, Ken Grant, and two anonymous reviewers for comments on an earlier version of the paper, and to David Freedman for numerous helpful conversations. The research has been supported by Grant No. R01-AG07998 from the National Institute on Aging and by the Veterans Affairs Biomedical and Laboratory Research and Development.

[1]The compression of the words was accomplished using the computer program ProAudio9, by the Cakewalk software company.

[2]Because performance decreased monotonically with increasing reverberation time, we used linear regression to interpolate or extrapolate to find **RT75**.

[3]The target sentences and babble were recorded digitally using an artificial head in anechoic space with five identical loudspeakers, one for the speech signal and four for the babble sources, positioned according to the babble span condition. The babble used in our spatial SPIN tests differed from that of the monaural SPIN tests: the total number of talkers was 16 instead of 12, to keep the number of male and female talkers balanced at each babble location.

[4]The SPIN test was presented at a speech-to-babble ratio lower than the recommended 8 dB S/B, to increase the intersubject performance variability.

[5]This refers to energetic masking. Informational masking is a different issue; if the definitions advocated by several authors (e.g., Durlach *et al.*, 2003) were adopted, all speech understanding decline shown in the present paper could be assigned to informational masking.

[6]This test is analogous to the sign test for the median (see, e.g., Lehmann and D'Abrera, 1988).

[7]The *P*–value is the probability that the test statistic would have a value as extreme as or more extreme than its observed value if the null hypothesis is true. Small *P*–values are evidence against the null hypothesis—they indicate that the data would be unlikely if the null hypothesis is true.

[8]*P*–values for the *t*–test were computed using Student's *t* distribution. *P*–values for the permutation test were computed using the binomial distribution (for dichotomous data) or approximated by simulation using 10 000 pseudorandom permutations of the data (for data that can take more than two values).

[9]We tested the null hypothesis that the variability did not change across testing phases using a permutation test. To calculate the test statistic, we subtracted from the measurements of each variable at each phase the median of those measurements. For each variable, we divided the measurements at both phases by the median absolute deviation of the measurements at the first phase, so that all sets of first-phase measurements have median absolute deviation equal to unity. We then calculated the ratio of the sum (across variables) of the median absolute deviations of the second-phase measurements to the sum of the median absolute deviations of the first-phase measurements. This ratio will tend to be larger than unity if the variability is higher in the second phase, which is the alternative hypothesis of interest. We estimated the *P*–value of the observed test statistic using a randomization: For each individual and each variable, we randomly assigned one of the two values to the first phase and one to the second phase, then recalculated the test statistic for that random permutation, repeating the process 1000 times.

[10]We did not interpolate measures to standardize the time lapse, as we did in the previous section.

[11]The assumption that the subjects are like a random sample is implausible, even though it is common in audiological research. The assumption that the relationship between the explanatory and speech variables is linear might be a reasonable approximation; more data are required to test it.

[12]The ratio (3) would have an *F* distribution under the null hypothesis if all the data were drawn independently from normal distributions. This assumption can be avoided by using resampling to estimate critical values for the resulting tests, the approach taken here.

[13]It also would be reasonable to use the ratio of the sum of normalized numerators to the sum of normalized denominators in the ratio (3).

[14]The other approaches to combining the Chow statistics gave similar results.

[15]The bootstrap approximation of the distribution of an estimator is the distribution of the estimator when it is applied to random samples with replacement from the data; see Efron (1982).

[16]This last assertion depends on the assumption that the relationship is linear in the principal components, as described more fully in Sec. III C of the results.

[17]For a review, see Divenyi (2004b).

Avendano, C., and Hermansky, H. (**1996**). "Study on the dereverberation of speech based on temporal envelope filtering," Proc. ICSLP'96, pp. 889–892.

Beasley, D. S., Schwimmer, S., and Rintelmann, W. F., (**1972**). "Intelligibility of time-compressed CNC monosyllables," J. Speech Hear. Res. **15**, 340–350.

Bergman, M. (**1980**). *Aging and the Perception of Speech* (University Park, Baltimore, MD).

Bergman, M., Blumfield, V. G., Cascardo, D., Dash, B., Levitt, H., and Margulies, M. K. (**1976**). "Age-related decrement in hearing for speech. Sampling and longitudinal studies," J. Gerontol. **31**, 533–538.

Bilger, R. C., Nuetzel, J. M., Rabinowitz, W. M., and Rzeczkowski, C. (**1984**). "Standardization of a test of speech perception in noise," J. Speech Hear. Res. **27**, 32–48.

Blumenfield, V. G., Bergman, M., and Millner, E. (**1969**). "Speech discrimination in an aging population," J. Speech Hear. Res. **12**, 210–217.

Brant, L. J., and Fozard, J. L. (**1990**). "Age changes in pure-tone hearing thresholds in a longitudinal study of normal human aging," J. Acoust. Soc. Am. **88**, 813–820.

Cheesman, M. F., Hepburn, D., Armitage, J. C., and Marshall, K. (**1995**). "Comparison of growth of masking functions and speech discrimination abilities in younger and older adults," Audiology **34**, 321–333.

Chmiel, R., and Jerger, J. (**1996**). "Hearing aid use, central auditory disorder, and hearing handicap in elderly persons," J. Am. Acad. Audiol **7**, 190–202.

Chmiel, R., Jerger, J., Murphy, E., Pirozzolo, F., and Tooley-Young, C. (**1997**). "Unsuccessful use of binaural amplification by an elderly person," J. Am. Acad. Audiol **8**, 1–16.

Davis, A. C., Ostri, B., and Parving, A. (**1990**). "Longitudinal study of hearing," Acta Oto-Laryngol., Suppl. **476**, 12–22.

Davison, A. C., and Hinkley, D. V. (**1997**). *Bootstrap Methods and Their Application* (Cambridge U.P., Cambridge, UK).

Divenyi, P. (**2004a**). "Masking the feature-information in multi-stream

speech-analogue displays," in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Academic, New York), pp. 269–281.
Divenyi, P. L. (**2004b**). (ed.) *Speech Separation by Humans and Machines* (Kluwer Academic, New York).
Divenyi, P. L., and Haupt, K. M. (**1997a**). "Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. I. Age and laterality effects," Ear Hear. **18**, 42–61.
Divenyi, P. L., and Haupt, K. M. (**1997b**). "Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. II. Correlation analysis," Ear Hear. **18**, 100–113.
Divenyi, P. L., and Haupt, K. M. (**1997c**). "Audiological correlates of speech understanding deficits in elderly listeners with mild-to-moderate hearing loss. III. Factor representation," Ear Hear. **18**, 189–201.
Divenyi, P. L., and Simon, H. J. (**1999**). "Hearing in aging: Issues old and new," Curr. Opin. Otolaryngol. **7**, 282–289.
Dubno, J. R., Dirks, D. D., and Morgan, D. E. (**1984**). "Effects of age and mild hearing loss on speech recognition in noise," J. Acoust. Soc. Am. **76**, 87–96.
Dubno, J. R., Lee, F., Matthews, L. J., and Mills, J. H. (**1997**). "Age-related and gender-related changes in monaural speech recognition," J. Speech Lang. Hear. Res. **40**, 444–452.
Duquesnoy, A. J., and Plomp, R. (**1980**). "Effect of reverberation and noise on the intelligibility of sentences in cases of presbyacusis," J. Acoust. Soc. Am. **68**, 537–544.
Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., and Kidd, G., Jr. (**2003**). "Informational masking: counteracting the effects of stimulus uncertainty by decreasing target-masker similarity," J. Acoust. Soc. Am. **114**, 368–379.
Efron, B. (**1982**). *The Jackknife, the Bootstrap, and other Resampling Plans* (Society of Industrial and Applied Mathematicians (SIAM), Philadelphia).
Fitzgibbons, P. J., and Gordon-Salant, S. (**1995**). "Age effects on duration discrimination with simple and complex stimuli," J. Acoust. Soc. Am. **98**, 3140–3148.
Fitzgibbons, P. J., and Gordon-Salant, S. (**1996**). "Auditory temporal processing in elderly listeners," J. Am. Acad. Audiol **7**, 183–189.
Fitzgibbons, P. J., and Gordon-Salant, S. (**2004**). "Age effects on discrimination of timing in auditory sequences," J. Acoust. Soc. Am. **116**, 1126–1134.
Frisina, D. R., and Frisina, R. D. (**1997**). "Speech perception and presbycusis: relations to possible neural mechanisms," Hear. Res. **106**, 95–104.
Gates, G. A., and Cooper, J. C., Jr. (**1991**). "Incidence of hearing decline in the elderly," Abstracts of the 14th Midwinter Research Meeting, Assoc Res Otolaryng Vol. **34**.
Gates, G. A., Cooper, J. C., Jr., Kannel, W. B., and Miller, N. J. (**1990**). "Hearing in the Elderly: The Framingham Cohort, 1983-1985. Part I. Basic Audiometric Test Results," Ear Hear. **11**, 247–256.
Gelfand, S. A., Piper, N., and Silman, S. (**1985**). "Consonant recognition in quiet as a function of aging among normal hearing subjects," J. Acoust. Soc. Am. **78**, 1198–1206.
Gelfand, S. A., Piper, N., and Silman, S. (**1986**). "Consonant recognition in quiet and in noise with aging among normal hearing subjects," J. Acoust. Soc. Am. **80**, 1589–1598.
Gelfand, S. A., Ross, L., and Miller, S. (**1988**). "Sentence reception in noise from one versus two sources: Effect of aging and hearing loss," J. Acoust. Soc. Am. **83**, 248–256.
Gelfand, S. A., Hoffman, S., Waltzman, S. B., and Piper, N. (**1980**). "Dichotic CV recognition at various interaural temporal onset asynchronies: effect of age," J. Acoust. Soc. Am. **68**, 1258–1261.
Gordon-Salant, S., and Fitzgibbons, P. J. (**1993**). "Temporal factors and speech recognition performance in young and elderly listeners," J. Speech Hear. Res. **36**, 1276–1285.
Gordon-Salant, S., and Fitzgibbons, P. J. (**1995**). "Recognition of multiply degraded speech by young and elderly listeners," J. Speech Hear. Res. **38**, 1150–1156.
Gordon-Salant, S., and Fitzgibbons, P. J. (**1997**). "Selected cognitive factors and speech recognition performance among young and elderly listeners," J. Speech Lang. Hear. Res. **40**, 423–431.
Gordon-Salant, S., and Fitzgibbons, P. J. (**1999**). "Profile of auditory temporal processing in older listeners," J. Speech Lang. Hear. Res. **42**, 300–311.
Gordon-Salant, S., and Fitzgibbons, P. J. (**2004**). "Effects of stimulus and noise rate variability on speech perception by younger and older adults," J. Acoust. Soc. Am. **115**, 1808–1817.
Greenwald, R. R., and Jerger, J. (**2001**). "Aging affects hemispheric asymmetry on a competing speech task," J. Am. Acad. Audiol **12**, 167–173.
Helfer, K. S. (**1992**). "Aging and the binaural advantage in reverberation and noise," J. Speech Hear. Res. **35**, 1394–1401.
Helfer, K. S., and Huntley, R. A. (**1991**). "Aging and consonant errors in reverberation and noise," J. Acoust. Soc. Am. **90**, 1786–1796.
Herman, G. E., Warren, L. R., and Wagener, J. W. (**1977**). "Auditory lateralization: Age differences in sensitivity to dichotic time and amplitude cues," J. Gerontol. **32**, 187–191.
Humes, L. E. (**1991**). "Understanding the speech-understanding problems of the hearing impaired," J. Am. Acad. Audiol **2**, 59–69.
Humes, L. E., and Christopherson, L. (**1991**). "Speech identification difficulties of hearing-impaired elderly persons: The contributions of auditory processing deficits," J. Speech Hear. Res. **34**, 686–693.
Humes, L. E., and Roberts, L. (**1990**). "Speech recognition difficulties of the hearing-impaired elderly: The contributions of audibility," J. Speech Hear. Res. **33**, 726–735.
Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C., and Lee, L. (**1994**). "Factors associated with individual differences in clinical measures of speech recognition among the elderly," J. Speech Hear. Res. **37**, 465–474.
Jerger, J., and Chmiel, R. (**1997**). "Factor analytic structure of auditory impairment in elderly persons," J. Am. Acad. Audiol **8**, 269–276.
Jerger, J., Chmiel, R., Allen, J., and Wilson, A. (**1994**). "Effects of age and gender on dichotic sentence identification," Ear Hear. **15**, 274–286.
Jerger, J., Silman, S., Lew, H. L., and Chmiel, R. (**1993**). "Case studies in binaural interference: Converging evidence from behavioral and electrophysiologic measures," J. Am. Acad. Audiol **4**, 122–131.
Jerger, J., Alford, B., Lew, H. L., Rivera, V., and Chmiel, R. (**1995**). "Dichotic listening, event-related potentials, and interhemispheric transfer in the elderly," Ear Hear. **16**, 482–498.
Jerger, J., Stach, B. A., Johnson, K., Loiselle, L. H., and Jerger, S. (**1990**). "Patterns of abnormality in dichotic listening in the elderly," in *Presbycusis and other age related aspects—14th Danavox Symposium*, edited by J. H. Jensen (Danavox, Copenhagen), pp. 143–150.
Jerger, J. F., and Hayes, D. (**1977**). "Diagnostic speech audiometry," Arch. Otolaryngol. **103**, 216–222.
Jerger, J. F., and Jordan, C. (**1992**). "Age-related asymmetry on a cued-listening task," Ear Hear. **13**, 272–277.
Jerger, J. F., Jerger, S., Oliver, T., and Pirozzolo, F. (**1989**). "Speech understanding in the elderly," Ear Hear. **10**, 79–89.
Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337–1351.
Kennedy, P. (**1998**). *A Guide to Econometrics* (MIT, Cambridge, MA).
Konkle, D. F., Beasley, D. S., and Bess, F. M. (**1977**). "Intelligibility of time-altered speech in relation to chronological aging," J. Speech Hear. Res. **20**, 108–115.
Lehmann, E. L., and D'Abrera, H. J. M. (**1988**). *Nonparametrics: Statistical Methods Based on Ranks*, 2nd ed. (McGraw Hill, New York).
Lynn, G. E., and Gilroy, J. (**1972**). "Neuro-audiological abnormalities in patients with temporal lobe tumors," J. Neurol. Sci. **17**, 167–184.
Lynn, G. E., and Gilroy, J. (**1975**). "Effects of brain lesions on the perception of monotic and dichotic speech stimuli," in *Central Auditory Processing Disorders*, edited by M. D. Sullivan (Univ. of Nebraska Medical Center, Omaha), pp. 47–83.
Marshall, L. (**1981**). "Auditory processing in aging listeners," J. Speech Hear Disord. **46**, 226–240.
Martini, A., Bovo, R., Agnoletto, M., Da Col, M., Drusian, A., Liddeo, M., and Morra, B. (**1988**). "Dichotic performance in elderly Italians with Italian stop consonant-vowel stimuli," Audiology **27**, 1–7.
Morrell, C. H., Gordon-Salant, S., Pearson, J. D., Brant, L. J., and Fozard, J. L. (**1996**). "Age- and gender-specific reference ranges for hearing level and longitudinal changes in hearing level," J. Acoust. Soc. Am. **100**, 1949–1967.
Nabelek, A. K., and Dagenais, P. (**1986**). "Vowel errors in noise and in reverberation by hearing-impaired listeners," J. Acoust. Soc. Am. **80**, 741–748.
Nabelek, A. K., and Robinson, P. K. (**1982**). "Monaural and binaural speech perception in reverberation for listeners of various ages," J. Acoust. Soc. Am. **71**, 1242–1248.
Ostri, B., and Parving, A., (**1991**). "A longitudinal study of hearing impairment in male subjects—an 8-year follow-up," Br. J. Audiol. **25**, 41–48.
Pearson, J. D., Morrell, C. H., Gordon-Salant, S., Brant, L. J., Metter, E. J., Klein, L. L., and Fozard, J. L. (**1995**). "Gender differences in a longitudi-

nal study of age-associated hearing loss," J. Acoust. Soc. Am. **97**, 1196–1205.
Pedersen, K. E., Rosenhall, U., and Møller, M. B. (**1991**). "Longitudinal study of changes in speech perception between 70 and 81 years of age," Audiology **30**, 201–211.
Pichora-Fuller, M. K. (**1997**). "Language comprehension in older listeners," J. Speech Lang. Pathol. Audiol. **21**, 125–142.
Pichora-Fuller, M. K., Schneider, B. A., and Daneman, M. (**1995**). "How young and old adults listen and remember speech in noise," J. Acoust. Soc. Am. **97**, 593–608.
Salthouse, T. A. (**1991**). *Theoretical perspectives on cognitive aging* (Erlbaum, Hillsdale, NJ).
Schneider, B., Speranza, F., and Pichora-Fuller, M. K. (**1998**). "Age-related changes in temporal resolution: Envelope and intensity effects," Can. J. Exp. Psychol. **52**, 184–191.
Shinn-Cunningham, B. G., Schickler, J., Kopco, N., and Litovsky, R. (**2001**). "Spatial unmasking of nearby speech sources in a simulated anechoic environment," J. Acoust. Soc. Am. **110**, 1118–1129.
Snell, K. B., and Frisina, D. R. (**2000**). "Relationships among age-related differences in gap detection and word recognition," J. Acoust. Soc. Am. **107**, 1615–1626.
Stach, B. A., Jerger, J. F., and Fleming, K. A. (**1985**). "Central presbyacusis: a longitudinal case study," Ear Hear. **6**, 304–306.
Takahashi, G. A., and Bacon, S. P. (**1992**). "Modulation detection, modulation masking, and speech understanding in noise in the elderly," J. Speech Hear. Res. **35**, 1413–1429.
Tun, P. A. (**1998**). "Fast noisy speech: age differences in processing rapid speech with background noise," Psychol. Aging **13**, 424–434.
van Rooij, J. C. G. M., and Plomp, R. (**1990**). "Auditive and cognitive factors in speech perception by elderly listeners. II: Multivariate analyses," J. Acoust. Soc. Am. **88**, 2611–2624.
Willeford, J. A. (**1985**). "Sentence tests of central auditory dysfunction," in *Handbook of Clinical Audiology*, 3rd edition, edited by J. Katz (Williams & Wilkins, Baltimore), pp. 404–420.

Vowel perception by noise masked normal-hearing young adults[a)]

Carolyn Richie,[b)] Diane Kewley-Port, and Maureen Coughlin
Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 29 July 2003; revised 2 March 2005; accepted 29 April 2005)

This study examined vowel perception by young normal-hearing (YNH) adults, in various listening conditions designed to simulate mild-to-moderate sloping sensorineural hearing loss. YNH listeners were individually age- and gender-matched to young hearing-impaired (YHI) listeners tested in a previous study [Richie *et al.*, J. Acoust. Soc. Am. **114**, 2923–2933 (2003)]. YNH listeners were tested in three conditions designed to create equal audibility with the YHI listeners; a low signal level with and without a simulated hearing loss, and a high signal level with a simulated hearing loss. Listeners discriminated changes in synthetic vowel tokens /ɪ e ɛ ʌ æ/ when F1 or F2 varied in frequency. Comparison of YNH with YHI results failed to reveal significant differences between groups in terms of performance on vowel discrimination, in conditions of similar audibility by using both noise masking to elevate the hearing thresholds of the YNH and applying frequency-specific gain to the YHI listeners. Further, analysis of learning curves suggests that while the YHI listeners completed an average of 46% more test blocks than YNH listeners, the YHI achieved a level of discrimination similar to that of the YNH within the same number of blocks. Apparently, when age and gender are closely matched between young hearing-impaired and normal-hearing adults, performance on vowel tasks may be explained by audibility alone. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1944053]

PACS number(s): 43.71.Ky, 43.71.Es, 43.71.Lz [KWG] Pages: 1101–1110

I. INTRODUCTION

Listeners with mild-to-moderate sensorineural hearing loss may perform more poorly than normal-hearing listeners on tests of speech perception. However, the specific causes for differences in performance between hearing-impaired listeners and normal-hearing listeners are not well understood. There is some evidence for both cognitive and auditive factors affecting listeners' performance on auditory tests of speech recognition (e.g., Era *et al.*, 1986; Rakerd *et al.*, 1996). However, converging evidence suggests that auditive factors are the primary contributor to results in tests of auditory performance (e.g., Humes *et al.*, 1994; Jerger *et al.*, 1991; Nábèlek, 1988; van Rooij and Plomp, 1992). While age is also a factor known to affect performance on tests of speech perception, the general reduction in auditory sensitivity associated with aging is primarily responsible for poor performance of elderly listeners, on tests of speech recognition (e.g., Nábèlek, 1988; Richie *et al.*, 2003).

The effects of sensorineural hearing loss on speech perception may include more than the reduced or attenuated sensitivity of the auditory system (cf. Plomp, 1978). As Van Tasell *et al.* (1982) point out, sensorineural hearing impairment may distort suprathreshold speech in a number of ways. Speech perception may be distorted through abnormal growth of loudness (i.e., recruitment), impaired frequency resolution (due to broadened auditory filters; see Dubno and Dirks, 1989; Peters and Moore, 1992; Sommers and Humes, 1993), and poor temporal resolution (Moore, 1985).

Research to equate listening conditions between normal-hearing and hearing-impaired listeners, for purposes of speech perception tasks, has used a number of different techniques. Some studies have attempted to spectrally attenuate speech according to properties of an impaired audiogram, in order to simulate a hearing loss in normal-hearing listeners. Others have attempted to present spectrally shaped signals to hearing-impaired listeners in order to simulate the gain function of a hearing aid (for example, see Baer and Moore, 1997). However, most studies have presented some form of noise masking to normal-hearing listeners, in order to simulate a hearing loss. All three methods are a means of equating audibility with the latter method simulating additional characteristics of sensorineural hearing loss such as loudness recruitment and reduced dynamic range.

As Humes and Roberts (1990) point out, the speech recognition difficulties of hearing-impaired listeners are most accurately simulated by the introduction of a spectrally shaped masking noise into the ears of normal-hearing listeners. This procedure produces normal-hearing masked thresholds similar to those of hearing-impaired listeners, and loudness recruitment. Sommers and Humes (1993) demonstrate that noise masking in normal-hearing listeners may also mimic reduced frequency selectivity, another characteristic exhibited by listeners with sensorineural hearing loss.

Previous work on speech perception by hearing-impaired and noise masked normal-hearing listeners matched for audibility, however, has yielded mixed results. Leek *et al.* (1987) determined the minimum difference in amplitude between vowel spectral peaks that was required for 20 normal-

[a)]Portions of these data were presented at the June 2001 and June 2002 meetings of the Acoustical Society of America.

[b)]Current address: Communication Disorders, Butler University, Indianapolis, Indiana 46208; electronic mail: crichie@butler.edu

hearing (age range 20–39 years) and 6 hearing-impaired listeners (age range 25–74 years). The amplitude of three formants in four synthetic vowels was varied 1–8 dB over the other spectral components. In this task, subjects were asked to identify the vowels. Results showed that normal-hearing listeners could detect peak-to-trough differences of 1–2 dB with more than 75% accuracy. Hearing-impaired listeners with a flat moderate hearing loss of about 50 dB HL required a 6–7 dB peak-to-trough amplitude difference for similar identification performance. When normal-hearing listeners were noise masked to raise their pure tone thresholds to simulate the hearing loss, peak-to-trough differences of 4 dB were required for similar identification. This result suggests hearing-impaired listeners may have difficulty using closely spaced formants in vowel identification, due to smoothing of the internal spectrum by broadened auditory filters associated with sensorineural hearing loss.

Humes *et al.* (1987) measured nonsense syllable recognition by four hearing-impaired (17–56 years of age) and 12 normal-hearing listeners (19–32 years of age, three matched to each hearing-impaired listener). Results indicated that when audibility was matched between groups, two hearing-impaired listeners performed better than their noise masked controls and two hearing-impaired listeners' performance was similar to that of the noise masked listeners. More typically, however, Humes and colleagues have observed close agreement between the speech recognition performance of hearing-impaired listeners and normal-hearing subjects with simulated hearing loss (e.g., Humes and Christopherson, 1991).

Dubno and Schaefer (1992) compared frequency selectivity and consonant recognition in six hearing-impaired (58–73 years of age) and 18 noise masked normal-hearing listeners (18–38 years of age, three matched to each hearing-impaired listener). Masked thresholds in notched noise and narrow-band noise were obtained and compared between groups. Measurements of consonant recognition were obtained at various speech presentation levels. Results indicated that frequency selectivity is poorer for the hearing-impaired listeners than for masked normal-hearing listeners (i.e., auditory filters are broader), even when pure-tone thresholds were closely matched between groups. However, no consistent differences were found between the hearing-impaired and masked normal-hearing listeners in terms of consonant recognition. Though hearing-impaired listeners showed poorer than normal frequency selectivity, their speech recognition was equivalent to normal when speech-spectrum audibility was equated across listeners. These findings suggest that speech recognition by hearing-impaired listeners is not dependent upon reduced frequency selectivity (these findings were further supported in Dubno and Schaefer, 1995).

Results from these studies, which created conditions of similar audibility between noise masked normal-hearing and hearing-impaired listeners, while not entirely consistent, have produced some useful insights into consonant recognition. However, the implications for understanding the effects of sensorineural hearing loss generally on speech perception are unclear, and few speech perception tasks, to date, have focused on vowel perception. Since vowel perception abilities are reported to remain relatively intact, but may be degraded in hearing-impaired listeners, vowel perception is the focus of this study (e.g., Burkle *et al.*, 2004; Dorman *et al.*, 1985; Godfrey and Millay, 1980; Owens *et al.*, 1968; Richie *et al.*, 2003; Van Tasell *et al.*, 1987).

The current work attempts to determine whether young normal-hearing (YNH) listeners perform similarly to young hearing-impaired (YHI) listeners in conditions of equivalent audibility, for a vowel discrimination task. Specifically, noise masking was used to match YNH adult listeners to YHI listeners with mild-to-moderate sloping sensorineural hearing loss tested in a previous study (Richie *et al.*, 2003). Also, a frequency-shaped gain was applied to the speech stimuli for the YHI listeners to ensure audibility. The purpose was to examine basic speech-processing abilities in well-trained listeners in conditions of matched audibility, using two different techniques, to reveal whether both groups' performance is due mainly to audibility, or whether additional pathology associated with sensorineural hearing loss may affect vowel perception in YHI listeners.

II. VOWEL DISCRIMINATION UNDER DIFFERENT LISTENING CONDITIONS

A. Purpose

This study represents an extension of previous work in this lab. Coughlin *et al.* (1998) examined the relation between vowel discrimination and identification abilities of young normal-hearing, elderly normal-hearing, and elderly hearing-impaired listeners (EHI). Results suggested that hearing impairment and age contributed to decreased vowel perception performance in the EHI group. However, in a subsequent study by Richie *et al.* (2003), young hearing-impaired performance was shown to be essentially the same as that of elderly hearing-impaired performance reported in Coughlin *et al.* (1998), on both discrimination and identification tasks. This result suggests that factors related to audibility, not age, are the primary contributors to decreased vowel perception. The present study was designed to examine factors of audibility versus pathology in greater detail under three listening conditions, for young adult listeners. Young normal-hearing listeners were carefully matched to the hearing-impaired listeners in Richie *et al.* (2003) for audibility, age, and gender. They were then tested on the same protocol for vowel discrimination to further investigate performance in terms of audibility and pathology.

B. Method

1. Participants

Five YNH adults participated in this study. They were closely matched to five YHI adults with sloping, mild-to-moderate sensorineural hearing loss tested in an earlier study (results from the YHI listeners were reported in Richie *et al.*, 2003). YNH listeners ranged from 21 to 41 years of age (mean=31 years), and were matched by gender (two women and three men) and age (±1 year) to the YHI listeners on an individual basis. All listeners were native speakers of American English, and were paid for their participation in the

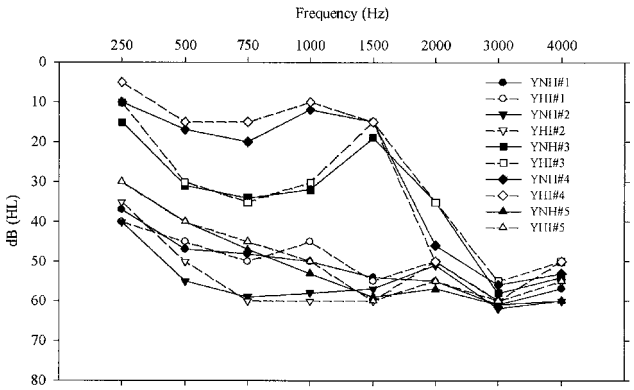


FIG. 1. Audiograms for the YNH–YHI pairs. Thresholds were obtained under noise masking conditions for the YNH listener and in quiet for the YHI listener of each pair. Thresholds for the YNH listeners represent the average of two tests. The YNH listeners' responses are indicated with closed symbols, while the YHI listeners' responses are indicated with open symbols.

study. Their audiograms tested better than 20 dB HL at octave frequencies from 250 to 4000 Hz. There was no evidence of middle-ear pathology at time of testing, as determined by normal tympanometric results.

For each listener, only one ear was tested in this study. The YNH subject listened with the same ear as the YHI listener of that matched pair. The YHI listeners in the previous study were selected using the following criteria: hearing threshold levels between 30 and 60 dB HL (ANSI, 1996) at 2000 Hz, and otherwise no threshold greater than 65 dB HL between 250 and 4000 Hz (see Fig. 1). In the earlier study, the YHI ear with the pattern of hearing loss most representative of selection criteria was selected; in some cases this was the listener's better hearing ear and in some cases the opposite was true.

2. Stimuli

The stimuli were five steady-state English vowels, /ɪ e ɛ ʌ æ/. These monophthongal vowels were selected based on previous work because they were acoustically similar with a wide range of frequencies across F1 and F2 (e.g., Coughlin *et al.*, 1998; Nábèlek *et al.*, 1992). The stimuli were synthesized based on spectrographic measurements of a female speaker, using the cascade branch of the KLTSYN synthesizer (Klatt, 1980). Stimuli were synthesized at 10 000 Hz. These five vowels were called the "standard" vowels used in the discrimination experiment, with formant frequencies shown in Table I.

In addition to the five standard vowels, vowel test stimuli were synthesized in sets of 14, for discrimination purposes. In each test set, either the first or second formant (F1 or F2) was incremented from the standard value, according to the ratio seen in Table I. For all stimuli, the fundamental frequency fell in a linear manner from 220 to 180 Hz, over a period of 175 ms (the average of the speaker's long and short vowels). Bandwidths for the formants were the same for all vowels, with BW1=70 Hz, BW2=90 Hz, and BW3=170 Hz, for the first, second, and third formants, respectively. In order to minimize inherent intensity differences between the five vowels, their amplitudes were then equated as follows: the rms energy over a 30 ms window for the five standard vowels was measured at the maximum amplitude (approximately 20–50 ms after vowel onset). The median rms value was chosen and all five vowels were then adjusted to be within ±1 dB of this median value.

3. Noise masking for normal-hearing listeners

YNH pure-tone thresholds were elevated so that they were similar to those of the YHI listeners, on a pairwise basis. Uniform noise (generated by a Tucker Davis Technologies waveform generator, model WG2) was "shaped" with cascaded IIR and FIR filters designed for these purposes. The shaped noise was lowpass filtered at 4300 Hz, with somewhat different roll-offs for different listeners (dependent upon threshold masked at 4000 Hz), but always greater than 44 dB/octave. Two audiograms were taken from each YNH listener while listening in the filtered noise, the average of which was compared on a frequency-by-frequency basis with the audiogram of the YHI pair (at 250, 500, 750, 1000, 1500, 2000, 3000, and 4000 Hz). The filters were adjusted until noise masked YNH pure-tone thresholds were found to be within ±5 dB of YHI thresholds at all the frequencies, as seen in Fig. 1.

To ensure listening levels were within safe limits, the masking noises designed for each of the five normal-hearing listeners were measured in a 6-cc coupler with a sound-level meter (Larson-Davis, model 800B) using the linear setting. Noise levels were in the range 72–77 dB SPL, dependent on hearing loss of the YHI matched listener, and thus determined to be safe for purposes of this experiment (see Melnick, 1992, p. 523).

TABLE I. Frequencies in hertz for formants F1–F3 used in synthesizing vowel stimuli. The range of values for the vowels indicates the formant values for the test stimuli, when F1 or F2 was incremented from the standard.

	Standard vowels			Ratios		Range values	
Vowel	F1	F2	F3	F1	F2	F1	F2
ɪ	450	2300	3000	1.010 592	1.006 338	455–522	2315–2513
e	550	2500	3100	1.008 739	1.005 844	555–621	2515–2712
ɛ	600	2200	3000	1.008 036	1.004 455	605–671	2210–2341
ʌ	700	1400	2600	1.006 923	1.006 923	705–771	1410–1542
æ	1000	1950	3000	1.007 258	1.005 014	1007–1107	1960–2091

TABLE II. Listening conditions for the discrimination experiment. Columns indicate conditions of similar audibility for the YNH and YHI listeners. Both groups listened in the Soft condition.

	Listening condition		
YNH	Soft+Noise Soft signal (~60 dB SPL) plus noise	Loud+Noise Loud signal (95 dB SPL) plus noise	Soft Soft signal (~60 dB SPL)
YHI	Soft Soft signal (~60 dB SPL)	Loud Loud signal (95 dB SPL)	Gain Soft signal (~60 dB SPL) plus gain

4. Listening conditions

Three conditions were designed to match the simulated everyday listening conditions for the YHI listeners in the previous study, from low to amplified levels. An outline of the listening conditions for the YNH listeners in this study, and YHI listeners tested in the previous study are described in the following, and can be seen in Table II. Although the YHI listeners did not hear noise masking in any condition, comparisons between these groups were designed to be equated for audibility.

For the YNH listeners, the Soft+Noise condition simulated the YHI listening condition for speech signals at an average conversational level. The Soft signal level used for the YHI listeners in Richie *et al.* (2003) was used, with the addition of the noise masking for the YNH listeners. The intent was that vowels would be at least partially audible at the Soft signal level for the YHI listeners such that the average maximum spectral level of the vowel stimuli was 10 dB above each YHI listener's pure tone average threshold. For the Soft signal level, 60 dB SPL was the nominal signal level. If the five-frequency average (500, 750, 1000, 1500, 2000 Hz) spectrum level of the vowel stimuli exceeded a YHI listener's equivalent five-frequency average pure tone threshold by at least 10 dB, 60 dB SPL remained the Soft level. Otherwise, higher signal levels were used; 74 (YNH–YHI pair 1), 75 (YNH–YHI pair 5) and 81 (YNH–YHI pair 2) dB SPL.

The second listening condition simulated listening to a more audible speech level where the signal presentation level was 95 dB SPL for all listeners. The YNH subjects also had the masking noise to simulate hearing loss, a condition called Loud+Noise.

In the third comparison shown in Table II, the Soft listening condition stimuli were presented to YNH listeners without masking noise, in order to parallel the Gain condition for YHI listeners in the previous study. The comparison here is between the effects of listening at a Soft signal level with normal hearing, or with hearing impairment aided by adding a shaped gain. This was intended to create conditions of adequate audibility between the two groups, in terms of ensuring the signal was at least 15 dB supra-threshold for all listeners, across the frequencies: 250, 500, 750, 1000, 1500, 2000, 3000, and 4000 Hz but without the noise masked elevated thresholds that simulate a sensorineural pathology. The Gain condition for YHI listeners was designed to simulate the frequency response provided by a linear gain hearing aid. The amount of gain for the YHI listeners was determined so that the five-vowel minimum spectral level at 60 dB SPL was at least 15 dB above the YHI listener's hearing threshold in dB SPL, from 250 to 4000 Hz. This approach was patterned after the Desired Sensation Level [(DSL); Cornelisse *et al.*, 1995] approach for the prescription of hearing aid gain.

Listeners heard stimuli under headphones (model TDH-39), with sounds presented monaurally to the ear previously selected. The vowel stimuli and noise were mixed using a summer (Tucker-Davis Technologies, TDT, model SM3). Note that in the previous study YHI listeners performed substantially better in the Loud and Gain conditions than in the Soft condition, as might be expected. However, there was no significant difference in YHI performance between the Loud and Gain conditions.

5. Calibration

All vowel stimuli were lowpass filtered with a cut-off frequency of 4300 Hz, preserving the first three formant peaks. The 4300 Hz cut-off was incorporated into FIR filters designed to achieve the specific signal levels previously described. Stimuli were output through a 16 bit D/A converter (Tucker-Davis Technologies, model DA1) followed by a digital filter (Tucker-Davis Technologies, model PF1), and a headphone buffer (Tucker-Davis Technologies, model HB6).

For calibration, the previously designed lowpass filters for each condition were loaded in the TDT modules. The calibration vowel (the standard /æ/ stimulus) was then measured through a headphone in a 6-cc coupler with a sound-level meter (Larson-Davis, model 800B) using the linear setting. The desired signal levels were then checked for presentation accuracy within ±0.5 dB and values were subsequently monitored throughout the experiment.

6. Task

The procedures for the YNH listeners in this study are the same as those described for the YHI listeners in the previous study (Richie *et al.*, 2003). A modified two-alternative, forced-choice procedure was implemented. The standard vowel was always presented first, followed by two more vowel stimuli. The two following vowels included the standard vowel and the test stimulus that differed from the standard in terms of a formant frequency increment. The test stimuli formants were incremented following the ratios expressed in Table I. The listener's task was to determine

whether the second or third vowel in the triplet differed from the first. The formant frequency increment was dictated by an adaptive-tracking algorithm designed to track 70.7% correct on the psychometric function (Levitt, 1971), via a two-down one-up rule. The adaptive tracks were separately controlled for each formant and listening condition, but were maintained across blocks. That is, the end points of one block of tokens provided the starting points of the next block of tokens. Listeners were tested individually in a sound-treated room. A message on the computer screen provided feedback about correct or incorrect responses.

7. Training

Since adequate training is necessary to establish stable threshold estimates (Kewley-Port, 2001), all listeners completed two 90 min training sessions. This training was designed to allow listeners to become accustomed to the experimental setup, synthetic stimuli, and response procedures. Training was focused on F1 for the /ɪ/ vowel, at a 95 dB SPL signal level, without masking noise. Feedback was given during training. All listeners' performance approached asymptote over two 90 min training sessions.

8. Testing

Testing consisted of five 90 min sessions following training. In each block, all ten types of vowel stimuli (5 vowels × 2 formant conditions) were randomized over 80 trials. The three listening conditions were fixed within a block. YNH listeners cycled through the listening conditions (Soft+Noise, Loud+Noise and Soft), at their own pace, completing 6, 9, or 12 blocks per test session. A measure of the just discriminable formant frequency difference between standard and test stimuli was calculated as ΔF in hertz. ΔF was obtained based on an average of the mean reversals from each listener's best four consecutive blocks within a particular listening condition (generally the last four). The mean of the reversals that each formant threshold was based on was thus calculated across 4 blocks of 8 trials (32 trials total). It should be noted that each listener's performance was seen to approach asymptote, over the course of testing.

C. Results

1. YNH and YHI pairs

Although a great deal of effort was taken to ensure conditions of equal audibility in terms of pure-tone thresholds (see Fig. 1), it was unknown whether equivalent performance on a vowel discrimination task would be observed. In order to quantify the degree of match between the YNH–YHI pairs, the following analyses were used. Performance was evaluated for the Soft+Noise (YNH) versus Soft (YHI), and Loud+Noise (YNH) versus Loud (YHI) conditions. The Soft (YNH) and Gain (YHI) conditions were not included in this analysis because YNH listeners were not exposed to noise masking in the Soft condition. Using ΔF in hertz, difference scores for all vowels were calculated between the YNH and the YHI listeners in the Soft+Noise versus Soft conditions, and in the Loud+Noise versus Loud conditions. On a pair-by-pair basis, mean difference scores (over all vowels in all listening conditions, $N=200$) were calculated. The mean differences on a pairwise basis (YNH-YHI pair 1 through YNH-YHI pair 5) were: 29.9, −2.5, −29.7, −0.6, and −17.8 Hz. The average over the five pairs was only 4.1 Hz. An average difference score of zero would indicate no relative difference in pairwise performance. This suggests a very small relative difference between the pairs.

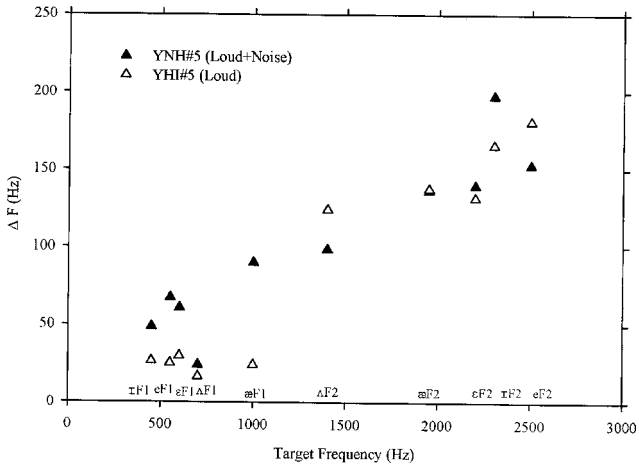


FIG. 2. An example of the typical degree of match in performance between the members of a YNH–YHI pair in the discrimination task. These results are for YNH–YHI pair 5 in the Loud+Noise and Loud conditions, respectively.

A histogram of the difference scores was made in order to examine their distribution. The mode of the difference scores was 0 Hz, with 73% of the distribution occurring between differences of ±30 Hz. This, too, indicates that for the five YNH–YHI pairs, most differences in performance were minimal. An example of the typical degree of match in performance between the members of YNH–YHI pair 5 in the Loud+Noise and Loud conditions, respectively, can be seen in Fig. 2.

Finally, the Pearson product moment correlation coefficient for the YNH and YHI discrimination scores in these two listening conditions was found to be $r=0.80$ ($N=200$), indicating a significant, strong relation between the two data sets ($p<0.05$). Because the YNH–YHI listeners were closely matched in terms of hearing threshold, age, and sex, the correlation between YNH and YHI performance is expectedly high.

2. YNH and YHI discrimination performance in conditions of similar audibility

The discrimination results for the YNH listeners in all listening conditions can be seen in Table III. These discrimination thresholds were measured in terms of ΔF in hertz (for F1 and F2) for the five vowels (/ɪ e ɛ ʌ æ/) in the three level conditions as shown in Fig. 3. As would be expected, thresholds are nearly constant in the F1 region and increase in the F2 region, for all conditions.

Performance was compared between the pairs in conditions of similar audibility in order to determine whether a significant difference existed between groups due to the presence of a real or simulated hearing loss. The first comparison made was between the YNH listeners in the Soft+Noise con-

TABLE III. Means and standard deviations of (a) F1 discrimination thresholds (ΔF in Hz) for YNH listeners, (b) F2 discrimination thresholds (ΔF in Hz) for YNH listeners.

(a)			ɪ F1 (450 Hz)	e F1 (550 Hz)	ɛ F1 (600 Hz)	ʌ F1 (700 Hz)	æ F1 (1000 Hz)
	Soft+Noise	Mean	52.4	55.4	47.2	37.6	77.3
		s.d.	10.1	10.1	18.0	21.4	25.4
	Loud+Noise	Mean	33.8	37.7	39.9	25.9	49.7
		s.d.	10.3	25.1	14.8	14.2	33.5
	Soft	Mean	37.9	40.5	33.1	39.4	45.0
		s.d.	18.4	22.8	18.2	19.4	24.1
(b)			ʌ F2 (1400 Hz)	æ F2 (1950 Hz)	ɛ F2 (2200 Hz)	ɪ F2 (2300 Hz)	e F2 (2500 Hz)
	Soft+Noise	Mean	115.6	110.6	124.0	176.4	169.7
		s.d.	34.4	36.1	34.2	68.9	60.0
	Loud+Noise	Mean	98.7	88.3	98.2	110.0	110.2
		s.d.	32.0	56.4	44.1	60.7	35.2
	Soft	Mean	73.7	102.2	100.3	104.8	116.4
		s.d.	35.1	45.0	44.5	56.7	56.2

dition and the YHI listeners in the Soft condition. The difference in ΔF thresholds was calculated between the performance of members of each YNH-YHI pair, in the Soft +Noise and the Soft conditions, respectively, to yield difference scores in hertz for each vowel formant tested. Difference scores (between YNH Soft+Noise and YHI Soft performance on the ten formants) were then subjected to a repeated measures multivariate analysis of variance (MANOVA) test of significance to see if difference scores across the ten formants differed significantly from zero. The MANOVA statistical method was adopted for the following reasons. The listeners in this study were tested on the same dependent measure (vowel formant difference limen in hertz) in different listening conditions. However, it is possible that the contrasts involved in testing these effects are not independent of each other, i.e., compound symmetry and sphericity might be violated (see Statsoft, Inc., 2002). If multivariate criteria are used to simultaneously test the significance of two or more repeated measures contrasts, they do not need to be independent of each other. The MANOVA, a more stringent statistical test, bypasses the assumptions of the univariate analysis of variance, related to compound symmetry and sphericity. This test seemed appropriate for these data, and the results of the MANOVA may be interpreted in basically the same way as an ordinary analysis of variance (Aron and Aron, 1999, p. 535). As expected from Fig. 3(a), this test did not reveal a significant result [$F(1,4)=4.805, p=0.328$]. Performance of the YNH listeners in the Loud+Noise and the YHI listeners in the Loud conditions was compared using the same procedures. This test also did not reveal a significant result [Fig. 3(b); $F(1,4)=69.978$, $p=0.089$]. Finally, performance of the YNH listeners in the Soft condition and YHI listeners in the Gain condition was compared. A significant result was not obtained [Fig. 3(c); $F(1,4)=1.348$, $p=0.562$].

Taken together, these comparisons showed no significant differences between the YNH and YHI groups of listeners, in conditions of similar audibility using two different methods for equating audibility, for this vowel discrimination task. This suggests that additional pathological factors surrounding the sensorineural hearing loss, if they exist, do not adversely affect YHI listeners' vowel formant discrimination abilities when compared to YNH listeners tested with a simulated hearing loss. The YNH listeners with noise masking did not perform in a manner significantly different from that of a group of YHI listeners, when age, gender, and audibility were equated.

III. DISCUSSION AND CONCLUSIONS

A. Summary of results

The results from the discrimination experiment indicate that in equal audibility listening conditions, the young noise masked normal-hearing adult listeners tested in this study did not perform significantly differently from the young hearing-impaired adults tested in an earlier study (Richie *et al.*, 2003). It was also found that performance between the young normal-hearing listeners listening in quiet at a soft conversation level performed similarly to the young hearing-impaired listeners listening to a soft conversation level through a simulated hearing aid. These results serve to strengthen the argument that factors related to audibility likely play a key role in determining performance on speech perception tasks.

Three further analyses along these lines are presented next. These include an examination of the relation between hearing threshold and formant frequency discrimination ability, a comparison of the effects of listening with a real versus a simulated hearing loss on vowel discrimination, and a discussion of the time course of learning during the vowel discrimination task for normal-hearing versus hearing-impaired listeners.

B. The relation between hearing threshold and formant frequency discrimination performance

The relation between hearing threshold and formant frequency discrimination ability can be quantified, in order to

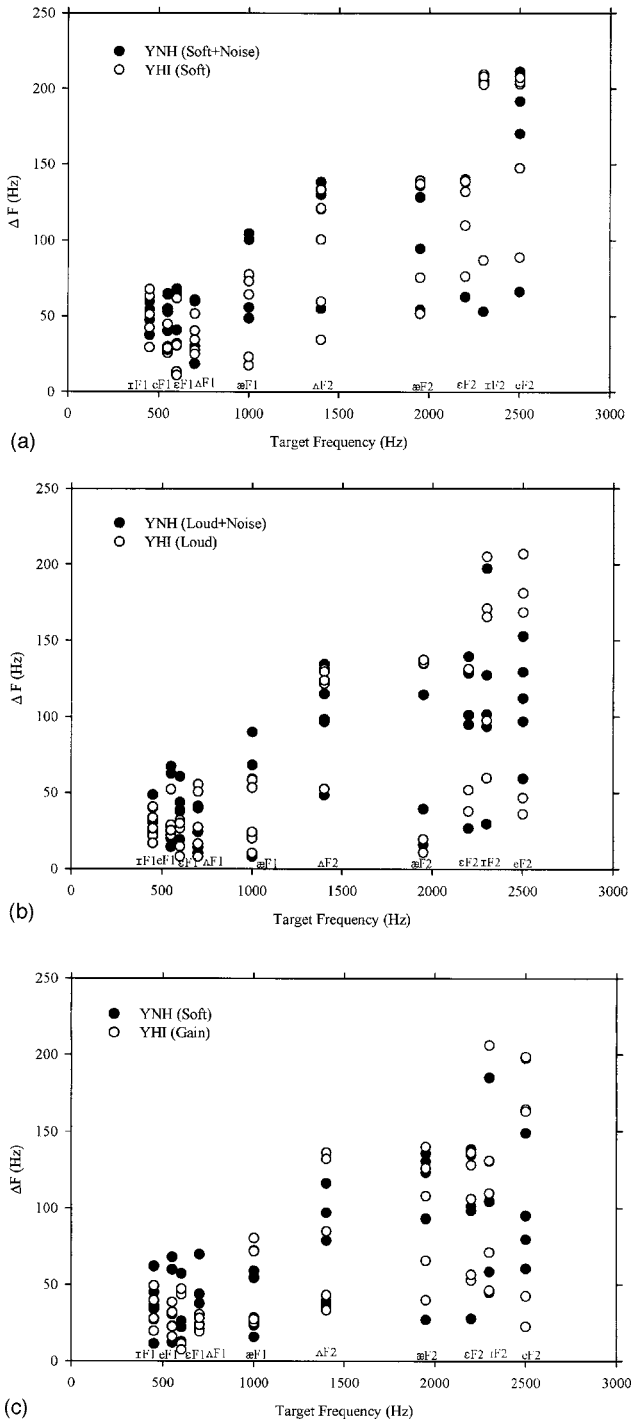


FIG. 3. Discrimination results for the five pairs of YNH and YHI listeners (from Richie *et al.*, 2003) in the (a) Soft+Noise and Soft listening conditions, (b) Loud+Noise and Loud listening conditions, and (c) Soft and Gain listening conditions.

further strengthen the argument that factors related to audibility affect performance on speech perception tasks such as the vowel formant frequency discrimination task in this study. In order to illustrate this relation, pure tone thresholds for the YHI listeners tested were compared to formant frequency discrimination abilities on a frequency-by-frequency basis. Results from the YHI listeners tested in the Soft listening condition were used for this posthoc analysis. Pure tone audiometric thresholds measured as 30 dB HL and

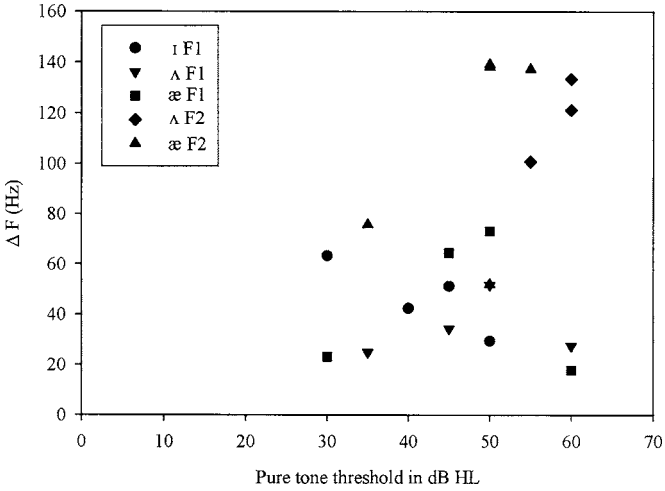


FIG. 4. The relation between hearing threshold (in dB HL) and formant frequency discrimination threshold (ΔF in Hz) for the YHI listeners in the Soft listening condition, for a subset of formants.

greater were selected (i.e., hearing thresholds greater than normal), at the following five frequencies; 500, 750, 1000, 1500, and 2000 Hz. Five of the standard vowel formants used in the discrimination experiment roughly correspond to those five frequencies; /ɪ/ F1 (450 Hz), /ʌ/ F1 (700 Hz), /æ/ F1 (1000 Hz), /ʌ/ F2 (1400 Hz), and /æ/ F2 (1950 Hz). The formant frequency discrimination thresholds, measured as ΔF in hertz, were compared to hearing thresholds, measured as dB HL, for the five correspondent pairs of audiometric/formant frequencies listed above. A scattergram for the five YHI listeners is shown in Fig. 4.

A moderate, significant correlation between hearing threshold and formant frequency discrimination threshold at those frequencies was obtained, $r=0.37$ ($p<0.05$). This result suggests that hearing ability, as measured via pure tone thresholds, is related to formant frequency discrimination, measured as ΔF thresholds.

C. Comparison of real and simulated hearing losses

This study presents a unique opportunity to compare vowel perception between groups with and without a real hearing loss. It also affords a within-group comparison of the effects of simulated hearing loss on vowel perception. Comparison of the YNH and YHI listeners in the Soft listening condition shows the between-groups effects of listening with a *real hearing loss*, while comparison of the YNH listeners in the Soft and Soft+Noise conditions shows the within-group effects of listening with a *simulated hearing loss*. Comparison of the real and the simulated hearing losses reveals marked similarities between their effects. As can be seen in Fig. 5, performance in the F1 region is essentially the same and relatively flat at 45 Hz between groups differing in real hearing loss, and within the YNH group in conditions differing in simulated hearing loss. However, in the F2 region, the effects of the real and simulated losses can be seen in apparently elevated formant discrimination thresholds for the YHI listeners with real hearing loss, and YNH listeners with simulated hearing loss. The elevation of F2 thresholds was on average 135% for the hearing-impaired listeners and

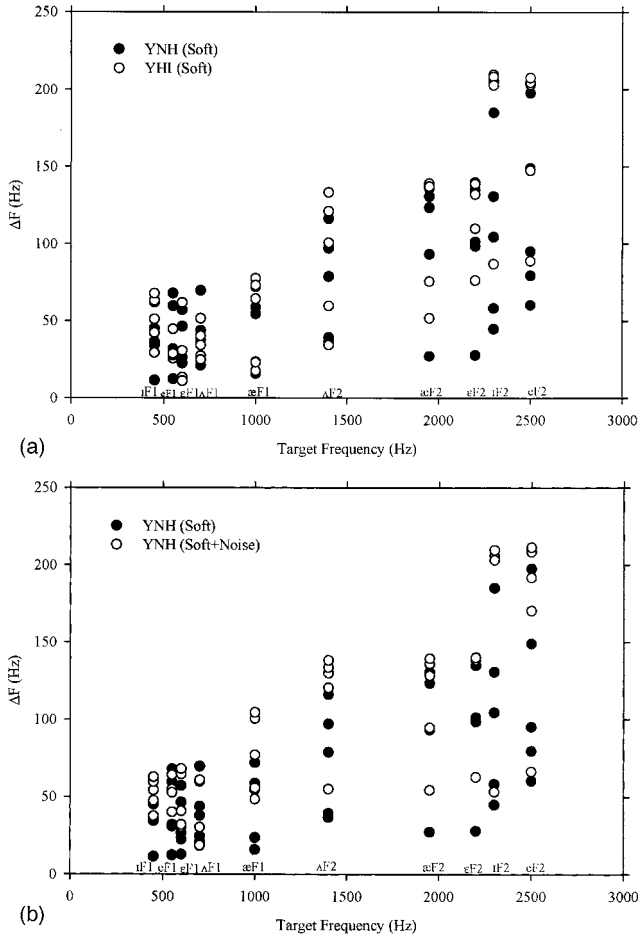


FIG. 5. Discrimination performance in various listening conditions, showing the effects of (a) the real hearing loss, and (b) the simulated hearing loss.

140% for the normal-hearing listeners with simulated hearing loss, over the normal-hearing listeners in the Soft listening condition in quiet. The median test was performed, in order to contrast the F2 discrimination performance between the YNH and YHI listeners in the Soft condition [Fig. 5(a)]. This test revealed that a significant difference exists between the two groups' performance, due to the presence of real hearing loss [Fig. 5(a); $\chi^2(1)=3.920$, $p=0.048$]. Similarly, results of a median test contrasting the F2 discrimination performance between the YNH listeners in the Soft and Soft+Noise conditions revealed a significant difference between normal-hearing listeners' performance, due to the simulated hearing loss [Fig. 5(b); $\chi^2(1)=6.480$, $p=0.011$].

The similarity in the effect on vowel discrimination performance by the real and simulated losses can be further quantified given that the basis of comparison for both is the normal-hearing in the Soft condition. The difference between the YNH and YHI listeners in the Soft condition was calculated (real loss), and the difference between the YNH in Soft+Noise and Soft was calculated (simulated loss). These difference scores ($N=100$) were found to have a significant correlation of $r=0.73$ ($p<0.05$). The correlation is unexpectedly high given that it is on a formant-by-formant comparison over two groups representing a real versus a simulated hearing loss. Further, the simulated hearing loss resulted in only minimally higher thresholds in the F2 region than those resulting from the real hearing loss, by an average of 5 Hz.

TABLE IV. Number of blocks completed in each listening condition.

Pair No.	YNH	YHI	% YHI increase over YNH
1	15	17	13
2	13	19	46
3	12	25	108
4	11	14	27
5	13	18	38
			Average=46% (s.d. 37%)

These facts together indicate that the real and simulated hearing losses had very similar effects on vowel formant discrimination performance, for the hearing-impaired and noise masked normal-hearing listeners, respectively. While it is possible that loudness cues could have affected formant thresholds if comparison stimuli were presented on the sloping part of the real or simulated hearing loss, it was not the intent of this study to eliminate those potential loudness cues. Rather, this study attempted to examine the effects of real and simulated hearing loss on vowel formant frequency discrimination, and found no significant differences between groups.

D. The time course of learning during the vowel discrimination task

A posthoc observation of data from this study and the previous study (Richie *et al.*, 2003) indicated that the hearing-impaired listeners completed many more blocks of testing than the normal-hearing listeners reported here. Specifically, the YHI listeners completed an average of 46% more blocks of testing than the YNH listeners, in the same self-paced, fixed time period (see Table IV). It is possible that the YNH listeners exposed to the masking noise took longer to formulate their responses. While the reasons for this difference are beyond the scope of this study, the time course of learning during the vowel discrimination task was compared between the YNH and YHI listeners, in conditions of matched audibility.

Other work has illustrated the relation between training and threshold estimates (e.g., Kewley-Port, 2001), whereby increased training leads to lower threshold values. Thus, a valid question concerning these data is whether the YHI listeners seem to perform as well as the YNH listeners due to the increased number of trials they completed. The patterns of results apparent in the learning curves for these two groups of listeners were examined, to determine when performance levels were similar and to determine whether different cognitive processes might be implied for the two listener groups.

A representative subset of the vowels tested was selected for analysis that had similar formant frequency thresholds for both members of a YNH–YHI pair. For each listening condition, thresholds for one vowel in the F1 region and one vowel in the F2 region were analyzed for each of the five listener pairs. Learning was assessed at three points on the

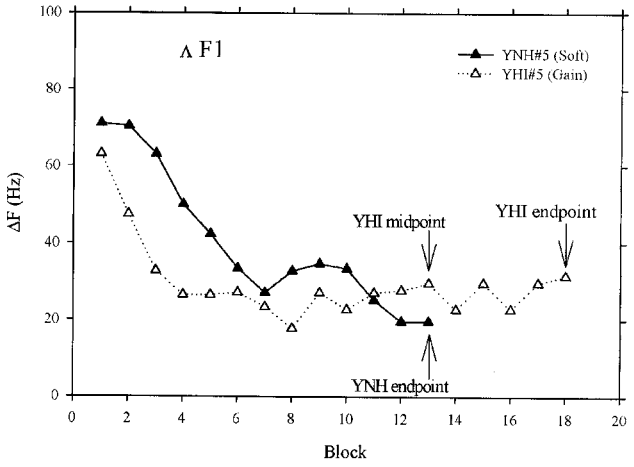


FIG. 6. An adaptive track for a pair of YNH and YHI listeners, while learning to discriminate fine changes in the F1 frequency of /ʌ/. Arrows point to where learning was assessed at three points; the YNH end point, YHI midpoint (block corresponding to the YNH end point), and YHI end point.

adaptive tracks, on a vowel-by-vowel basis (see Fig. 6); the YNH end point, YHI midpoint (corresponding to the YNH end point), and YHI end point. Thresholds (ΔF in hertz averaged over the four preceding blocks) were compared at the three tracking points described.

Correlated t-tests ($N=104$) showed no significant differences between thresholds across all YNH–YHI pairs and vowels, for any possible comparison (YNH end point vs YHI midpoint, YNH end point vs YHI endpoint, and YHI midpoint vs YHI end point). The "extra" blocks of testing completed by the YHI listeners were therefore not necessary for them to achieve a level of performance similar to the YNH listeners. Though the YHI listeners completed an average of 46% more blocks of testing than their YNH matches, these were not necessary to achieve similar performance.

E. Conclusions

Studies that measure speech-recognition performance when conditions of audibility are matched between normal-hearing and hearing-impaired listeners have yielded somewhat mixed results. This study, like most studies that compare performance in conditions of similar audibility, matched hearing-impaired listeners with noise masked normal-hearing listeners in order to simulate the effects of sensorineural hearing loss in the latter group. Spectrally shaped noise may produce masked thresholds that resemble the audiometric contours of hearing-impaired listeners, loudness recruitment, and reduced frequency selectivity (Humes and Roberts, 1990; Sommers and Humes, 1993). In addition, however, this study also examined a condition of similar audibility by providing gain to the speech signal for the hearing-impaired group and comparing performance to the young normal-hearing listeners in quiet. By doing so it eliminated the simulated effects of sensorineural hearing loss that occur with noise masked threshold elevation techniques.

However, generalizations from speech perception studies when conditions of similar audibility have been created have been difficult to achieve. In some cases, the hearing-impaired listeners performed worse than the noise masked controls, in some cases they performed at a level similar to masked normal-hearing listeners, and in some cases they performed consistently better than the noise-masked controls. These results may depend heavily on the type of speech perception task employed. In more psychophysical tasks such as masked threshold estimation in noise (Dubno and Schaefer, 1992), or determination of the minimum difference in amplitude between spectral peaks and troughs necessary for vowel identification (Leek *et al.*, 1987), noise masked normal-hearing listeners seemed to outperform their hearing-impaired matches. In higher-level speech-related tasks such as nonsense syllable identification and consonant recognition, hearing-impaired listeners were seen to perform similarly to or better than their noise masked normal-hearing matches (e.g., Dubno and Schaefer, 1992; Humes *et al.*, 1987).

It could be argued that the discrimination experiment in this study reflects psychophysical abilities related to just-noticeable change in formant frequency. In this case, it might be expected that the noise masked normal-hearing listeners would outperform the hearing-impaired listeners in the discrimination task. However, results from the present vowel discrimination experiment have failed to reveal differences between YNH and YHI pairs closely matched for age, gender, and pure-tone thresholds, in conditions of similar audibility. An examination of the relation between pure tone thresholds and vowel formant frequency discrimination performance revealed a moderate correlation for those factors. An analysis of the degrading effects of the real and simulated hearing losses on vowel formant discrimination performance also failed to reveal differences between pairs of listeners. Finally, an analysis of the time course of learning in these two groups of listeners also failed to reveal differences between the groups of listeners. Though the YHI listeners completed an average of 46% more blocks of testing than the YNH listeners, these blocks were not necessary for the YHI listeners to achieve similar discrimination performance.

In sum, performance on this vowel discrimination task failed to reveal substantial differences between pairs of listeners matched for hearing thresholds, age, and gender, but different in terms of hearing status. This suggests that the performance of young adult hearing-impaired listeners is in fact very similar to that of their normal-hearing peers, when they are tested in conditions of similar audibility either by noise masking to elevate normal hearing thresholds or by amplifying the speech signal for the impaired hearing thresholds.

ACKNOWLEDGMENTS

The authors wish to thank Larry Humes for contributions to experimental design, and providing feedback on earlier drafts of this manuscript. This research was supported by NIHDCD-02229.

ANSI (**1996**). ANSI S3.6-1996, "Specification for Audiometers," American National Standards Institute, New York.

Aron, A., and Aron, E. (**1999**). *Statistics for Psychology* (Prentice–Hall, Upper Saddle River, NJ).

Baer, T., and Moore, B. (**1997**). "Evaluation of a scheme to compensate for reduced frequency selectivity in hearing-impaired subjects," in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Englewood Cliffs, NJ), pp. 329–341.

Burkle, T., Kewley-Port, D., Humes, L., and Lee, J. (**2004**). "Contribution of consonant versus vowel information to sentence intelligibility by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **115**, 2601.

Cornelisse, L., Seewald, R., and Jamieson, D. (**1995**). "The input/output formula: A theoretical approach to the fitting of personal amplification devices," J. Acoust. Soc. Am. **97**, 1854–1864.

Coughlin, M., Kewley-Port, D., and Humes, L. (**1998**). "The relation between identification and discrimination of vowels in young and elderly listeners," J. Acoust. Soc. Am. **104**, 3597–3607.

Dorman, M., Marton, K., Hannley, M., and Lindholm, J. (**1985**). "Phonetic identification by elderly normal and hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 664–670.

Dubno, J., and Dirks, D. (**1989**). "Auditory filter characteristics and consonant recognition for hearing-impaired listeners," J. Acoust. Soc. Am. **85**, 1666–1675.

Dubno, J., and Schaefer, A. (**1992**). "Comparison of frequency selectivity and consonant recognition among hearing-impaired and masked normal-hearing listeners," J. Acoust. Soc. Am. **91**, 2110–2121.

Dubno, J., and Schaefer, A. (**1995**). "Frequency selectivity and consonant recognition for hearing-impaired and normal-hearing listeners with equivalent masked thresholds," J. Acoust. Soc. Am. **97**, 1165–1174.

Era, P., Jokela, J., Qvarnberg, Y., and Heikkinen, E. (**1986**). "Pure-tone thresholds, speech understanding, and their correlates in samples of men of different ages," Audiology **25**, 338–352.

Godfrey, J., and Millay, K. (**1980**). "Perception of synthetic speech sounds by hearing-impaired listeners," J. Aud. Res. **20**, 187–203.

Humes, L., and Christopherson, L. (**1991**). "Speech identification difficulties of hearing-impaired elderly persons: The contributions of auditory processing deficits," J. Speech Hear. Res. **34**, 686–693.

Humes, L. E., Dirks, D., Bell, T., and Kincaid, G. (**1987**). "Recognition of nonsense syllables by hearing-impaired listeners and by noise-masked normal hearers," J. Acoust. Soc. Am. **81**, 765–773.

Humes, L. E., and Roberts, L. (**1990**). "Speech recognition difficulties of the hearing-impaired elderly: The contributions of audibility," J. Speech Hear. Res. **33**, 726–735.

Humes, L. E., Watson, B. U., Christensen, L. A., Cokely, C. G., Halling, D. C., and Lee, L. (**1994**). "Factors associated with individual differences in clinical measures of speech recognition among the elderly," J. Speech Hear. Res. **37**, 465–474.

Jerger, J., Jerger, S., and Pirozzolo, F. (**1991**). "Correlational analysis of speech audiometric scores, hearing loss, age, and cognitive abilities in the elderly," Ear Hear. **12**, 103–109.

Kewley-Port, D. (**2001**). "Vowel formant discrimination. II. Effects of stimulus uncertainty, consonantal context, and training," J. Acoust. Soc. Am. **110**, 2141–2155.

Klatt, D. (**1980**). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. **67**, 971–995.

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.

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

Melnick, W. (**1992**). "Occupational noise standards: Status and critical issues," in *Noise-Induced Hearing Loss*, edited by A. Dancer, D. Henderson, R. Salvi, and R. Hamernik (Mosby-Year Book, St. Louis, MO), pp. 521–530.

Moore, B. (**1985**). "Frequency selectivity and temporal resolution in normal and hearing-impaired listeners," Br. J. Audiol. **19**, 189–201.

Nábèlek, A. (**1988**). "Identification of vowels in quiet, noise, and reverberation: Relationships with age and hearing loss," J. Acoust. Soc. Am. **84**, 476–484.

Nábèlek, A., Czyzewski, Z., and Krishnan, L. (**1992**). "The influence of talker differences on vowel identification by normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **92**, 1228–1246.

Owens, E., Talbott, C., and Schubert, E. (**1968**). "Vowel discrimination of hearing-impaired listeners," J. Speech Hear. Res. **11**, 648–655.

Peters, R., and Moore, B. (**1992**). "Auditory filter shapes at low centre frequencies in young and elderly hearing-impaired subjects," J. Acoust. Soc. Am. **91**, 256–266.

Plomp, R. (**1978**). "Auditory handicap of hearing impairment and the limited benefit of hearing aids," J. Acoust. Soc. Am. **63**, 533–549.

Rakerd, B., Seitz, P., and Whearty, M. (**1996**). "Assessing the cognitive demands of speech listening for people with hearing losses," Ear Hear. **17**, 97–106.

Richie, C., Kewley-Port, D., and Coughlin, M. (**2003**). "Discrimination and identification of vowels by young, hearing-impaired adults," J. Acoust. Soc. Am. **114**, 2923–2933.

Sommers, M., and Humes, L. (**1993**). "Auditory filter shapes in normal-hearing, noise-masked normal, and elderly listeners," J. Acoust. Soc. Am. **93**, 2903–2914.

StatSoft, Inc. (**2002**). *Electronic Statistics Textbook*, Tulsa, OK: StatSoft. WEB: http://www.statsoft.com/textbook/stathome.html.

van Rooij, J., and Plomp, R. (**1992**). "Auditive and cognitive factors in speech perception by elderly listeners. III. Additional data and final discussion," J. Acoust. Soc. Am. **91**, 1028–1033.

Van Tasell, D., Fabry, D., and Thibodeau, L. (**1987**). "Vowel identification and vowel masking patterns of hearing-impaired subjects," J. Acoust. Soc. Am. **81**, 1586–1597.

Van Tasell, D., Hagen, L., Koblas, L., and Penner, S. (**1982**). "Perception of short-term spectral cues for stop consonant place by normal and hearing-impaired subjects," J. Acoust. Soc. Am. **72**, 1771–1780.

Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners

Belinda A. Henry[a)]
Department of Communicative Disorders, University of Wisconsin—Madison, Madison, Wisconsin 53706

Christopher W. Turner
Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242

Amy Behrens
Department of Speech Pathology and Audiology, The University of Iowa, Iowa City, Iowa 52242

(Received 10 May 2004; revised 8 May 2005; accepted 9 May 2005)

Spectral peak resolution was investigated in normal hearing (NH), hearing impaired (HI), and cochlear implant (CI) listeners. The task involved discriminating between two rippled noise stimuli in which the frequency positions of the log-spaced peaks and valleys were interchanged. The ripple spacing was varied adaptively from 0.13 to 11.31 ripples/octave, and the minimum ripple spacing at which a reversal in peak and trough positions could be detected was determined as the spectral peak resolution threshold for each listener. Spectral peak resolution was best, on average, in NH listeners, poorest in CI listeners, and intermediate for HI listeners. There was a significant relationship between spectral peak resolution and both vowel and consonant recognition in quiet across the three listener groups. The results indicate that the degree of spectral peak resolution required for accurate vowel and consonant recognition in quiet backgrounds is around 4 ripples/octave, and that spectral peak resolution poorer than around 1–2 ripples/octave may result in highly degraded speech recognition. These results suggest that efforts to improve spectral peak resolution for HI and CI users may lead to improved speech recognition. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1944567]

PACS number(s): 43.71.Ky, 43.66.Ts, 43.66.Sr, 43.71.Es [KWG] Pages: 1111–1121

I. INTRODUCTION

Speech recognition in listeners with sensorineural hearing loss varies widely, both among those with impaired acoustic hearing [hearing impaired (HI) listeners] and among those using cochlear implants (CI listeners). While some HI and CI listeners achieve high levels of audition-alone open-set speech recognition, at the other end of the range are some listeners who must rely on supplementary visual cues in order to understand speech. The loss of absolute sensitivity in HI listeners is the primary factor affecting speech perception, and amplification via the use of hearing aids compensates for this to some extent. However, for those HI listeners with hearing losses in the moderate to profound range, audibility does not account for the entire deficit in speech perception (e.g., Ching *et al.*, 1998; Dubno *et al.*, 1989; Hogan and Turner, 1998; Humes *et al.*, 1986; Pavlovic, 1984; Pavlovic *et al.*, 1986; Skinner, 1980). In these cases, variability in performance is likely to be related not only to the audibility of speech cues, but also to abnormalities in the perceptual analysis of sound at suprathreshold levels. In CI listeners, it is this latter factor that is thought to be related to performance variability. Acoustic signals are transformed by the speech processor, and speech cues are represented in the pattern of electrical stimulation across the electrode array. Audibility is not the primary factor contributing to performance variability, since the audibility of conversational-level speech is determined primarily by the input dynamic range and sensitivity setting of the speech processor (provided that the electrical stimulation levels and sensitivity control are set optimally). Rather, it is the ability to extract speech cues from the audible patterns of electrical stimulation that is the most important factor limiting performance in CI listeners.

One perceptual factor that is likely to limit speech perception in both HI and CI listeners is reduced spectral resolution. Accurate speech recognition depends partly on the ability to perceive the spectral shapes of speech sounds, and, in particular, to identify the frequencies of spectral peaks. In normal hearing (NH) listeners, the frequency selectivity of the auditory system is thought to underlie the process of resolving spectral peaks in the speech signal. Impaired frequency selectivity has been demonstrated in HI listeners in both physiological (e.g., Dallos *et al.*, 1977; Liberman and Dodds, 1984) and psychophysical studies (e.g., Dubno and Dirks, 1989; Glasberg and Moore, 1986; Trees and Turner, 1986; Wightman *et al.*, 1977). The bandwidth of auditory filters has been shown to be up to three to four times greater than normal in listeners with cochlear hearing loss (Glasberg and Moore, 1986). Spectral resolution is also reduced in CI listeners, although for different underlying reasons than in HI listeners. Multichannel CIs replace the peripheral frequency selectivity of the normal auditory system with multiple intra-

[a)]Electronic mail: bahenry@wisc.edu

cochlear electrodes. Spectral cues are coded by resolving the frequency components of the signal using bandpass filtering, and mapping the outputs of these bands onto the intracochlear electrodes in a tonotopically assigned manner. Spectral resolution is limited in CI listeners by the number of stimulating channels (currently between 6 and 22, depending on the device and speech processing strategy) and the ability of the individual to resolve the spectral cues that are provided.

What is the effect of impaired spectral resolution on speech recognition? One might hypothesize that if spectral resolution is reduced, either due to an impaired auditory system or the use of a CI, the spectral envelope may be "blurred," making it difficult for a listener to identify the frequency locations of spectral peaks in speech. It is of theoretical importance, as well as significant practical importance in designing improved hearing aids and CIs and optimizing these devices for individuals, to determine the relationship between spectral peak resolution and speech recognition, and the degree of spectral peak resolution that is required for accurate speech recognition. In other words, a question of particular importance is as follows: What is the minimum requirement for spectral peak resolution, below which speech recognition becomes degraded? These questions have been investigated by numerous authors using at least two different approaches.

The first approach involves manipulating the spectral resolution available in the speech signal and examining the effects of this processing on speech recognition both in NH listeners as well as in HI and CI listeners. The results of studies on the effects of simulated reduced spectral resolution in NH listeners indicate that speech recognition in quiet is highly resistant to degraded spectral resolution. Several authors have investigated the effects of simulated broadened auditory filters by measuring speech recognition in NH listeners under conditions of spectral smearing. Baer and Moore (1993) and ter Keurs *et al.* (1992) showed that simulating auditory filters up to six times broader than those of NH listeners has little effect on speech recognition in quiet. Furthermore, Boothroyd *et al.* (1996) showed that in order to reduce phoneme recognition in quiet by 50%, the spectral information needed to be smeared by as much as 1400 Hz. Simulations of CI processing, in which multiple bands of speech-modulated noise are presented to NH listeners, show high levels of speech recognition for NH listeners in quiet with between 4 and 12 spectral channels, depending on the degree of difficulty of the speech materials (Dorman *et al.*, 1997; Friesen *et al.*, 2001; Shannon *et al.*, 1995; Turner *et al.*, 1995). These findings further indicate that fine spectral resolution is not required for speech recognition in quiet.

While the CI simulation studies indicate that CIs can ideally present sufficient spectral detail for accurate speech recognition in quiet, studies on the effect of the number of channels on speech recognition in CI recipients indicate that the effective number of channels perceived by these listeners is lower than the physical number of channels provided. CI users show an asymptote in speech recognition on average across listeners with between two and seven channels, depending on the degree of difficulty of the speech material presented (Dorman and Loizou, 1997; Fishman *et al.*, 1997; Friesen *et al.*, 2001). Furthermore, the effect of the number of channels varies widely, with better-performing CI listeners generally able to use more channels than those with poorer overall performance (Friesen *et al.*, 2001). The effects of limiting the spectral resolution provided in the speech signal on performance in HI listeners has also been investigated (Turner *et al.*, 1999). The HI listeners in that study performed equivalently to NH listeners for single band speech. However, as the number of bands of speech-modulated noise presented was increased, the performance of HI listeners did not increase to the same extent as for NH listeners, indicating that some HI listeners cannot utilize all the spectral information in speech.

The second approach is to attempt to relate performance on psychophysical measures of spectral resolution to speech recognition in HI and CI listeners. If reduced frequency resolution is associated with poorer speech recognition, a statistical relationship between these two measures may be expected. However, in HI listeners strong correlations between speech recognition in quiet and frequency selectivity, as measured in psychoacoustic masking experiments for instance, have been difficult to establish and findings vary across studies (Dreschler and Plomp, 1980, 1985; Festen and Plomp, 1983; Lutman and Clark, 1986; Glasberg and Moore, 1989; Stelmachowicz *et al.*, 1985; Tyler *et al.*, 1982). In studies where correlations were found, it was difficult to separate the roles of frequency selectivity and audibility in speech recognition since both frequency selectivity and speech recognition were correlated with absolute hearing thresholds. Correlations between speech recognition and frequency selectivity were often reduced or eliminated after the effect of absolute threshold was statistically partialled out. In CI listeners, place of stimulation perception, as determined using psychophysical measures such as electrode discrimination and pitch ranking, is generally assumed to underlie to some extent the ability to resolve the spectral aspects of the speech signal. While several authors have reported a relationship between speech recognition and place of stimulation perception (Collins *et al.*, 1997; Donaldson and Nelson, 2000; Nelson *et al.*, 1995; Throckmorton and Collins, 1999), Zwolan *et al.* (1997) showed no correlation between these two measures, and Henry *et al.* (2000) showed a correlation for the low- to mid-frequency regions only, and only when there was random level variation between stimuli. These results generally indicate that those CI listeners who are more sensitive to the place of stimulation in the cochlea, particularly in the presence of random level variation, are better at recognizing speech.

The traditional measures of frequency resolution in HI listeners and place of stimulation perception in CI listeners, which typically require a listener to detect a signal in the presence of a masker (HI listeners) or to discriminate between stimulation on different electrodes activated individually (CI listeners), are indirect measures of spectral peak resolution for complex broadband acoustic signals. The distinct methodologies that have been employed in these studies do not allow the comparison of spectral peak resolution abilities across NH, HI, and CI listeners. Consequently, the gen-

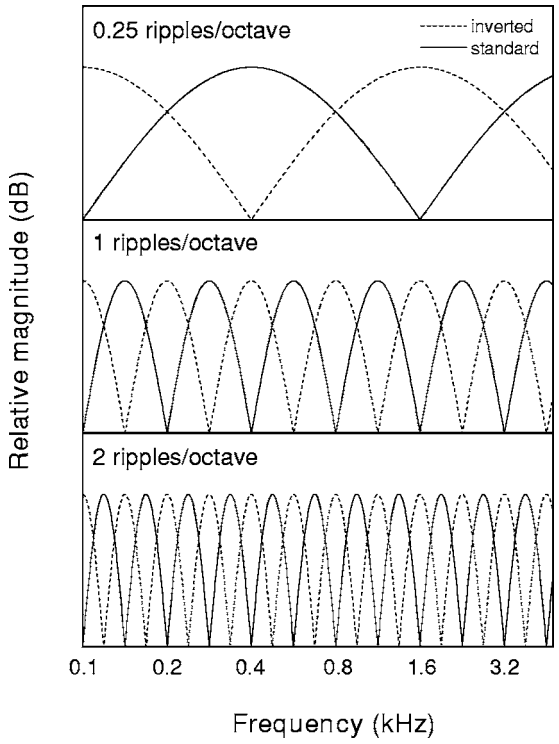


FIG. 1. Rippled noise spectra. Standard and inverted peak positions for ripple frequencies of 0.25, 1 and 2 ripples/octave are shown.

eral relationship between spectral peak resolution and speech recognition and the degree of spectral peak resolution that is required for accurate speech recognition currently remains somewhat unclear.

A more direct measure of spectral peak resolution for acoustic signals was recently developed by Henry and Turner (2003) and applied to CI listeners. The method is based on the "ripple phase reversal test" (Supin *et al.*, 1994). In Henry and Turner (2003) the stimuli were rippled noise signals, which are broadband noise signals with spectral ripples spaced on a linear scale. The task involves discriminating between two rippled noise stimuli in which the frequency positions of the peaks and valleys are interchanged. The ripple spacing is varied (with the ripple depth held constant), and the minimum ripple spacing at which a reversal in peak and trough positions can be detected is determined as the threshold for spectral peak resolution. Examples of rippled noise stimuli (with a logarithmic spacing of ripples in this case; see below) are shown in Fig. 1. This test is hypothesized to provide a direct measure of the ability of listeners to perceive the frequency locations of spectral peaks in a broadband acoustic signal. The results showed a significant relationship between spectral peak resolution and vowel recognition in CI listeners, indicating that listeners who can resolve more closely spaced peaks are better at recognizing vowels. This measure has potential applications in predicting and optimizing speech recognition in CI listeners. Furthermore, it is hypothesized that this test may also be applicable to listeners with acoustic hearing, and may therefore provide a measure of spectral peak resolution in HI listeners. As such, this test provides the opportunity to directly compare spectral peak resolution abilities among listeners with acoustic hearing (normal or impaired) and listeners with electric hearing.

In the present study, the spectral peak resolution test was adapted for the investigation of spectral peak resolution in NH and HI listeners, and for a further investigation of this ability in CI listeners. The spectral peak resolution test was implemented in this study using logarithmically spaced ripples, instead of using linear-spaced ripples, as in Henry and Turner (2003). There are two reasons for using a logarithmic spacing of ripples. First, it is hypothesized that such a spacing would more closely approximate the properties of the normal auditory system as well as the acoustics of speech, and may therefore be more strongly correlated with speech recognition. Second, since it is thought that the processing of linear rippled noises may involve a time-domain waveform analysis in the acoustic auditory system (e.g., Fay *et al.*, 1983; Yost *et al.*, 1996), log-rippled spectra are better suited to the specific examination of spectral peak resolution in listeners with acoustic hearing.

The modified spectral peak resolution test was used to investigate the differences in spectral peak resolution between and among NH, HI, and CI listeners. In addition, the possible relationship between spectral peak resolution and speech recognition both across the NH, CI, and HI listener groups, and within each of the CI and HI listener groups was examined. The present experiments, by measuring these relations across the wide range of listeners, can therefore test whether the ability to perceive the locations of spectral peaks is a requirement for speech recognition in general, and if there is some minimum requirement in this ability, below which speech perception is highly degraded.

II. METHODS

A. Subjects

Three subject groups participated in this study: (1) NH listeners, (2) HI listeners, and (3) CI listeners. All participants were native American English speaking adults. There were 12 young adult NH subjects. Normal hearing was defined as having pure-tone air conduction thresholds $\leqslant$15 dB HL at octave frequencies from 125 to 8000 Hz in the tested ear.

Thirty-two HI listeners ranging in age from 29 to 83 years participated. The hearing losses were diagnosed as sensorineural (and assumed to be of cochlear origin) based on the lack of an air-bone gap and tympanograms consistent with normal middle ear function. The ear with the better pure tone thresholds was selected as the test ear. The degree of hearing loss ranged from mild to profound, and the audiometric configurations (flat or sloping) varied across the HI listeners. Pure-tone thresholds for the test ear for each subject, along with the ages of each subject, are shown in Table I. In Table I, the downward-pointing arrows indicate that the thresholds were higher than could be measured with the audiometer.

The 23 CI subjects were users of the Cochlear Ltd. Nucleus 24M (CI24M) or Nucleus 24 Contour (CI24R) implant and had a minimum of 6 months experience with their

TABLE I. Individual subject details: Hearing impaired subjects. Audiometric thresholds for the test ear are shown in dB HL.

Subject	Age (yrs)	Frequency (Hz)								
		250	500	1000	1500	2000	3000	4000	6000	8000
HI1	55	35	40	50		50		40		50
HI2	69	25	20	25	35	55		65		85
HI3	55	10	10	25	30	45	60	70		55
HI4	64	10	15	20	55	85		80		75
HI5	75	40	60	50		35	45	65		65
HI6	47	15	15	45		35	20	15		15
HI7	81	20	25	25	25	65	70	70		80
HI8	55	15	10	15	60	70		70	85	95
HI9	29	60	70	100		100		↓		↓
HI10	71	40	40	55		45		40		65
HI11	63	60	55	45		40		45		60
HI12	59	35	35	30	45	55		55	80	85
HI13	76	15	15	25	25	45		60	60	80
HI14	69	65	60	65		70		55		85
HI15	57	35	25	20		25	60	65		55
HI16	61	15	20	35	50	60		70		65
HI17	61	35	30	55		60	90	105		↓
HI18	75	25	35	40		55	75	80		95
HI19	75	40	35	50		60		65		75
HI20	75	45	45	45		45		50		65
HI25	62	10	5	15	30	35	70	80		90
HI26	53	35	50	45		45		55		55
HI27	79	35	35	40		45		55		55
HI28	79	50	60	75		75		70		65
HI29	65	55	50	50		45		55		60
HI30	61	20	30	30	35	50		55		50
HI31	79	35	40	35		70		75		90
HI32	83	35	40	35		40	45	65		70
HI33	79	20	30	35		50	55	55	45	50
HI34	71	50	50	60		80	105	↓		↓
HI35	73	25	65	100		95		85		75
HI36	76	15	25	40		55	65	75		70

implant. Nineteen subjects used the CI24M implant, which has 22 intracochlear and 2 extracochlear electrodes, while the remaining subjects used the CI24R implant, which has a preformed (curved) perimodiolar electrode array instead of a straight array, and is designed to be positioned adjacent to the modiolar wall, decreasing the distance to the target neurons. Three subjects used the Continuous Interleaved Sampling (CIS) strategy (Wilson *et al.*, 1991), 5 used the SPEAK strategy (Seligman and McDermott, 1995), and 15 used the Advanced Combination Encoder (ACE) strategy (Skinner *et al.*, 2002; Vandali *et al.*, 2000). In the CIS strategy implemented with the Nucleus device, the amplitude envelope is estimated within each of typically 6–12 channels during each stimulation period. These amplitudes are converted to electrical stimulation levels, and stimulus pulses representing each band are presented sequentially on the associated electrodes at a rate between 740 and 2400 pulses per second (pps)/channel. SPEAK and ACE are both "peak-picking" strategies that estimate the amplitude envelope in up to 20 (SPEAK) or 22 (ACE) channels, each assigned in a tonotopic order to an equal number of implanted electrodes. In each analysis cycle, the channels with the largest amplitudes (maxima) are selected, and stimulus pulses are then presented sequentially on the associated electrodes. The number of maxima selected is 6 on average for SPEAK and between 1 and 20 for ACE (typically between 8 and 12), and the rate of stimulation is approximately 250 pps/channel for SPEAK and is between 250 and 2400 pps/channel (limited by the maximum rate of 14 400 pps across all channels) for ACE. Individual subject details are shown in Table II, including the parameters used for each subject's map. Also shown in Table II are audition-alone word recognition scores for CNC words, measured in the University of Iowa Cochlear Implant Clinic during the most recent speech processor mapping session with the subject's clinical map.

B. Stimuli

Speech recognition was assessed using vowel and consonant stimuli. The consonant test used a closed-set 16-alternative identification paradigm for consonants presented in an /a/-consonant-/a/ context (Turner *et al.*, 1995). The tokens were produced by four talkers (2 female and 2 male). Each talker produced one token of each of the /aCa/syllables,

TABLE II. Individual subject details: Cochlear implant subjects. Prog.=progressive; ACE used 8 maxima unless otherwise noted.

Subject	Age (yrs)	Duration of profound deafness (yrs)	CI experience (yrs)	Etiology	Implant type, processor type	Processing strategy, # maxima	Pulse rate (pps/ch), # channels	CNC word score (% correct)	Average dynamic range (dB)
CI1	74	12	6	Infection	CI24M	ACE	720, 22	86	8.6
CI2	47	13	3	Congenital, prog.	CI24M	ACE	900, 22	39	12.5
CI3	64	4	5	Unknown	CI24M	ACE	720, 18	56	9.7
CI5	73	8	4	Congenital, prog.	CI24M	ACE	900, 20	64	7.6
CI6	73	1	2	Meniere's disease	CI24M	SPEAK	250, 18	54	4.9
CI7	44	0.5	3	Autoimmune disease	CI24M	ACE	1200, 22	72	15.5
CI10	75	25	5	Congenital, prog.	CI24M	ACE	720, 20	68	7.7
CI11	77	40	3	Unknown	CI24M	CIS	900, 6	4	6.9
CI13	49	2	2	Unknown	CI24M	CIS	2400, 6	74	12.0
CI14	55	5	4	Unknown	CI24M	ACE, 10	1200, 22	18	4.0
CI15	81	3	4	Unknown	CI24M	SPEAK	250, 19	54	4.9
CI16	37	2	3	Unknown	CI24R	ACE, 12	720, 22	42	8.8
CI18	57	36	6	Unknown	CI24M	ACE, 12	1200, 20	50	9.0
CI19	79	0.5	5	Viral infection	CI24M	SPEAK	250, 20	66	2.5
CI20	47	7	4	Unknown	CI24M	ACE	900, 20	22	16.5
CI22	63	0.3	3	Infection	CI24M	ACE	720, 22	82	10.9
CI23	75	8	3	Unknown	CI24M	SPEAK	250, 20	42	4.9
CI24	85	11	2	Unknown	CI24R	SPEAK	250, 18	24	5.8
CI25	76	10	4	Unknown, prog.	CI24M	ACE	900, 22	68	11.6
CI26	62	1	3	Meniere's disease	CI24M	ACE	720, 22	54	9.5
CI27	47	28	0.5	Infection	CI24R	ACE	900, 22	64	19.2
CI28	41	3	2	Hereditary	CI24R	ACE	900, 22	84	9.5
CI29	49	8	3	Unknown, prog.	CI24M	CIS	900, 6	58	7.7

and each token was repeated in random order 3 times in the test, for a total of 192 test items. Vowel recognition was measured using a closed-set 12-alternative identification procedure. Medial vowel tokens produced by 10 male and 10 female talkers were selected from the materials recorded by Hillenbrand *et al.* (1995), and presented in a /h/-vowel-/d/ context, for a total of 240 test items. The speech stimuli were stored in digital form on a Macintosh G4 computer.

Rippled noise stimuli of 100–5000 Hz bandwidth and with peak-to-valley ratios of approximately 30 dB were synthesized on an Apple Macintosh G4 computer by algebraically summing 200 pure-tone frequency components with amplitudes determined by a sinusoidal envelope with ripples spaced on a logarithmic frequency scale. The starting phases of the individual frequency components were randomized for each stimulus to avoid fine structure pitch cues that may be perceptible to listeners. The frequency of the spectral envelope of the stimulus complex was varied in 14 steps: 0.125, 0.176, 0.250, 0.354, 0.500, 0.707, 1.000, 1.414, 2.000, 2.828, 4.000, 5.657, 8.000, and 11.314 ripples per octave (ripples/octave). The spectral envelope phase of the stimulus complex was set to zero at the low-frequency edge of the complex for the standard (reference) stimulus, and the inverted (test) stimulus had a reversed phase. Examples of standard and inverted 0.25, 1 and 2 ripples/octave rippled noise spectra are shown in Fig. 1. The stimuli were of 500 ms duration and had 150 ms rise/fall times, and were shaped with a filter that approximated the long-term speech spectrum (Byrne *et al.*, 1994). The overall levels of the rippled noise sound files were then approximately equalized.

C. Procedures

All subjects were tested in a double-walled sound treated room. The speech and rippled noise stimuli were output via custom software routines through a 16-bit DigiDesign (Audio Media III) digital-to-analog converter at a sampling rate of 44.1 kHz. These stimuli were presented to NH and HI subjects monaurally through Sennheiser HD 25-SP1 circumaural headphones. The presentation level for both speech and rippled noise stimuli was 65 dB SPL for the normal-hearing listeners. Stimuli were presented to HI subjects through an analog high-pass emphasis spectrum shaper (Altec-Lansing 1753), which provided approximately 20 dB of relative gain with a transition slope of 40 dB/octave, starting at 1000 Hz. High-frequency emphasis was not provided to either the NH or CI subjects. The presentation level was set on an individual basis for each of the HI listeners at the highest possible level that was acceptable for each subject, as determined in pilot test sessions. The aim was to optimize the audibility of the signals across the wide range of frequencies for the HI listeners, although it should be noted that the high-pass emphasis would not have been sufficient to provide signal audibility in the high frequencies for some subjects. The chosen level for the speech materials was then also used for the rippled noise stimuli.

The speech and rippled noise stimuli were presented to the CI subjects using the SPrint speech processor in the free field, positioned approximately 1 m from a loudspeaker (Cerwin-Vega Model E712), at an average level of 65 dB SPL. The laboratory SPrint speech processor was pro-

grammed with the clinical map used by each subject (see Table II). While it seems highly likely that spectral resolution may be affected by electrode array design, speech processing strategy, and processing parameters, our purpose in this study was not to directly investigate these effects. Instead, these potential sources of variability in spectral resolution across CI listeners could be exploited to test if spectral resolution abilities were predictive of speech recognition. Therefore, various processing strategies and electrode designs were specifically included in this study in order to assess the relationship between spectral peak resolution and speech recognition with the everyday map used by each subject. In all cases, stimulus pulses of 25 μs duration were presented using the monopolar M1+2 electrode configuration, where current flows between the active intracochlear electrode and both extracochlear electrodes. Prior to commencing the experiment, threshold (T) and comfortably loud (C) levels were measured for each electrode, using standard clinical procedures. In order to minimize as much as possible individual variation in the audibility of acoustic signals, the same speech processor sensitivity (set to 8) was used for all subjects. The sensitivity was set so that peaks in the stimulus resulted in electrical stimulation at approximately 90% of the dynamic range for the 65 dB SPL acoustic input signal. This setting was determined by measuring the speech processor output using the SCILAB (Swiss Cochlear Implant Laboratory software) program (Lai *et al.*, 2003), which records the RF transmissions from the speech processor and provides the current levels for all activated electrodes. Reference was also made to the published sound pressure levels that result in electrical stimulation for the range of sensitivity settings for the SPrint processor (Nucleus Technical Reference Manual, Fig. 3.12).

A single run of the speech tests consisted of 240 trials for vowels and 192 items for consonants. The test items (12 words containing the medial vowels for the vowel test; 16 individual consonants for the consonant test) were displayed as buttons on a touchscreen (MicroTouch). On each trial, a stimulus token was chosen randomly, without replacement, and following the presentation of each token the subject responded by pressing one of the buttons on the touch screen. Two runs (a practice and a test run) of each test were administered. Correct-answer feedback was provided during the practice run only.

Prior to speech testing, training was provided for both the vowel and consonant identification tasks. Subjects were instructed to press the button on the touch screen corresponding to the phoneme they wanted to hear, and five examples of that phoneme (i.e., the phoneme spoken by five different talkers) were then presented. Fifty trials of the training task were conducted (or more as desired by the individual subject) in order to allow the subjects to familiarize themselves with the phonemes and their associated touch screen labels.

Ripple resolution thresholds were determined using a three interval forced-choice adaptive procedure, based on the method developed by Henry and Turner (2003). For each set of three intervals, two intervals contained the standard or reference stimulus, and the test interval, chosen at random, contained the inverted stimulus. There was an interstimulus interval of 500 ms. In order to minimize the use of loudness cues, the presentation level was varied randomly from trial to trial within an 8 dB range in 1 dB steps, using a Tucker-Davis Technologies programmable attenuator. Three numerically labeled buttons were displayed on the touch screen, corresponding to the three intervals, and subjects were instructed to press the button corresponding to the interval that sounded "different" (i.e., that contained the test stimulus), ignoring any loudness variation between intervals. Correct answer feedback was provided throughout the experiment. Each test run commenced at a ripple frequency of 0.176 ripples/octave, and the ripple frequency was varied in a two-down, one-up procedure. After each incorrect response the ripple frequency was decreased by a step, and it was increased after two correct responses. This procedure converged on the 70.7% correct point (Levitt, 1971) for ripple resolution. The threshold was estimated for each run as the geometric mean of the ripple frequencies for the final 8 of 12 reversals. Based on previous results using the linear rippled noise stimuli in this laboratory (Henry and Turner, 2003), and pilot testing for these logarithmic rippled noise stimuli, only a few practice runs are necessary to achieve asymptotic performance. Therefore, four practice runs were completed for each subject. Following the practice runs, three test runs were obtained for each subject, and the final threshold value for each subject was recorded as the arithmetic mean of the thresholds across these three test runs.

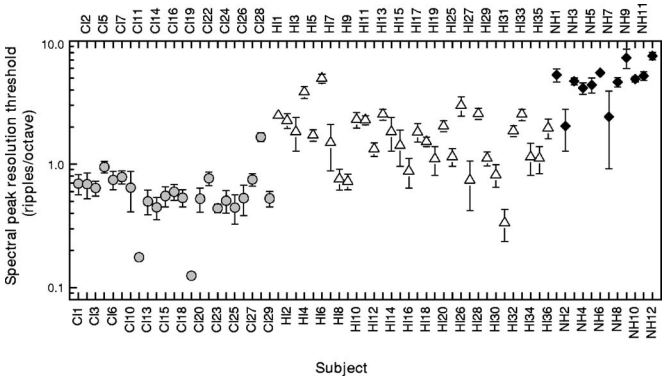


FIG. 2. Thresholds for spectral peak resolution for NH, HI, and CI subjects. Error bars represent ± one standard deviation.

III. RESULTS

The mean spectral resolution thresholds are shown for each subject in Fig. 2. Higher thresholds (more ripples per octave) indicate a better spectral peak resolution ability. Spectral peak resolution varied from 0.13 to 7.55 ripples/octave across all listeners. NH listeners had the best spectral peak resolution, with an average threshold across listeners of 4.84 ripples/octave and a range of 2.03–7.55 ripples/octave, while CI listeners had the poorest spectral peak resolution, with an average threshold across listeners of 0.62 ripples/octave, and a range of 0.13–1.66 ripples/octave. The average spectral peak resolution threshold of 1.77 ripples/octave for the HI listeners was between those of the NH and the CI

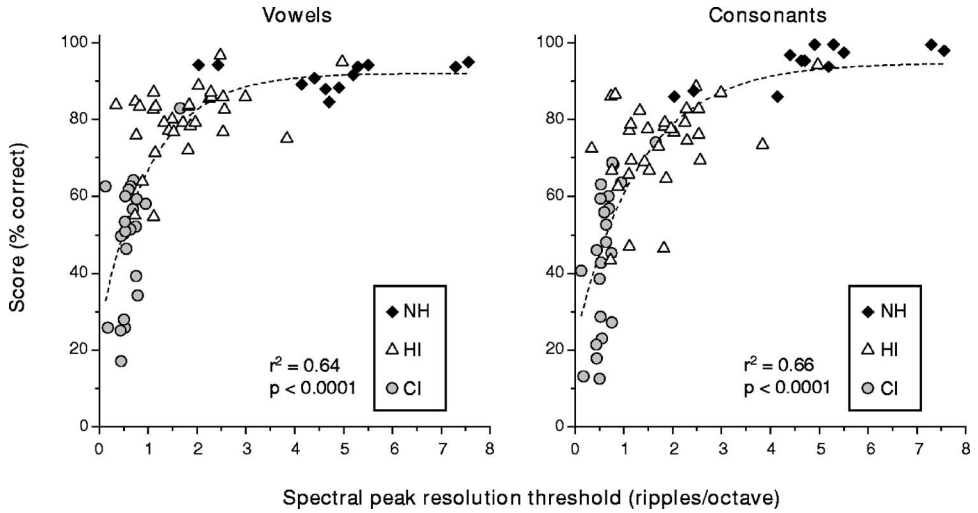


FIG. 3. The relationship between spectral peak resolution and vowel recognition (left panel) and consonant recognition (right panel) across NH, HI, and CI subjects. The dashed curves represent the functions of best fit to the data [Eq. (1)].

groups, and the individual thresholds of 0.33–4.97 ripples/octave essentially spanned the range of thresholds of the NH and CI groups.

The relationship between spectral peak resolution and both vowel and consonant recognition across the three listener types (NH, HI, and CI) is shown in Fig. 3. The following function was found to provide the best fit to both the vowel and consonant data:

$$P = ae^{-S/b} + c, \tag{1}$$

where P is the percent correct score, S is the spectral peak resolution threshold, and a, b, and c are fitting parameters. For the vowel data, $a=-66.88$, $b=1.03$, and $c=92.12$, and for the consonant data, $a=-72.24$, $b=1.33$, and $c=94.76$. Nonlinear regression analysis based on the fitted functions indicated a significant relationship between spectral peak resolution and both vowel recognition ($r^2=0.64, p<0.0001$) and consonant recognition ($r^2=0.66, p<0.0001$). These results suggest that the ability to resolve spectral peaks in a complex acoustic spectrum may be associated with accurate speech recognition.

For vowel recognition, there was an asymptote in performance at a score of approximately 92% correct, which corresponded to a spectral peak resolution threshold of approximately 4 ripples/octave, and for consonant recognition there was an asymptote in performance at a score of approximately 94%, which corresponded to a spectral peak resolution threshold of approximately 4.5 ripples/octave. There was a rapid deterioration in performance for both vowel and consonant recognition when spectral peak resolution fell below approximately 1–2 ripples/octave.

The relationship between spectral peak resolution and speech recognition was also examined for the CI and HI listener groups individually. The relationship between spectral peak resolution and vowel recognition (left panel) and consonant recognition (right panel) for the CI listener group is shown in Fig. 4 and for the HI listener group in Fig. 5. Regression analyses showed a significant moderate linear correlation between spectral peak resolution thresholds and both vowel recognition ($r^2=0.27, p=0.01$) and consonant

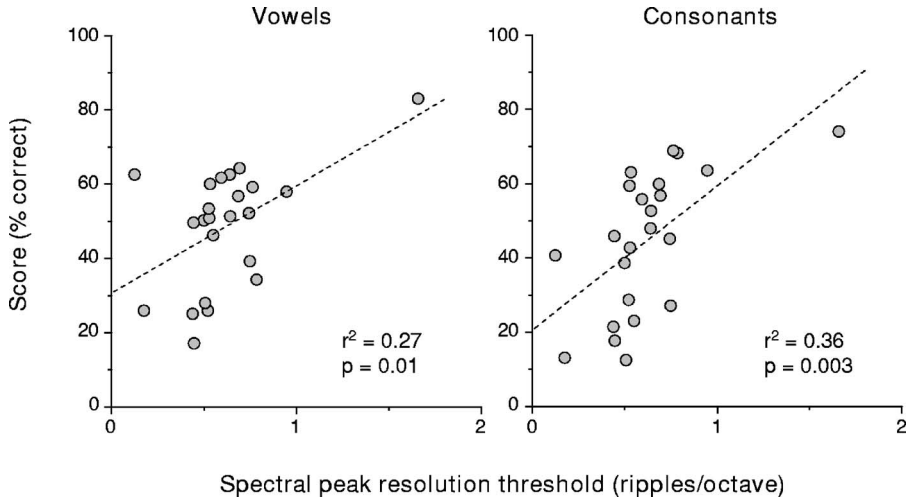


FIG. 4. The relationship between spectral peak resolution and vowel recognition (left panel) and consonant recognition (right panel) for CI subjects. Linear regressions are represented by the dashed lines.

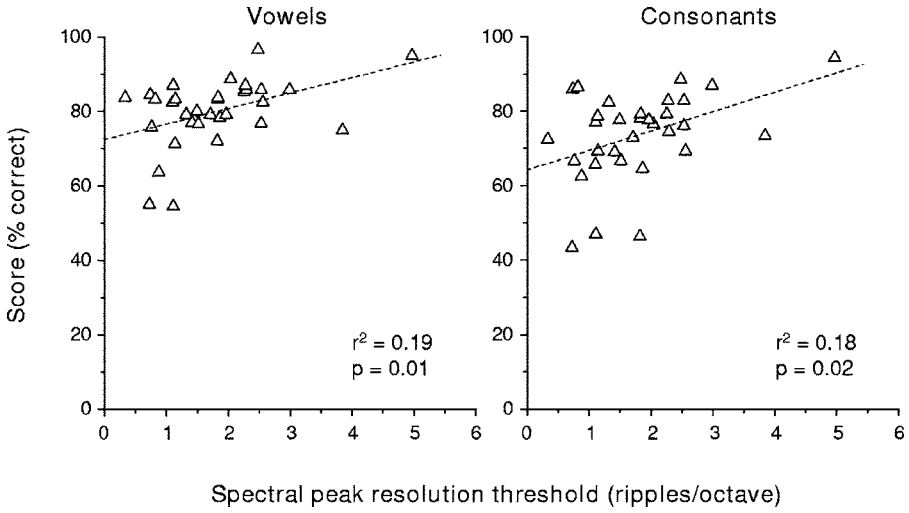


FIG. 5. The relationship between spectral peak resolution and vowel recognition (left panel) and consonant recognition (right panel) for HI subjects. Linear regressions are represented by the dashed lines.

recognition ($r^2=0.36, p=0.003$) for the CI listeners and a significant but quite weak linear correlation between spectral peak resolution thresholds and both vowel recognition ($r^2=0.19, p=0.01$) and consonant recognition ($r^2=0.18, p=0.02$) for the HI listeners.

For the HI listener group, while there was a significant correlation between absolute threshold (calculated as the average of thresholds at 0.5, 1 and 2 kHz) and speech recognition (consonants: $r^2=-0.22$, $p=0.006$; vowels: $r^2=-0.24$, $p=0.004$), there was no correlation between absolute threshold and spectral peak resolution threshold ($r^2=-0.03, p=0.35$). Therefore it seems unlikely that the observed relationship between speech recognition and spectral peak resolution ability was due to underlying relationships between each of these variables and absolute threshold. This finding is in contrast to previous studies that have generally shown a correlation of both frequency selectivity and speech recognition with absolute hearing thresholds in HI listeners, and a reduction or elimination of correlations between frequency selectivity and speech recognition when the effect of absolute threshold was statistically partialed out (see the Introduction). The reason for the lack of a correlation between spectral peak resolution and absolute threshold in this study is not clear. It may suggest that absolute threshold may not be associated with the ability to perform this spectral peak resolution task, possibly due to the fact that the audibility of the rippled noise signals was optimized for individual listeners, or may arise from the fact that broadband stimuli were used to assess spectral peak resolution, where listeners could use the region in which spectral resolution was best, while absolute threshold was averaged across frequency.

For both the CI and HI listeners there were no significant correlations between age and either speech recognition (averaged consonant and vowel score, CI: $r^2=-0.11$, $p=0.12$; HI: $r^2=0.01$, $p=0.55$) or spectral peak resolution threshold (CI: $r^2=-0.17$, $p=0.07$; HI: $r^2=-0.01$, $p=0.56$).

IV. DISCUSSION

The research reported in this study suggests a possible relationship between spectral peak resolution and speech recognition across the NH, HI, and CI listener groups (Fig. 3). The regression analyses between spectral peak resolution and speech recognition across all listeners accounted for 64% of the variance in vowel scores and 66% of the variance in consonant scores, indicating that there may be an underlying dependence for speech recognition in general upon spectral peak resolution. These relationships appear to be stronger than those demonstrated in previous studies that have attempted to link speech recognition and spectral resolution in HI listeners (e.g., Dreschler and Plomp, 1980, 1985; Festen and Plomp, 1983; Lutman and Clark, 1986; Stelmachowicz *et al.*, 1985; Tyler *et al.*, 1982) and speech recognition and place of stimulation perception in CI listeners (e.g., Collins *et al.*, 1997; Donaldson and Nelson, 2000; Henry *et al.*, 2000; Nelson *et al.*, 1995; Throckmorton and Collins, 1999; Zwolan *et al.*, 1997). Stronger relationships in the present study may result from assessing spectral resolution ability in a wide range of individuals across the clinical populations. The spectral peak resolution measure used in this study provides the opportunity to investigate spectral peak resolution in NH, HI, and CI listeners, which has not been possible in the previous studies since the methodologies used to assess spectral resolution in the acoustic hearing (such as psychoacoustic masking measures) and electric hearing (such as electrode discrimination measures) listener groups individually cannot be applied to the assessment of spectral resolution across all groups. While the overall relationship between spectral peak resolution and speech recognition is nonlinear [Fig. 3 and Eq. (1)], reference to the data subsets for the isolated CI and HI groups (Figs. 4 and 5, respectively) shows that the overall nonlinear relationship seen across the listener groups is reduced and is linear. This indicates that restricting the examination of possible relationships between spectral resolution and speech recognition to individual listener groups, as has been done in previous studies, may obscure the overall nonlinear form of the relation and reduce its strength.

What degree of spectral peak resolution is required for accurate speech recognition, and below what degree of spectral peak resolution does speech recognition become highly

degraded? While the ability to resolve a higher number of ripples per octave was generally associated with better speech recognition, there was a plateau in multitalker vowel and consonant recognition at around four ripples/octave (Fig. 3). This indicates that spectral peak resolution better than around four ripples/octave is probably not necessary for the accurate identification of vowels and consonants produced by multiple talkers in quiet backgrounds. The relationship illustrated in Fig. 3 also indicates that spectral peak resolution poorer than around one to two ripples/octave, as seen in many CI listeners and some HI listeners, may result in substantially reduced speech recognition. This finding of an association between severely reduced spectral peak resolution and degraded speech intelligibility in this study is broadly consistent with studies by Baer and Moore (1993), ter Keurs *et al.* (1992), Shannon *et al.* (1995), and others (see the Introduction), who have shown that frequency resolution must be highly impaired in order to severely degrade speech recognition in quiet, and, further, it provides a quantification of the limits of spectral peak resolution below which significant degradation in speech recognition may occur.

While the stimulus used in the spectral peak resolution task was broadband, accurate performance did not require broadband analysis of the signal. Listeners may have used a specific region in which their ability to resolve spectral peaks was particularly good. The specific region of the frequency band the listeners were using for their discrimination was not determined in this study. In addition, listeners may have potentially used level cues at the lower or upper spectral edges of the rippled noise stimuli, since the stimuli were not tapered at the spectral edges (apart from the speech spectrum shaping), and the random variation in the presentation level of the stimuli within an 8 dB range (see Sec. II) may not have been sufficient to eliminate these cues. Despite these limitations, the task provided a strong prediction of speech recognition, with some implant listeners behaving as if they were limited essentially to a single channel, and normal-hearing listeners showing much finer spectral resolution. Further research is required to determine which frequency region(s) are used by individual listeners to perform the task, and the potential perception of level cues at the spectral edges of the stimuli.

These results may have important implications for speech recognition in both HI and CI listeners. Current CI devices and speech processing strategies preserve only crude spectral information. Indeed, while spectral peak resolution varied among CI listeners, and some CI listeners showed performance in the range of the better HI subjects, spectral resolution was poorest in CI listeners on average (see Fig. 2). In addition, while many HI listeners showed spectral peak resolution within the normal range, some HI listeners showed substantially reduced spectral peak resolution. These results indicate that efforts to improve spectral resolution in HI listeners via improved hearing aids, and in CI listeners via improved electrode arrays and speech processing strategies, may result in improved speech recognition.

Turning to the relationships between spectral peak resolution and speech recognition for the individual clinical groups, there was a moderate correlation between spectral peak resolution and both vowel and consonant recognition within the CI listener group, and a fairly weak but significant correlation between spectral peak resolution and vowel and consonant recognition within the HI listener group. The spectral peak resolution test has potential clinical applications in the prediction of speech recognition in HI and CI listeners, as well as in optimizing speech recognition in CI listeners by using the test to determine in a time-efficient manner which speech processing strategy and particular speech processing parameters may provide the best spectral peak resolution for an individual listener. Further research is required to determine whether optimization of the spectral peak resolution test, for instance, by measuring spectral peak resolution in different frequency regions, may improve the predictive value of this test. For CI listeners, while a variety of different speech processing strategies and parameter settings were included in this study in order to specifically assess the relationship between spectral peak resolution and speech recognition with each subject's everyday map, further research is required to examine the effects of these strategies and parameters on both spectral peak resolution and speech recognition. Finally, it seems likely that perceptual factors in addition to spectral peak resolution may contribute to deficits in speech recognition. For example, in cases of poor spectral peak resolution, listeners may rely more on temporal aspects of the speech signal. Modeling other perceptual factors, such as temporal resolution together with spectral peak resolution, may account for a higher amount of variance in speech recognition. Such modeling should explore the inclusion of audibility measures for HI listeners, since audibility is a primary factor related to performance variability in these listeners.

It is important to consider the fact that the findings reported in the present study apply to speech recognition in quiet backgrounds. It is well known, however, that speech recognition in HI and CI listeners is highly susceptible to the effects of competing backgrounds, and it is likely that this is due to reduced spectral resolution. Research suggests that spectral smearing has a more detrimental effect on speech recognition in NH listeners when the processed speech signal is presented in competing backgrounds compared to quiet listening conditions (Baer and Moore, 1993, 1994; Boothroyd *et al.*, 1996; ter Keurs *et al.*, 1992, 1993). In addition, CI simulations in NH listeners indicate that a higher number of spectral channels is required when listening in competing backgrounds compared to quiet listening conditions. Friesen *et al.* (2001) showed an increase in performance in competing backgrounds as the number of channels was increased to 20 channels, which was the highest number tested. Importantly, however, CI listeners in that study did not show a similar increase in performance as the number of channels was increased to 20, but rather showed an asymptote in performance on average with between 4 and 7 channels (depending on the speech material). Further research is required to quantify the role of reduced spectral peak resolution in speech recognition in competing backgrounds.

V. CONCLUSIONS

A direct method of measuring the ability to resolve spectral peaks in complex acoustic spectra was applied to NH, HI, and CI listeners in this study. This method enabled a comparison of spectral peak resolution between NH, HI, and CI listener groups, and an investigation of the relationship between spectral peak resolution and speech recognition across a wide range of perceptual abilities.

The principal findings were as follows.

(1) Spectral peak resolution varied widely among listeners, from 0.13 to 7.55 ripples/octave. The average spectral peak resolution was 4.84 ripples/octave in NH listeners (2.03–7.55 ripples/octave), 1.77 ripples/octave in HI listeners (0.33–4.97 ripples/octave), and 0.62 ripples/octave in CI listeners (0.13–1.66 ripples/octave).
(2) There was a significant relationship between spectral peak resolution and both vowel and consonant recognition across the NH, HI, and CI listener groups, suggesting that the ability to resolve spectral peaks in a complex acoustic spectrum is associated with accurate speech recognition.
(3) There was a plateau in vowel and consonant recognition at around four ripples/octave, indicating that spectral peak resolution better than around four ripples/octave is probably not necessary for the accurate identification of vowels and consonants produced by multiple talkers in quiet backgrounds.
(4) Both vowel and consonant recognition performance deteriorated rapidly when spectral peak resolution fell below one to two ripples/octave.
(5) There was a significant but quite weak linear correlation between spectral peak resolution thresholds and both vowel and consonant recognition for the HI listeners and a significant moderate linear correlation for the CI listeners.

ACKNOWLEDGMENTS

We are grateful to the subjects for their considerable time and effort in participating in this research. We wish to thank Ken Grant, Brian Moore, and Monita Chatterjee for their very helpful comments on previous versions of this paper. Thanks also to Norbert Dillier and Waikong Lai for use of the SCILAB program, and also to Arik Wald and Stephanie Schultz for their assistance. Funding for this research was provided in part by Grant No. 1RO1DC000377 from the National Institutes on Deafness and Other Communicative Disorders, National Institutes of Health and Grant No. RR00059 from the General Clinical Research Centers Program NCRR, National Institutes of Health.

Baer, T., and Moore, B. C. J. (**1993**). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am. **94**, 1229–1241.

Baer, T., and Moore, B. C. J. (**1994**). "Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech," J. Acoust. Soc. Am. **95**, 2277–2280.

Boothroyd, A., Mulhearn, B., Ging, J., and Ostroff, J. (**1996**). "Effects of spectral smearing on phoneme and word recognition," J. Acoust. Soc. Am. **100**, 1807–1818.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Kotby, M. N., Nasser, N. H. A., El Kholy, W. A. H., Nakanishi, Y., Oyer, H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavartkiladze, G., Frolenkov, G. I., Westerman, S., and Ludvigsen, C. (**1994**). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. **96**, 2108–2120.

Ching, T., Dillon, H., and Byrne, D. (**1998**). "Speech recognition of hearing impaired listeners: Predictions from audibility and the limited role of high frequency amplification," J. Acoust. Soc. Am. **103**, 1128–1140.

Collins, L. M., Zwolan, T. A., and Wakefield, G. H. (**1997**). "Comparison of electrode discrimination, pitch ranking, and pitch scaling data in postlingually deafened adult cochlear implant subjects," J. Acoust. Soc. Am. **101**, 440–455.

Dallos, P., Ryan, A., Harris, D., McGee, T., and Ozdamar, O. (**1977**). "Cochlear frequency selectivity in the presence of hair cell damage," in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic Press, New York), pp. 249–258.

Donaldson, G. S., and Nelson, D. A. (**2000**). "Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies," J. Acoust. Soc. Am. **107**, 1645–1658.

Dorman, M. F., and Loizou, P. C. (**1997**). "Speech intelligibility as a function of the number of channels of stimulation for normal-hearing listeners and patients with cochlear implants," Am. J. Otolaryngol. **18**, S113–S114.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997**). "Speech understanding as a function of the number of channels of stimulation for processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**, 2403–2411.

Dreschler, W. A., and Plomp, R. (**1980**). "Relations between psychophysical data and speech perception for hearing-impaired subjects. I," J. Acoust. Soc. Am. **68**, 1608–1615.

Dreschler, W. A., and Plomp, R. (**1985**). "Relations between psychophysical data and speech perception for hearing-impaired subjects. II," J. Acoust. Soc. Am. **78**, 1261–1270.

Dubno, J. R., and Dirks, D. D. (**1989**). "Auditory filter characteristics and consonant recognition for hearing-impaired listeners," J. Acoust. Soc. Am. **85**, 1666–1675.

Dubno, J. R., Dirks, D. D., and Ellison, D. E. (**1989**). "Stop-consonant recognition for normal-hearing listeners and listeners with high-frequency hearing loss. II: Articulation index predictions," J. Acoust. Soc. Am. **85**, 355–364.

Fay, R. R., Yost, W. A., and Coombs, S. (**1983**). "Psychophysics and neurophysiology of repetition noise processing in a vertebrate auditory system," Hear. Res. **12**, 31–55.

Festen, J. M., and Plomp, R. (**1983**). "Relations between auditory functions in impaired hearing," J. Acoust. Soc. Am. **73**, 652–662.

Fishman, K., Shannon, R. V., and Slattery, W. H. (**1997**). "Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor," Hear. Res. **40**, 1201–1215.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

Glasberg, B., and Moore, B. C. J. (**1986**). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," J. Acoust. Soc. Am. **79**, 1020–1033.

Glasberg, B. R., and Moore, B. C. J. (**1989**). "Psychoacoustic abilities of subjects with unilateral and bilateral cochlear impairments and their relationship to the ability to understand speech," Scand. Audiol. Suppl. **32**, 1–25.

Henry, B. A., and Turner, C. W. (**2003**). "The resolution of complex spectral patterns in cochlear implant and normal hearing listeners," J. Acoust. Soc. Am. **113**, 2861–2873.

Henry, B. A., McKay, C. M., McDermott, H. J., and Clark, G. M. (**2000**). "The relationship between speech perception and electrode discrimination in cochlear implantees," J. Acoust. Soc. Am. **108**, 1269–1280.

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.

Hogan, C. A., and Turner, C. W. (**1998**). "High-frequency audibility: Benefits for hearing-impaired listeners," J. Acoust. Soc. Am. **104**, 432–441.

Humes, L. E., Dirks, D. D., Bell, T. S., Ahlstrom, C., and Kincaid, G. E. (**1986**). "Application of the articulation index and the speech transmission index to the recognition of speech by normal-hearing and hearing-impaired listeners," J. Speech Hear. Res. **29**, 447–462.

Lai, W. K., Bogli, H., and Dillier, N. (**2003**). "A software tool for analyzing multichannel cochlear implant signals," Ear Hear. **24**, 380–391.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
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.
Lutman, M. E., and Clark, J. (**1986**). "Speech identification under simulated hearing-aid frequency response characteristics in relation to sensitivity, frequency resolution, and temporal resolution," J. Acoust. Soc. Am. **80**, 1030–1040.
Nelson, D. A., Van Tassell, D. J., Schroder, A. C., Soli, S., and Levine, S. (**1995**). "Electrode ranking of 'place pitch' and speech recognition in electrical hearing," J. Acoust. Soc. Am. **98**, 1987–1999.
Pavlovic, C. V. (**1984**). "Use of the articulation index for assessing residual auditory function in listeners with sensorineural hearing impairment," J. Acoust. Soc. Am. **75**, 1253–1258.
Pavlovic, C. V., Studebaker, G. A., and Sherbecoe, R. L. (**1986**). "An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals," J. Acoust. Soc. Am. **80**, 50–57.
Seligman, P. M., and McDermott, H. J. (**1995**). "Architecture of the Spectra 22 speech processor," Ann. Otol. Rhinol. Laryngol. Suppl. 104, 139–141.
Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.
Skinner, M. W. (**1980**). "Speech intelligibility in noise-induced hearing loss: Effects of high-frequency compensation," J. Acoust. Soc. Am. **67**, 306–317.
Skinner, M. W., Arndt, P. L., and Staller, S. J. (**2002**). "Nucleus 24 advanced encoder conversion study: Performance versus preference," Ear Hear. **23**, 2S–16S.
Stelmachowicz, P. G., Jesteadt, W., Gorga, M. P., and Mott, J. (**1985**). "Speech perception ability and psychophysical tuning curves in hearing-impaired listeners," J. Acoust. Soc. Am. **77**, 620–627.
Supin, A., Popov, V. V., Milekhina, O. N., and Tarakanov, M. B. (**1994**). "Frequency resolving power measured by rippled noise," Hear. Res. **78**, 31–40.
ter Keurs, M., Festen, J. M., and Plomp, R. (**1992**). "Effect of spectral envelope smearing on speech reception. I," J. Acoust. Soc. Am. **91**, 2872–2880.
ter Keurs, M., Festen, J. M., and Plomp, R. (**1993**). "Limited resolution of spectral contrast and hearing loss for speech in noise," J. Acoust. Soc. Am. **94**, 1307–1314.
Throckmorton, C. S., and Collins, L. M. (**1999**). "Investigation of the effects of temporal and spatial interactions on speech-recognition skills in cochlear-implant subjects," J. Acoust. Soc. Am. **105**, 861–873.
Trees, D. A., and Turner, C. W. (**1986**). "Spread of masking in normal subjects and in subject with hearing loss," Audiology **25**, 70–83.
Turner, C. W., Chi, S., and Flock, S. (**1999**). "Limiting spectral resolution in speech for listeners with sensorineural hearing loss," J. Speech Lang. Hear. Res. **42**, 773–784.
Turner, C. W., Souza, P. E., and Forget, L. N. (**1995**). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," J. Acoust. Soc. Am. **97**, 2568–2576.
Tyler, R. S., Wood, E. J., and Fernandez, M. (**1982**). "Frequency resolution and hearing loss," Br. J. Audiol. **16**, 45–83.
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.
Wightman, F. L., 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. F. Evans and J. P. Wilson (Academic, London), pp. 295–306.
Wilson, B. S., Finley, C. F., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). "Better speech recognition with cochlear implants," Nature (London) **352**, 236–238.
Yost, W. A., Patterson, R., and Sheft, S. (**1996**). "A time domain description for the pitch strength of iterated rippled noise," J. Acoust. Soc. Am. **99**, 1066–1078.
Zwolan, T. A., Collins, L. M., and Wakefield, G. H. (**1997**). "Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects," J. Acoust. Soc. Am. **102**, 3673–3685.

Using auditory-visual speech to probe the basis of noise-impaired consonant–vowel perception in dyslexia and auditory neuropathy

Joshua Ramirez and Virginia Mann
Cognitive Science, University of California, Irvine, 3151 Social Science Plaza, Irvine, California 92697

(Received 15 July 2004; revised 4 May 2005; accepted 4 May 2005)

Both dyslexics and auditory neuropathy (AN) subjects show inferior consonant–vowel (CV) perception in noise, relative to controls. To better understand these impairments, natural acoustic speech stimuli that were masked in speech-shaped noise at various intensities were presented to dyslexic, AN, and control subjects either in isolation or accompanied by visual articulatory cues. AN subjects were expected to benefit from the pairing of visual articulatory cues and auditory CV stimuli, provided that their speech perception impairment reflects a relatively peripheral auditory disorder. Assuming that dyslexia reflects a general impairment of speech processing rather than a disorder of audition, dyslexics were not expected to similarly benefit from an introduction of visual articulatory cues. The results revealed an increased effect of noise masking on the perception of isolated acoustic stimuli by both dyslexic and AN subjects. More importantly, dyslexics showed less effective use of visual articulatory cues in identifying masked speech stimuli and lower visual baseline performance relative to AN subjects and controls. Last, a significant positive correlation was found between reading ability and the ameliorating effect of visual articulatory cues on speech perception in noise. These results suggest that some reading impairments may stem from a central deficit of speech processing. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1940509]

PACS number(s): 43.71.−k, 43.71.Es, 43.71.Ky, 43.71.Ma [ALF] Pages: 1122–1133

I. INTRODUCTION

Developmental dyslexia has been characterized as a specific learning disability that accounts for difficulties in acquiring basic proficiency in reading and writing. This reading disorder impacts many reading and reading-related behaviors—(including but not limited to) single word decoding, phonological processing (Torgesen and Wagner, 1998; Torgesen, 1999), and phonological awareness (for reviews, see Mann, 2002; Torgesen and Wagner, 1998; Torgesen, 1999)—and cannot be attributed to experiential factors (i.e., impoverished educational opportunity), personal motivation, or diminished intellectual capacity (Vellutino, Scanlon, and Lyon, 2000). Over the course of life-long development, reading disorders like dyslexia share a definite continuity from childhood to adulthood (Scarborough, 1984). Dysfunctions of a temporal processing nature—auditory, visual, or both—have been proposed as a basis for the phonological disorders that underlie poor reading (Tallal, Fitch, and Miller, 1993; Tallal and Piercy, 1973a, 1973b; Livingstone *et al.*, 1991). However, these temporal processing accounts remain controversial (Studdert-Kennedy and Mody, 1995, 1997; Brady, Shankweiler, and Mann, 1983; Breier *et al.*, 2002) as there is alternative evidence suggesting that the phonological difficulties associated with developmental reading disabilities like dyslexia are specific to the domain of speech perception and not generally related to the temporal processing of sensory information.

In Tallal *et al.*'s (1993) account, the phonological difficulties of poor readers are regarded as a symptom of an underlying deficit in the rapid temporal processing of acoustic events. Furthermore, Tallal *et al.* surmised that differences between good and poor readers, in terms of their perception of speech and nonspeech stimuli, may be related to (1) the duration of formant transitions, and (2) the rate at which nonspeech acoustic stimuli are presented (Tallal and Piercy, 1973a, 1973b). To substantiate the former case, Schwartz and Tallal (1980) point to the results of a dichotic listening study in which extending the formant transition durations (from 40 to 80 ms) of synthetic consonant–vowel (CV) syllables seemingly reduces the degree to which speech/language specific areas of the brain are engaged by speech stimuli that normally possess brief formant transitions. Additionally, a number of studies by Tallal and Piercy (1974, 1975) have consistently shown that developmental aphasic children are relatively worse than controls in discriminating synthetic CV syllables with short duration formant transitions and in temporal order judgments for pairs of stimuli presented at short (8–305 ms) interstimulus intervals (ISI), but show improvement when the transitions and ISIs are extended. One may infer from these findings that (1) the course of normal speech/language and reading development is inexorably bound to the rapid temporal processing of auditory stimuli, and (2) an inability to recognize the short-duration sounds of speech (on the order of milliseconds) may lie at the root of some poor readers' phonological difficulties. While the aforementioned claims are based on tests wholly within the auditory domain, there are a number of temporal processing accounts which extend these findings to other sensory modalities.

A pervasive system for temporal processing in the auditory domain also speaks to findings (Livingstone *et al.*, 1991) which demonstrate various deficits among dyslexics for pro-

 0001-4966/2005/118(2)/1122/12/$22.50

cessing rapidly presented visual information—a function that is normally mediated by the magnocellular pathways of the lateral geniculate nucleus (LGN). A growing body of literature (Talcott *et al.*, 1998; Demb, Boynton, and Heeger, 1998a; Demb *et al.*, 1998b; Klein and McMullen, 1999), for example, purports that poor readers have impaired contrast sensitivity to transient low-luminance visual stimuli (e.g., flickering low-contrast gratings, moving sinusoidal gratings). Unlike the visual system, the auditory system possesses no clearly defined magnocellular pathways. However, Stein and Talcott (1999) have argued that the auditory system does incorporate an analogous set of large auditory neurons specialized for registering changes in the frequency of sounds. Therefore, a temporal processing deficit which (1) pervades across visual and auditory sensory modalities, and (2) manifests in a decreased sensitivity to dynamic visual stimuli as well as diminished temporal acuity for rapidly changing acoustic stimuli, could account for dyslexics' visual problems when reading printed text and their phonological deficits in a number of speech-related tasks.

An alternative to these temporal processing positions suggests that the problems of poor readers reflect a deficient speech system in particular as opposed to a pervasive system for temporal processing. Brady *et al.* (1983), and recently, Breier *et al.* (2002), provide evidence for this alternative approach in their studies of reading ability in relation to the effect of acoustic noise masking on the perception of speech and environmental sounds. Their results indicated that poor readers are less able to identify spoken words under "noisy" listening conditions, but perform just as well as controls when the masked stimuli are environmental sounds. Thus, it appears that some poor readers may possess inferior phonological perception skills or phonological categories that are more prone to disruption during perception in noise. Brady *et al.* and Tallal *et al.* (1993) agree that poor readers possess a degraded ability to identify speech sounds (relative to controls) that is not simply a hearing problem, *per se*. Indeed, poor readers from both Brady *et al.*'s and Tallal *et al.*'s studies displayed normal audiometry scores in the face of marked deficiencies in identifying speech sounds. However, since the poor readers in Brady *et al.*'s study were just as capable as normal readers at identifying stimuli that were environmental sounds, it appears then that their problems may be limited to the perception of speech. As Brady *et al.*, and subsequently, Studdert-Kennedy and Mody (1995) surmised, the aforementioned perceptual deficits in noise are specific to speech perception and not necessarily reflective of a general deficit in poor readers' rate of auditory processing.

The research of Brady *et al.* and Breier *et al.* suggests that the phonological difficulties associated with disorders of reading like dyslexia are specific to the "speech mode" of perception, implicating a potential dysfunction of the phonetic module discussed by Liberman and his colleagues (Liberman and Mattingly, 1985; Whalen and Liberman, 1996). As these authors note, one special proclivity of the phonetic module is its ability to recruit and integrate both auditory and visual information. For example, past accounts of experiments which present subjects with audio-visual speech signals have shown that, in constructing a speech percept, listeners are able to combine seen and heard information about the perceived articulation of speech events (as illustrated in Fowler, Brown and Mann, 2000, for example). This integration of auditory and visual articulatory (lip movement) cues serves to help us distinguish between general auditory accounts and speech-specific accounts of poor reading ability and its association with problems identifying speech elements in noise. For now, we wish to make the point that Brady *et al.*'s view of reading impairments raises the possibility of finding not only intact nonspeech acoustic percepts among poor readers but also some disruptions of nonacoustic speech mechanisms such as those involved in lip-reading or auditory-visual integration. This possibility is consistent with de Gelder and Vroomen's (1998) previous work, which shows that poor readers of Dutch are less influenced by visual articulatory cues in the presence of ambiguous auditory speech stimuli than controls. Furthermore, given only visual articulatory cues, poor readers are generally worse than controls at lip-reading.

Such disruptions of general speech processing should stand in contrast to modality specific disruptions that emerge from more peripheral, sensory origins. As a case in point, auditory neuropathy (AN) can be considered. Like poor readers, AN subjects typically exhibit speech perception deficits, especially under noisy listening conditions. However, in sharp contrast to the speech processing deficits associated with developmental reading impairments, the deficits associated with AN manifest from a hearing disorder involving disruptions of the normal synchronous activities of the auditory nerve that are critical to the representation of speech sounds. As Zeng *et al.* (1999) have observed, AN subjects often complain that they can hear sounds, but cannot understand speech without the assistance of visual articulatory cues. This difficulty in understanding acoustic speech is especially apparent in cases where speech stimuli are embedded in noise.

Although AN is a relatively new diagnosis that likely reflects more than a single etiology (and is perhaps more accurately conceived of as one of a collection of "neuropathies" that may or may not be entirely circumscribed to the auditory system), patients of various ages, nevertheless, show a consistent set of symptoms: auditory characteristics consistent with normal outer hair cell function, coupled with abnormal (in some cases, absent) auditory brainstem responses (ABRs). To date, research has been unable to identify, decisively, the actual sites that have been compromised in cases of AN, but converging lines of evidence point to potential demyelination and axonal loss at or near the VIIIth cranial nerve (including, but not limited to, the connections between the inner hair cells and the cochlear branch of the VIIIth cranial nerve, and perhaps ascending auditory pathways of the brainstem; Zeng *et al.*, 1999) affecting both the synchronous firing of nerve fibers and their capacity to achieve high rates of discharge (Starr *et al.*, 2003). This results in disruptions of the precise encoding of temporal cues, leading to a wide variety of impairments in speech and nonspeech tasks (for reviews of speech comprehension and gap detection tasks, see Starr *et al.*, 1996). However this demyelination occurs (e.g., whether through exposure to ototoxic

drugs, inherited degenerative axonal diseases similar to leukodystrophy, or point mutations of genes implicated in congenital hypomyelination disorders), it has been implicated as a catalyst for disruptions in the normal synchronous activity of afferent fibers that encode the temporal characteristics of speech sounds.

We find AN an interesting foil to reading impairments like dyslexia because temporal aspects of auditory perception are known to be severely abnormal in AN subjects. It has been suggested that disruptions of the normal synchronous activity of the auditory nerve may result in a "smearing" of the temporal representations of acoustic stimuli. For example, consistent with Zeng *et al.*'s arguments, a case study by Kraus *et al.* (2000) reported that, in a gap detection experiment, their AN subject possessed poor temporal processing abilities, characterized by a threshold of 100 ms required to detect a gap between tones. In terms of speech processing, Kraus *et al.*'s work exemplifies how "dys-synchronous" activity at the VIIIth cranial nerve adversely affects the representation of critical features that occur at the onset of certain speech stimuli. Specifically, discrimination thresholds for stop versus glide CV syllables revealed that Kraus *et al.*'s AN subject had very good discrimination for speech sounds along a /ba/–/wa/ continuum where the speech sounds differ in terms of the manner of articulation. However, the AN subject also displayed very poor discrimination for sounds on a /da/–/ga/ continuum (with both /da/ and /ga/ sharing similar transition duration and rate of spectral change, but differing in frequency and direction of spectral change in *F*3 at stimulus onset, whereas /ba/ and /wa/ are more discernible via *F*1 and *F*2 durational cues). A theory emphasizing temporal processing deficits would anticipate this result.

Herein lies a potential parallel between the speech perception deficits associated with reading impairments like dyslexia and those that associate with AN. Both populations are reported to have inordinate difficulty perceiving speech in noise. Various accounts of these disorders implicate a deficit of temporal processing, although the focal point of the disruption is clearly different (more peripheral in AN, and, likely, more central in dyslexia). Temporal processing may, indeed, be critical to peripheral auditory processing in cases of AN, but whether or not it is a general multimodal property underlying reading impairments is debatable. A reevaluation of the aforementioned auditory and visual temporal processing accounts yields a number of problems with this assertion.

First, Studdert-Kennedy and Mody (1995, 1997) have found little evidence to support Tallal and Piercy's (1975) findings of impaired perception among poor readers for speech stimuli that possess fast formant transitions. There have been a number of attempts to generalize temporal processing deficits among other, often very distinct, reading impaired populations (e.g., developmental aphasia) to dyslexia (Tallal, 1980, 1984). Consistent with Tallal and Piercy's (1973a, 1973b) work with developmental aphasic children, Farmer and Klein (1995a, 1995b) have reported that dyslexics require an increased ISI compared to controls in simple gap-detection tasks. Although these results are consistent with a view that some poor readers possess an abnormally low limit on the rate at which they can process the rapid acoustic shifts inherent in strings of speech sounds, the link between dyslexics' temporal acuity in nonspeech tasks and general speech perception is tenuous at best. Studdert-Kennedy and Mody have noted a number of difficulties in replicating Tallal *et al.*'s previous finding that extending the duration of formant transitions improves perception of synthetic CV syllables among poor readers.[1] Furthermore, there is no substantial evidence that the inordinate effects of noise on poor readers (Brady *et al.*, 1983; Breier *et al.*, 2002) can be explained by reference to temporal processing.

For the most part, visual processing accounts of dyslexia have met with falsification (for a review, see Mann 2002). It is not clear whether impairments in the processing of low-luminance stimuli or transient (i.e., rapidly flickering) visual events might be concomitant factors as opposed to direct causal factors in poor reading and poor speech perception problems. Much of the evidence tying reading difficulties to magnocellular deficits is still circumstantial. For example, the flickering conditions of low illumination under which dyslexic individuals "see" less well are not characteristic of printed text read under normal conditions (lines of text on a page, for example, do not usually flicker, and the contrast between printed text and background is typically high; Stein and Talcott, 1999). In light of the methodological and conceptual shortcomings of the aforementioned auditory and visual temporal processing accounts, we elect for an alternative approach that contrasts the speech perception deficits inherent to cases of AN with those found in dyslexia. One may aptly compare and contrast the speech perception deficits underlying AN and dyslexia by focusing on speech perception that draws upon both auditory and visual sensory modalities instead of merely tapping into either sensory modality exclusively. In a similar vein to de Gelder and Vroomen's previous study, a reasonable step in speech-based investigations of dyslexia and AN would be to examine how both auditory and visual systems operate independently, and how they interact in tasks that require subjects to draw upon all relevent sensory information in identifying speech elements. There is clear behavioral and neuropsychological evidence that viewing visual articulatory cues can enhance auditory perception, especially under conditions in which it is difficult to identify speech sounds (Grant, Walden, and Seitz, 1993; Sumby and Pollack, 1954). However, it is unclear how efficiently dyslexics can utilize visual articulatory cues.

In normal speech perception, it has been shown that auditory speech recognition is significantly correlated with visual speech reading (Watson *et al.*, 1996), which suggests that, to some degree, speech processing is a modality-independent function. This, in turn, suggests that failures of speech processing among poor readers would pervade across both auditory and visual tasks as de Gelder and Vroomen have demonstrated. If normal speech perception proceeds with at least some implicit reference to the manner in which it is produced by the vocal tract (i.e., a tacit understanding of the mechanics of speech production; Liberman, 1970), and if disorders of reading reflect a central deficit of speech processing, one might expect the use of visual articulatory cues to be problematic for dyslexics in tasks that require an identification of either place or manner of articulation. Con-

versely, if a speech perception deficit reflects difficulties with auditory processing exclusively, then the provision of visual articulatory cues would likely yield an improvement in subject performance relative to baselines for acoustic speech in noise. Furthermore, if a temporal processing limitation truly underlies deficits of speech perception, then dyslexics should also have greater difficulties with the more rapid formant transitions inherent to stop CV syllables than with nasal, liquid, and glide CV syllables (in line with Tallal and Piercy, 1973a, 1973b).

In this study, we have compared the extent to which adult dyslexic and AN subjects are able to use visual cues to help identify noise-masked CV speech stimuli. We expected that our AN subjects would encounter problems with auditory signals—particularly under noisy conditions—but show normal peripheral processing of visual articulatory cues and normal central processing as needed for cross-modal integration and speech perception. We thus expected that the presence of visual articulatory cues would improve AN and control subjects' performance with masked acoustic speech stimuli. In sharp contrast, depending on the extent to which developmental reading impairments reflect a dysfunction of speech processing as opposed to a disorder of auditory temporal processing, we hypothesized that dyslexics would not similarly benefit from the introduction of visual articulatory cues. We shall have more to say about whether impairments in visual temporal processing could play a role in dyslexics' limited use of visual articulatory cues later in Sec. IV.

II. METHOD

A. Participants

In seeking to compare dyslexic and AN subjects, we were mindful of the fact that most studies of neuropathy have involved adults, whereas most of the dyslexia studies cited above have involved children. Admittedly, had we been able to use children as subjects we might have minimized developmental and educational differences that would no doubt be present in an adult sample of dyslexics, AN subjects, and controls. However, the AN subjects who were available for study as part of a routine visit to UC Irvine for other research purposes were all adults; hence, we employed the tactic of studying adult dyslexics and controls. Scarborough (1984) has demonstrated that adults who were previously diagnosed as reading impaired during childhood remained poor readers in adulthood, sometimes performing more than two standard deviations (S.D.) below levels predicted by their IQ. It was, thus, our intention to study adult dyslexics who were formally diagnosed as "dyslexic" or reading impaired during childhood.

Fifteen control subjects (five males, ten females) who possessed neither reading nor hearing deficits were used as a control group and were recruited from a pool of UC Irvine undergraduate students. They were compensated with course credit for their particpation. Control subjects were screened for hearing ability via audiometric exams using a GSI audiometer (testing for hearing losses of 20 dB or more from 250 to 8000 Hz in both ears). They were also tested for level of reading ability using the Woodcock Word ID test (consisting of 106 printed words, each of which subjects must identify verbally within 10 s) and Word Attack test (consisting of 45 printed pseudowords, each of which subjects must also identify verbally within 10 s). The Word ID test is used to determine subjects' word reading mastery. The Word Attack test, however, is used to determine subjects' ability to decode novel pseudowords that are normally not encountered in everyday reading but still follow lawful constructions of the speech sounds in the English language (Woodcock, 1998). Typically, the normal reading range expected for controls is a raw score of at least 85 (typical of a high school senior) on the Woodcock Word ID test. Normal readers were also required to have a raw score of at least 37 on the Word Attack test to be assigned to the control group.

Ten self-reported reading impaired adults (three males, seven females) who were previously diagnosed during childhood as dyslexic were recruited from local community colleges and outreach groups throughout Central and Southern California and served as paid volunteers. Typically, those diagnosed with dyslexia during their elementary school years (with one of our subjects having been identified as reading impaired, at the latest, during his 7th grade year) read two or more levels lower than predicted by their chronological age with no other signs of neurological impairment and performed on par with their peers on measures of nonverbal intelligence. Audiometry results showed that the dyslexic subjects used in this study had normal hearing thresholds. Ranging in age from 19 to 37 years old (with three subjects undergoing speech and language training to help remediate their reading difficulties at the time of testing), the dyslexic subjects had completed at least a high school education (or a general education equivalent), but scored 1 to 2 S.D. below the sample mean on Word ID and Word Attack tests (with raw scores falling between 60 and 80 on Word ID measures and between 20 and 35 on Word Attack measures).

Four AN subjects were recruited through UC Irvine's Department of Otolaryngology as part of a routine visit for research purposes for which they were compensated. In addition to being given Word ID and Word Attack tests, the four AN subjects (all female) were given a thorough clinical evaluation of their respective hearing impairments. Our subjects were diagnosed with neuropathy relatively late in life, and at the time of testing, all AN subjects had achieved at least a high school education, performed on par with their control counterparts on measures of reading ability, and showed strong proficiency with spoken language (e.g., did not speak with the "deaf speech" characteristics of hearing-impaired individuals who have been clinically deaf since birth/early childhood).

AN subjects were given air-conduction audiometry tests using a GSI G1 Clinical Audiometer and gap detection tests for pairs of pure tones. AN subject 1 (age 18) was identified as hearing impaired in her early teens and was recently fitted with a cochlear implant in her right ear. When tested without the assistance of the implant, AN subject 1 showed a relatively normal audiogram with 200-ms tones but exhibited relatively poor gap-detection thresholds (normally ranging between 5–10 ms at 10–40 dB SPL among controls) of 30, 15, and 10 ms at 10, 30, and 40 db SPL, respectively. AN

subject 2 (age 21; identified at the age of 5 as being hearing impaired, and in 1997, given a more formal diagnosis of neuropathic hearing loss) had bilateral, moderate to severe hearing loss that was especially pronounced at low frequencies. Hearing thresholds for AN subject 2 were similar between the ears (except at 2000 Hz, showing a difference of about 15 dB), but gap-detection thresholds were relatively poor at 62.00 and 29.63 ms in the left and right ear, respectively. Diagnosed with a hearing impairment in her early 30's, and finding that conventional assistive listening devices did not alleviate her symptoms, AN subject 3 (age 36) was referred to UC Irvine for potential outfitting with a cochlear implant. Although audiometry results showed similar hearing thresholds between the left and right ears at 250, 1000, and 4000 Hz for tones at 108 dB SPL, AN subject 3 exhibited poor gap-detection skills (with thresholds of 22.00 and 29.75 ms in the left and right ear, respectively). Last, AN subject 4 (age 55) was referred to UC Irvine for treatment for both auditory and potential peripheral neuropathies. Audiometry results revealed a bilateral AN with mild to severe hearing loss that was more pronounced in the left ear (a 15–45–dB loss between 500 and 4000 Hz). AN subject 4 also showed poor gap detection with a threshold of 30, 25, and 10 ms at 10, 20, and 40 dB SPL, respectively.

All subjects had normal to corrected-to-normal vision and no subjects reported any difficulties with viewing the visual stimuli presented in this study.

B. Apparatus and stimuli

A Compaq Presario laptop equipped with a high-resolution screen (width: 11 in.; height: 8 in.) and standard headphones (sensitivity/SPL: 97 dB/mW; impedance: 32 ohm) were used to present subjects with .avi files of digital audio-video stimuli. A female native speaker of English served as the talker for all stimuli. The talker was viewed (at eye level) in a playback video window (width: 6 in.; height: 5 in.), face forward and from the chin up against a neutral background, lighted above by fluorescent light bulbs. Because the playback window did not encompass the entire screen, a black background was used on the Windows XP Desktop and all other on-screen icons and tool bars were rendered invisible. Audio CV components and noise were sampled at 44.1 kHz and video components were recorded using a standard Intel PC camera at a 4:1:1 sampling rate. The audio CV stimuli were samples of the woman's recorded voice for each of the following ten CV syllables: /ba/, /da/, /ga/, /ka/, /ma/, /na/, /pa/, /ra/, /ta/, and /wa/. Our visual stimuli included digital movie files depicting scenes of the female speaker articulating these same ten CV syllables. Ten audio-visual CV stimuli in .avi format were created by carefully combining the corresponding audio and visual files and matching their onsets (e.g., audio /ba/ was combined with visual /ba/, audio /ma/ with visual /ma/, etc.).

In addition to presenting audio CV syllables in quiet (at a comfortable listening level of approximately 65 dB SPL), noise-masked stimuli were developed by using DIGITAL VIDEO PRODUCER 4.0 software to splice .wav files of samples of speech-spectrum shaped noise with .avi sound files to create the following signal-to-noise ratios (SNR): 7 dB ("low noise"), −2 dB ("moderate noise;" where in a pilot study a 0-dB SNR did not sufficiently mask audio CV stimuli), and −7 dB ("high noise;" +/−0.15 dB).

For presentation, each .avi file was stored numerically and randomly selected to form a playlist. A Voyetra media player was used to present preset playlists of the randomized .avi files for each session: the audio CV stimuli alone (masked and unmasked), the audio-visual stimuli (masked and unmasked), and the visual articulatory cues alone.

C. Design

Two conditions in which an audio CV stimulus was either presented alone or with visual articulatory cues (the "audio only" and "audio+visual cues" conditions, respectively) were factorially combined with four different noise levels ("quiet," "low noise," "moderate noise," and "high noise") and presented to our three groups (controls, AN subjects, and dyslexics). A visual baseline condition in which only the visual articulatory cues were given ("visual only") was also presented to subjects. Subjects' percent-correct identification of the CV syllable presented in each of these conditions served as the dependent measure (more on scoring issues in Sec. II E).

D. Procedure

Subjects were first seated in front of a laptop and fitted with headphones. Subjects were then given five practice trials. They were given verbal and written instructions to watch a computer monitor, and listen for speech sounds that would be heard over headphones. They were then instructed to say what they thought the speaker had said and to input their response in a data entry box in the upper right-hand corner of the screen. The stimuli in the practice sessions were representative of the conditions they would experience in the actual experimental trials. Subjects' responses were open-ended. For each of these practice trials, subjects were provided with feedback regarding the correct response. Before beginning the experimental trials, subjects were asked to repeat the instructions given to them for the practice trials. They were also told in advance that they would not receive feedback on the experimental trials.

Following practice, subjects were presented with CV syllables (i.e., /ba/, /da/, /ga/, /ka/, /ma/, /na/, /pa/, /ra/, /ta/, and /wa/) in two test sessions of 90 randomized trials (i.e., 10 syllables, each presented twice × (4 "audio only" conditions +1 "visual only" condition +4 "audio+visual cues" conditions). On some trials (the "audio only" conditions), the audio CV syllable was either masked by noise at one of the three intensities or was presented in quiet. Sometimes, the subjects would be given an appropriate visual articulatory cue of the female speaker's lip movements accompanied by the audio CV stimulus (the "audio+visual cues" condition). Sometimes, they were presented only with the visual articulatory cues (i.e., the "visual only" baseline condition). Again, for each trial, subjects were asked to indicate what they thought the speaker said by inputting their responses into a data entry box in the upper right-hand corner of the screen.

Additionally, all subjects were required to say aloud their responses for each trial, and the experimenter would record each of these responses in order to circumvent any keyboard input problems or potential discrepancies (however uncommon) between the subjects' verbal and typed responses. In the event of a discrepancy between the typed and verbal responses, the subject's verbal response served as the default answer.

E. Scoring

We were interested in determining whether visual articulatory cues would help subjects determine the place and/or manner of articulation of the initial consonant in each CV stimulus. In scoring place of articulation for a particular CV syllable, for a subject's answer to be considered correct, the subject needed to respond with a consonant that had the same place of articulation, even if the voicing and/or manner characteristics of the subject's response failed to match the presented stimulus. Similarly, in scoring manner of articulation, for a subject's answer to be considered correct the subject needed to respond with a consonant that had the same manner of articulation (e.g., stop, nasal, liquid, or glide) regardless of voicing or place characteristics. Although not very common, if the vocalic portion of each CV syllable—a low, back, nonrounded, vowel sound [a]—was not identified correctly, the subject's response was considered an error. Likewise, if only the vowel sound was identified, the subject's response was considered an error.

III. RESULTS

The data consist of the accuracy of speech perception under varying conditions created by the factorial combination of noise level and modality of presentation. Figures 1(a)–1(c) plot out a summary of mean accuracy in identifying CV syllables across noise levels and conditions, for controls, AN, and dyslexic subjects, respectively. For the following analyses, CV syllables were first scored in terms of their initial consonant's place of articulation, since that is where one would expect to find marked influences of the visual articulatory cues.

A repeated measures ANOVA was conducted to test group effects in the "audio only" versus "audio+visual cues" conditions. Our analyses showed a main effect for condition, where subject performance in the "audio only" condition was generally inferior to that of the "audio+visual cues" condition $[F(1,27)=81.004, p<0.001]$. There was a clear main effect for noise level $[F(3,25)=142.603, p<0.001]$; accuracy generally fell as noise intensity increased. There was also an interaction between condition and noise level $[F(3,27)=70.964, p<0.05]$; noise had less of an effect on perception in the "audio+visual cues" condition. More importantly, there was also a main effect for group $[F(2,27)=25.338, p<0.05]$, and there were significant two-way interactions between condition and group $[F(2,27)=86.792, p<0.05]$, and noise level and group $[F(6,52)=15.234, p<0.05]$. These will be discussed below.

As expected, control subjects performed at or near ceiling when the CV syllables were presented in the absence of noise. They showed steady declines in performance as noise increased with little difference in their performance across both the "audio only" and "audio+visual cues" conditions until noise levels increased to "high noise." On average, the control subjects showed a 15% gain in accuracy at high lev-

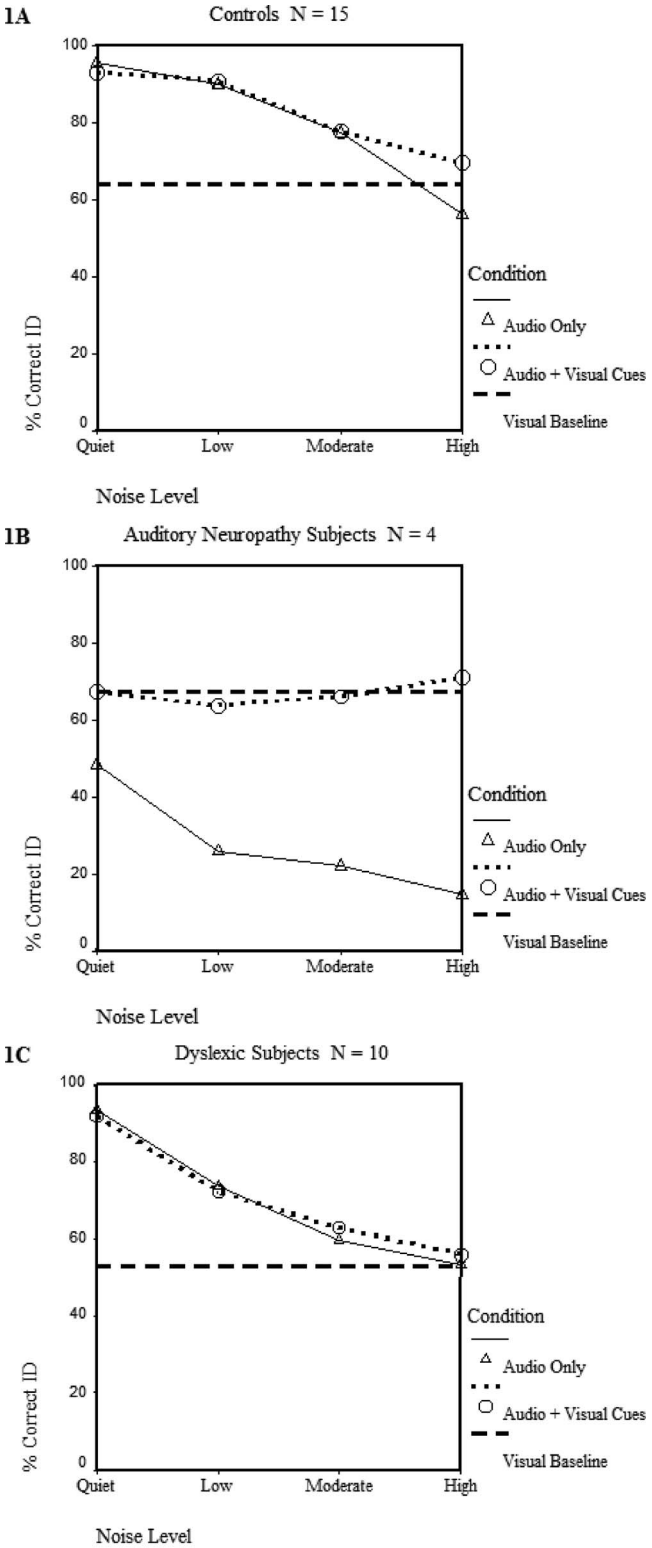


FIG. 1. (A), (B), and (C) Group percent-correct identification of CV syllables (scored by initial consonant place of articulation) at various noise levels in the "audio only," "Audio+Visual Cues," and "visual only" conditions.

els of noise with the introduction of visual articulatory cues, but show no significant difference between conditions at lower noise levels. A comparison of simple effects contrasting the two conditions at high noise levels ($F=18.447, p<0.05$) versus moderate noise levels ($F=0.802, p=0.394$) confirmed this.

For "audio only" perception in quiet, the AN subjects performed the worst of the three groups, largely accounting for the significant main effect for group. Their performance dropped sharply at low levels of noise and showed steadier, less drastic declines at higher levels. They showed a larger drop between the "quiet" and "low noise" levels than our controls when only auditory cues were given. While both AN and dyslexic subjects showed comparable declining slopes in their perception of CV sounds when they were presented in noise, upon the introduction of visual articulatory cues, AN subject performance showed definite improvement in identifying place of articulation over their performance with auditory cues alone. It should be noted, however, that while they showed impressive recoveries of consonant place of articulation (as high as a 50% gain under conditions where the CV sounds are heavily masked) when they were provided with visual articulatory cues, the combination of auditory cues and visual articulatory cues did not boost the AN subjects' performance above the visual baseline except at the "high noise" level. This difference in AN subjects' performance in the "visual only" baseline condition versus the "audio +visual cues" condition at maximum noise levels was not found to be significant $[t(3)=-1.57, p<2.15]$, suggesting that AN subjects relied almost exclusively on the visual articulatory cues.

The pattern of results for the dyslexic subjects departs from both the controls and the AN subjects. Although our dyslexic subjects performed nearly as well as controls in "quiet" (90% versus 95%, respectively), in the absence of visual cues they showed a decrement in accuracy between the "quiet" and "high noise" levels comparable to that exhibited by the AN subjects (a difference of roughly 40% among dyslexics and 30% among AN subjects, although they clearly start at different thresholds to begin with). *Post hoc* analyses comparing group performance at the "visual only" baseline condition show that dyslexics were less accurate than either controls or AN subjects (Fisher's LSD; $p<0.05$ for either comparison) in identifying isolated visual articulatory cues. Furthermore, when noise levels were at their highest, dyslexics showed no significant differences between their performance in the "audio only" and "audio+visual cues" conditions $[t(10)=-1.491, p=0.167]$, unlike the control $[t(14)=-5.29, p<0.001]$ and AN subjects $[t(3)=-11.89, p<0.001]$.

Thus, Figs. 1(a)–1(c) show obvious group differences in how our subjects were affected by acoustic noise and in how effectively they used visual cues to recover consonant place of articulation under conditions where CV sounds were masked by noise. This is consistent with our finding of a significant three-way interaction between condition, noise level, and group $[F(12,46)=10.012, p<0.001]$. It appears that the AN subjects relied heavily on the visual articulatory cues regardless of noise levels, even when given both auditory and visual information. However, regardless of noise level, the dyslexic group performed as if they exclusively relied on the auditory cues. They did not perceive the visual articulatory cues as effectively as our other subjects, nor did they use these visual cues to improve their performance in the "audio+visual cues" condition.

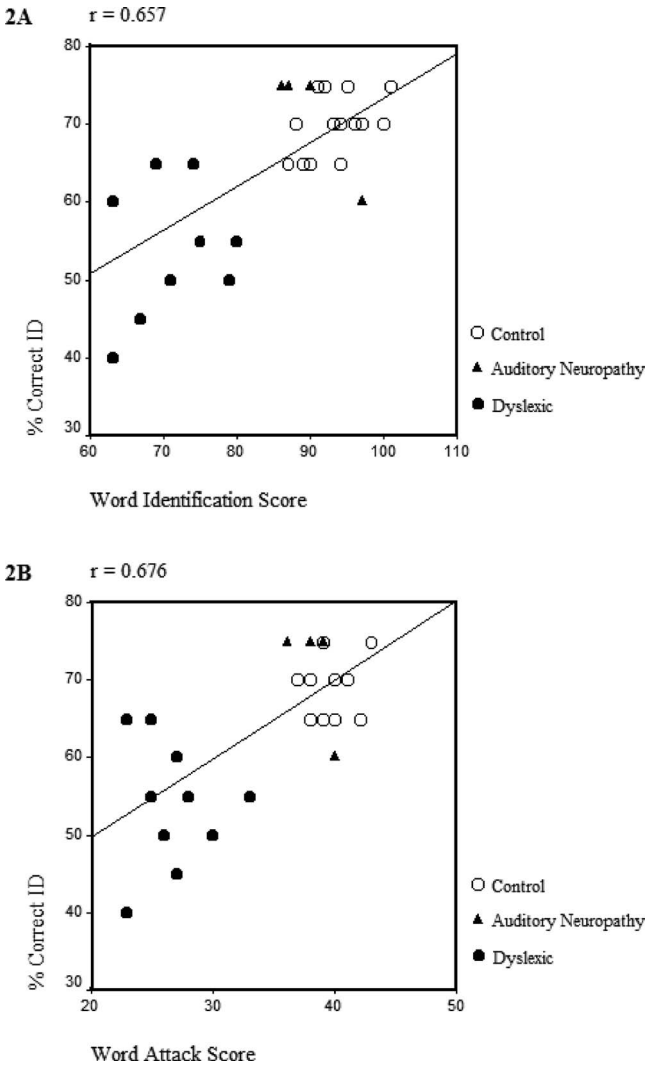


FIG. 2. (A) and (B) Word ID and Word Attack scores plotted against subjects' percent-correct CV identification (scored in terms of initial consonant place of articulation) at high noise levels.

As another illustration of the relationship between reading ability and perception of auditory-visual stimuli, Figs. 2(a) and 2(b) show subjects' Word ID and Word Attack (pseudoword identification) scores, respectively, plotted against subjects' percent-correct CV identification (again, scored in terms of initial consonant place of articulation) in the "audio+visual cues" condition with "high noise" masking. Subjects' percent-correct CV identification scores were positively correlated with both Word ID scores ($r=0.657, p<0.01$) and, similarly, Word Attack scores ($r=0.676, p<0.01$).

In the Introduction, we raised the question of whether perception of stop consonant articulation would be markedly inferior to other consonants (i.e., if stressing perception of rapid acoustic changes is a factor in performance). Figures

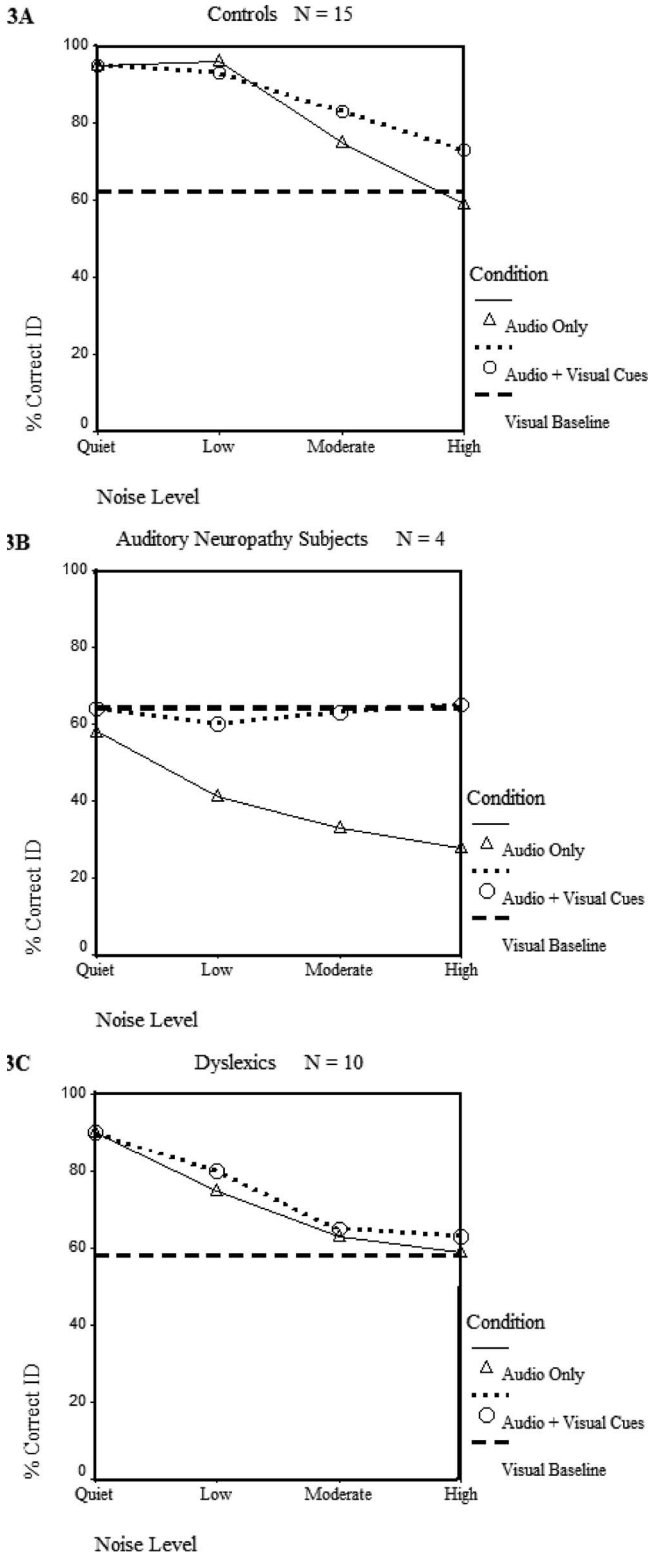


FIG. 3. (A), (B), and (C) Group percent-correct identification of CV syllables (scored by initial consonant manner of articulation) at various noise levels in the "audio only," "audio+visual cues," and "visual only" conditions.

3(a)–3(c) plot the mean accuracy of each group's CV identification. This time, subjects' responses were scored according to each CV syllable's initial consonant manner of articulation. A test of within-subject effects showed a main effect for condition, where again subject performance in the "audio only" condition was inferior to that of the "audio+visual cues" condition [$F(1,27)=85.342, p<0.001$]. Accuracy for identifying manner across noise levels fell with increases in noise intensity, yielding a main effect for noise level [$F(3,25)=139.323, p<0.001$]. Likewise, a test of between-subject effects also showed a significant main effect for group [$F(2,27)=22.837, p<0.05$]. Additionally, significant two-way interactions between condition and noise level [$F(3,27)=71.002, p<0.05$], condition and group [$F(2,27)=82.664, p<0.05$], and noise level and group [$F(6,52)=15.592, p<0.05$] were also apparent in our analysis of manner. Last, our finding of a significant three-way interaction between condition, noise level, and group [$F(12,46)=14.213, p<0.001$] is consistent with the obvious group differences in how effectively the introduction of visual cues helped subjects recover consonant manner of articulation across various noise levels.

A more fine-grained analysis of the data was conducted to discern whether there were different patterns of accuracy for stops, nasals, liquids, and glides.[2] At the 95% CI, under "high noise" masking, subjects from all three groups found the glide CV syllables to be the least difficult to identify. The stop CV syllables proved the most problematic for controls and dyslexic subjects, but stop, nasal, and liquid CV syllables presented the same level of difficulty for AN subjects. When visual articulatory cues were present, the performance of the controls and AN subjects rose to about 70% or higher for all four manner classes. However, the scores of the dyslexic subjects showed little change regardless of the introduction of visual articulatory cues, and tended to average below 70% [see Figs. 4(a)–4(c)]. Control subjects found stop CV syllables difficult in noise but showed recovery when visual articulatory cues were present; AN subjects found stop, nasal, and liquid CV syllables difficult but also showed recovery. Dyslexics found stop CV syllables most difficult and showed little recovery across the four manner classes. It appears that, in the auditory domain, CV syllables that contain longer durational cues (like those in glide CV syllables like /wa/) at stimulus onset are more discernible under masking than those that possess rapid transitions (like those found in stop CV syllables like /ba/ and /da/). However, it is not evident whether the same can be said about the visual signal itself.

IV. DISCUSSION

Based on the available literature (Brady *et al.* 1983, Zeng *et al.*, 1999), we expected to find that both dyslexic and AN subjects would have problems identifying acoustic speech stimuli that are masked by noise. We have replicated this result for both populations, extending prior work with children who are poor readers to our present population of adult dyslexics. The fact that adult dyslexics in our study exhibit speech perception deficits in noise similar to reading disabled children is consistent with Scarborough's (1984) assertion that adult and childhood reading impairments share a continuity. This stands in sharp contrast to developmental lag theories proposing that dyslexic children follow a slow but

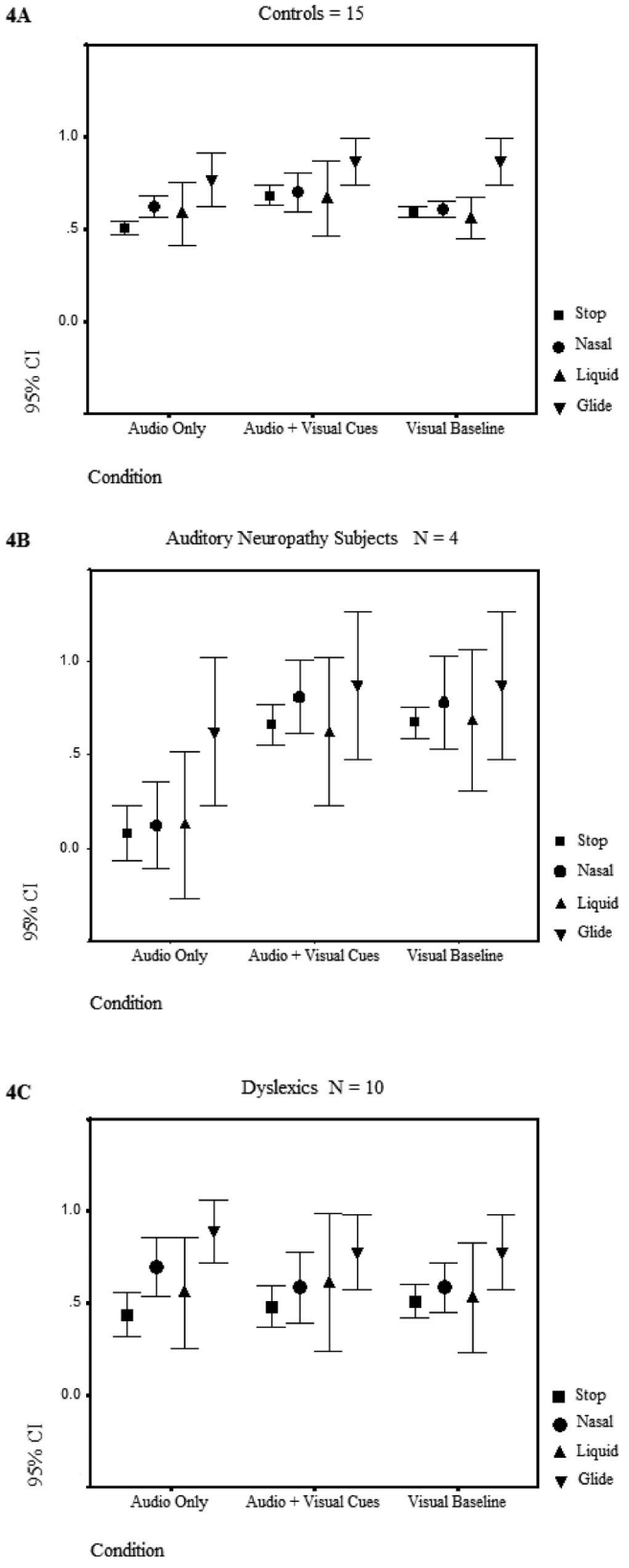


FIG. 4. (A), (B), and (C) 95% CI of group mean performance across manner of articulation at the high noise levels.

steady course of reading skill development and eventually "catch up" with their adult peers (Satz and Van Nostrand, 1973).

This matter aside, our primary question dealt with whether the presence of visual articulatory cues would lead to an improvement in dyslexic and AN subjects' performance with impoverished audio CV stimuli. In the case of AN, there exists a well-accepted explanation of the basis of impaired identification of acoustic stimuli in noise. Converging evidence on the etiology of AN characterizes it as a peripherally based auditory problem, located at or near the VIIIth cranial nerve, that smears the representation of speech. We therefore expected that our AN subjects would show normal peripheral processing of the visual articulatory cues and normal central processing as needed for cross-modal integration and perception of speech. This led us to predict that the presence of visual articulatory cues would improve performance with masked acoustic speech stimuli, and this is as we have observed. This improvement in performance pervades across perception of place but appears to be relatively pronounced for stop consonants.

With regard to our dyslexic subjects, our manipulation of masking and stimulus modality permitted us to test two slightly different explanations of the speech problems that typify poor readers: Tallal *et al.*'s temporal processing account and Brady *et al.*'s speech processing account. We hypothesized that a centralized speech processing impairment could prevent dyslexics from benefiting from the introduction of visual articulatory cues where a temporal impairment in the auditory processing domain could be partially circumvented by reliance on visual speech. Our experimental data confirm that the dyslexics show less effective use of visual articulatory cues in identifying masked CV sounds than either controls or AN subjects and show less accurate perception of visual speech signals in the absence of acoustic speech cues. The data further show that noise adversely affects their perception of both place and manner, though it may be particularly detrimental to stop consonant perception.

The data also suggest that reading ability (as measured through Woodcock Word ID and Word Attack scores) and the degree of auditory-visual speech perception in noise are generally correlated. We do, however, concede that this correlation should be interpreted with some caution, given that our dyslexic subjects' utilization of visual articulatory cues at maximum noise levels was more variable than that demonstrated by either the controls or AN subjects [see Figs. 2(a) and 2(b)]. Indeed, at least a few dyslexic individuals showed a similar perceptual pattern to controls and AN subjects in terms of their performance with visual articulatory cues at maximum noise levels. As we have noted in Sec. II, a number of our dyslexic subjects were undergoing speech and language training at the time of testing, but we can only speculate as to whether or not this may account for some of the individual differences among our dyslexic subjects. A long-term study on how speech and language remediation training can help dyslexics more effectively utilize visual articulatory cues in identifying impoverished speech elements may help address this issue better. That matter aside, our other analyses still effectively show that, under conditions where visual articulatory cues are matched with heavily masked audio CV stimuli, both AN subjects and controls made significantly more use of the visual cues than dyslexics, and they did so to improve perception of both place and manner of articulation. One may interpret these results as

suggesting that some forms of reading impairment are accompanied by central deficits of speech processing.

Our findings of impaired perception of consonants—particularly for stops—in noise are anticipated by Tallal *et al.*'s view that reading impaired individuals have difficulties processing auditory stimuli whose properties change rapidly over a short time course. However, the fact that dyslexic subjects also make less effective use of visual articulatory cues may not be similarly anticipated unless the "temporal processing" deficit is pervasive across modalities. This seems dubious, for, as noted in the Introduction, the visual temporal impairments heretofore reported among dyslexics have involved magnocellular processes and threshold-level stimuli quite different from the video clips we employed. There are three reasons why the visual impairments ensuing from magnocellular layer disorders would not likely be involved in our dyslexics' visual speech performance. First, our visual stimuli were presented at normal luminance values well above those employed by Klein and McMullen (1999). Second, the time course of articulatory events is well in excess of the threshold levels studied by Demb *et al.* (1998a, 1998b). Third, it is not clear why visual speech perception would be impaired across all manner classes while acoustic stop consonants are particularly problematic for dyslexics.

We interpret our behavioral data as demonstrating a potential dissociation between the speech perception deficits of AN subjects and dyslexics. In the former case we assume that a more peripherally based disorder underlies speech processing deficits in noise that could be partially circumvented by the introduction of visual articulatory cues. However, in the latter case, we propose that dyslexics possess a central deficit in speech perception that hinders them from effectively utilizing these visual articulatory cues. Although behavioral data presented by Hayes *et al.* (2003) is partially anticipated by our notion of a central speech deficit, their examination of the underlying neurology associated with learning disabilities converges toward an alternative theoretical explanation of our data emphasizing brainstem mechanisms.

In a similar auditory-visual experiment, Hayes *et al.* have shown that learning disabled (LD) individuals with a history of reading and spelling problems exhibit relatively poor perception of visual speech in quiet. Furthermore, in classic McGurk-type tasks (in which incongruent visual articulatory and auditory cues are matched, typically leading to a fusion of the two percepts; for a review, refer to seminal work by McGurk and MacDonald, 1976), overall, LD subjects report a lower proportion of fusions. Interestingly, these drop-offs in reporting fusions are dramatically pronounced in those LD subjects that exhibited delayed auditory brainstem responses (ABRs). Contrary to what we have found in the present study, under high noise conditions in McGurk-type tasks, Hayes *et al.* found that LD subjects with delayed ABRs reported the visual speech components more often than the auditory components despite initial difficulties with perceiving visual speech in quiet. Assuming that the neocortical representation of auditory speech characteristics (e.g., onset and formant structure; for reviews see Wible, Nicol and Kraus, 2004) is tightly bound to the integrity of the brainstem, one may surmise that the integrity of the brainstem sets the bar on how auditory percepts are effectively recruited and integrated with visual speech elements. Given that LD subjects utilized visual speech in the presence of impoverished auditory cues, Hayes *et al.*'s LD subjects may have adopted compensatory behaviors not unlike those exhibited by the AN subjects in our study to circumvent problems with auditory perception.

These findings, at a first glance, may appear problematic for our attempts to outline a dissociation between the speech deficits inherent in cases of dyslexia and AN, insofar as they suggest a potential overlap between the two disorders at the level of the brainstem. Indeed, AN subjects also exhibit abnormal, sometimes even absent ABRs. However, it is not entirely clear from Hayes' *et al.*'s data that timing deficits at the level of the brainstem play a role in speech perception impairments that associate with disorders of reading. First, an interpretation of both the behavioral and electrophysiological evidence presented by Hayes *et al.* is far from conclusive, as LD subjects who exhibited normal ABRs also reported a reduced proportion of fusions, albeit less dramatically so than those who showed delayed ABRs. Second, the behavioral data alone do not present a wholly consistent picture of the speech perception deficits exhibited by Hayes' *et al.*'s LD subjects. Although one could anticipate that problems in processing basic visual speech would yield difficulties in fusing auditory and visual stimuli in classic McGurk-type tasks, it is not clear why, under heavy noise conditions, LD subjects would *more* heavily report the presence of the visual component if they had marked difficulties in processing visual speech to begin with. The fact that AN subjects show abnormal ABRs and exhibit poor perception of auditory speech would likely not show a larger proportion of fusions relative to controls in McGurk tasks and could circumvent their speech perception deficits via heavy dependence on visual articulatory cues is not at all surprising. That LD subjects with abnormal ABRs would similarly rely on visual speech under impoverished auditory conditions *despite* showing poor perception of visual speech in quiet is not only perplexing, but also not anticipated by de Gelder and Vroomen's previous work, and not supported by the data presented in our study.

Klingberg, Hedehus, and Temple (2000) present a somewhat different perspective that may offer a better reconciliation of the present study and other work. Their strategy applied diffusion tensor imaging (DTI) of the left cerebral hemisphere as a means of studying the integrity of the microstructure of white matter in dyslexic individuals. Klingberg *et al.* argued that dyslexic adults show evidence of decreased myelination within temporo-parietal white matter tracts—the anterior–posterior circuits that potentially link auditory and visual cortices to the primary speech areas of the left cerebral hemisphere. Furthermore, Klingberg *et al.* found that the extent of decrease in myelination correlates with measures of reading ability, whereas our study demonstrated a correlation between reading ability and the degree to which visual articulatory cues are utilized. This appears to be consistent with more recent findings from a functional magnetic resonance imaging (fMRI) study by Klingberg

et al.'s associates (Temple *et al.*, 2003), who have shown that reading-impaired individuals exhibit a decreased blood oxygen level-dependent (BOLD) effect, relative to controls, in the left temporal-parietal cortex and left inferior frontal gyrus (both of which have been previously associated with phonological processing). It would seem, then, that a potential impairment in the strength of communication between cortical areas that subserve speech perception may better explain why the dyslexic subjects in our study demonstrated little effective use of visual articulatory cues compared to their AN and control counterparts.

The findings of the current study add support to the notion that reading impairments co-occur with a central deficit of speech processing that may not necessarily reflect a disruption of a pervasive system specialized for rapid temporal processing. To date, a growing body of literature has suggested that training with audio-visual tools can help advance efforts to remediate learning disabilities that are accompanied by problems with reading and writing. Although our control and AN subjects, not surprisingly, showed effective use of visual speech in the presence of impoverished auditory cues, our dyslexic subjects did not similarly benefit, indicating a need to reexamine efforts to employ visual speech as a medium for improving dyslexics' phonological skills. The present findings also offer some new insights that may help expand the repertoire of diagnostic tools currently available for identifying dyslexic individuals. Extensions of our study that may best complement and extend the current findings should further examine the proportion of fusions that dyslexic adults exhibit relative to normal reading controls in McGurk-type tasks. Additionally, given that much of the research on dyslexia has focused on the reading proficiency and deficits of children, a developmental approach to the study of reading disorders would stand to benefit from an examination of how dyslexic adults compare with younger reading-level matched controls in their ability to integrate auditory and visual information. If reading-impaired adults show a reduced proportion of fusions relative to reading-level matched controls, we may be able to speculate on whether the ability to integrate auditory and visual percepts follows a general developmental course that is somehow arrested early in speech and language training among dyslexics, or if it is the integrity of each modality that truly underlies problems of reading.

ACKNOWLEDGMENTS

The authors wish to thank Dr. Arnold Starr, Dr. Fan-Gang Zeng, and staff members of UC Irvine's Department of Otolaryngology for providing access to the auditory neuropathy subjects involved in this study. This research was generously supported by their grant funding from the National Institutes of Health (Grant DC-02618). Special thanks also go to the Orange County Branch of the Orton Society for providing access to reading-impaired adults and vital contact information for the Southern California area.

[1]Manipulating formant transitions in this manner alters the intelligibility of CV syllables. As Studdert-Kennedy and Mody (1995) have noted, extending the transition of the CV syllable /ba/, for example, alters the phonological contrast such that the initial stop consonant [b] approximates a more "glide-like" sound. At the time that Studdert-Kennedy and Mody offered this critique (and to our current knowledge), no one has been able to replicate Tallal *et al.*'s (1993) findings.

[2]Admittedly, while the results of this analysis of manner class may be of interest to some readers, the authors admit that, as a whole, they are far from conclusive as the experimental design (and indeed, the very nature of the stimuli themselves) does not allow a fair assessment of how our subjects process different manner classes. The English language, itself, will never produce as much variability among glide consonants ([w]) as stop consonants ([b], [d], [g], [p], [t], [k]). Nevertheless, had we limited our study to the six stop consonants in our repertoire of speech stimuli, it would still be interesting to note that dyslexics showed little effective recovery of manner of articulation even when given visual articulatory cues. In future studies, a more balanced design focusing on the use of either stop or fricative consonants would be appropriate.

Brady, S., Shankweiler, D., and Mann, V. A. (**1983**). "Speech perception and memory coding in relation to reading ability," J. Exp. Child Psychol. **35**, 345–367.

Breier, J. I., Gray, L. C., Fletcher, J. M., Foorman, B., and Klaas, P. (**2002**). "Perception of speech and nonspeech stimuli by children with and without reading disability and attention deficit hyperactivity disorder," J. Exp. Child Psychol. **82**, 226–250.

de Gelder, B., and Vroomen, J. (**1998**). "Impaired speech perception in poor readers: Evidence from hearing and speech reading," Brain Lang **64**, 269–281.

Demb, J. B., Boynton, G. M., and Heeger, J. G. (**1998a**). "Functional magnetic resonance imaging of early visual pathways in dyslexia," J. Neurosci. **18**, 6939–6951.

Demb, J. B., Boynton, G. M., Heeger, J. G., and Best, M. (**1998b**). "Psychophysical evidence for a magnocellular pathway deficit in dyslexia," Vision Res. **38**, 1555–1559.

Farmer, M. E., and Klein, R. M. (**1995a**). "The evidence for a temporal processing deficit linked to dyslexia," Psychonomic Bulletin Review **2**, 460–493.

Farmer, M. E., and Klein, R. M. (**1995b**). "Dyslexia and temporal processing deficit: A reply to the commentaries," Psychonomic Bulletin Review **2**, 515–526.

Fowler, C. A., Brown, J. M., and Mann, V. A. (**2000**). "Contrast effects do not underlie effects of preceding liquid consonants on stop identification in humans," J. Exp. Psychol. Hum. Percept. Perform. **26**, 877–888.

Grant, K. W., Walden, B. E., and Seitz, P. F. (**1998**). "Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory visual integration," J. Acoust. Soc. Am. **103**, 2677–2690.

Hayes, E., Tiippana, K., Nicol, T., Sams, M., and Kraus, N. (**2003**). "Integration of heard and seen speech: A factor in learning disabilities in children," Neurosci. Lett. **351**, 46–50.

Klein, R. M., and McMullen, P. A. (**1999**). *Converging Methods for Understanding Reading and Dyslexia* (MIT Press, Cambridge, MA).

Klingberg, T., Hedehus, M., and Temple, E. (**2000**). "Microstructure of temporo-parietal white matter as a basis for reading ability: Evidence from diffusion tensor magnetic resonance imaging," Neuron **25**, 493–500.

Kraus, N., Bradlow, A. L., Cheatham, M. A., Cunningham, J., King, C. D., Koch, D. B., Nicol, T. G., McGee, T. J., Stein, L. K., and Wright, B. A. (**2000**). "Consequences of neural asynchrony: A case of auditory neuropathy," J. Assoc. Res. Otolaryngol. **10**, 1007–1016.

Liberman, A. M. (**1970**). "The grammars of speech and language," Cogn. Psychol. **1**, 301–323.

Liberman, A. M., and Mattingly, I. G. (**1985**). "The motor theory of speech perception revised," Cognition **21**, 1–36.

Livingstone, M., Rosen, G., Drislane, F., and Galaburda, A. (**1991**). "Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia," Proc. Natl. Acad. Sci. U.S.A. **88**, 7943–7947.

Mann, V. A. (**2002**). "Developmental reading disorders," in *Encyclopedia of the Brain*, edited by V. S. Ramachandran (Academic Press, San Diego) **4**, 141–154.

McGurk, H., and MacDonald, J. (**1976**). "Hearing lips and seeing voices," Nature (London) **264**, 746–748.

Satz, P., and Van Nostrand, G. K. (**1973**). "Developmental dyslexia: An evaluation of a theory," in *The Disabled Learner. Early Detection and*

Intervention, edited by P. Satz and J. Ross (Rotterdam University Press, Rotterdam).
Scarborough, H. (**1984**). "Continuity between childhood dyslexia and adult reading," Br. J. Psychol. **75**, 329–348.
Schwartz, J., and Tallal, P. (**1980**). "Rate of acoustic change may underlie hemispheric specialization for speech perception," Science **207**, 1380–1381.
Starr, A., Picton, T., Sininger, Y., Hood, L. J., and Berlin, C. I. (**1996**). "Auditory neuropathy," Brain **119**, 741–753.
Starr, A., Michalewski, H. J., Zeng, F-G., Fujikawa-Brooks, S., Linthicum, F., Kim, C. S., Winnier, D., and Keats, B. (**2003**). "Pathology and physiology of auditory neuropathy with a novel mutation in the MPZ gene (Tyr145→Ser)," Brain **125**, 1604–1619.
Stein, J., and Talcott, J. (**1999**). "Impaired neuronal timing in developmental dyslexia—the magnocelluar hypothesis," Dyslexia: An International Journal of Research Practice. **5**, 59–77.
Studdert-Kennedy, M., and Mody, M. (**1995**). "Auditory temporal perception deficits in the reading impaired: A critical review of the evidence," Psychonomic Bulletin Review **4**, 508–514.
Studdert-Kennedy, M., and Mody, M. (**1997**). "Speech perception deficits in poor readers: Auditory processing or phonological conding?" J. Exp. Child Psychol. **64**, 199–231.
Sumby, W., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**, 212–215.
Talcott, J., Hansen, P., Willis-Owen, C., McKinnell, I., Richardson, A., and Stein, J., (**1998**). "Visual magnocellular impairment in adult developmental dyslexics," Neuro-ophthalmology **20**, 187–201.
Tallal, P. (**1980**). "Auditory temporal perception, phonics and reading disabilities in children," Brain Lang **9**, 182–198.
Tallal, P. (**1984**). "Temporal or phonetic processing deficit in dyslexia? That is the question," Applied Psycholinguistics **5**, 167–169.
Tallal, P., and Piercy, M. (**1973a**). "Defects of non-verbal auditory perception in children with developmental aphasia," Nature **241**, 468–469.
Tallal, P., and Piercy, M. (**1973b**). "Developmental aphasia: Impaired rate of non-verbal processing as a function of sensory modality," Neuropsychologia **11**, 389–398.
Tallal, P., and Piercy, M. (**1974**). "Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception," Neuropsychologia **13**, 83–93.
Tallal, P., and Piercy, M. (**1975**). "Developmental aphasia: The perception of brief vowels andextended stop consonants," Neuropsychologia **13**, 69–74.
Tallal, P., Fitch, R. H., and Miller, S. (**1993**). "Neurobiological basis of speech: A case for the preeminence of temporal processing," Ann. N.Y. Acad. Sci. **682**, 27–47.
Temple, E., Deutsch, G. K., Poldrack, R. A., Miller, S. L., Tallal, P., Merzenich, M. M., and Gabrieli, J. D. E. (**2003**). "Neural deficits in children with dyslexia ameliorated by behavioral remediation: Evidence from functional MRI," Neuroscience **5**, 2860–2865.
Torgesen, J. K. (**1999**). "Reading Disabilities," in *Developmental Perspectives on Children with High Incidence Disabilities*. The LEA series on special education and disability, edited by R. Gallimore and L. P. Bernheimer (Lawrence Erlbaum Associates, Mahwah, NJ), 157–181.
Torgesen, J. K., and Wagner, R. (**1998**). "Alternative diagnostic approaches for specific developmental reading disabilities," Learning Disabilities Research Practice **13**, 220–232.
Vellutino, F. R., Scanlon, D. M., and Lyon, G. R. (**2000**). "Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ achievement discrepancy definition of reading disability," J. Learn Disabil **33**, 223–238.
Watson, C. S., Qiu, W. W., Chaberlain, M. M., and Li, X. (**1996**). "Auditory and visual speech perception: Confirmation of a modality-independent source of individual differences in speech recognition," J. Acoust. Soc. Am. **100**, 1153–1162.
Whalen, D. H., and Liberman, A. M. (**1996**). "Limits on phonetic integration in duplex Perception," Percept. Psychophys. **56**, 857–870.
Wible, B., Nicol, T., and Kraus, N. (**2004**). "Atypical brainstem representation of onset and formant structure of speech sounds in children with language-based learning problems," Biol. Psychol. **67**, 299–317.
Woodcock, R. W. (**1998**). *Woodcock Reading Mastery Tests—Revised, Normative Update* (American Guidance Services, Inc., Circle Pines, MN).
Zeng, F-G., Oba, S., Garde, S., Sininger, Y., and Starr, A. (**1999**). "Temporal and speech processing deficits in auditory neuropathy," NeuroReport **10**, 3429–3435.

Predicting fundamental frequency from mel-frequency cepstral coefficients to enable speech reconstruction

Xu Shao and Ben Milner
School of Computing Sciences, University of East Anglia, Norwich, NR4 7TJ, United Kingdom

(Received 22 September 2004; revised 24 May 2005; accepted 24 May 2005)

This work proposes a method to reconstruct an acoustic speech signal solely from a stream of mel-frequency cepstral coefficients (MFCCs) as may be encountered in a distributed speech recognition (DSR) system. Previous methods for speech reconstruction have required, in addition to the MFCC vectors, fundamental frequency and voicing components. In this work the voicing classification and fundamental frequency are predicted from the MFCC vectors themselves using two maximum *a posteriori* (MAP) methods. The first method enables fundamental frequency prediction by modeling the joint density of MFCCs and fundamental frequency using a single Gaussian mixture model (GMM). The second scheme uses a set of hidden Markov models (HMMs) to link together a set of state-dependent GMMs, which enables a more localized modeling of the joint density of MFCCs and fundamental frequency. Experimental results on speaker-independent male and female speech show that accurate voicing classification and fundamental frequency prediction is attained when compared to hand-corrected reference fundamental frequency measurements. The use of the predicted fundamental frequency and voicing for speech reconstruction is shown to give very similar speech quality to that obtained using the reference fundamental frequency and voicing.
 [DOI: 10.1121/1.1953269]

PACS number(s): 43.72.Ar, 43.70.Ep [DOS] Pages: 1134–1143

I. INTRODUCTION

A significant advance in speech recognition performance from mobile devices has been achieved through the development of distributed speech recognition (DSR) systems.[1,2] These offer increased robustness by replacing the low bit-rate speech codec in the mobile handset with the front-end processing component of the speech recognizer. This removes codec distortion, which is particularly noticeable in the presence of acoustic noise or channel errors, and enables the feature vectors to be transmitted directly into the back-end of the recognizer for decoding. The European Telecommunications Standards Institute (ETSI) Aurora standard for DSR specifies an MFCC-based static feature vector which is augmented with temporal derivatives at the back-end.[2] However, because feature vectors are designed to be a compact representation of the speech signal, typically based on the spectral envelope and optimized for discriminating between different speech sounds, they contain insufficient information to be inverted back into an intelligible representation of the time-domain signal. In particular, valuable excitation information, such as the fundamental frequency, is lost.

The first version of the ETSI Aurora DSR standard was designed to provide an MFCC feature vector stream at a bit rate of 4800 bps.[1] However, it was subsequently considered important to enable a time-domain speech signal to be reconstructed from the received MFCC vectors. This requirement was motivated by legal and security requirements where it may be necessary to listen to the speech being input into the automated system in case of disputes arising from possible speech recognition errors. Several schemes have been proposed to enable speech reconstruction, and these use the MFCC vector to supply vocal-tract information with additional fundamental frequency and voicing components being transmitted to provide the necessary excitation information.[3–6] These schemes require modification to the feature extraction on the terminal device such that fundamental frequency estimation is included, and they also need additional bandwidth to transmit the fundamental frequency and voicing components. Such a system is included in the latest version of the ETSI Aurora standard[2] and is based upon the sinusoidal model of speech.[7] This delivers reasonable quality, intelligible speech at a bit rate of 5600 bps which includes an additional 800 bps for the fundamental frequency and voicing components.

The aim of this work is to predict the voicing and fundamental frequency associated with a frame of speech from its MFCC representation. In a DSR environment this will enable speech to be reconstructed solely from the stream of MFCC vectors, and therefore avoid the need for modification to feature extraction and increased transmission bandwidths. Such a technique will also allow an audio speech signal to be reconstructed from MFCC-parametrized utterances that have no time-domain signal associated with them. Figure 1 illustrates the general operation of the proposed system in the context of a DSR system.

Fundamental frequency prediction is motivated by several studies which have indicated that class-dependent correlation exists between the spectral envelope, or formants, and the fundamental frequency.[8,9] In particular, it was observed that the first formant tended to increase in response to increases in the fundamental frequency. Knowledge of this correlation has been exploited in speech recognition by a phoneme-based normalization of spectral features by the fun-

 0001-4966/2005/118(2)/1134/10/$22.50

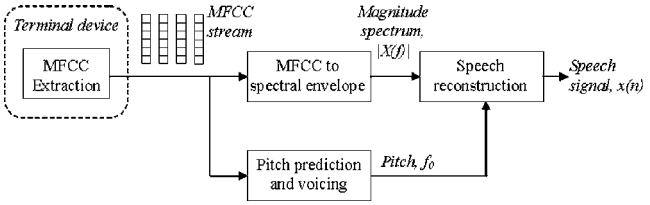


FIG. 1. Distributed speech-recognition-based speech reconstruction from MFCC vectors with fundamental frequency prediction.

damental frequency.[10] This has reduced spectral variations in vowel sounds that are introduced as a result of different fundamental frequencies, and leads to an overall reduction in intervowel confusions. Further work has also reported both improved phoneme and isolated digit accuracy by exploiting this correlation to adapt the observation probabilities within an HMM-based speech recognizer according to the associated fundamental frequency.[11] The correlation has also been utilized in concatenative text-to-speech (TTS) synthesis systems to adjust the spectral envelope of speech units in response to large differences between the measured and target fundamental frequency contours. Adjustments to the spectral envelope have been computed through both codebook mappings[12] and a Gaussian mixture model (GMM).[13] Listening tests indicate that the more realistic speech units, in terms of fundamental frequency and spectral envelope correlations, lead to higher quality synthesized speech. A voice conversion application has also utilized this correlation to determine the most appropriate fundamental frequency for a target frame of speech from the spectral envelope of the converted spectral envelope.[14] These observations of correlation between fundamental frequency and spectral envelope lead in part to the proposed exploitation of correlation between MFCC vectors and fundamental frequency. In addition, it is interesting to consider the mel-spacing of filterbank channels used in MFCC extraction, as these are relatively closely spaced at low frequencies (around 100-Hz spacing up to 1 kHz). This allows the low-frequency regions of the filterbank to preserve part of the harmonic structure of the speech spectrum, although not sufficiently to allow direct fundamental frequency estimation.

The remainder of this work is arranged as follows. Section II gives a brief description into reconstructing a time-domain speech signal from MFCC vectors and their associated fundamental frequency estimates, using a sinusoidal model. Section III introduces two methods of predicting fundamental frequency from MFCC vectors. The first uses a GMM to model the joint density of fundamental frequency and MFCCs, while the second extends this to also model the temporal correlation of fundamental frequency through a set of combined HMM-GMMs. Voicing classification methods are developed in Sec. IV to determine whether an MFCC vector represents voiced (in which case fundamental frequency prediction is employed) or unvoiced speech. Section V evaluates the accuracy of fundamental frequency prediction and voicing classification on speaker-independent male and female speech datasets. Spectrograms of reconstructed speech using the predicted fundamental frequency and MFCC vectors are also shown. Finally, some conclusions are drawn in Sec. VI.

II. SPEECH RECONSTRUCTION FROM MFCC VECTORS AND FUNDAMENTAL FREQUENCY

This section briefly demonstrates how a time-domain speech signal can be reconstructed from a set of MFCC vectors and their associated fundamental frequency. Too much information is lost in the feature extraction process to simply invert the MFCC vectors back into a time-domain signal. For example, the magnitude operation applied to the complex frequency spectrum output of the Fourier transform loses phase information. Similarly, the mel-filterbank analysis and truncation of higher-order coefficients following the discrete cosine transform (DCT) both introduce spectral smoothing (the effect of this is demonstrated in Sec. V). However, it is possible to recover a reasonable, although smoothed, estimate of the magnitude spectrum by first zero padding the K' dimensional MFCC vector, $\mathbf{x}$, to the dimensionality of the filterbank, K, and then applying an inverse DCT followed by an exponential operation to give estimates of the K mel-filterbank channels, X_k,

$$X_k = \exp\left[\frac{2}{K}\sum_{i=0}^{K'} x(i)\cos\left(\frac{\pi i(k-0.5)}{K}\right)\right]. \tag{1}$$

An N-dimensional magnitude spectrum estimate, $|\hat{X}(f)|$, can be obtained from the K mel-spaced filterbank channels (where $N \gg K$) using interpolation techniques. However, the magnitude spectrum estimate is distorted by high-frequency spectral tilt which has been introduced by the pre-emphasis stage of MFCC extraction and also from the increasing mel-filterbank channel bandwidths. The distortion has a multiplicative effect in the frequency domain and can be removed in the cepstral domain through subtraction of the cepstral representations of the pre-emphasis filter, $\mathbf{w}_{\text{pe}}$, and mel-filterbank bandwidths, $\mathbf{w}_{\text{mfb}}$, prior to computation of Eq. (1),[4] as

$$\mathbf{x} = \mathbf{x}' - \mathbf{w}_{\text{pe}} - \mathbf{w}_{\text{mfb}}, \tag{2}$$

where $\mathbf{x}'$ is the original MFCC vector and $\mathbf{x}$ is the equalized vector used in Eq. (1).

From the resulting smoothed magnitude spectrum estimate, $|\hat{X}(f)|$, and a measure of the fundamental frequency, f_0, the sinusoidal model of speech can synthesize a time-domain speech signal, $x(n)$, by the summation of L sinusoids with amplitudes, A_l, frequencies, f_l, and phases, θ_l,[7]

$$x(n) = \sum_{l=1}^{L} A_l \cos(f_l n + \theta_l). \tag{3}$$

In the sinusoidal model the frequencies of the sinusoids correspond to the frequencies of the fundamental frequency and its harmonics. Knowing only the fundamental frequency, the frequencies of the sinusoids, f_l, can be approximated as multiples of the fundamental frequency

$$f_l = lf_0, \quad 1 \leqslant l \leqslant L. \tag{4}$$

The amplitude, A_l, of each sinusoid can be computed from the smoothed magnitude spectrum estimate

$$A_l = |\hat{X}(lf_0)|. \tag{5}$$

The phase offset, θ_l, is calculated as the sum of phase components from the speech excitation, φ_l, and the vocal tract, ϕ_l,[15]

$$\theta_l = \varphi_l + \phi_l. \tag{6}$$

The phase component from the excitation signal at the fundamental frequency is estimated using a linear phase model and aims to keep the phase continuous from one frame to the next. The phase at the harmonics frequencies is calculated through multiplication of the harmonic number by the phase at the fundamental frequency. The phase component from the vocal tract is computed by assuming a minimum phase system. This allows the phase at each harmonic frequency to be computed from the spectral envelope, $|\hat{X}(f)|$, through a Hilbert transform.[15]

Therefore, for each MFCC vector and fundamental frequency estimate, a frame of reconstructed speech can be generated. For unvoiced frames of speech the frequencies of the sinusoids are selected randomly to provide a suitable wideband excitation source. To reconstruct an entire utterance each reconstructed frame of speech is extended by half a frame width either side of its center point, and a triangular windowing function is applied. This allows the overlap-and-add algorithm to combine the individual frames of speech and smooth discontinuities at frame boundaries.

III. FUNDAMENTAL FREQUENCY PREDICTION

This section proposes two methods for predicting the fundamental frequency associated with a frame of speech from its MFCC vector representation. The idea behind both methods is to model the joint density of the MFCCs and fundamental frequency of a frame of speech in order to enable a statistical prediction of the fundamental frequency. Previous studies have indicated that correlation does exist between the spectral envelope and fundamental frequency, although not enough to formulate a generic relation. Instead, this work proposes two methods which make a localized, class-dependent prediction of the fundamental frequency of a frame of speech from its MFCC representation. The first method is based on the unsupervised creation of a Gaussian mixture model (GMM), while the second uses a supervised approach through a combined hidden Markov model (HMM)-GMM approach. The GMM-only system provides a relatively simple method of predicting the fundamental frequency, whereas incorporating the HMM framework increases complexity but allows a more localized prediction of the fundamental frequency.

A. GMM-based fundamental frequency prediction

To model the joint density of the MFCC vector, $\mathbf{x}$, and fundamental frequency, f, an augmented feature vector, $\mathbf{y}$, is defined

$$\mathbf{y} = [\mathbf{x}, f]^T. \tag{7}$$

From a set of training data utterances, the augmented feature vector is extracted with the MFCC component comprising static coefficients 0 to 12 (as in the ETSI Aurora standard[2]). The fundamental frequency is computed using a comb function[16] applied to the original frame of time-domain speech samples, and is subsequently manually corrected where necessary. To signify unvoiced frames or nonspeech, the fundamental frequency value is set to zero. A detailed description of the database and the feature extraction is given in Sec. V A.

From the training set of augmented vectors, unsupervised clustering is implemented using the expectation-maximization (EM) algorithm[17] to produce a GMM which comprises a set of K clusters which localizes the correlation between the fundamental frequency and MFCCs in the joint feature vector space[18]

$$p(\mathbf{y}) = \sum_{k=1}^{K} \alpha_k f(\mathbf{y}; \boldsymbol{\mu}_k^y, \boldsymbol{\Sigma}_k^y). \tag{8}$$

Initialization of the EM algorithm was performed using the Linde–Buzo–Gray (LBG) algorithm followed by k-means clustering.[18] The EM clustering was terminated either when no change occurred in successive iterations or when the number of iterations exceeded 120. The number of clusters, K, was determined experimentally and was based on maximizing the accuracy of fundamental frequency prediction—this is discussed in Sec. V A.

Each of the K clusters is represented by a Gaussian probability density function (PDF) with prior probability, α_k, and mean vector and covariance matrix

$$\boldsymbol{\mu}_k^y = \begin{bmatrix} \boldsymbol{\mu}_k^x \\ \boldsymbol{\mu}_k^f \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_k^y = \begin{bmatrix} \boldsymbol{\Sigma}_k^{xx} & \boldsymbol{\Sigma}_k^{xf} \\ \boldsymbol{\Sigma}_k^{fx} & \boldsymbol{\Sigma}_k^{ff} \end{bmatrix}, \tag{9}$$

$$\sum_{k=1}^{K} \alpha_k = 1. \tag{10}$$

This set of K clusters enables a prediction to be made of the fundamental frequency of the ith frame of speech, $\hat{f}_i$, from the MFCC vector representation of that frame, $\mathbf{x}_i$. Prediction can be made from the closest cluster, in some sense, to the-input MFCC vector or from a weighted contribution from all K clusters.

The closest cluster, k^*, to the input MFCC vector, $\mathbf{x}_i$, is given

$$k^* = \arg\max_k \{p(\mathbf{x}_i | c_k^x)\alpha_k\}, \tag{11}$$

where $p(\mathbf{x}_i | c_k^x)$ is the marginal distribution of the MFCC vector for the kth cluster, c_k^x, with prior probability α_k. Using the joint density of the fundamental frequency and MFCC vector from the kth cluster, a maximum *a posteriori* (MAP)[19] prediction of the fundamental frequency can be made

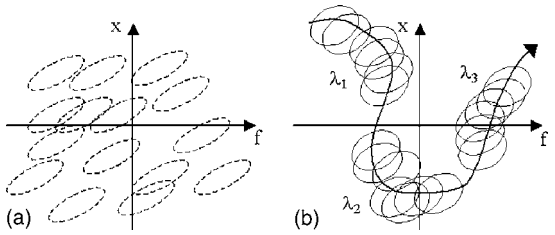


FIG. 2. Modeling of the joint MFCC and fundamental frequency feature space using (a) GMM clustering; (b) A series of GMMs, each located within the state of a set of HMMs.

$$\hat{f}_i = \boldsymbol{\mu}^f_{k^*} + \boldsymbol{\Sigma}^{fx}_{k^*}(\boldsymbol{\Sigma}^{xx}_{k^*})^{-1}(\mathbf{x}_i - \boldsymbol{\mu}^x_{k^*})^T. \tag{12}$$

To avoid making a hard decision in terms of identifying which cluster to predict the fundamental frequency from, an alternative is to combine the MAP fundamental frequency prediction from all K clusters in the GMM according to the posterior probability, $h_k(\mathbf{x}_i)$, of MFCC vector, $\mathbf{x}_i$, belonging to the kth cluster

$$\hat{f}_i = \sum_{k=1}^{K} h_k(\mathbf{x}_i)(\boldsymbol{\mu}^f_k + \boldsymbol{\Sigma}^{fx}_k(\boldsymbol{\Sigma}^{xx}_k)^{-1}(\mathbf{x}_i - \boldsymbol{\mu}^x_k)^T). \tag{13}$$

The posterior probability, $h_k(\mathbf{x}_i)$, of MFCC vector $\mathbf{x}_i$, belonging to the kth cluster is given

$$h_k(\mathbf{x}_i) = \frac{\alpha_k p(\mathbf{x}_i|c^x_k)}{\sum_{k=1}^{K}\alpha_k p(\mathbf{x}_i|c^x_k)}, \tag{14}$$

where α_k and $p(\mathbf{x}_i|c^x_k)$ are as defined for Eq. (11).

B. HMM-based fundamental frequency prediction

The unsupervised training used to create the GMM does not fully exploit the class-based correlation between the MFCC vector and fundamental frequency, nor does it satisfactorily model the temporal correlation within the fundamental frequency contour. No account is made during the EM training of whether feature vectors occur adjacently when deciding upon cluster allocation. Similarly, in testing no account is made of the previous frame's fundamental frequency when determining the fundamental frequency of the current frame. To model the inherent correlation within the feature vector stream, and to therefore select a more appropriate region, or subspace, from which to predict the fundamental frequency, a combined HMM-GMM method is proposed. This method utilizes a set of left–right HMMs which have associated with them a series of state-dependent GMMs which models the local joint density of MFCCs and fundamental frequency. Fundamental frequency is predicted from a stream of MFCC vectors by first computing the model and state sequence through the set of HMMs. This allows a more localized prediction of the fundamental frequency to be made from the model and state-dependent GMMs associated with the states of the HMMs.

To show the differences between the GMM and HMM-GMM prediction methods, Fig. 2 illustrates a conceptual two-dimensional joint MFCC and fundamental frequency feature space. For illustration purposes, the multiple MFCC dimensions are represented as a single dimension along the ordinate, and the fundamental frequency is shown along the abscissa. Figure 2(a) shows the joint MFCC and fundamental frequency feature space which is populated by a set of clusters forming a single GMM. As discussed in Sec. III A, prediction is made from either the closest cluster to the MFCC vector or from a combination of all clusters. Figure 2(b) illustrates the same feature space but now modeled by a set of HMMs each containing state-dependent GMMs. The solid line illustrates the trajectory of an example stream of vectors passing through the states (indicated by the circles) of three example models—λ_1, λ_2, and λ_3. Within each state of these models a GMM provides a localized prediction of the fundamental frequency.

Training begins with the creation of a set of HMM-based speech models, $\boldsymbol{\Lambda}=\{\lambda_1,\lambda_2,\ldots,\lambda_W\}$. This work is currently based upon the ETSI Aurora connected digit database (see Sec. V A), and therefore the model set comprises 11-digit models together with a silence model to give $W=12$ HMMs in total. These models are trained on the MFCC component, $\mathbf{x}$, of the augmented vector, $\mathbf{y}$, using standard Baum–Welch training to produce a set of 16 emitting state, single mode, diagonal covariance matrix digit HMMs—in accordance with the ETSI Aurora guidelines. The set of training data utterances is then realigned to the speech models using Viterbi decoding[20] to provide a model and state allocation for every feature vector used in training. Therefore, for a training data utterance which comprises N MFCC vectors, $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N]$, an associated model allocation, $\mathbf{m}=[m_1,m_2,\ldots,m_N]$, and state allocation, $\mathbf{q}=[q_1,q_2,\ldots,q_N]$, are computed. This indicates the state, q_i, and model, m_i, to which the ith MFCC vector, $\mathbf{x}_i$, is allocated, where $m_i=\{1,\ldots,W\}$ and $q_i=\{1,\ldots,S_{m_i}\}$, with W indicating the number of models and S_{m_i} the number of states in model m_i. Voiced vectors (as indicated by the fundamental frequency component in the augmented feature vector) belonging to each state, s, of each model, w, are then pooled together to form state- and model-dependent subsets of feature vectors, $\Omega_{s,w}$, from the overall set of feature vectors, Z, made up of all training data utterances

$$\Omega_{s,w} = \{\mathbf{y}_i \in Z: f_i \neq 0, q_i = s, m_i = w\}, \quad 1 \leqslant s \leqslant S_w, \quad 1 \leqslant w \leqslant W. \tag{15}$$

Similarly, subsets of unvoiced vectors belonging to each state and model can be created, $\Psi_{s,w}$

$$\Psi_{s,w} = \{\mathbf{y}_i \in Z: f_i = 0, q_i = s, m_i = w\}, \quad 1 \leqslant s \leqslant S_w, \quad 1 \leqslant w \leqslant W. \tag{16}$$

EM clustering is applied to each subset of voiced augmented vectors to create a series of model- and state-dependent GMMs which are represented by mean vectors, $\boldsymbol{\mu}^y_{k,s,w}$, covariance matrices, $\Sigma^y_{k,s,w}$, and prior probabilities, $\alpha_{k,s,w}$, corresponding to the kth cluster of state s of speech model w. Some states have very few voiced vectors associated with them, and this forms the basis of voicing classification which is discussed in Sec. IV. At this stage the joint feature vector space is modeled by a series of GMMs which are linked together by the states of a set of HMMs and

provide localized regions from which fundamental frequency can be predicted.

Prediction of the fundamental frequency, for voiced frames, is made from the MFCC vectors in a speech utterance, $\mathbf{X}=[\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_N]$, by first determining their model allocation, $\mathbf{m}=[m_1,m_2,\ldots,m_N]$, and state allocation, $\mathbf{q}=[q_1,q_2,\ldots,q_N]$, from the set of HMMs using Viterbi decoding. For each MFCC vector, $\mathbf{x}_i$, in the utterance, this provides the model, m_i, and state, q_i, to which it is allocated. This information localizes the region from which fundamental frequency is to be predicted to the particular GMM associated with state q_i of model m_i. The model- and state-dependent MAP prediction of the fundamental frequency, $\hat{f}_i$, associated with MFCC vector $\mathbf{x}_i$ is then computed as

$$\hat{f}_i = \sum_{k=1}^{K} h_{k,q_i,m_i}(\mathbf{x}_i)(\boldsymbol{\mu}^f_{k,q_i,m_i} + \boldsymbol{\Sigma}^{fx}_{k,q_i,m_i}(\boldsymbol{\Sigma}^{xx}_{k,q_i,m_i})^{-1}(\mathbf{x}_i - \boldsymbol{\mu}^x_{k,q_i,m_i})^T), \quad (17)$$

where $h_{k,qi,mi}(\mathbf{x}_i)$ is calculated from Eq. (14), with $p(\mathbf{x}_i|c^x_k)$ being made specific to state q_i of model m_i.

IV. VOICED/UNVOICED CLASSIFICATION

This section introduces two methods to determine whether an MFCC vector represents a voiced frame of speech. Accurate identification of voiced frames is important as this information is used to decide whether a prediction of fundamental frequency is subsequently made. The discussion of these two techniques is based upon the combined HMM-GMM system of Sec. III B and relies on the model and state sequence of the stream of MFCC vectors determined by Viterbi decoding. Voicing classification from the GMM-only system is discussed at the end of this section.

A. Voicing classification using prior voicing probabilities

The first method of voicing classification is based on the computation of a prior voicing probability for each state of each HMM. The prior voicing probability, $v_{s,w}$, of state s of model w is calculated from the proportion of vectors allocated to it during training which are voiced

$$v_{s,w} = \frac{n(\Omega_{s,w})}{n(\Omega_{s,w}) + n(\Psi_{s,w})}, \quad 1 \leqslant s \leqslant S_w, \quad 1 \leqslant w \leqslant W. \quad (18)$$

To illustrate how the prior voicing probability changes across models and states, Fig. 3 shows the prior voicing probability of the 16 emitting states of digit models "*six*" and "*three*." Overlaid on the two figures are the state boundaries for the phonemes which make up the digits.

Considering the digit *six* (which comprises ARPAbet phonemes /s/ /ih/ /k/ /s/), the first few and last few states contain relatively few voiced vectors which correspond to the unvoiced phonemes /s/ and /k/ /s/. The central states of the model are associated with the vowel /ih/ and comprise nearly all voiced vectors which give a high prior voicing probability. The state occupancy for the model *three* (/th/ /r/ /iy/) shows similar behavior. Initial states have low prior voicing probabilities due to relatively few voiced vectors corresponding to the unvoiced /th/ phoneme. The remaining states have high prior voicing probabilities due to the domination of voiced vectors from the voiced phonemes /r/ and /iy/. It is interesting to observe that the first state in both models has a broadly midvalued prior voicing probability. This can be attributed to the state being on the transition from one model to the next, meaning it has relatively unstable voicing characteristics. In fact, for most models the initial state has a similar prior voicing probability and also relatively fewer vectors allocated to it during training compared to the remaining states.

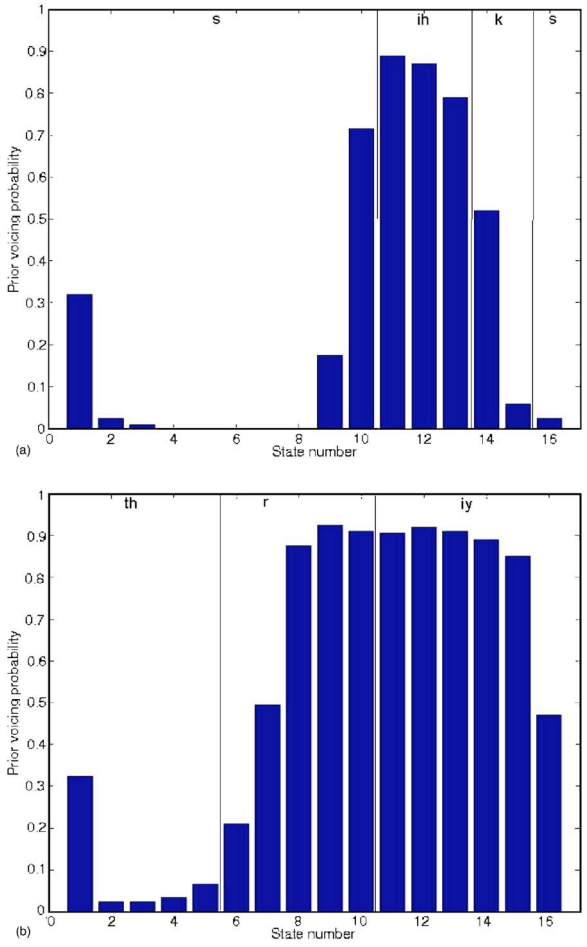


FIG. 3. Examples of prior voicing probabilities for the digits (a) six; (b) three, computed from the proportion of voiced vectors allocated to each state within the respective HMMs.

The voicing associated with an input MFCC vector, $\mathbf{x}_i$, can now be determined from the prior voicing probability of the state, q_i, of model, m_i, to which it is aligned during Viterbi decoding

$$\text{voicing}_i = \begin{cases} \text{voiced} & v_{q_i,m_i} > \theta, \\ \text{unvoiced} & v_{q_i,m_i} \leqslant \theta. \end{cases} \quad (19)$$

The threshold, θ, has been determined through a combination of minimizing the voicing classification error and then maximizing the resulting speech reconstruction quality over a number of utterances, with a suitable value found to be θ

=0.2. The listening tests revealed that it is preferable to have a relatively low threshold value, as this will cause more voicing errors to arise from unvoiced frames being classified as voiced. As the energy of unvoiced frames is usually low, the voicing errors make little perceptible sound. Conversely, if more errors were made when classifying voiced frames as unvoiced, their higher energy would cause more noticeable noise-like errors.

B. Voicing classification using posterior voicing probabilities

An analysis of the prior voicing probabilities for the set of HMMs revealed that some states are strongly voiced (for example, 81 states out of a total of 176 have a prior voicing probability over 0.9) while for other states the distinction is not so clear. This section discusses an alternative to the threshold-based voicing decision by making a voicing decision using posterior probabilities of voicing.

For the HMM-GMM fundamental frequency prediction, described in Sec. III B, a GMM with K clusters has been created in each state, s, of each model, w, from the subset of voiced vectors allocated to that state, $\Omega_{s,w}$. For voicing classification an additional, $K+1$th, cluster is now created from the set of vectors allocated to that state which were labeled as unvoiced, $\Psi_{s,w}$. The probability of an input MFCC vector, $\mathbf{x}_i$, being allocated to any one of the $K+1$ clusters is given

$$p(c_{k,q_i,m_i}|\mathbf{x}_i) = \frac{p(\mathbf{x}_i|c^x_{k,q_i,m_i})P(c_{k,q_i,m_i})}{\sum_{k=1}^{K+1} p(\mathbf{x}_i|c^x_{k,q_i,m_i})P(c_{k,q_i,m_i})}, \quad 1 \leqslant k \leqslant K+1. \tag{20}$$

$p(\mathbf{x}_i|c^x_{k,q_i,m_i})$ is the marginal distribution of the MFCC vector for the kth cluster in state q_i and model m_i, and the prior probability, $P(c_{k,q_i,m_i})$, for each cluster is given

$$P(c_{k,q_i,m_i}) = \begin{cases} \alpha_{k,q_i,m_i} v_{q_i,m_i} & 1 \leqslant k \leqslant K, \\ u_{q_i,m_i} & k = K+1, \end{cases} \tag{21}$$

where $\alpha_{k,qi,mi}$ is as defined for Eq. (11) for state q_i and model m_i, and $u_{s,w}$ is given

$$u_{s,w} = \frac{n(\Psi_{s,w})}{n(\Omega_{s,w}) + n(\Psi_{s,w})} = 1 - v_{s,w}. \tag{22}$$

The decision as to whether the MFCC vector is voiced or unvoiced can be computed

$$\text{voicing}_i = \begin{cases} \text{voiced} & \sum_{k=1}^{K} p(c_{k,q_i,m_i}|\mathbf{x}_i) > p(c_{K+1,q_i,m_i}|\mathbf{x}_i), \\ \text{unvoiced} & \sum_{k=1}^{K} p(c_{k,q_i,m_i}|\mathbf{x}_i) \leqslant p(c_{K+1,q_i,m_i}|\mathbf{x}_i). \end{cases} \tag{23}$$

This method of voicing classification can also be applied to the GMM-only scheme described in Sec. III A. In this case classification is no longer made from the state-specific GMMs but instead uses the single GMM which models the entire feature space.

V. EXPERIMENTAL RESULTS

The aim of these experiments is to first measure the accuracy of fundamental frequency prediction and voicing classification using the techniques discussed in Secs. III and IV. Second, the quality of reconstructed speech using the predicted fundamental frequency is compared to that reconstructed using the reference fundamental frequency within the framework of the sinusoidal model of speech reviewed in Sec. II.

A. Fundamental frequency prediction accuracy

This section measures the accuracy of fundamental frequency prediction and voicing classification using a subset of the ETSI Aurora connected digits database which is sampled at a rate of 8 kHz. A set of 1266 utterances has been used for training, and comprises 633 male utterances and 633 female utterances taken from noise-free speech utterances. Each set is made up of 50 male speakers and 50 female speakers. A separate set of 1000 utterances, spoken by different talkers from those used in training, is used for testing and comprises 501 male utterances and 499 female utterances. Each set uses 50 male and 50 female speakers. In accordance with the ETSI Aurora standard, 13-D MFCC vectors are extracted from the speech at a rate of 100 vectors per second using a window width of 25 ms. The reference fundamental frequency associated with each MFCC vector is made from the time-domain signal using a comb function[16] and subsequent manual correction. The fundamental frequency prediction methods are evaluated on both their classification of MFCC vectors as voiced or unvoiced and also on the percentage fundamental frequency prediction error for voiced frames.[21] Classification error, E_c, is measured as

$$E_c = \frac{N_{V/U} + N_{U/V} + N_{>20\%}}{N_{\text{Total}}} \times 100\ \%\ , \tag{24}$$

where $N_{V|U}$ is the number of unvoiced frames classified as voiced, $N_{U|V}$ is the number of voiced frames classified as unvoiced, and $N_{>20\%}$ is the number of frames in which the fundamental frequency error is greater than 20%. N_{Total} is the total number of frames, which was about 36 000 for the female speakers and 25 000 for the male speakers. For frames correctly classified as voiced, the percentage fundamental frequency error, $E_{\%}$, is computed

$$E_{\%} = \frac{1}{N}\sum_{i=1}^{N} \frac{|\hat{f}_i - f_i|}{f_i} \times 100\ \%\ , \tag{25}$$

where $\hat{f}_i$ is the predicted fundamental frequency from the ith frame and f_i is the reference fundamental frequency for the ith frame.

Tables I and II show the classification error, E_c, and percentage fundamental frequency prediction error, $E_{\%}$, for male speech and female speech using gender-dependent models. In each table results are presented first using the prior voicing probability method of determining voiced frames (Sec. IV A) and then for the posterior probability method (Sec. IV B). Results are shown for the two GMM methods which use either the closest cluster to the input

TABLE I. Classification accuracy and percentage fundamental frequency error for male speech.

	Prior voicing probability		Posterior voicing probability	
	Classification error, E_c	Fundamental frequency error, $E_{\%}$	Classification error, E_c	Fundamental frequency error, $E_{\%}$
GMM—closest	22.5%	9.1%	22.0%	9.2%
GMM—*posteriori*	22.5%	9.0%	22.0%	9.1%
HMM—1 cluster	17.7%	7.8%	13.4%	7.8%
HMM—2 clusters	16.5%	7.3%	12.1%	7.2%
HMM—3 clusters	15.9%	6.8%	11.7%	6.7%
HMM—4 clusters	16.0%	6.8%	11.7%	6.7%
HMM—5 clusters	16.1%	6.8%	11.9%	6.8%

MFCC vector [Eq. (12)], or the *posteriori* weighted MAP prediction [Eq. (13)]. For both GMM methods it was found that using $K=64$ clusters gave the best performance. Results for HMM-based prediction are shown using from one to five clusters within each state with *posteriori*-weighted MAP prediction of the fundamental frequency [Eq. (17)]. Fewer clusters are used with the HMM-based GMMs due to the reduction in training data allocated to each state of each HMM used to train the GMM.

Comparing first the performance differences between using the prior voicing probability and the posterior voicing probability, for determining MFCC voicing classification, shows that the second method consistently outperforms the first. On average, classification errors are reduced by about 3.5% using the posterior voicing probability, while errors made in the fundamental frequency prediction remain unchanged.

Examining the performance of GMM-based prediction for the male speech reveals a slight improvement in fundamental frequency prediction accuracy when taking the *posteriori* weighted prediction from all clusters [Eq. (13)] over using only the closest cluster [Eq. (12)]. For female speech the classification error is significantly better when using the weighted prediction, although fundamental frequency prediction error is slightly worse. Investigation into the reason for this drop in prediction accuracy revealed that the fundamental frequency prediction error was actually worse when using the closest cluster, as some of the errors were greater than 20%, which meant they were labeled as classification errors, E_c [see Eq. (20)], rather than being included in the fundamental frequency error measurement, $E_{\%}$.

The HMM-GMM prediction gives considerably more accurate frame classification and lower fundamental frequency errors in comparison to the GMM, although at higher computational complexity. This can be attributed to the better localization for fundamental frequency prediction that the HMM gives. Increasing the number of clusters in each state of the HMM enables more detailed modeling of the joint distribution of MFCCs and fundamental frequency; this results in a general reduction of frame classification error to a minimum of 11.7% for male speech and 11.2% for female speech. Percentage fundamental frequency prediction error also reduces as the number of clusters increases to a minimum of 6.7% for male speech and 5.4% for female speech. For large numbers of clusters the amount of training data from which to estimate the cluster statistics is reduced, which leads to a slight decrease in accuracy. Using more training data should allow larger numbers of clusters to be reliably created and lead to further reductions in error.

It is interesting to note that the accuracy of the speech recognizer was 97%, which means that 3% of digits were aligned to incorrect models from which voicing and fundamental frequency were predicted. Analysis of predicted values also revealed that the significant majority of frame classification errors arises from incorrect voiced/unvoiced decisions which occur in low-energy regions at the start and end of speech.

Figure 4 compares the predicted fundamental frequency

TABLE II. Classification accuracy and percentage fundamental frequency error for female speech.

	Prior voicing probability		Posterior voicing probability	
	Classification error, E_c	Fundamental frequency error, $E_{\%}$	Classification error, E_c	Fundamental frequency error, $E_{\%}$
GMM—closest	19.7%	5.6%	13.4%	5.7%
GMM—*posteriori*	18.5%	5.9%	12.2%	5.9%
HMM—1 cluster	15.1%	7.0%	11.7%	7.0%
HMM—2 clusters	14.8%	6.1%	11.3%	6.1%
HMM—3 clusters	14.6%	5.6%	11.2%	5.6%
HMM—4 clusters	14.6%	5.5%	11.2%	5.5%
HMM—5 clusters	14.7%	5.4%	11.5%	5.4%

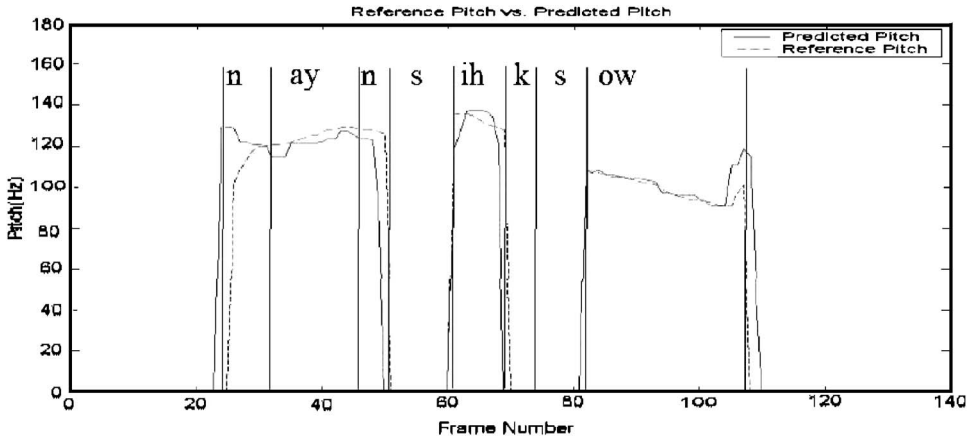


FIG. 4. Comparison of the predicted fundamental frequency contour (solid line) and reference fundamental frequency contour (dashed line) for the utterance "*nine six oh*." A value of zero indicates unvoiced speech or nonspeech.

contour made from the five-cluster HMM-GMM (solid line) with the reference fundamental frequency (dashed line) for the digit sequence "*nine-six-oh*" (comprising ARPAbet phonemes /n/ /ay/ /n/ /s/ /ih/ / k/ / s/ / ow/).

Comparing the two fundamental frequency contours shows that the classification of frames as voiced is effective and follows closely the voicing associated with the reference fundamental frequency. For example, accurate voicing classification can be observed for the digit *six*, where the central /ih/ phoneme is correctly classified as voiced in contrast to the unvoiced phonemes /s/ and /k/, /s/ at the start and end of the digit. For voiced frames the predicted fundamental frequency tracks closely the reference fundamental frequency, although some fluctuations can be observed. For example, the predicted fundamental frequency overshoots at the beginning of the digit *nine* as a result of the voicing classification error, and then stabilizes further into the digit.

Most voicing classification errors occur at voiced/nonspeech boundaries and voiced/unvoiced boundaries which correspond to relatively low-energy regions of the signal. This is illustrated at the start of the digit *nine* (frame 22), where frames are incorrectly labeled as voiced and at the end of the digit (frame 50), where voiced frames are labeled as unvoiced. Voicing errors are also observed at the boundary of phonemes /ih/ and /k/ in the digit *six* (frame 70), where voiced frames are labeled as unvoiced. As will be discussed in the next section, the effect of these errors in reconstructed speech is generally not severe due to the lower energy associated with these regions making the voicing errors less audible.

B. Speech reconstruction results

The motivation behind fundamental frequency prediction in this work is to enable an acoustic speech signal to be reconstructed from a stream of MFCC vectors with no additional fundamental frequency information necessary. To illustrate the effectiveness of this approach, Fig. 5(a) shows the narrow-band spectrogram of the original speech utterance *nine-six-oh*—as used in Fig. 4. Figure 5(b) shows the spectrogram of the speech signal reconstructed from MFCC vectors and the reference fundamental frequency using the sinusoidal model described in Sec. II. Figure 5(c) shows the spectrogram of the speech signal reconstructed solely from MFCC vectors with the fundamental frequency and voicing predicted using the five-cluster HMMs.

Comparing Figs. 5(a) and 5(b) shows that formant peaks become broader as a result of the spectral smoothing which the mel-filterbank analysis and truncation of DCT coefficients impart on the magnitude spectrum in the MFCC extraction process. Only slight differences are observed between Figs. 5(b) and 5(c), which arise from voicing classification errors and fundamental frequency prediction errors. It is interesting to note that the voiced/unvoiced classification errors observed in Fig. 4 have little effect on the reconstructed speech as they are associated with very low-energy regions of the speech. This can be observed at the start of phoneme /n/ at the beginning of the digit *nine*, where the low energy makes the extra frames labeled as voiced almost unnoticeable. A similar effect occurs at the end of phoneme /ow/ as the energy of the digit *oh* falls away. Differences between the original speech and reconstructed speech can be observed for the two occurrences of the phoneme /s/. These are both correctly identified as being unvoiced in Fig. 5(c), but in both reconstructed speech signals [Figs. 5(b) and 5(c)] they appear to have less energy than in the original speech. This can be attributed to the sinusoidal model reconstructing these unvoiced sounds using a random distribution of sinusoids frequencies which does not reproduce the resonances which can be observed in the original speech at 1800 and 2800 Hz.

Listening tests were performed to compare the intelligibility and quality of speech reconstructed using the reference fundamental frequency and voicing with that predicted from the MFCC vectors. In general, little difference could be heard between speech reconstructed from the reference fundamental frequency and that predicted. Differences were more likely to be heard as short-duration bursts in the speech signal arising from either voicing errors or abrupt fundamental frequency errors. Unvoiced frames which were incorrectly classified as voiced led to almost inaudible errors in the reconstructed speech due to their relatively low energy. However, voiced frames incorrectly labeled as unvoiced

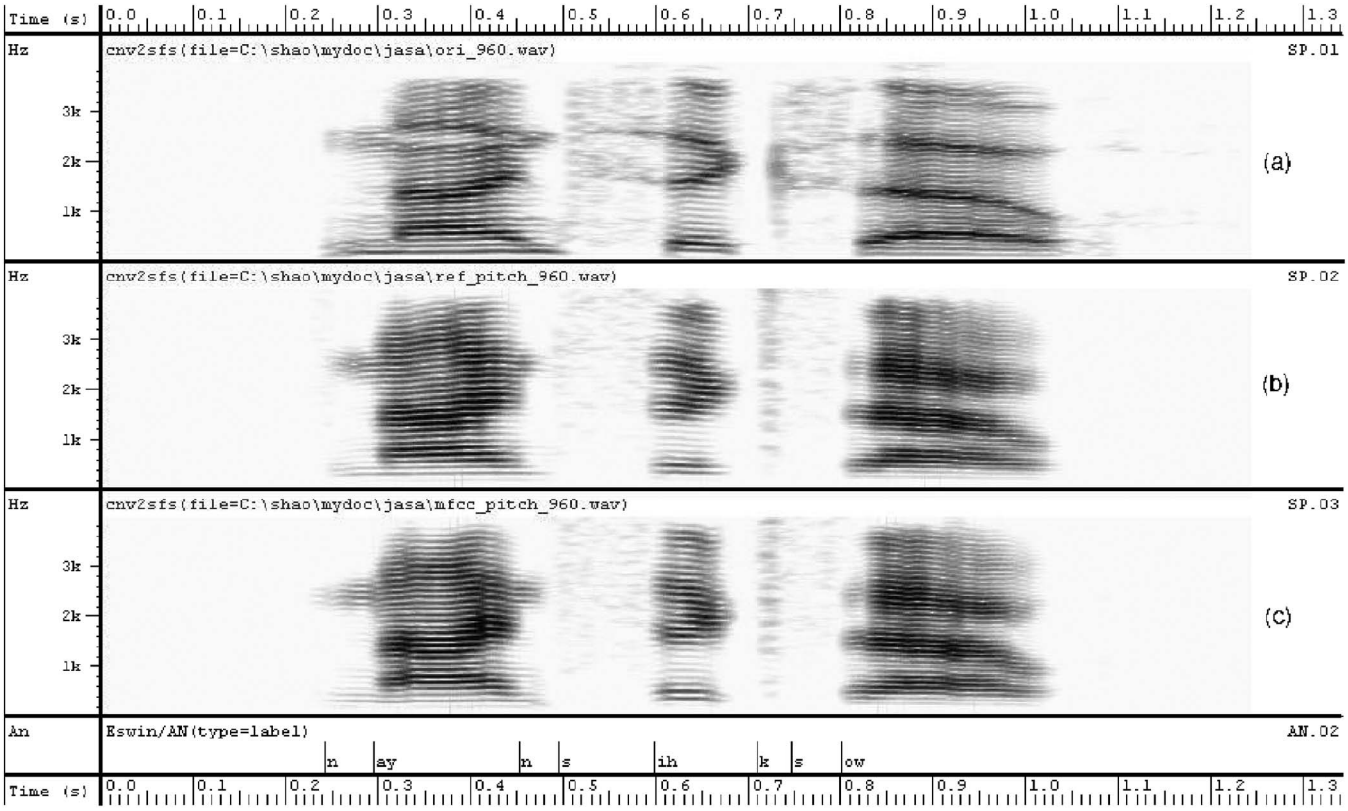


FIG. 5. Comparison of narrow-band spectrograms of the utterance "*nine six oh*" for (a) original speech signal; (b) reconstructed speech using the reference fundamental frequency; (c) reconstructed speech using the predicted fundamental frequency.

could be heard in the reconstructed speech as short-duration noise-like sounds. Errors in the predicted fundamental frequency which lead to abrupt changes in the fundamental frequency contour could be heard as distortions in the reconstructed speech. However, fundamental frequency errors that were small and consistent caused a general upward or downward shift in the contour over a longer duration, did not reduce speech quality or intelligibility, and generally could not be detected. In some cases it was observed that the fundamental frequency contour became flattened in comparison to the reference, and this gave the speech a more monotone sound to it. Overall, it was observed that fundamental frequency prediction errors had generally less effect on the reconstructed speech than voicing classification errors in terms of speech quality and intelligibility.

VI. CONCLUSION

This work has shown that it is possible to predict the voicing and fundamental frequency of a frame of speech from its MFCC representation. Using this information, it has been possible to reconstruct an intelligible speech signal solely from a sequence of MFCC vectors. A fundamental frequency prediction method using a GMM to model the joint density of MFCCs and fundamental frequency was introduced in Sec. III A and gave reasonably accurate voicing classification and fundamental frequency prediction accuracy. This was extended in Sec. III B with a set of combined HMM-GMMs which used the HMMs to localize the region from which fundamental frequency is predicted through a series of state-dependent GMMs. This led to significant improvements in both voicing classification and fundamental frequency prediction accuracy over the GMM-only system. For speech reconstruction the use of the predicted voicing classification and fundamental frequency gave similar speech quality to that obtained when speech was reconstructed using the reference fundamental frequency.

At present the prediction and reconstruction systems have been evaluated on speaker-independent speech, but the vocabulary has been restricted to a connected-digits task. Further investigation needs to be undertaken to extend the task to unconstrained speech input. The application of fundamental frequency prediction to this task is likely to be more difficult but is necessary to enable reconstruction from any free-speech input. Reductions in accuracy are also expected when predicting fundamental frequency from noise-contaminated speech due to the mismatch between the clean models and the input speech.

[1]*ETSI ES 201 108—STQ: DSR*—Front-end feature extraction algorithm; compression algorithm, European Telecommunications Standards Institute (ETSI) (2000).

[2]*ETSI ES 202 212—STQ: DSR*; Extended advanced front-end feature extraction algorithm; Compression algorithms; Back-end speech reconstruction algorithm, European Telecommunications Standards Institute (ETSI) (2003).

[3]D. Chazan, R. Hoory, G. Cohen, and M. Zibulski, "Speech reconstruction from mel-frequency cepstral coefficients and pitch," Proceedings of ICASSP (2000).

[4]T. Ramabadran, J. Meunier, and D. Pearce, "Enhancing distributed speech recognition with back-end speech reconstruction," Proceedings of Eurospeech (2001).

[5]B. P. Milner and X. Shao, "Speech reconstruction from MFCCs using a source-filter model," Proceedings of ICSLP (2002).
[6]X. Shao and B. P. Milner, "Integrated pitch and MFCC extraction for speech recognition and speech reconstruction applications," Proceedings of Eurospeech (2003).
[7]R. J. McAulay and T. F. Quatiery, "Speech analysis/synthesis based on a sinusoidal representation," IEEE Trans. Acoust., Speech, Signal Process. **34**, 744–754 (1986).
[8]T. Hirahara, "On the role of fundamental frequency in vowel perception," The Second Joint Meeting of ASA and ASJ, November, 1988.
[9]A. K. Syrdal and S. A. Steele, "Vowel *F*1 as a function of speaker fundamental frequency," 110th Meeting of JASA, Vol. **78** (1985).
[10]H. Singer and S. Sagayama, "Pitch-dependent phone modeling for HMM based speech recognition," Proceedings of ICASSP (1992).
[11]K. Fujinaga, M. Nakai, H. Shimodaira, and S. Sagayama, "Multiple regression hidden Markov model," Proceedings of ICASSP (2001).
[12]K. Tanaka and M. Abe, "A new fundamental frequency modification algorithm with transformation of spectrum envelope according to F_0," Proceedings of ICASSP (1997).
[13]A. Kain and Y. Stylianou, "Stochastic modeling of spectral adjustment for high quality pitch modification," Proceedings of ICASSP (2000).
[14]T. En-Najjary, O. Rosec, and T. Chonavel, "A new method for pitch prediction from spectral envelope and its application in voice conversion," Proceedings of Eurospeech (2003).
[15]R. McAulay and T. Quatieri, *Speech Coding and Synthesis* (Elsevier, Amsterdam, 1995), Chap. 4.
[16]D. Chazan, M. Zibulski, R. Hoory, and G. Cohen, "Efficient periodicity extraction based on sine-wave representation and its application to pitch determination of speech signals," Proceedings of Eurospeech (2001).
[17]C. M. Bishop, *Neural Networks for Pattern Recognition* (Clarendon, Oxford, 1995).
[18]C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1992).
[19]B. R. Ramakrishnan, "Reconstruction of incomplete spectrograms for robust speech recognition," Ph.D. thesis, Carnegie Mellon University, 2000.
[20]L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proc. IEEE **77**, No. 2, 257–286 (1989).
[21]L. Van Immerseel and J. P. Martens, "Pitch and voiced/unvoiced determination with an auditory model," J. Acoust. Soc. Am. **91**, 3311–3526 (1992).

Analysis and synthesis of the three-dimensional movements of the head, face, and hand of a speaker using cued speech

Guillaume Gibert, Gérard Bailly,[a] Denis Beautemps, and Frédéric Elisei
Institut de la Communication Parlée (ICP), UMR CNRS 5009, INPG/U3, 46, av. Félix Viallet—38031 Grenoble, France

Rémi Brun
Attitude Studio SA, 50, avenue du Président Wilson—93214 St Denis-la-Plaine, France

(Received 19 July 2004; revised 9 May 2005; accepted 9 May 2005)

In this paper we present efforts for characterizing the three dimensional (3-D) movements of the right hand and the face of a French female speaker during the audiovisual production of cued speech. The 3-D trajectories of 50 hand and 63 facial flesh points during the production of 238 utterances were analyzed. These utterances were carefully designed to cover all possible diphones of the French language. Linear and nonlinear statistical models of the articulations and the postures of the hand and the face have been developed using separate and joint corpora. Automatic recognition of hand and face postures at targets was performed to verify *a posteriori* that key hand movements and postures imposed by cued speech had been well realized by the subject. Recognition results were further exploited in order to study the phonetic structure of cued speech, notably the phasing relations between hand gestures and sound production. The hand and face gestural scores are studied in reference with the acoustic segmentation. A first implementation of a concatenative audiovisual text-to-cued speech synthesis system is finally described that employs this unique and extensive data on cued speech in action. [DOI: 10.1121/1.1944587]

PACS number(s): 43.72.Ja [DOS] Pages: 1144–1153

I. INTRODUCTION

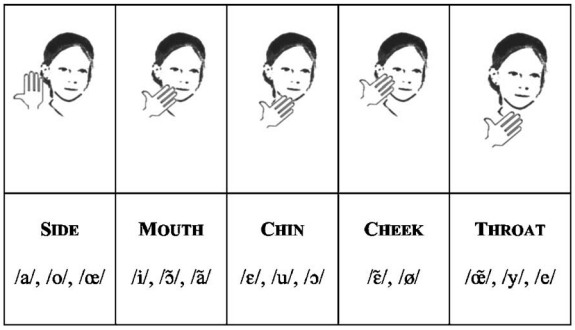
Speech articulation has clear visible consequences. When a person speaks, the movements of the jaw, the lips, and the cheeks are immediately visible. However, the movements of the underlying organs that shape the vocal tract and the sound structure (larynx, velum, and tongue) are not so visible: tongue movements are correlated with visible movements ($R \sim 0.7$) (Kuratate *et al.*, 1999; Yehia *et al.*, 1998; Jiang *et al.*, 2000), but this correlation is insufficient for recovering essential phonetic cues such as place of articulation (Bailly and Badin, 2002; Engwall and Beskow, 2003).

People with hearing impairment typically rely heavily on speech reading based on visual information of the lips and face. However, speech reading alone is not sufficient due to the lack of information on the place of tongue articulation and the mode of articulation (nasality or voicing) as well as to the ambiguity of the lip shapes of some speech units (visemes as [u] versus [y]). Indeed, even the best speech readers do not identify more than 50 percent of phonemes in nonsense syllables (Owens and Blazek, 1985) or in words or sentences (Bernstein *et al.*, 2000). This performance depends on various factors but remains quite far from hearing subjects. The highly trained deaf subjects in the Uchanski *et al.* experiments (1994) obtained mean scores varying from 21% to 62% with lip reading alone, depending on sentence predictability, whereas scores from 78% to 97% are obtained with the help of Cued Speech.

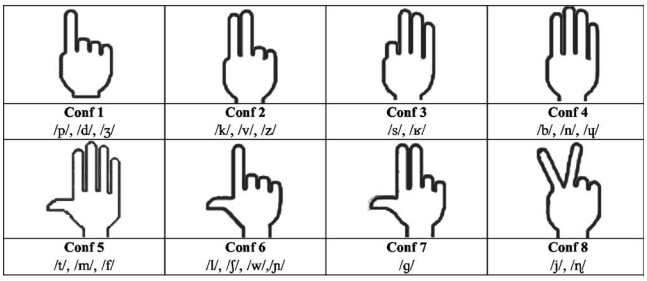
Cued Speech (CS) was designed to complement speech reading. Developed by Cornett *et al.* (1967; 1992) and adapted to more than 50 languages (Cornett, 1988), this system is based on the association of speech articulation with cues formed by the hand. While speaking, the cuer[1] uses one of his/her hand to point out specific positions on the face (indicating a subset of vowels) with a hand shape (indicating a subset of consonants). The French CS (FCS) system is described in Fig. 1. Numerous studies have demonstrated the drastic increase of intelligibility provided by CS compared to speech reading alone (Nicholls and Ling, 1982; Uchanski *et al.*, 1994) and the effective facilitation of language learning using FCS (Leybaert, 2000; Leybaert 2003).

A large amount of work has been devoted to CS perception, but few works have provided insights in the CS production. Attina and colleagues (2002; 2004, 2003a) studied the hand movements of a FCS cuer and their phasing relations with visible (notably lip area) and audible speech. They used a corpus of nonsense words ($[\mathrm{CaCV_1CV_2CV_1}]$ sequences with $\mathrm{C} \in [\mathrm{m,p,t}]$ and V_1 and $V_2 \in [\mathrm{a,i,u,}\phi\mathrm{,e}]$) and they observed an average advance of 200 ms for the beginning of the hand gesture with respect to the acoustic realization of the CV syllable, and they also observed the hand target position was reached quasisynchronically with the acoustic consonantal onset of the CV syllable. This confirmed the *ad hoc* rules retained by Duchnovski *et al.* (2000) for their system of automatic generation

[a]Contact Gérard Bailly, ICP, INPG, 46 Avenue Félix Viallet, 38031 Grenoble Cédex 01, France. Electronic mail: bailly@icp.inpg.fr; Phone: 33 + 476 57 47 11; fax: 33 + 476 57 47 10.



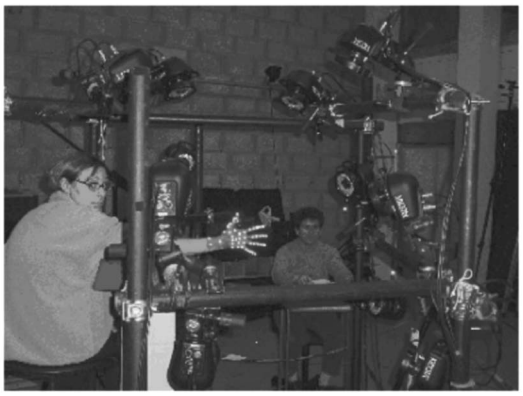
(a) the 5 hand placements. Side is also used for a consonant followed by another consonant or a schwa.



(b) hand shapes. Conf 5 is also used for a vowel not preceded by a consonant.

FIG. 1. French cued speech system.

(a) Position of the retroreflective markers



(b) Experimental setting for capturing the hand motion in free space

FIG. 2. Motion capture experiment.

of CS for English. The authors concluded in a "topsy-turvy vision of Cued Speech" that hypothesizes that the hand placement (i.e., the reached hand target) first gives a set of possibilities for the vowel, the lips then delivering the uniqueness of the solution. Rules for hand movement based on this principle were integrated in a first 2-D audiovisual Cued Speech synthesizer delivering Cued Speech from text (Attina *et al.* 2003b).

In the study we extend this pioneering work to the complete characterization of the 3-D movements of the head, face, and hand. We characterize here the cued speech production of complete utterances. Section II is dedicated to the description of our experimental design for collecting massive motion capture data with high temporal and spatial precision. In Sec. III we describe how motion capture data are regularized using statistical shape models for the face and hand built using selected data. In Sec. IV we further describe the gestural scores we built from motion capture data in order to (a) verify *a posteriori* that the cuer has effectively produced the hand shapes and placements she has to do for complementing speech production (b) study phasing relations of hand placements and hand shapes with the speech signal. In Sec. V we sketch the first version of an audiovisual concatenative text-to-cued speech synthesis system built using the resources of this study.

II. MOTION CAPTURE DATA

We recorded the 3-D positions of 113 retroreflective markers glued on the hands and face of the subject, a skilled cuer who has a daily practice of FCS with relatives, using a Vicon® motion capture system with 12 cameras [see Fig. 2(a)]. The system delivers the 3-D positions of candidate markers at 120 Hz. Recordings and ground truth data processing were performed at Attitude Studio. Further software was provided to assist users in collecting coherent 3-D trajectories and deleting outliers. Note that this tedious semiautomatic task was not error-free. Two different settings of the cameras enabled us to record three corpora.

(i) Corpus 1—*hand only*: the cuer produced all possible transitions between eight hand shapes in free space, with each hand shape corresponding to a subset of consonants [see Fig. 2(b)].
(ii) Corpus 2—*face and audio*: the cuer uttered without cueing, visemes of all isolated French vowels and all consonants in symmetrical context VCV, where V is one of the extreme vowels [a], [i], or [u]. This corpus in similar to the one usually used at ICP for facial cloning (Badin *et al.* 2002)
(iii) Corpus 3—*hand, face, and audio*: the cuer uttered 238 sentences, carefully selected to contain all French diphones. This corpus has more than 200 000 frames in total and was used for FCS recognition and synthesis.

All productions were also videotaped using a camera placed approximately 4 m in front of the cuer. The content of the corpus was delivered sentence by sentence to the cuer by a prerecorded acoustic prompt. This content was available to her several weeks in advance. She was instructed to listen to the acoustic prompt and repeat the utterances aloud as if she was cueing them for a deaf partner standing just behind the video recorder.

For corpus 1, cameras were placed all around the hand in order to gather 3-D positions of the markers even at extreme retracted positions of the fingers. Another camera setting—with a larger working space—was used for the second and third corpora. Corpora 1 and 2 were used to build statistical models of the hand and face movements separately. The models were then used to recover missing data in corpus 3, where the face was partially blocked by the hand, and *vice versa*.

III. ARTICULATORY MODELS OF THE FACE AND HAND

Building statistical models from raw motion capture data has several technological and scientific motivations.

The first technological motivation was to provide a way to clean up the data automatically: the semiautomatic labeling of 3-D trajectories is usually very costly and quite time consuming. The second technological motivation was to ease the work of the infographist, who will be responsible for adjusting the movements of a predefined character or avatar to these raw data: inverse kinematics is more effective when dealing with clean target trajectories.

The scientific motivations concern (1) the study of production of FCS and (2) the coordination between acoustics, face, and hand movements during cued speech production.

A. Face

The basic methodology developed at ICP for cloning speech articulation has already been applied to different speech organs such as the face (Revéret *et al.*, 2000) and the tongue (Badin *et al.*, 2002) and to different speakers (Bailly *et al.*, 2003). From raw motion data, we estimated and subtracted iteratively the elementary movements of segments (lips, jaw,…) known to drive facial motion. These elementary movements were estimated using a Principal Component Analysis (PCA) performed on pertinent subsets of flesh points (e.g., points on the jaw line for the estimation of jaw rotation and protrusion, lip points for lip protrusion/retraction gesture).

This basic methodology was previously applied to quasistatic heads. Since the head was free to move in corpora 2 and 3, we need to solve the problem of the repartition of the variance of the positions of the markers placed on the throat between head and face movements. This problem was solved in three steps.

(1) An estimation of the head movement using the hypothesis of a rigid motion of markers placed on the nose and forehead. A principal component analysis of the 6 parameters of the rototranslation extracted for corpus 3 was then performed and the *nmF* (stands for *number of free movements of the head*) first components were retained as control parameters for the head motion.
(2) Facial motion cloning subtracting the inverse rigid motion of the full data. Only *naF* (stands for *number of free articulatory movements of the face*) components were retained as control parameters for the facial articulation.
(3) Throat movements were considered to be equal to head movements weighted by factors (parameters *wmF* below) less than one. A joint optimization of these weights and the directions of the throat deformations was then performed keeping the same values for the *nmF* and *naF* predictors for each frame.

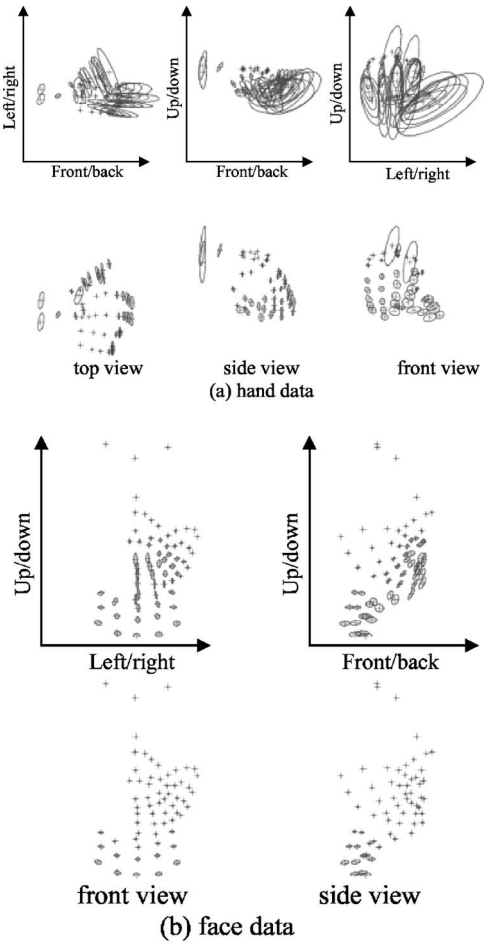


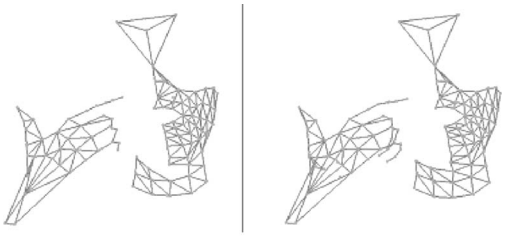
FIG. 3. Dispersion ellipses of the hand and face data (displayed relative to the mean configuration). Top: raw data. Bottom: residual. Both residuals are computed, taking into account both the motion of the segment in space and intrinsic motion.

These operations were performed using facial data from corpus 2 and 3 with all markers visible. A simple vector quantization that guaranteed a minimum 3-D distance between selected training frames (2 mm) was performed before modeling. This pruning step provided statistical models with conditioned data.

The final algorithm for computing the 3-D positions *P3DF* of the 63 face markers of a given frame is:

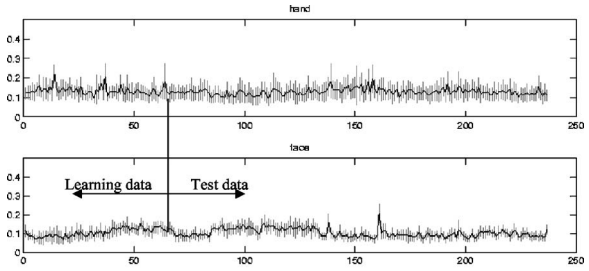
```
mvt = mean_mF+pmF*eigv_mF;

P3D = reshape(mean_F+paF*eigv_F,3,63);

for i=1:63

 M = mvt.*wmF(:,i);

 P3DF(:,i) = Rigid_Motion(P3D(:,i),M);
end
```

where *mvt* is the head movement controlled by the *nmF*



(a) incomplete motion capture data *(b) reconstruction*

FIG. 4. Reconstruction of a FCS frame. Part of the throat and fingers have not been captured by the motion tracking system but have been reconstructed properly by the face and hand models.



(a) learning the models with only the first 68 sentences

FIG. 5. Mean and standard deviation of the mean reconstruction error for each sentence processed by the hand (top) and face (bottom) models. The models were built using the 68 first sentences as learning data. The peak errors for reconstructions are mainly due to false labeling of raw motion data.

parameters *pmF*, *M* is the movement weighted for each marker (equal to 1 for all face markers, less than 1 for markers on the throat) and *P3D* are the 3-D positions of the markers without head movements controlled by *naF* parameters *paF*.

B. Hand

Building a statistical model of the hand articulation was more complex. If we consider the palm as the carrier of the hand (the 50 markers undergo a rigid motion that is computed as the optimal translation and rotation of the 11 markers glued on the back of the hand, a reference configuration of the markers being chosen with all fingers out), the movements of the wrist, the palm and the phalanges of the fingers have quite a complex nonlinear influence on the 3-D positions of the markers. These positions also reflect poorly on the underlying rotations of the joints: skin deformations induced by the muscle and skin tissues produce very large variations of the distances between markers glued on the same phalange.

The model of hand articulation was built in four steps.

(1) An estimation of the hand movement using the hypothesis of a rigid motion of markers placed on the back of the hand in corpus 3. A principal component analysis of the six parameters of this hand motion was then performed and the *nmH* (stands for *number of free movements of the hand*) first components were retained as control parameters for the hand motion.
(2) For all frames from corpus 1 and 3, where the 50 markers were all visible, all possible angles between each hand segment and the back of the hand as well as between successive phalanges were computed (rotation, twisting, spreading,…).
(3) A principal component analysis of these *nH* angles was then performed and the *naH* (stands for *number of free articulatory movements of the hand shape*) first components were retained as control parameters for the hand shaping.
(4) The sin( ) and cos( ) of these *predicted* values were computed and a linear regression between these $2*nH+1$ values (see vector P below) and the 3-D coordinates of the hand markers (see matrix *Xang* below) was performed (subtracting the inverse rigid motion of the full hand data).

Step (4) made the hypothesis that the displacement induced by a pure joint rotation may produce an elliptic movement on the skin surface (together with a scaling factor).

The final algorithm for computing the 3-D positions *P3DH* of the 50 hand markers for a given frame is as follows:

```
mvt = mean_mH+pmH*eigv_mH;

ang = mean_A+paH*eigv_A;

P = [1 cos(ang) sin(ang)];

P3DH = Rigid_Motion
       x(reshape(P*Xang,3,50),mvt);
```

where *mvt* is the movement of the back of the hand controlled by the *nmH* parameters *pmH* and *ang* is the set of angles controlled by the *naH* parameters *paH*.

C. Modeling results

After pruning corpus 1 and 3, the training data for constructing the hand shape model consisted of 8446 frames. After pruning corpora 2 and 3, the training data for facial movements consisted of 4938 frames. The number of elementary angles nH is equal to 23. Figure 3 shows the reduction of variance obtained by keeping $naH=12$ hand shape parameters and $naF=7$ face parameters. Figure 4 shows an example of a raw motion capture frame and the predicted hand and face shapes.

We retained $nmF=5$ and $nmH=5$ parameters for the head and hand movements.

Using the first 68 utterances of corpus 3 as training data (68641 frames) and a joint estimation of hand motion and hand shaping (resp., head motion and facial movements), the resulting average absolute modeling error for the position of the visible markers was 1.2 mm for the hand and 1 mm for the face (see Fig. 5). Regularization of the test data (the next 170 utterances) by the hand and face models do not lead to a substantial increase of the mean reconstruction error.

IV. FURTHER DATA ANALYSIS

Further data analysis was performed in order to verify that the cuer had realized the recommended hand shapes and

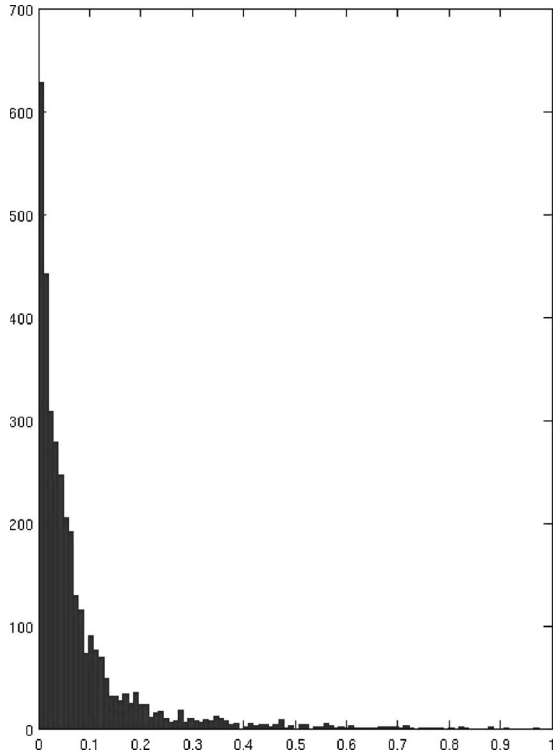


FIG. 6. Histogram of the percentage of distance accomplished by the head of our speech cuer for producing hand/face constrictions.

hand positions with the consonants and vowels effectively. In the following, all available frames were considered. Movements and articulations of hand and face were regularized and reconstructed using the hand and face models described above.

A. The constriction model

Globally the FCS functions as a constriction model: with a certain shape of the final effector (i.e., the hand), a constriction—most of the time a full contact, i.e., an occlusion—is made between the hand and the face. The *place of constriction* is determined by the vowel and the shape of the effector is determined by the consonant. Contrary to vocal tract constrictions where the walls are almost rigid—with the exception of the lips and the velum—cued speech constrictions concern two movable body segments, i.e., the head and the hand. If head movements were known to contribute to the encoding of the linguistic structure of the utterance and signals cognitive activities of the speaker, the head movements here also participated in the realization of hand–face constrictions: during speech, the head and the hand both move toward each other (see Fig. 10, later) and the chin or the throat to the hand according to the required hand placements, i.e., finger/head constrictions. We computed the relative displacements of the head and hand for producing hand/face constrictions by considering the maximum distance between the hand and the face in the interval between successive target hand shapes and placements. Figure 6 shows the histogram of the percentage of this distance accomplished by the head. The mean contribution is 7.7%.

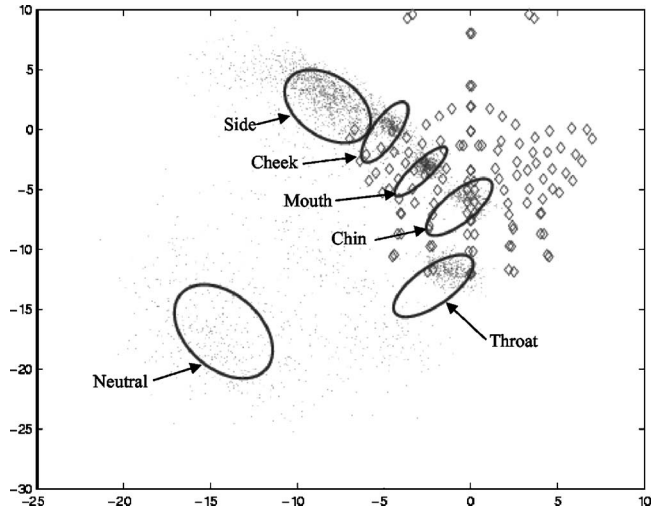


FIG. 7. Data and dispersion ellipsis of the position of finger tip of the longest finger for each targeted hand placements. Please note the main orientation of the four dispersion ellipsis for throat, chin, mouth and cheek. As expected, neutral and side hand placements exhibit the larger dispersion ellipsis (with the main axis perpendicular to the others).

B. Recognizing hand shapes and consonants

According to our previous experience (Attina *et al.*, 2002), the maximal extension/retraction of the fingers—i.e., the hand shape target—was roughly synchronized with the acoustic onset of the consonant (and of the vowel in case of a vowel not preceded by a consonant). We thus selected target frames in the vicinity of this relevant acoustic event and labeled them with the appropriate key value, i.e., a number between 0 and 8 (see Fig. 1): 0 was dedicated to the rest

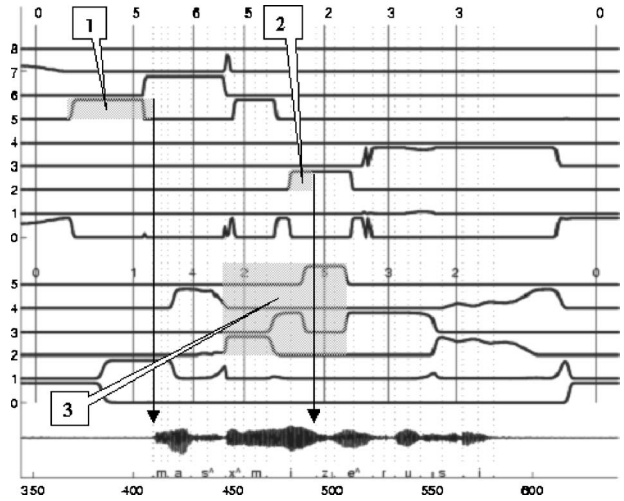


FIG. 8. Recognition of the hand shapes (top) and hand placements (bottom) by simple Gaussian models. The vertical lines show hand targets together with the required hand shapes and placements. Note that intermediate models may be triggered by a movement between hand targets. See, for example, the transition (zone 3 enlightened with translucent gray) between the hand placements 2 and 5: the model for hand placements 3 is naturally triggered (the hand goes nears the chin while moving from the mouth to the throat). Consonants are often cued well in advance of their acoustic onset (see zones 1 and 2): for example, the hand shape 5 for the first consonant [m] is deployed 50/.12=416 ms before its acoustic onset. Similarly, the hand shape 2 that signals the fourth consonant [z] is deployed 15/.12=125 ms before its acoustic onset.

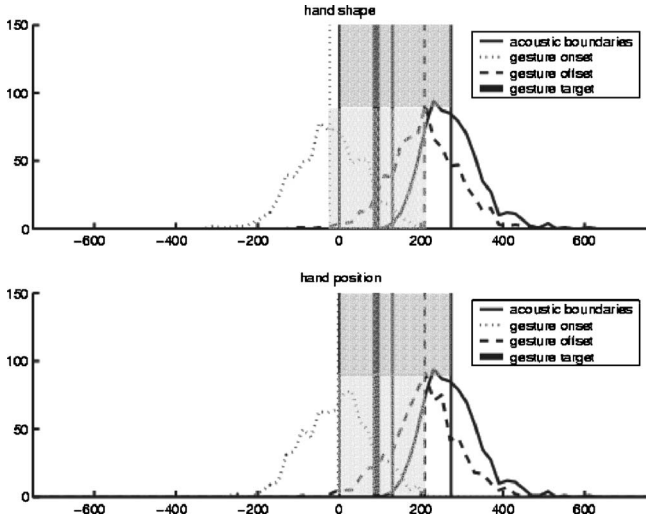


FIG. 9. Phasing gestures with reference to the different acoustic segments they are cueing. Distributions of absolute time difference of different events with reference to the acoustic onset of the segment. Both hand shapes and hand placements start well before the acoustic onset of the speech segment they are supposed to disambiguate.

position chosen by the cuer with a closed knuckle. These target frames were carefully chosen by plotting the values of seven parameters against time.

(i) For each finger, the absolute distance between the flesh point of the first phalange closest to the palm and that closest to the finger tip: a maximal value indicated an extension whereas a minimal value cued a retraction.
(ii) The absolute distance between the tips of the index and middle finger in order to differentiate between hand shapes 2 vs 8.
(iii) The absolute distance between the tip of the thumb and the palm in order to differentiate between hand shapes 1 vs 6 and 2 vs 7.

The 4114 hand shapes were identified and labeled. The seven characteristic parameters associated with these target hand shapes were then collected and simple Gaussian models were estimated for each hand shape. The *a posteriori* probability for each frame belonging to each of the eight hand shape models can then be estimated. We exhibit in Fig. 8 an example of the time course of these probability functions over the first utterance of the corpus together with the acoustic signal. Large anticipatory patterns revealed that the lip shape effectively acts as a *complementary information* to the hand shape (see Sec. IV A).

The recognition rate was quite high: there were only 4 omissions and 50 errors for a recognition rate of 98.78% (see the detailed confusion matrix in Table I). The errors involved mainly confusions between the coding of mid-vowels (/e/ vs /ɛ/ /o/ vs /ɔ/) and omissions of the coding of glides in complex CCCV sequences (such as /ʁ/ in /stʁɔ/).

C. Recognizing hand placements and vowels

Not only did the maximal extension/retraction of the fingers coincide most of the time with the acoustic onset of the consonant, but also the hand placement: CS provided both

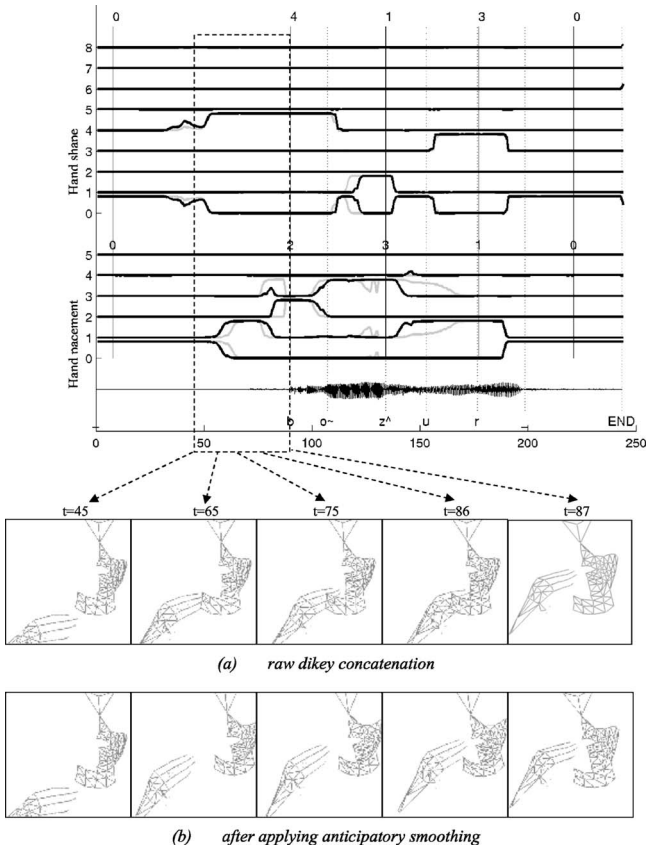


FIG. 10. Synthesis of the word "Bonjour!" ([_bɔ̃ʒuʁ_] means Hello!) by the audiovisual text-to-cued speech system. Two chronographs of the hand and face gestures generated for sequence [_bɔ̃] are shown (a) after raw concatenation and (b) after applying the smoothing procedure. The interval on the score is evidenced by a dotted rectangle. Since the dikey [00-24], is not in the dikey dictionary, the system has selected the nearest dikey in the dictionary, i.e., [00-34]. The raw dikey concatenation produces the expected movement discontinuity between frame 86 and 87 while the anticipatory smoothing procedure corrects nicely both the hand and face movements. Please check the effects of this procedure on the recognition of the hand shapes and placements (top caption): raw concatenation (data in light gray) triggers consecutively the hand placements 0, 1, 3, and 2, whereas the smoothing procedure (data in back) restores the good sequence 0, 1, and 2. Also see the increased anticipation for cueing the final [r].

the upcoming vowel and the consonant *together*, far ahead of the actual realization of the segments.

We thus added to the labels of nine hand shape targets set by the procedure described previously, the appropriate hand placement value, i.e., a number between 0 and 5 (0 for the rest position, i.e., the same as above, the cueing hand with the close knuckle being far from the face). Additional targets were also added for single vowels and start/end hand positions. Targets for single vowels were labeled with hand shape 5 while the rest position was labeled with hand placement 0.

We characterized the hand placement for these target configurations in a 3-D referential linked to the head: the 3-D position of the longest finger (index for hand shape 1 and 6 and middle finger for the others) were collected and simple Gaussian models were estimated for each hand placement.

Of the 4114 hand placements, 96.76% were identified with a total of 133 errors (Table I). There were two main sources of incorrect identifications.

TABLE I. Confusion matrices. Left: for hand shapes; right: for hand placements.

		Expected hand shapes								
		0	1	2	3	4	5	6	7	8
Recognized	0	462	1	1	4	4	15	0	0	3
	1	0	482	0	1	0	0	1	0	0
	2	1	0	419	0	0	1	0	0	0
	3	0	0	0	598	0	0	0	0	0
	4	2	0	2	0	357	0	0	0	0
	5	9	0	0	1	0	957	0	1	0
	6	2	1	0	0	0	0	545	1	0
	7	0	0	0	0	0	0	0	82	0
	8	0	0	0	0	0	0	0	0	162

		Expected hand placements					
		0	1	2	3	4	5
Recognized	0	475	27	6	5	3	8
	1	1	1647	2	2	3	0
	2	0	14	565	3	7	0
	3	0	9	6	361	0	3
	4	0	9	9	1	359	0
	5	0	1	4	8	2	574

(i) Hand placement 1 (side). This hand placement was used for a consonant followed by another consonant or a *schwa* and an undershoot of this short target occurs very often (i.e., the cuer only points to the side but does not reach it). The tendency to undershoot is clearly shown in Fig. 7.

(ii) Hand placement 0 displays a large variance (see Fig. 7) and hand placements 1 (side) and 4 (cheeks) realized too far away from the face were sometimes captured by the Gaussian model as hand placement 0.

D. Phasing speech and gestures

In the present study, the data also gathered valuable information on phasing relations between speech and hand gestures and confirmed the advance of the hand onset gesture already hypothesized by Attina *et al.* (2003a). On the basis of the gestural scores provided by the cued speech decoder presented in Sec. IV B, we analyzed the profile of hand shape and hand placement gestures with reference to the acoustic realization of the speech segment they are related to (hand shape for consonants and hand placement for vowels). The extension of a gesture was defined as the time interval where the probability of the appropriate key (shape or placement) dominates the other competing keys. We excluded from the analysis the segments that required the succession of two identical keys. A sketch of the profiles for CV, V-only, and C-only sequences is presented in Fig. 9 and Table II). These results extended the data provided by Attina et al: despite an important dispersion of the data, both the hand shape and hand placement were realized well before the acoustic onset of the speech segment they relate to. Furthermore, the hand shape and hand placement gestures were highly synchronized since they participated both in the hand/head constriction, as amplified above. We tested several hypothesizes on these phasing profiles. The conclusions are as follows.

(i) For CV segments: hand placement onsets are significantly synchronized with the acoustic onset of the segment ($p<0.05$), hand shape onsets being notably in advance. Their offsets are within the second part of the vowel ($p<0.01$). The target (labeled by hand at the center of the holding of both the hand shape and placement) remain within the consonant (a mean delay of 89 ms for an average consonantal closure of 129 ms): This result is an important one because it validates the conclusions of Attina and colleagues on a more extensive corpus, i.e., the synchronization of the hand with the beginning of the CV syllable, the duration of the hand placement until the beginning of the vowel then its move during the vowel; towards the placement of the next CV syllable.

(ii) For C-only segments, hand shape and position onsets are significantly in advance of the acoustic onset of the consonant ($p<10e-9$). Their targets are synchronized with the acoustic onset of the consonant ($p<0.05$). The offset of the hand shape is within the consonant. The offset of the hand position is synchronized with the acoustic offset of the consonant ($p<0.01$).

(iii) For V-only segments, hand shape and position onsets and targets are significantly in advance of the acoustic onset of the vowel ($p<10e-9$). Their targets are synchronized with the acoustic onset of the vowel ($p<0.05$). Their offsets are realized within the vowel.

TABLE II. Average delay (ms) between the acoustic onset of CV, C-only or V-only segments and the onset, offset, and target position of the hand shape and position gestures. Remark: only the segments whose hand shapes and placements differ from their neighbors are taken into account.

Nb	CV 1027		C 474		V 182	
	Shape	Position	Shape	Position	Shape	Position
Onset	−191	−189	−234	−386	−517	−324
Offset	174	218	24	248	62	15
Target	−24	−24	−110	−110	−177	−177

TABLE III. Number of «dikeys»: transitions between two hand targets. Left: for hand shapes; right: for hand placements.

		Next hand shape								
		0	1	2	3	4	5	6	7	8
Previous	0	0	36	12	26	11	80	69	1	3
	1	27	38	55	99	49	118	62	14	22
	2	27	44	45	65	41	95	69	7	29
	3	47	89	68	74	61	157	67	11	30
	4	28	43	38	47	23	74	75	13	20
	5	47	123	94	183	105	244	130	15	31
	6	37	83	87	71	48	138	41	17	24
	7	7	5	9	20	10	13	16	3	1
	8	18	23	14	19	13	53	17	3	5

		Next hand placement					
		0	1	2	3	4	5
Previous	0	0	70	34	27	51	56
	1	127	689	305	172	168	246
	2	32	306	77	47	39	91
	3	14	288	22	21	10	25
	4	19	142	78	46	42	47
	5	46	212	76	67	64	120

V. TOWARD AN AUDIOVISUAL TEXT-TO-CUED SPEECH SYNTHESIS SYSTEM

This corpus provided an extensive coverage of the movements implied by FCS and we have designed a first audiovisual text-to-cued speech synthesis system using concatenation of multimodal speech segments. Concatenative synthesis using a large speech database and multirepresented speech units has been largely used for acoustic synthesis (Campbell, 1997; Hunt and Black, 1996) and more recently for facial animation (Minnis and Breen, 1998). This system is—to our knowledge—the first system attempting to generate hand and face movements and articulations together with speech using the concatenation of gestures and acoustics. Two units will be considered below: diphones for the generation of the acoustic signal and facial movements; and dikeys[2] for the generation of head motion as well as for hand movements and articulations.

A. Coverage of the corpus: towards text-to-cued speech synthesis

This corpus was designed initially for acoustic concatenative speech synthesis. The coverage of polysounds (the part of speech comprised between successive stable allophones, i.e., similar to diphones but excluding glides as stable allophones) was quasioptimal: we collected a minimum number of two occurrences of each polysound with a small number of utterances.

Although not quite independent (see Sec. IV), hand placements, and hand shapes were almost orthogonal. The coverage of the corpus in terms of successions of hand placements and hand shapes was quite satisfactory (see Table III): no succession of hand shapes nor hand placements was missing.

A first text-to-cued speech system had been developed using these data. This system proceeds in three steps.

(i) A prosodic model (Bailly, 2004; Bailly and Holm, submitted) trained using corpus 3 computes phoneme durations from the linguistic structure of the sentences.
(ii) Sound and facial movements are handled by a first concatenative synthesis using polysounds (and diphones if necessary) as basic units.
(iii) Head movements, hand movements, and hand shaping movements are handled by a second concatenative synthesis using dikeys (see below) as basic units.

A key (hand and head gesture) will be referenced in the following by two numbers representing the hand placement and hand shape. For example, the key 24 (hand placement 2 together with hand shape 4) will be selected for cueing the CV sequence [bõ]. The so-called dikeys are part of movements comprised between two successive keys. For example, the dikey [00-24] stores the hand and head movements from the rest position 00 toward the key 24. Once selected, the onsets of these dikeys were further aligned with the acoustic C mid-point for full CV realizations, vocalic onsets for "isolated" vowels (not immediately preceded by a consonant) and consonantal onsets for "isolated" consonants (not immediately followed by a vowel). This phasing relation is in accordance with the data presented in Sec. IV D. If the full dikey does not exist, replacement dikeys are found by replacing the second hand placement of the dikey by the closest one that does exist in the dikey dictionary. The proper dikey will then be realized through the application of an *anticipatory* smoothing procedure (Bailly, 2002) that considered the onset of each dikey as the intended target of the first key: a linear interpolation of the parameters for the hand model is gradually applied within the preceding dikey in order that its final target coincide with the onset of the current dikey.

An example of a sequence generated by the proposed system is shown in Fig. 10. The sequence results from the concatenation and smoothing of four dikeys selected from four different utterances. Automatic recognition of the hand shapes and placements (similar to Fig. 8) is provided in order to demonstrate the ability of the system to generate the appropriate gestures. This two-step procedure generated acceptable synthetic cued speech. It, however, considers the head movements to be entirely part of the realization of hand-face constrictions and uses, for now a crude approximation of the speech/gesture coordination (although being a very satisfactory first-order approximation as suggested by Attina *et al.*, 2002).

B. From gestures to appearance

The text-to-cued speech synthesis system sketched above delivered trajectories of a few flesh points placed on

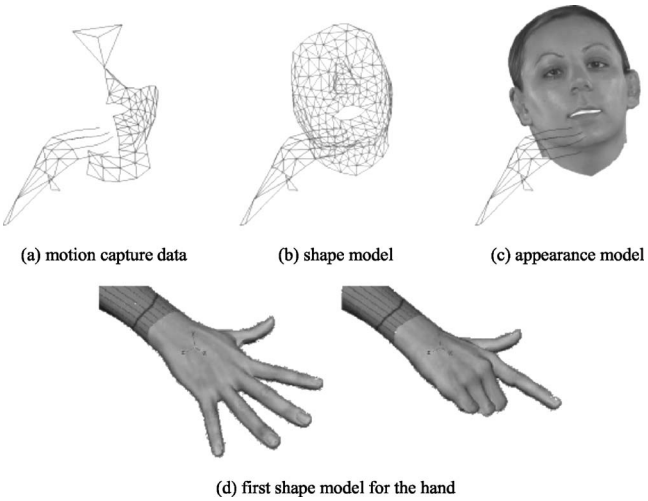


FIG. 11. Driving a complete shape and video-realistic appearance model of the cuer from the parameters of the face model. (a) Initial facial flesh points; (b) high-definition mesh; (c) textured mesh. High-definition meshes and video-realistic textures for the hand [see (d)], the teeth and the inner mouth are currently under development.

the surface of the right hand and face. We plan to evaluate the benefits brought by this system in speech understanding using the point-light paradigm we already used for face only (Bailly *et al.*, 2002; Odisio and Bailly, 2004).

We are currently interfacing this trajectory planning with a detailed shape and appearance model of the face and hand of the original speaker. High definition models of these organs—comprising several hundreds of vertices and polygons—are first mapped onto the existing face and hand parameter space. A further appearance model using video-realistic textures is then added (Elisei *et al.*, 2001; Bailly *et al.*, 2003). Figure 11 illustrates our ongoing effort toward the animation of a videorealistic virtual cuer. Applying the same procedure to a high definition model of the hand is currently under study.

C. Comments

An important challenge of the phonetic description of FCS is clearly understanding the constraints acting on the multisegment gestural planning: several segments are, in fact, recruited in addition to the usual speech articulators (jaw, lips, tongue, larynx,…). Head and hand movements should be appropriately phased in order to ease lip reading. We suggest here that this planning be done in terms of constrictions made while recruiting specific hand configurations. This planning ensures that cued speech information will be delivered well in advance of the lip reading information. This gestural score has interesting similarities with data on synchronization between deictic pointing and speech that evidences also an important anticipatory pointing gesture in relation with acoustic onset (as large as 300 ms in Castiello *et al.*, 1991). For an optimal processing of the message, the "*where*" component of the multimodal deictic gesture should precede the "*who*" component.

Movement execution recruits segments in an optimal manner, but all segments actually participate in the realization of the series of constrictions. A major challenge for movement analysis and generation will be to separate out prosodic—or literally suprasegmental—movements of the head from movements of the head contributing to the correct decoding/encoding of the speech segments.

Another important issue will be to understand what the influence of FCS is, on the temporal organization of the speech gestures and more specifically on the rhythmical organization of the speech stream.

VI. CONCLUSIONS AND PERSPECTIVES

The immense benefits of Cued Speech in terms of giving access to language structure and speech comprehension to deaf people should be grounded in a deep understanding of its implementation by actual speakers. Although precise qualitative guidelines have been specified by Cornett, the FCS is an evolving system whose phonetic structure is constantly enriched by cuers.

We analyzed here the live recordings of the hand, face and head gestures with reference to the phonetic structure of the speech sounds produced by a user of FCS. When compared to the expected hand shapes and hand placement targets positioned by the hand, the automatic cued speech recognizer operating on hand gestural scores identified respectively 98.14% and 95.52% of the hand shapes and placements. The errors could clearly be interpreted in terms of undershoot or true phonetic confusions, mostly involving confusions between vowel apertures or consonant devoicing/voicing.

The study of the phasing relation between these gestural scores and the phonetic structure of the sound produced confirmed the empirical rules used by Duchnowski *et al.* (2000) for automatically superimposing virtual hand shapes on a prerecorded video of a speaking person and the motion capture data analyzed by Attina *et al.* (2002; 2003a; 2004): the hand movements provided phonetic cues for the decoding of the incoming speech well before any other acoustic or facial cues; typically more than 200 ms before the acoustic onset of the sound the gesture related to. Further perception experiments involving gating experiments and reaction times will be required to test if this cued speech advantage is actually used by observers and to test the sensitivity of their performance when this anticipatory coarticulation is altered.

The observation of cuers in action is thus a prerequisite for developing technologies that will assist deaf people in learning FCS. The database recorded, analyzed, and characterized here is currently exploited within a multimodal text-to-FCS system that will supplement or replace on-demand subtitling by a virtual FCS cuer for TV broadcasting or home entertainment. With the ARTUS project, ICP and Attitude Studio collaborate with academic and industrial partners in order to provide the French–German TV channel ARTE with the possibility of broadcasting programs dubbed with virtual CS. The movements of which are computed from existing subtitling or captured life on a FCS interpreter, watermarked within the video and acoustic channels and rendered locally by the TV set. The low transmission rate of CS as required

by watermarking should also benefit from a better understanding of the kinematics of the different segments involved in the production of CS.

ACKNOWLEDGMENTS

Many thanks to Yasmine Badsi, our CS speaker for having accepted the recording constraints. We thank Christophe Corréani, Xavier Jacolot, Jeremy Meunier, Frédéric Vandenbergh, and Franck Vayssettes for the processing of the raw motion capture data. We also acknowledge Virginie Attina for providing her cued speech expertise when needed. We thank Siyi Wang for helping us to correct the English. This work is financed by the RNRT ARTUS. We acknowledge three anonymous reviewers for thoughtful remarks on the earlier versions of this paper.

[1]Throughout this article, the use of the term cuer refers to a user of cued speech.

[2]A dikey is defined as the part of movements comprised between two successive keys (see Sec. V A).

Attina, V., Beautemps, D., and Cathiard, M.-A. (**2002**). "Coordination of hand and orofacial movements for CV sequences in French Cued Speech," *International Conference on Speech and Language Processing*, Boulder, pp. 1945–1948.

Attina, V., Beautemps, D., and Cathiard, M.-A. (**2003a**). "Temporal organization of French Cued Speech production," *International Conference of Phonetic Sciences*, Barcelona, Spain.

Attina, V., Beautemps, D., Cathiard, M.-A., and Odisio, M. (**2003b**). "Towards an audiovisual synthesizer for Cued Speech: rules for CV French syllables," *Auditory–Visual Speech Processing*, St Jorioz, France, pp. 227–232.

Attina, V., Beautemps, D., Cathiard, M.-A., and Odisio, M. (**2004**). "A pilot study of temporal organization in Cued Speech production of French syllables: rules for a Cued Speech synthesizer," Speech Commun. **44**, 197–214.

Badin, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C., and Savariaux, C. (**2002**). "Three-dimensional linear articulatory modeling of tongue, lips and face based on MRI and video images," J. Phonetics **30**(3) 533–553.

Bailly, G., and Badin, P. (**2002**). "Seeing tongue movements from outside," *International Conference on Speech and Language Processing*, Boulder, Colorado, pp. 1913–1916.

Bailly, G., Gibert, G., and Odisio, M. (**2002**). "Evaluation of movement generation systems using the point-light technique," *IEEE Workshop on Speech Synthesis*, Santa Monica, CA, pp. 27–30.

Bailly, G., Holm, B., and Aubergé, V. (**2004**). "A trainable prosodic model: learning the contours implementing communicative functions within a superpositional model of intonation," *International Conference on Speech and Language Processing*, Jeju, Korea, pp. 1425–1428.

Bailly, G., and Holm, B. (to be published). "SFC: a trainable prosodic model," Speech Commun..

Bailly, G., Bérar, M., Elisei, F., and Odisio, M. (**2003**). "Audiovisual speech synthesis," International Journal of Speech Technology **6**(4) 331–346.

Bernstein, L. E., Demorest, M. E., and Tucker, P. E. (**2000**). "Speech perception without hearing," Percept. Psychophys. **62** 233–252.

Campbell, N. (**1997**). "Computing prosody: Computational models for processing spontaneous speech," *Synthesizing Spontaneous Speech*, edited by Y. Sagisaka, N. Campbell, and N. Higuchi (Springer-Verlag, Berlin), pp. 165–186.

Castiello, U., Paulignan, Y., and Jeannerod, M. (**1991**). "Temporal dissociation of motor responses and subjective awareness," Brain **114**, 2639–2655.

Cornett, R. O. (**1967**). "Cued Speech," Am. Ann. Deaf **112**, 3–13.

Cornett, R. O. (**1988**). "Cued Speech, manual complement to lipreading, for visual reception of spoken language. Principles, practice and prospects for automation," Acta Otorhinolaryngol. Belg. **42**, 375–384.

Cornett, R. O., and Daisey, M. E. (**1992**). *The Cued Speech Resource Book for Parents of Deaf Children*. Raleigh, NC, The National Cued Speech Association, Inc.

Duchnowski, P., Lum, D. S., Krause, J. C., Sexton, M. G., Bratakos, M. S., and Braida, L. D. (**2000**). "Development of speechreading supplements based on automatic speech recognition," IEEE Trans. Biomed. Eng. **47**(4): 487–496.

Elisei, F., Odisio, M., Bailly, G., and Badin, P. (**2001**). "Creating and controlling video-realistic talking heads," *Auditory-Visual Speech Processing Workshop*, Scheelsminde, Denmark, pp. 90–97.

Engwall, O., and Beskow, J. (**2003**). "Resynthesis of 3D tongue movements from facial data," *EuroSpeech*, Geneva.

Hunt, A. J., and Black, A. W. (**1996**). "Unit selection in a concatenative speech synthesis system using a large speech database," *International Conference on Acoustics, Speech and Signal Processing*, Atlanta, GA, pp. 373–376.

Jiang, J., Alwan, A., Bernstein, L., Keating, P., and Auer, E. (**2000**). "On the Correlation between facial movements, tongue movements and speech acoustics," *International Conference on Speech and Language Processing*, Beijing, China, pp. 42–45.

Kuratate, T., Munhall, K. G., Rubin, P. E., Vatikioti-Bateson, E., and Yehia, H. (**1999**). "Audio-visual synthesis of talking faces from speech production correlates," *EuroSpeech*, pp. 1279–1282.

Leybaert, J. (**2000**). "Phonology acquired through the eyes and spelling in deaf children," J. Exp. Child Psychol. **75**, 291–318.

Leybaert, J. (**2003**). "The role of Cued Speech in language processing by deaf children: An overview," *Auditory-Visual Speech Processing*, St Jorioz, France, pp. 179–186.

Minnis, S., and Breen, A. P. (**1998**). "Modeling visual coarticulation in synthetic talking heads using a lip motion unit inventory with concatenative synthesis," *International Conference on Speech and Language Processing*, Beijing, China, pp. 759–762.

Nicholls, G., and Ling, D. (**1982**). "Cued Speech and the reception of spoken language," J. Speech Hear. Res. **25**, 262–269.

Odisio, M., and Bailly, G. (**2004**). "Shape and appearance models of talking faces for model-based tracking," Speech Commun. **44**, 63–82.

Owens, E., and Blazek, B. (**1985**). "Visemes observed by hearing-impaired and normal-hearing adult viewers," J. Speech Hear. Res. **28**, 381–393.

Revéret, L., Bailly, G., and Badin, P. (**2000**). "MOTHER: a new generation of talking heads providing a flexible articulatory control for video-realistic speech animation," *International Conference on Speech and Language Processing*, Beijing, China, pp. 755–758.

Uchanski, R., Delhorne, L., Dix, A., Braida, L., Reed, C., and Durlach, N. (**1994**). "Automatic speech recognition to aid the hearing impaired: Prospects for the automatic generation of cued speech," J. Rehabil. Res. Dev. **31**, 20–41.

Yehia, H. C., Rubin, P. E., and Vatikiotis-Bateson, E. (**1998**). "Quantitative association of vocal-tract and facial behavior," Speech Commun. **26**, 23–43.

Touch and temporal behavior of grand piano actions

Werner Goebl
Austrian Research Institute for Artificial Intelligence (OFAI), Freyung 6/6, 1010 Vienna, Austria

Roberto Bresin
Department of Speech, Music, and Hearing (TMH), Royal Institute of Technology (KTH), Lindstedtsvägen 24, 10044 Stockholm, Sweden

Alexander Galembo
Sechenov Institute of Evolutionary Physiology and Biochemistry, Russian Academy of Sciences, M. Toreza av. 44, St. Petersburg 194223, Russia

(Received 18 January 2005; revised 10 May 2005; accepted 10 May 2005)

This study investigated the temporal behavior of grand piano actions from different manufacturers under different touch conditions and dynamic levels. An experimental setup consisting of accelerometers and a calibrated microphone was used to capture key and hammer movements, as well as the sound signal. Five selected keys were played by pianists with two types of touch ("pressed touch" versus "struck touch") over the entire dynamic range. Discrete measurements were extracted from the accelerometer data for each of the over 2300 recorded tones (e.g., finger-key, hammer-string, and key bottom contact times, maximum hammer velocity). Travel times of the hammer (from finger-key to hammer-string) as a function of maximum hammer velocity varied clearly between the two types of touch, but only slightly between pianos. A travel time approximation used in earlier work [Goebl W., (2001). J. Acoust. Soc. Am. **110**, 563–572] derived from a computer-controlled piano was verified. Constant temporal behavior over type of touch and low compression properties of the parts of the action (reflected in key bottom contact times) were hypothesized to be indicators for instrumental quality.
[DOI: 10.1121/1.1944648]

PACS number(s): 43.75.Mn, 43.75.St [NHF] Pages: 1154–1165

I. INTRODUCTION

The universe of expressive music to be played on the modern grand piano[1] is produced by sophisticated acceleration of the (usually) 88 keys, none of which travels through a distance greater than one centimeter, combined with the use of the pedals. The piano action provides the pianist the only point of contact to the strings; it is therefore both an extremely important as well as a highly elaborate and complex mechanical interface. It allows accurate control over the speed at which the hammer arrives at the strings over a vast dynamical range from the very pianissimo to the ultimate fortissimo. Since not only the intensity of tone, but also the precise onset timing of the outcoming sound is crucial to expressive performance, it can be assumed that trained pianists are intuitively well acquainted with the temporal behavior of a piano action, and that they take it into account while performing expressively.

The piano action functions as follows: The movement of the key is transferred to the hammer via the whippen, on which the jack is positioned so that it touches the roller (knuckle) of the hammer shank. During a keystroke, the tail end of the jack is stopped by the escapement dolly (let-off button, jack regulator) causing the jack to rotate away from the roller, and thus breaking the contact between key and hammer. From this moment, the hammer travels to the strings with a small deceleration due to gravitation and friction, strikes them, and rebounds from them ("free flight of the hammer"). The roller falls back to the repetition lever, while the hammer is caught by the back check. For a fast repetition, the jack slides back under the roller when the key is only released half-way, and the action is ready for another stroke (Askenfelt and Jansson, 1990b; Fletcher and Rossing, 1998, pp. 354–358).

A. Temporal properties of the piano action

Temporal aspects of the piano action have been investigated recently by Askenfelt and Jansson (1990a,b, 1991).[2] The time interval from the key's initial position to key bottom contact ranges from about 25 ms at a *forte* keystroke (approximately 5 m/s final hammer velocity FHV) to 160 ms at a *piano* tone (or 1 m/s FHV, Askenfelt and Jansson, 1991, p. 2385).[3] In a grand piano, the moments of hammer contact (when the hammer excites the strings) are temporally shifted in comparison to key bottom contact. Hammer contact occurs 12 ms before key bottom contact at a *piano* tone (1 m/s FHV), but 3 ms *after* the key bottom contact at a *forte* attack (5 m/s FHV, Askenfelt and Jansson, 1990a, p. 43). However, these studies provided measurement data solely for a few example keystrokes.

The timing properties of the piano action can be modified by changing the regulation of the action. Modifications, e.g., in the hammer-string distance or in the *let-off distance* (the distance of free flight of the hammer, after the jack is released by the escapement dolly), alter the timing relation between hammer-string contact and key bottom contact or the free flight time, respectively (Askenfelt and Jansson,

1990b, p. 57). Greater hammer mass in the bass range (Conklin, 1996, p. 3287) influences the hammer-string contact durations (Askenfelt and Jansson, 1990b), but not the timing properties of the action.

Data on timing properties of a baby grand piano action were provided by Repp (1996) who worked with a Yamaha Disklavier (Mark II series, similar to the one used in the present study) on which the "prelay function" was not working.[4] This gave him the opportunity to measure roughly a grand piano's travel time characteristics. He measured onset asynchronies at different MIDI velocities in comparison to a note with a fixed MIDI velocity in the middle register of the keyboard. The time deviations extended over a range of about 110 ms for MIDI velocities between 30 and 100 and were fit well by a quadratic function (Repp, 1996, p. 3920).

A function similar to the Disklavier's prelay function was obtained from a Bösendorfer SE290 computer-controlled grand piano by Goebl (2001). He accessed data from an internal memory chip of the SE system that presumably stored information on the travel time intervals for each of the 97 keys and seven selected FHVs. An average travel time function (against FHV) was derived from these data (Goebl, 2001, Fig. 5, p. 568). It was applied to predict the amounts of note onset asynchronies to be expected at given dynamic differences between the voices of a chord.

There have been several attempts to model piano actions (Gillespie, 1994; Hayashi *et al.*, 1999), also for a possible application in electronic keyboard instruments (Cadoz *et al.*, 1990; Van den Berghe *et al.*, 1995). Gillespie and colleagues developed a virtual keyboard that simulates the haptic feel of a real grand piano action using motorized keys (Gillespie, 1994). Van den Berghe *et al.* (1995) performed measurements on a grand piano key with two optical sensors for hammer and key displacement and a strain gauge for key force. Unfortunately, they provided only a single exemplary keystroke of their data. Hayashi *et al.* (1999) tested one piano key on a Yamaha grand piano. The key was hit with a specially developed key actuator able to produce different acceleration patterns. The displacement of the hammer was measured with a laser displacement gauge. They developed a simple model and tested it in two touch conditions (with constant key velocity and constant key acceleration). Their model predicted the measured data for both conditions accurately.

B. Different types of touch

While physicists and technicians argue that the sole factor that controls the sound and the timbre of the piano is the hammer velocity at which the hammer hits against the strings (Hart *et al.*, 1934; Seashore, 1937; White, 1930), it is of extraordinary importance for pianists, how they touch and accelerate the keys. As Ortmann (1929, p. 3) puts it: "The complex problem of physiological mechanics as applied to piano technique resolves itself finally, into one basic question: the variations of force produced at the key-surface by the player." And in order to produce the forces at the key surface that entail a particular desired musical outcome, pianists have to practice for decades. Over this time period, they develop a tacit tactile knowledge of how piano actions behave under the various physical forces they apply to it. Thus, an integral part of what pianists perceive from a piano is the haptic-tactile response of the keys (including particularly key resistance and inertia) in relation to the physical force they apply and to the acoustical result they hear (Galembo, 2001).

An article by Bryan (1913) was the starting point of a lively discussion on piano touch. Bryan puts the "single-variable hypothesis" (timbre of a piano tone determined solely by FHV) into question with rudimentary experiments performed with a player-piano. His contribution entailed a discussion of six Letters to the Editor and three further replies by Bryan [all to be found in Nature (London) **91–92**, 1913].

A first profoundly scientific investigation to this controversy contributed Otto Ortmann from the Peabody Conservatory of Music in Baltimore (Ortmann, 1925). He approached the "mystery of touch and tone" at the piano through physical investigation. With a piece of smoked glass mounted on the side of a piano key and a tuning fork, he was able to record and to study key depression under different stroke conditions. He investigated various kinds of keystrokes ("percussive" versus "nonpercussive," different muscular tensions, and positions of the finger). He found different acceleration patterns for nonpercussive (finger rests on the surface of the key before pressing it) and percussive touch (an already moving finger strikes the key). The latter starts with a sudden jerk, thereafter the key velocity decreases for a moment and increases again. During this period, the finger slightly rebounds from the key (or vice versa), then re-engages the key and "follows it up" (Ortmann, 1925, p. 23). On the other side, the nonpercussive touch caused the key to accelerate gradually.

Ortmann (1925) found that these different types of touch provide a fundamentally different kind of key control. The percussive touch required precise control of the very first impact, whereas with nonpercussive touch, the key depression needed to be controlled up to the very end. "This means that the psychological factors involved in percussive and nonpercussive touches are different" (Ortmann, 1925, p. 23). "In nonpercussive touches key resistance is a sensation, in percussive touches it is essentially an image" (Ortmann, 1925, p. 23, footnote 1). His conclusions were that different ways of touching the keys produced different intensities of tones, but when the intensity was the same, also the quality of the tone must be the same. "The quality of a sound on the piano depends upon its intensity, any one degree of intensity produces but one quality, and no two degrees of intensity can produce exactly the same quality" (Ortmann, 1925, p. 171).

The discussion was enriched by introducing the aspect of different noises that emerge with varying touch (Báron and Holló, 1935; Cochran, 1931). Báron and Holló (1935) distinguished between "Fingergeräusch" (finger noise) that occurs when the finger touches the key (which is absent when the finger velocity is zero as touching the key—in Ortmann's terminology "nonpercussive touch"), "Bodengeräusch" (keybed noise) that emerges when the key hits the keybed, and "Obere Geräusche" (upper noises) that develop when the key is released again (e.g., the damper hitting the

strings). As another (and indeed very prominent) source of noise they mentioned the pianist's foot hitting the stage floor (or the pedals) in order to emphasize a *fortissimo* passage. In a later study, Báron (1958) advocated a broader concept of tone quality, including all kinds of noise (finger-key, action, and hammer-string interaction), which he argued to be included into concepts of tone characterization of different instruments (Báron, 1958).

More recent studies investigated these different kinds of noise that emerge when the key is struck in different ways (Askenfelt, 1994; Koornhof and van der Walt, 1994; Podlesak and Lee, 1988). The hammer-impact noise ("string precursor") arrives at the bridge immediately after hammer-string contact and characterizes the "attack thump" of the piano sound without which it would not be recognized as such (Chaigne and Askenfelt, 1994a,b). This noise is independent of touch type. The hammer impact noises of the grand piano do not radiate equally strongly in all directions (Bork *et al.*, 1995). As three-dimensional measurements with a two-meter Bösendorfer grand piano revealed, increased noise levels were found horizontally towards the pianist and in the opposite direction, to the left (viewed from the sitting pianist), and vertically towards the ceiling.

Before the string precursor, another noise component could occur: the "touch precursor," only present when the key was hit from a certain distance above ("staccato touch," Askenfelt, 1994). It precedes the actual tone by 20 to 30 ms and was much weaker than the string precursor. Similar results were reported by Koornhof and van der Walt (1994). They called the noise prior to the sounding tone "early noise" or "acceleration noise;" it occurs closely in time with finger-key contact. They performed an informal listening test with four participants. The two types of touch (staccato touch with the early noise and "legato touch") could be easily identified by the listeners, but not anymore with the early noise removed. Unfortunately, no further systematic results were reported (Koornhof and van der Walt, 1994).

In a recent perception study (Goebl *et al.*, 2004), musicians could hardly identify what type of touch piano tone samples were played when the finger-key noises were included (only half of them rated significantly better than chance), but not at all, when finger-key noises were removed from the stimuli. This evidence suggests that finger-key noise, which occurs only with a percussive ("struck") touch, is responsible for pure aural touch recognition.

The different kinds of touch also produced different finger-key touch forces (Askenfelt and Jansson, 1992b, p. 345). A *mezzo forte* attack played with staccato touch typically has 15 N, very loud such attacks show peaks up to 50 N (*fortissimo*), very soft touches go as low as 8 N (*piano*). Playing with legato touch, finger-key forces of about one third of those obtained with staccato touch are found, usually having a peak when the key touches the keybed. At a very *pianissimo* tone, the force hardly exceeds 0.5 N.

In the literature, there are other ways of categorizing touch, as e.g., Suzuki (2003) who introduced a "hard-soft" antagonism, relying on a professional pianist's intuition how this distinction is realized on the piano. However, he did not control for this variable (human factor) in his experiments.

In the present study, two prototypical types of depressing the keys are used based on the criterion of the finger's speed when beginning the keystroke. These two types ("struck touch" and "pressed touch") are identical to the categories introduced by Askenfelt and Jansson (1991). However, the terminology was deliberately changed from "legato-staccato" (that was also used in earlier studies by the authors, i.e., Goebl and Bresin, 2003; Goebl *et al.*, 2003) to "pressed-struck" (Goebl *et al.*, 2004) in order to draw a clear distinction between terms referring to touch and those referring to articulation (that is the length of each tone relative to its nominal value in the score, thus referring to the connection of tones). Especially in conversations with performing musicians, they get very quickly confused by legato-staccato used in a double sense. Nevertheless, there might be parallels between these two meanings of legato-staccato. E.g., articulated and short tones may be more likely played with a struck touch and legato tones smoothly overlapping with each other might be more likely played from the key surface (pressed touch). However, these parallels occur only in very typical situations; pianists will have no difficulty in producing opposite examples, e.g., a short staccato tone played with a pressed touch and vice versa.

II. AIMS

The present study aimed to collect a large amount of measurement data from different grand pianos, different types of touch, and different keys, in order to determine and provide benchmark functions that may be useful in performance research as well as in piano pedagogy. The measurement setup with accelerometers was similar to that as used by Askenfelt and Jansson (1991). However, in order to obtain a large and reliable data set, the data processing procedure and the reading of discrete values was automated with purpose-made computer software. Each of the measured notes was equipped with two accelerometers monitoring key and hammer velocity. Additionally, a microphone recorded the sound of the piano tone. With this setup, various temporal properties were determined and discussed (travel time, key bottom time, time of free flight). Moreover, the speed histories of both key and hammer revealed essential insights into the fundamentally different nature of the two types of touch examined in this study.

In a study on tone onset asynchronies in expressive piano performance ("melody lead," Goebl, 2001), finger-key onset times were inferred from the hammer-string onset times through an approximation of the travel times of the hammer (from finger-key to hammer-string contact) as a function of FHV. This travel time function was obtained from data of an internal chip of a Bösendorfer SE290 reproducing system. The present study additionally aims to reconsider that approximation.

III. METHOD

A. Material

Three grand pianos by different manufacturers were investigated. Two of them were computer-controlled pianos, the same as in an earlier study (Goebl and Bresin, 2003).

(1) **Steinway grand piano**, (model C, 225 cm, serial number: 516000, built in Hamburg, Germany, in 1989),[5] situated at TMH–KTH in Stockholm, Sweden.
(2) **Yamaha Disklavier grand piano** (DC2IIXG, 173 cm, serial number: 5516392, built in Japan, approximately 1999), situated at the Dept. of Psychology at the University of Uppsala, Sweden.
(3) **Bösendorfer computer-controlled grand piano** (SE290, 290 cm, internal number: 290-3, built in Austria, 2000), situated at the Bösendorfer Company in Vienna.

Immediately before the experiments, the instruments were tuned, and the piano action and—in the case of the computer-controlled pianos—the reproduction unit serviced and regulated. The Steinway grand has been regularly maintained by a piano technician of the Swedish National Radio. At the Disklavier, this procedure was carried out by a specially trained Disklavier piano technician from the Stockholm "Konserthus." At the Bösendorfer company, the company's SE technician took care of this work.

B. Equipment and calibration

The tested keys were equipped with an accelerometer on the key[6] and another one on the bottom side at the end of the hammer shank.[7] The sound was picked up by a sound-level meter microphone[8] placed about 10 cm above the strings. The velocities of key and hammer and the sound signal were recorded on a multichannel DAT recorder (TEAC RD-200 PCM) with a sampling rate of 10 kHz and 16-bit word length. The data were transferred to a computer harddisk and analyzed with computer software written for this purpose. The recorded voltages were transformed to obtain required measures (m/s and dB SPL). The measuring equipment and the calibration procedure was identical as in Goebl and Bresin (2003), so we do not repeat further details here.[9]

C. Procedure

Five keys distributed over the whole range of the keyboard were tested: C1 (MIDI note number 24, 32.7 Hz), G2 (43, 98.0), C4 (60, 261.6), C5 (72, 523.3), and G6 (91, 1568.0).[10] The first two authors served as pianists to perform the recorded test tones. Each key was hit at as many different dynamic levels (hammer velocities) as possible, with two different kinds of touch: one with the finger resting on the key surface (*pressed touch*), the other hitting the key from a certain distance above, thus with the finger touching the key already with a certain speed (*struck touch*). Parallel to the accelerometer setting, the two computer-controlled grand pianos recorded these test tones with their internal device on computer hard disk (Bösendorfer) or floppy disk (Disklavier).

For each of the five keys, both players played in both types of touch from 30 to 110 individual tones, so that a sufficient amount of data was recorded. In case of the two computer-controlled devices (Bösendorfer and Yamaha), the internally recorded file was reproduced by the grand piano immediately after each recording of a particular key, and the accelerometer data was recorded again onto the multichannel DAT recorder. However, for the sake of clarity and due to limited space, we restricted this paper to the human data. For the Steinway, 595 individual attacks were recorded, for the Yamaha 996, and for the Bösendorfer 756 (not counting the keystrokes repeated by the reproducing devices).[11]

D. Data analysis

In order to analyze the three-channel data files, discrete measurement values had to be extracted from them. Several instants in time were defined as listed later and automatically read off with the help of Matlab scripts prepared by the first author for this purpose. This method allowed to obtain timing data without specially having to install additional sensors or contacts into the piano action (as, e.g., done by Askenfelt and Jansson, 1990b), only by processing the key and hammer trajectory and the sound information.

The **hammer-string contact** was defined as the moment of maximum deceleration (minimum acceleration) of the hammer shank (hammer accelerometer) which corresponded well to the physical onset of the sound, and conceptually with the "note on" command in the MIDI file.[12]

The **finger-key contact** was defined to be the moment when the key started to move. It was obtained by a simple threshold procedure applied on the key velocity track. In mathematical terms, it was the moment when the (slightly smoothed) key acceleration exceeded a certain threshold. Finding the correct finger-key point was not difficult for struck tones; they showed typically a very abrupt initial acceleration. However, automatically determining the right moment for soft pressed tones was more difficult and sometimes ambiguous. The threshold was optimized iteratively by hand. It was found that softer tones required a smaller threshold than louder ones; therefore it was coupled to the hammer velocity by a linear function. When the automatic procedure failed, it failed by several tens of milliseconds—an error easy to discover in explorative data plots.

The **key bottom contact** was the instant when the downwards travel of the key was stopped by the keybed. This point was defined as the maximum deceleration of the key (MDK). In some keystrokes, the MDK was not the actual keybed contact, but a rebound of the key after the first key bottom contact. For this reason, the time window of searching MDK was restricted to 7 ms before and 50 ms after hammer-string contact. The time window was iteratively modified depending on the maximum hammer velocity until the correct instant was found. The indicator MDK was especially clear and nonambiguous when the key was depressed in a range of medium intensity (see Fig. 1).

The **maximum hammer velocity** (MHV, in meters per second) was the maximum value in the hammer velocity track before hammer-string contact.

The **escapement point** was defined as being the instant after which the hammer travels freely (with no further acceleration) towards the strings. It was approximated by fitting a line onto the hammer velocity track between the point of MHV and hammer-string contact. The slope of this line was set to the theoretical deceleration caused by gravity

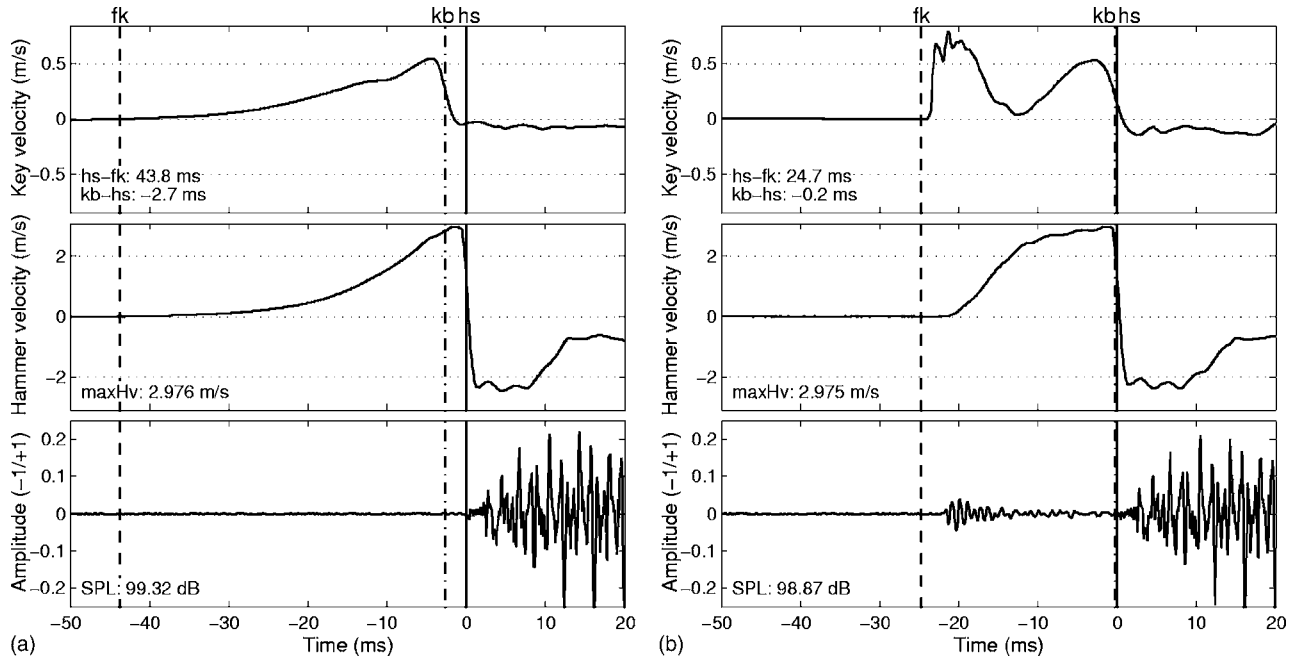


FIG. 1. Two *forte* keystrokes (C4, 60) played on the Yamaha grand piano with a pressed touch (left) and struck touch (right). The top panels show key velocity, the middle panel hammer velocity, and the bottom panel the amplitude of the sound signal. Both keystrokes exhibit similar MHVs and peak sound levels. Finger-key contact time ("fk"), hammer-string contact ("hs"), and key bottom contact ("kb") are indicated by vertical lines.

($-9.81\ \mathrm{m/s^2}$), disregarding any influence of friction. This instant in time was measurable only at soft and very soft touches. At MHVs exceeding approximately 1.5 m/s, it virtually coincided with the moment of MHV.

To inspect the recorded key and hammer velocity tracks and the sound signal, an interactive tool was created in order to display one keystroke at a time in three panels, one above the other. Screen shots of this tool are shown below (see Fig. 1). The data were checked and inspected for errors with the help of this tool.

IV. RESULTS AND DISCUSSION

In this section, measurement results of the three investigated pianos are presented and compared. Recall that these data apply to specific instruments and depend strongly on their regulation so that generalization to other instruments of the same brands may be problematic.

A. Two types of touch

The recorded three-channel data of two example keystrokes performed on the Yamaha are plotted in Fig. 1. They both exhibit an almost identical MHV (2.976 and 2.975 m/s, respectively) and a similar peak sound level (99.32 and 98.87 dB, respectively).[13] In musical terms this corresponds roughly to a *forte* dynamic. The first keystroke [Fig. 1(a)] was played from the key with a "pressed touch." From the beginning of the keystroke, the key velocity increases gradually; the hammer velocity grows in parallel. The hammer reaches its MHV immediately before it arrives at the strings. Hammer-string contact is characterized by a very sudden deceleration (Fig. 1, indicated by vertical solid lines). Key bottom contact shows a slightly less abrupt deceleration and occurs immediately before hammer-string contact. On the other hand, the keystroke produced with a struck touch [Fig. 1(b)] shows a very sudden jerk at the beginning of the key movement that has no correspondence in the hammer movement, but can be seen in the audio data ("touch precursor").[14] The hammer starts its travel to the strings with a delay of several milliseconds. It receives a first, larger acceleration by this initial blow applied to the key; later the key "catches up" (Ortmann, 1925, p. 23) and brings the hammer to its final speed. The whole striking procedure needs roughly 20 ms less time with a struck touch compared to the pressed touch, both with almost identical intensities.

B. Relation between key and hammer movement

In order to demonstrate the behavior of the hammer in relation to the key movement under two touch conditions, "touch trajectories" of pressed and struck keystrokes are plotted in Figs. 2–4. These plots depict the progression of hammer velocity against key velocity from finger-key contact to key bottom or hammer-string contact (depending on which of these two points was later). Each panel compares keystrokes with almost identical MHV values played at the C5 on all three pianos. Marked on the trajectories are escapement point (upward triangle), hammer-string contact (diamond), and key bottom contact (downward triangle), as well as elapsing time (filled circles every 2 ms). Figure 2 contains *mezzo-piano* keystrokes, Fig. 3 *forte*, and Fig. 4 *fortissimo*, which can only be achieved with a struck touch.

The pressed keystrokes develop fairly linearly until the escapement point (top panels in Figs. 2 and 3). The average slope of this part of the trajectory for all recorded pressed tones is 5.6 for Steinway and Yamaha, and 5.3 for the Bösendorfer. In the softer pressed example (Fig. 2 top), the key-bottom occurs after hammer-string contact (the trajectories

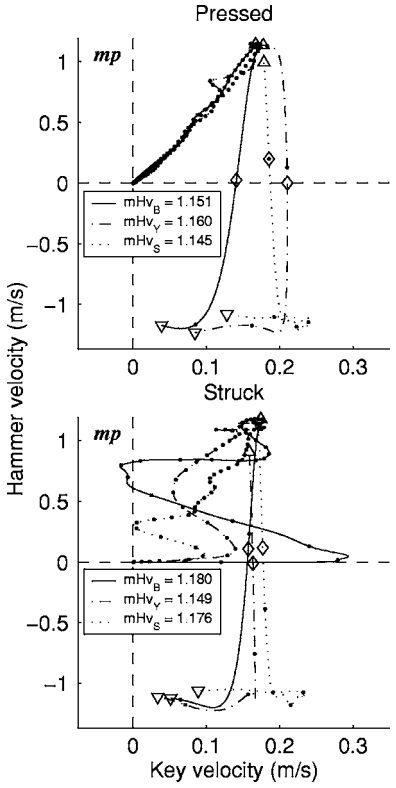


FIG. 2. Touch trajectories of *mezzo-piano* keystrokes at the C5 (72) on three pianos played with pressed touch (upper panel) and struck touch (lower). Upward-pointing triangles denote escapement point, diamonds hammer-string contact, and downward-pointing triangles key bottom contact times. Small filled circles are plotted on the trajectories every 2 ms to indicate time.

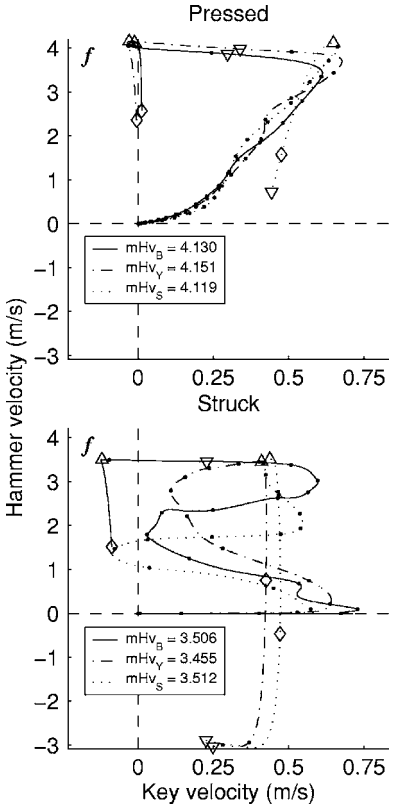


FIG. 3. Touch trajectories of *forte* keystrokes at the C5 (72) on three pianos played with pressed touch (upper panel) and struck touch (lower). Symbols and axes proportions as in Fig. 2.

drop at the right before going leftwards), while at the *forte* example (Fig. 3 top) the key-bottom is before hammer-string (trajectory moves left before dropping downwards). The exception here is the keystroke at the Steinway, at which the key-bottom contact is still after hammer-string contact (though the time difference is negligible) and therefore the trajectory still drops first.

Struck tones show very different trajectories (bottom panels in Figs. 2–4). They deviate clearly from the diagonal; the initial acceleration of the key pushes the trajectories rightwards, before the hammer starts to move. After this first blow, the key stops for a moment and reaccelerates again, while the hammer still gets faster. This pattern is quite consistent across pianos and intensities (see Figs. 2 and 3). However, at very loud keystrokes the second acceleration of the key does not occur anymore (Fig. 4) so that the whole keystroke consists of one strong impulse and the acceleration of the hammer during retardation of the key (diagonal trajectory to the upper right). The exception is the Steinway which still exhibits a second key acceleration phase.

The struck touches compared here display almost identical MHVs (bottom panels Figs. 2–4). However, the effort spent for the keystrokes does not appear to be similarly identical: the initial blow (maximum key velocity) at the Bösendorfer is larger than at the other pianos for all three intensities; relatively the most in Fig. 2. A larger initial amplitude to the right denotes a larger energy loss in a keystroke due to compression of the parts in the action (e.g., cushions, dunnage) and bending of the key and hammer shank. Therefore, the Bösendorfer action exhibits the largest degree of compression (especially with soft tones) and the Steinway the least. To draw more profound conclusions from this, measurements on the touch form would have to be performed that include monitoring of finger speed and force applied to the key. However, the present data suggests that the Bösendorfer requires more playing effort at struck touches to achieve the same dynamic level than the two other pianos.

In order to quantify the transformation effectivity of a keystroke, the correlation coefficient between the key and hammer velocity track (starting from finger-key contact until

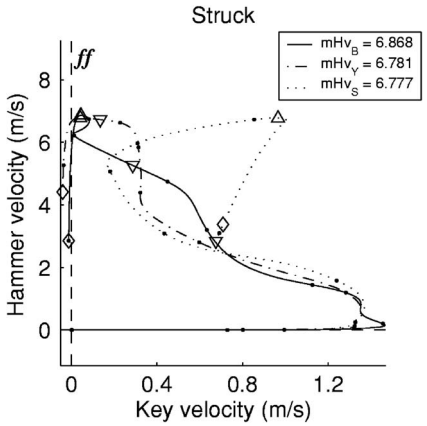


FIG. 4. Touch trajectories of *fortissimo* keystrokes at the C5 (72) on three pianos played with struck touch. Symbols and axes proportions as in Fig. 2.

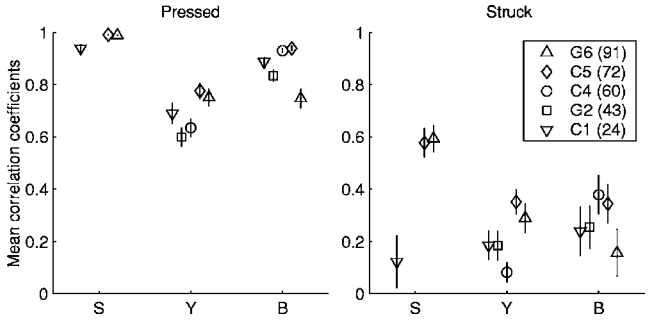


FIG. 5. Mean correlation coefficients of the touch trajectories (key and hammer velocity histories from finger-key through the escapement point) by touch (panels), piano (*x* axes: Steinway, Yamaha, Bösendorfer), and pitch (marker shape). Error bars denote 95% confidence intervals.

the escapement point) was introduced. The mean correlation coefficients for touch, pianos, and key are each plotted separately in Fig. 5. As this measure determines linearity between key and hammer movement, it may serve as a "touch index," distinguishing clearly between the two types of touch. All pressed touches display coefficients beyond approximately 0.6 and struck ones below that value. In this sense, pressed touch is a more effective way of transforming finger force into hammer velocity than playing with a struck touch.

Moreover, this index may also hold for a measure of tone control for the pianist. With a struck touch, the action decompresses after compression (relaxing of compressed cushions, bent key, and the hammer shank). At very loud keystrokes, this must be the reason of the high acceleration of the hammer (the key decelerates clearly while the hammer accelerates up to 7–8 m/s, see Fig. 3). Therefore by striking a key, the tone intensity is controlled through the initial key or finger velocity; by pressing a key, the tone intensity is controlled through the key or finger velocity until the escapement of the jack ("early versus late impulse," cf. Askenfelt and Jansson, 1991).

C. Travel time

The time interval between finger-key contact and hammer-string contact is defined here as the *travel time*.[15] The travel times of all recorded tones are plotted in Fig. 6 against MHV separately for the three grand pianos (different panels), different types of touch (filling of symbols), and different keys (denoted by symbol).

Some very basic observations can be drawn from this figure. The two pianists were able to produce much higher MHVs on all three pianos with a struck attack (almost 8 m/s), whereas with a pressed tone, the MHVs hardly exceeded 5 m/s. There was a small trend towards higher hammer velocities at higher pitches (due to smaller hammer mass, see Conklin, 1996). The highest velocities on the Yamaha and the Steinway were obtained at the G6, but at the middle C on the Bösendorfer. The lowest investigated key (C1) showed slightly lower MHVs by comparison to the fastest attacks (loudest attacks on the Steinway: C1: 6.8 m/s vs G6: 7.5 m/s, on the Yamaha: C1: 6.4 m/s vs G6: 7.8 m/s, and on the Bösendorfer: C1: 6.0 m/s vs G6: 6.6 m/s and C4: 7.6 m/s). The variability between the intensity distributions of the keys could be due to the fact that the tones were played by human performers.

The travel times ranged from 20 ms to around 200 ms (up to 230 ms on the Steinway) and depicted clearly different patterns for the two types of touch. The travel time curves were independent of pitch although hammer mass in the low register is greater (Conklin, 1996).

The data plotted in Fig. 6 were approximated by power curves of the form $tt = a \times HV^b$ separately for each type of touch ("pr," "st") and each of the three pianos. The results of these curve interpolations are shown in the legends of Fig. 6. Struck touch needed less time to transport the hammer to the strings than a pressed touch which smoothly accelerated the key (and thus the hammer). The travel times were more spread out when the tones were pressed, indicating that there was a more flexible control of touch in this way of actuating the keys (also reflected by the lower R^2 values of the curve fits). On the Steinway, the struck data showed higher variability, almost similar to the pressed data.

The present data were generally congruent with findings by Askenfelt and Jansson (1991) and Hayashi *et al.* (1999). The travel time approximations used in Goebl (2001, *tt*

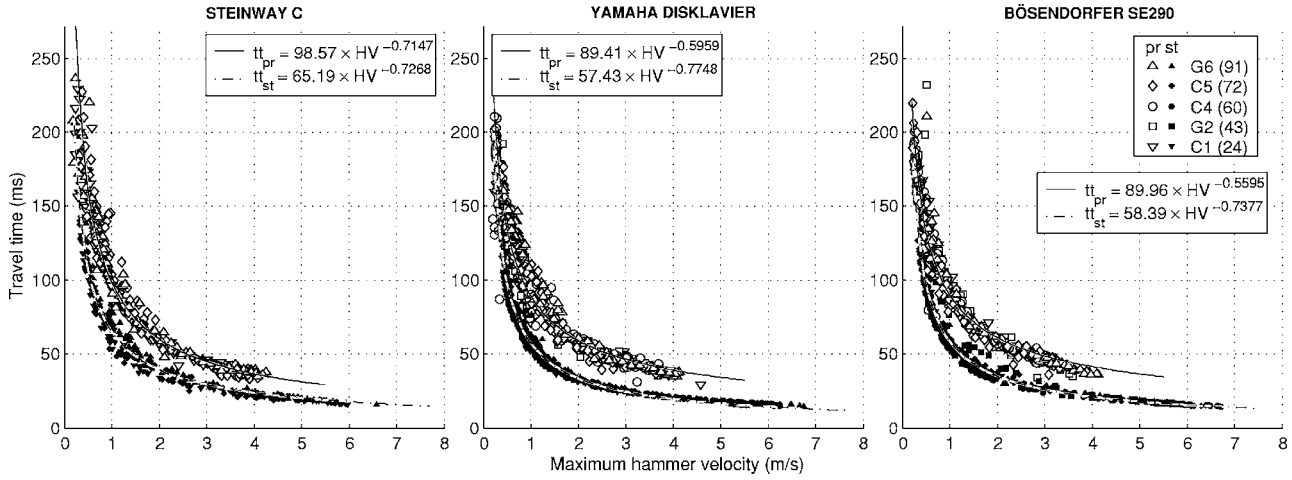


FIG. 6. **Travel time** (from finger-key to hammer-string contact) against MHV for the three grand pianos (three panels), different types of touch (*pressed* and *struck*), and different keys (from C1 to G6, see legend). The two types of touch ("pr" and "st") were approximated by power functions (see legends).

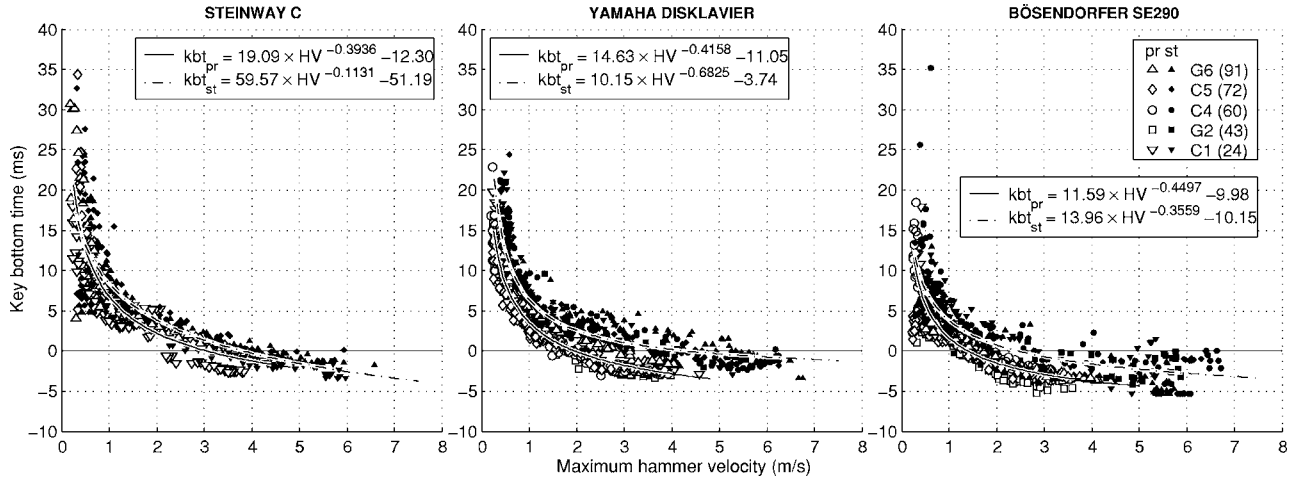


FIG. 7. **Key bottom time** relative to hammer-string contact against MHV. Negative key bottom values denote instants *preceding* hammer-string contact. Legends list power curve fits of the data separately for pressed touch ("kbt_{pr}") and struck touch ("kbt_{st}").

$=89.16\times HV^{-0.570}$) were very similar to the curve fit of the Bösendorfer's pressed data. The impact of this updated travel time function on the melody lead predictions was rather negligible; it is discussed elsewhere (Goebl, 2003, p. 74).

D. Key bottom time

Figure 7 displays the key bottom contact times relative to hammer-string contact ($t_{kb_{rel}}=t_{kb}-t_{hs}$). Negative values indicate key bottom contacts *before* hammer-string contact, positive values key bottom contacts *after* the hammer hits the strings (see overview display in Fig. 9). The keybed was reached by the key up to 35 ms *after* hammer-string contact in very soft tones (up to 39 ms at the Bösendorfer) and as early as 4 ms before in very strong keystrokes. This finding coincides with Askenfelt and Jansson's (1990a,b) results, but since much softer tones were measured in the present study (as low as 0.1 m/s), the key bottom times extended more after hammer-string contact.

Key bottom contact times varied with the type of touch. Keys played with a pressed touch tended to reach the keybed earlier than keys hit in a struck manner. This was especially evident for the Bösendorfer and for the Yamaha, but not for the Steinway. Askenfelt and Jansson (1992b, p. 345) stated that the interval between key bottom and hammer-string contact varies only marginally between legato and staccato touch. They obviously refer with this statement to an earlier study (Askenfelt and Jansson, 1990b), where the investigated grand piano was also a Steinway grand piano.[16]

Power functions were fitted to the data as depicted in Fig. 7, separately for the two types of touch and the different pianos (see legends). Since the data to fit contains also negative values on the y axis, power functions of the form $kbt=a\times HV^{b}+c$ were used. The data spread out more than in the travel time curves (reflected in smaller R^2 values) and showed considerable differences between types of touch, except for this Steinway, where touch did not divide the data visibly. This finding suggests that struck keystrokes tend to compress the parts of the action more than pressed ones that this behavior was least at the Steinway piano.

Askenfelt and Jansson (1990b) considered key bottom times as being sensed with the fingertips by pianists and thus as being important for the vibrotactile feedback in piano playing. Temporal asynchronies of the order of 30 ms are in principle beyond the temporal order threshold (Hirsh, 1959), so at very soft keystrokes key bottom contact and hammer-string contact could be perceived as two separate events by the pianists. But for the majority of keystrokes these time differences are not perceptually distinguishable; however, they may be perceived subconsciously and perhaps as part of the response behavior of a particular piano. Especially, the different key bottom behavior for the different kinds of touch might be judged by the pianists as part of the response behavior of the action (Askenfelt and Jansson, 1992b). Hammer-string contact occurs earlier relative to key bottom contact when the key was struck compared to when it was pressed. For a pianist, a struck touch produces a tone earlier than a pressed touch with comparable intensity, both relative to key bottom contact and relative to finger-key contact, and thus may be perceived as being louder and more direct.

E. Escapement point

Shortly before the hammer crown arrives at the strings, the tail end of the jack gets pushed away from under the roller by the escapement dolly and the pianist loses physical contact with and thus control over the hammer, which is then moving freely along a circular path to the strings. This measurement point was comparatively difficult to extract automatically from the data, since at many keystrokes this point was not obvious at all. The higher the hammer velocities, the more it tends to coincide with the instant of MHV. Only at soft and very soft dynamics, the hammer might reach its maximum speed considerably before escapement.

An example of a very soft keystroke is displayed in Fig. 8. In addition to the display in Fig. 1, the point of MHV ("Vmax") and the escapement point ("ep") are sketched, as well as the line of gravity fitted in the hammer track between escapement point and hammer-string contact. At this *pianissimo* tone, the hammer reaches its maximum speed quite soon after the begin of the keystroke (after 31.9 ms), travels

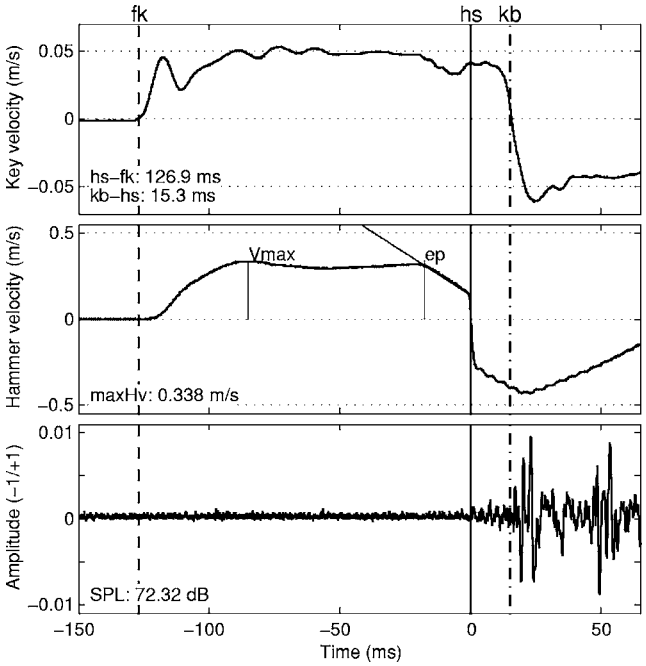


FIG. 8. A struck pianissimo keystroke at C2 (24) played on the Steinway grand piano. Additionally indicated are the point of MHV ("Vmax") and the estimated escapement point ("ep"). The skewed line in the middle panel denotes expected deceleration according to gravity.

another 78.7 ms with connection to the jack, and decelerates approximately by gravity for a period of 16.3 ms.

F. Free flight of the hammer

The time interval from the escapement point until hammer-string contact is called here "the free flight of the hammer."[17] The individual data points are not plotted here due to space limitations, but they were fitted by power curves separately for the type of touch. The formulas are provided in Table I. The free flight of the hammer ranges from almost zero at louder tones up to 20 ms at very soft keystrokes, with some outliers up to around 40 ms at the Yamaha piano (struck touch). Generally, pressed touches exhibit shorter free flight times than struck touches. This finding coincides with the earlier stated proposition that a pressed touch provides generally a better control over the tones than a struck touch. Of all three actions, the Steinway action showed the shortest free flight times (see Fig. 9).

G. Comparison among tested pianos

In Fig. 9, all power curve approximations reported above (Figs. 6 and 7, and Table I) are plotted in a single display, separately for the type of touch (panels) and the three tested piano actions (line style) against time (in seconds) relative to the hammer-string contact. The temporal differences between extremes in intensity were largest for the finger-key times and smallest for key bottom times (both relative to hammer-string contact). The differences of the curves between the pianos by different manufacturers were small compared to the differences introduced through the type of touch. The finger-key curve of this Steinway action was the left-most except for loud pressed tones. Also our

TABLE I. Power curve approximations of the form $fft = a \times HV^b$ for the free-flight time data, separately for type of touch and piano.

fft=	Pressed	Struck
Steinway	$1.63 \times HV^{-1.403}$	$3.04 \times HV^{-1.581}$
Yamaha	$2.78 \times HV^{-1.266}$	$5.27 \times HV^{-1.384}$
Bösendorfer	$3.21 \times HV^{-1.353}$	$4.32 \times HV^{-1.404}$

Steinway's key bottom curve was the right-most of the three actions. Thus, the Steinway action needed more time for the attack process than the other two pianos, except for very loud pressed tones. At the free flight time approximation, the Steinway showed the shortest of all tested actions.

These data apply only to the tested instruments and temporal behavior changes considerably with regulation (especially key-bottom contact and the time of free flight, see Askenfelt and Jansson, 1990b; Dietz, 1968). We do not know how different the temporal behavior of other instruments of these three manufacturers will be. Changes in regulation (hammer-string distance, let-off distance) resulted in changes of the key-bottom timing and the time interval of the hammer's free flight, respectively, of up to 5 ms (for a medium intensity, see Askenfelt and Jansson, 1990b, pp. 56–57). The differences between piano actions in the present data are approximately of the same order.[18]

It can be concluded that the temporal behavior of the tested piano actions by different manufacturers were similar. However, no definitive conclusions can be drawn whether or not these (comparably small) differences in temporal behavior were crucial for the pianist's estimation of the piano's quality and whether they apply also to other instruments of these manufacturers.

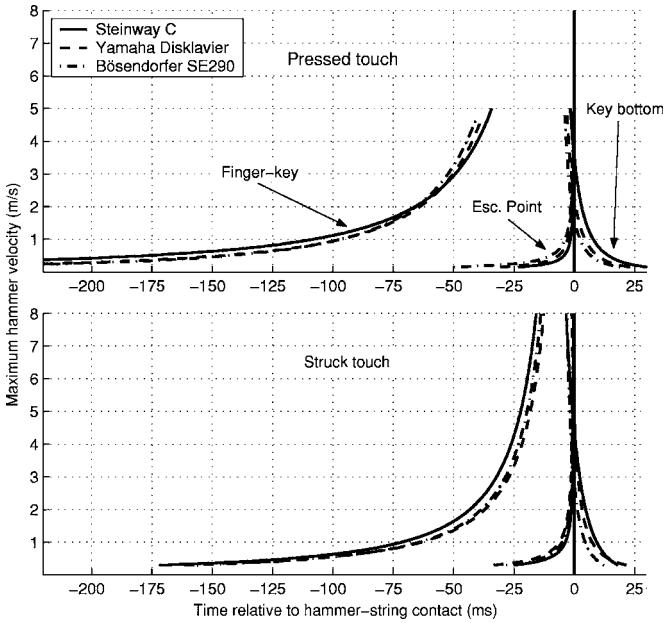


FIG. 9. Temporal properties of three grand piano actions. Power curve approximations for the three pianos (line style), the two types of touch (panels), and for finger-key (left), escapement point (middle), and key bottom times (right) relative to hammer-string contact.

V. GENERAL DISCUSSION

This study provides benchmark data on the temporal properties of three different grand pianos under two touch conditions (*pressed* and *struck touch*). Prototypical functions were obtained for travel time, key bottom time, and the time of the hammer's free flight by fitting power curves to measured data. The temporal properties varied considerably between type of touch, only marginally between pianos, and not at all between the different tested keys. The latter was not surprising, since piano technicians generally aim to adjust a grand piano action so that all keys show similar and consistent behavior over the whole range of the keyboard.

Different kinds of actuating the keys produced different ranges of hammer velocity. Very soft tones could only be achieved with a pressed touch (minimum 0.18 m/s or 50.0 dB-pSPL) and the extremely loud attacks only with a struck touch (maximum 6.8 m/s or 110.4 dB-pSPL). Playing from the keys (pressed) did not allow MHVs beyond around 4 m/s, thus for some very loud intensities hitting the keys from above was the only possible means. The free flight times were shorter for pressed touch than for struck touch which suggests a better tone control for the pianist when playing with pressed touch. Moreover, depressing a key caused less touch noise than striking a key which is commonly regarded as a desired aesthetic target in piano playing and teaching (cf., e.g., Gát, 1965).

The two types of touch (in the present terminology pressed and struck touch) do represent two poles of a variety of possible ways to actuate a piano key (i.e., late acceleration versus early, hesitating in between, or accelerating directly at the escapement point). It must be assumed that a professional pianist will (even unconsciously) be able to produce many different shades of touch between pressed and struck.

The travel times and the key bottom times changed considerably with intensity of key depression. A soft tone may take over 200 ms longer from the first actuation by the pianist's finger to sound production compared to a very sudden *fortissimo* attack. Moreover, travel times and key bottom times changed considerably with touch. A struck tone needed around 30–40 ms less from finger-key to hammer-string than a pressed tone with a similar MHV. These findings were not surprising (since they follow a very basic physical law), but the performing artist has to anticipate these changes in temporal behavior while playing in order to achieve the desired expressive timing of the played tones. The pianist not only has to estimate before playing a tone, how long the keystroke will take for what desired dynamic level, but also for what intended way of actuating the key. These complex temporal interactions between touch, intensity and the tone onset are dealt with and applied by the pianist unconsciously; they are established over years of intensive practising and extensive self-listening. Immediately, musical situations come to mind in which loud chords tend to come early with pianists at beginning or intermediate level; or that *crescendo* passages tend to accelerate in tempo as well, because each keystroke is performed with a harder blow and thus quicker in order to achieve the *crescendo*, but the time intervals between finger activity were not correspondingly increased.

A keystroke starts for the pianist kinesthetically with finger-key contact (the acceleration impulse by the finger)[19] and ends at key bottom, but it starts aurally for pianist and audience at (or immediately after) hammer-string contact. Typical intensities (at an intermediate level) in expressive piano performances (i.e., as measured in Goebl, 2001) fall between 40 and 60 MIDI velocity units (0.7–1.25 m/s) and thus typical travel times are between 80 and 108 ms, thus varying as much as about 30 ms. At such keystrokes, the key bottom times are between 3.5 and 0.5 ms before hammer-string contact, thus a range of the order of 3 ms. It can be assumed that with such moderate intensity levels (and a default touch which is likely to be pressed rather than struck), the changes in travel times due to varying intensity might not be directly relevant for the player since they are at the threshold of perceivability. Nevertheless, they are sufficiently large to produce the typical *melody lead* (Goebl, 2001).

At that typical dynamic range, key bottom times are even more unlikely to be perceived by the pianist separately from the sound (hammer-string), since those temporal differences are there of the order of a few milliseconds. However, the differences between key bottom and hammer-string can be up to 40 ms in extreme cases which is of the order of or just beyond *just noticeable differences* for perceiving two separate events (Askenfelt and Jansson, 1992b, p. 345). Also as Fig. 9 made visually evident, the travel times were far larger than the time differences of the other readings (escapement point, hammer-string contact, key bottom contact), so it can be assumed that the pianist (especially in the dynamic middle range) only senses two points in time: the start of the keystroke (finger-key) and its end which coincides with the beginning of the sound.

Conceptually, the key bottom contact has to be *after* hammer-string contact. If it were the other way round, no soft tones could be played at all. The fact that key bottom contact moves towards and beyond (that is *before*) hammer-string contact with increasing hammer velocity (cf. Fig. 7) was due to the bending of the hammer shank and the compression of various parts in the action (i.e., cushions, dunnage) and a later unbending and decompression of those. The three actions showed different behavior as to when the key bottom line crossed the hammer-string line with changing hammer velocity (Fig. 9). According to the power curve approximations of the data (Fig. 7), the Bösendorfer's crossed at 1.4 m/s (pr) and 2.5 m/s (st), the Yamaha's at 2 m/s (pr) and 4.3 m/s (st), and the Steinway's at 3 m/s (pr) and 3.8 m/s (st). If we considered these values to be a measure of compressivity of the action, the Steinway would have the least compressive action (of the three), and the Bösendorfer the most for pressed touches. At struck touches, the Yamaha showed the least compressive behavior. A smaller compression behavior might be considered a criterion for the subjective quality of a piano action (see discussion further below). However, further investigation would be necessary to verify this hypothesis (e.g., measuring static compression behavior of the investigated keys).

Furthermore, sensomotoric feedback is considered an utmost important factor for pianists not only for judging the action's response, but also to judge the piano's tone (Gale-

mbo, 1982, 2001). In an extended perception experiment, Galembo (1982) asked a dozen professors from the Leningrad Conservatory of Music to rate the instrumental quality of three grand pianos under different conditions. The participants agreed that the Hamburg Steinway grand piano was superior, followed by the Bechstein grand piano, while the lowest quality judgment received a grand piano from the Leningrad piano factory. In different discrimination tasks, the participants were not able to distinguish between the instruments (although all indicated to be able to) only by listening to them when played by some other person behind a curtain. But they could very well discriminate between instruments when they played on them blindly or deaf-blindly (Galembo, 1982, 2001). This study implied that the haptosensorial feedback of the piano action to the playing pianist is crucial for the estimation of the instrumental quality.

Another factor altering the haptosensorial feedback of the pianist is the room acoustics (Bolzinger, 1995; Galembo, 1987). A piano action might feel easily to handle in a room with reverberant acoustic, while the same action feels intractable and tiring in a room without any reverberation. Similarly, the timbre of that instrument might be judged differently with changing room acoustics. A pianist is usually not able to separate the influences of room acoustics from properties of the instrument and directly attributes room acoustics to instrumental properties (Galembo, 1987, 2001).

The reported temporal properties of the piano actions were derived from isolated piano tones (without pedal) such as they rarely occur in piano performances. For a new keystroke, the key does not necessarily have to come back to its resting position, but, due to the double repeating feature of modern grand piano actions, the hammer is captured by the check and the repetition lever stopped by the drop screw (Askenfelt and Jansson, 1990b). When the key is released approximately half way (of the approximately 10 mm touch depth), the jack is able to resile back underneath the roller and another keystroke can be performed. This occurs usually some 2–4 mm below the key surface. For such keystrokes, the key can travel only 6–8 mm, so the travel times can be expected to be shorter than with a pressed touch from the key's resting position. Also for such repeated keystrokes, it would be impossible to calculate or to determine a finger-key contact point in time. The study of such repeated keystrokes has to remain for future investigation.

An interesting issue with respect to the reported data is whether there is a relationship between the actions' temporal properties and the instrumental quality of the tested grand pianos. The authors' personal opinion as pianists was that from the three investigated grand pianos in this study the Steinway grand piano was qualitatively superior to the other two (in terms of the actions' responsiveness), although the Bösendorfer was a high-standard concert grand piano as well. The small Yamaha baby grand was the least interesting instrument also due to its size. However, all pianos were on a mechanically high standard and they were well maintained and tuned. It is assumed here that one of the most important features of a "good" piano is a precise and responsive action.

In the data reported earlier, some differences between the pianos could be observed that might influence the subjective judgment of instrumental quality and that support the authors' subjective preference for the Steinway. The Steinway showed (1) less difference in key bottom times due to touch than the other two pianos; (2) late crossing of the key bottom approximations and the hammer-string contact line, indicating a low compressivity of the parts of the action; and (3) shorter time intervals of free flight (almost zero at keystrokes with a MHV of more than 1.5 m/s, while for the Bösendorfer it was around 2.5 m/s, for the Yamaha above 3 m/s). Moreover, the Yamaha showed many very early hammer velocity maxima at velocities between about 1 and 2 m/s, the Bösendorfer some, the Steinway almost none.

Although further evaluative investigations would be required to be able to state more conclusively any hypotheses on the relation of temporal behavior of grand piano actions and instrumental quality, it seems likely that a constant behavior over type of touch and late hammer velocity maxima are crucial for precise touch control and a subjective positive appreciation of instrumental quality.

ACKNOWLEDGMENTS

This research was supported by a START Research Prize of the Austrian Science Fund (FWF Project No. Y99-INF), by the Viennese Science and Technology Fund (WWTF Project CI010), and by the European Union [Marie Curie Fellowship, HPMT-GH-00-00119-02, the *Sounding Object* project (SOb), IST-2000-25287, and the MOSART IHP network, HPRN-CT-2000-00115]. The Wenner-Gren Foundation provided a visiting professorship grant to the third author during 2001/02. The OFAI acknowledges support from BMBWK and BMVIT. Thanks to Anders Askenfelt, Erik Jansson, Simon Dixon, Friedrich Lachnit and the Bösendorfer company, Tore Persson, Alf Gabrielsson, and two anonymous reviewers for essential help during the experiments and valuable comments on earlier versions of this manuscript.

[1]The term "modern grand piano" refers to what is nowadays commonly used by pianists in concert halls. However, it is not modern anymore, because, e.g., the Steinway model D grand was introduced already in the second half of the 19th century and has remained essentially unchanged since then.

[2]The first three paragraphs are taken from Goebl (2001) and repeated here for the sake of completeness.

[3]Askenfelt and Jansson (1990b) measured the C4 on a Hamburg Steinway & Sons grand piano, model B (211 cm).

[4]The "prelay function" compensates for the different travel times of the action at different hammer velocities. In order to prevent timing distortions in reproduction, the MIDI input is delayed by 500 ms. The solenoids (the linear motors moving the keys) are then activated earlier for softer notes than for louder notes, according to a preprogrammed function.

[5]This particular piano was used in Askenfelt and Jansson (1992a).

[6]Brüel & Kjær accelerometer type 4393 (2.4 g).

[7]Brüel & Kjær ENDEVCO accelerometer model 22 (0.14 g).

[8]Brüel & Kjær sound level calibrator type 4230.

[9]In order to account for differences in radius between the accelerometer placement on the hammer shank and striking point at the hammer crown, the hammer velocity data was corrected for that (resulting in values increased by 14%). This correction was not applied in Goebl and Bresin (2003).

[10]Only three keys were tested at the Steinway piano (C1, C5, G6).

[11]The recordings were done between May 2001 and January 2002.

[12]This measurement was also used to find the individual attacks in a recorded file. All accelerations below a certain value were taken as onsets.

The very rare silent attacks were not captured with this procedure, as well as some very soft attacks.

[13]The peak sound level was calculated by taking the maximum of the sound energy (maximum root-mean-square with a 10 ms sliding window).

[14]The intensity of the touch precursor depends strongly on the way a particular keystroke was played. It is possible to produce loud struck tones without clearly visible touch precursors (see also Goebl *et al.*, 2004).

[15]This terminology might be misleading, because "time" refers to a point in time, although in this case a time duration is meant. Terms like "travel time" or "time of free flight" were used according to the term "rise time" that is commonly used in acoustic literature (see, e.g., Truax, 1978).

[16]Askenfelt and Jansson (1990b) used a Steinway model B, serial number 443001, built in Hamburg 1975.

[17]In Goebl *et al.* (2003) the time interval between the points of MHV and hammer-string contacts are plotted under the label "free flight of the hammer." We consider the present data display (escapement through hammer-string contact) to be more appropriate.

[18]Note that all three pianos were maintained and regulated by professional technicians before the measurement so that all pianos were in concert condition before the tests.

[19]Certainly the performing pianist may in some cases hear the finger-key noise as well.

Askenfelt, A. (**1994**). "Observations on the transient components of the piano tone," in *Proceedings of the Stockholm Music Acoustics Conference (SMAC'93), July 28–August 1, 1993*, edited by A. Friberg, J. Iwarsson, E. V. Jansson, and J. Sundberg (Royal Swedish Academy of Music, Stockholm), Vol. **79**, pp. 297–301.

Askenfelt, A., and Jansson, E. V. (**1990a**). "From touch to string vibrations," in *Five Lectures on the Acoustics of the Piano*, edited by A. Askenfelt (Royal Swedish Academy of Music, Stockholm), Vol. **64**, pp. 39–57.

Askenfelt, A., and Jansson, E. V. (**1990b**). "From touch to string vibrations. I. Timing in the grand piano action," J. Acoust. Soc. Am. **88**, 52–63.

Askenfelt, A., and Jansson, E. V. (**1991**). "From touch to string vibrations. II. The motion of the key and hammer," J. Acoust. Soc. Am. **90**, 2383–2393.

Askenfelt, A., and Jansson, E. V. (**1992a**). "From touch to string vibrations. III. String motion and spectra," J. Acoust. Soc. Am. **93**, 2181–2196.

Askenfelt, A., and Jansson, E. V. (**1992b**). "On vibration and finger touch in stringed instrument playing," Music Percept. **9**, 311–350.

Báron, J. G. (**1958**). "Physical basis of piano touch," J. Acoust. Soc. Am. **30**, 151–152.

Báron, J. G., and Holló, J. (**1935**). "Kann die Klangfarbe des Klaviers durch die Art des Anschlages beeinflußt werden?" Z. Sinnesphysiologie **66**, 23–32.

Bolzinger, S. (**1995**). "Contribution a l'étude de la rétroaction dans la pratique musicale par l'analyse de l'influence des variations d'acoustique de la salle sur le jeu du pianiste," Doctoral thesis, Institut de Mécanique de Marseille, Université Aix-Marseille II, Marseille (unpublished).

Bork, I., Marshall, H., and Meyer, J. (**1995**). "Zur Abstrahlung des Anschlaggeräusches beim Flügel," Acustica **81**, 300–308.

Bryan, G. H. (**1913**). "Pianoforte touch," Nature (London) **91**, 246–248.

Cadoz, C., Lisowski, L., and Florens, J.-L. (**1990**). "A modular feedback keyboard design," Comput. Music J. **14**, 47–51.

Chaigne, A., and Askenfelt, A. (**1994a**). "Numerical simulations of piano strings. I: A physical model for a struck string using finite differences methods," J. Acoust. Soc. Am. **95**, 1112–1118.

Chaigne, A., and Askenfelt, A. (**1994b**). "Numerical simulations of piano strings. II: Comparisons with measurements and systematic exploration of some hammer-string parameters," J. Acoust. Soc. Am. **95**, 1631–1640.

Cochran, M. (**1931**). "Insensitiveness to tone quality," Austral. J. Psychol. **9**, 131–134.

Conklin, H. A. (**1996**). "Design and tone in the mechanoacoustic piano. Part I. Piano hammers and tonal effects," J. Acoust. Soc. Am. **99**, 3286–3296.

Dietz, F. R. (**1968**). *Steinway Regulation. Das Regulieren von Flügeln bei Steinway* (Verlag Das Musikinstrument, Frankfurt am Main).

Fletcher, N. H., and Rossing, T. D. (**1998**). *The Physics of Musical Instruments*, 2nd ed. (Springer, New York,).

Galembo, A. (**1982**). "Quality evaluation of musical instruments (in Russian)," Tech. Aesthetics **5**, 16–17.

Galembo, A. (**1987**). *The Quality of Piano Tones (in Russian)* (Legkaya Industria, Moscow).

Galembo, A. (**2001**). "Perception of musical instrument by performer and listener (with application to the piano)," in *Proceedings of the International Workshop on Human Supervision and Control in Engineering and Music, September 21–24, 2001* (University of Kassel Press, Kassel, Germany), pp. 257–266; http://www.engineeringandmusic.de.

Gát, J. (**1965**). *The Technique of Piano Playing* 3rd ed. (Corvina, Budapest).

Gillespie, B. (**1994**). "The virtual piano action: Design and implementation," in *Proceedings of the 1994 International Computer Music Conference, Århus, Denmark* (International Computer Music Association, San Francisco), pp. 167–170.

Goebl, W. (**2001**). "Melody lead in piano performance: Expressive device or artifact?" J. Acoust. Soc. Am. **110**, 563–572.

Goebl, W. (**2003**). "The role of timing and intensity in the production and perception of melody in expressive piano performance," Doctoral thesis, Institut für Musikwissenschaft, Karl-Franzens-Universität Graz, Graz, Austria, available online at http://www.oefai.at/music.

Goebl, W., and Bresin, R. (**2003**). "Measurement and reproduction accuracy of computer-controlled grand pianos," J. Acoust. Soc. Am. **114**, 2273–2283.

Goebl, W., Bresin, R., and Galembo, A. (**2003**). "The piano action as the performer's interface: Timing properties, dynamic behaviour, and the performer's possibilities," in *Proceedings of the Stockholm Music Acoustics Conference (SMAC'03), August 6–9, 2003*, edited by R. Bresin (Department of Speech, Music, and Hearing, Royal Institute of Technology, Stockholm, Sweden), Vol. 1, pp. 159–162.

Goebl, W., Bresin, R., and Galembo, A. (**2004**). "Once again: The perception of piano touch and tone. Can touch audibly change piano sound independently of intensity?" in *Proceedings of the International Symposium on Musical Acoustics (ISMA'04)* (The Acoustical Society of Japan, Nara, Japan), pp. 332–335, CD-ROM.

Hart, H. C., Fuller, M. W., and Lusby, W. S. (**1934**). "A precision study of piano touch and tone," J. Acoust. Soc. Am. **6**, 80-94.

Hayashi, E., Yamane, M., and Mori, H. (**1999**). "Behavior of piano-action in a grand piano. I. Analysis of the motion of the hammer prior to string contact," J. Acoust. Soc. Am. **105**, 3534–3544.

Hirsh, I. J. (**1959**). "Auditory perception of temporal order," J. Acoust. Soc. Am. **31**, 759–767.

Koornhof, G. W., and van der Walt, A. J. (**1994**). "The influence of touch on piano sound," in *Proceedings of the Stockholm Music Acoustics Conference (SMAC'93), July 28–August 1, 1993*, edited by A. Friberg, J. Iwarsson, E. V. Jansson, and J. Sundberg (Publications issued by the Royal Swedish Academy of Music, Stockholm), Vol. 79, pp. 302–308.

Ortmann, O. (**1925**). *The Physical Basis of Piano Touch and Tone* (Kegan Paul, Trench, Trubner; J. Curwen; E. P. Dutton, London).

Ortmann, O. (**1929**). *The Physiological Mechanics of Piano Technique* (Kegan Paul, Trench, Trubner, E. P. Dutton, London), Paperback reprint: E. P. Dutton, New York, 1962.

Podlesak, M., and Lee, A. R. (**1988**). "Dispersion of waves in piano strings," J. Acoust. Soc. Am. **83**, 305–317.

Repp, B. H. (**1996**). "Patterns of note onset asynchronies in expressive piano performance," J. Acoust. Soc. Am. **100**, 3917–3932.

Seashore, C. E. (**1937**). "Piano touch," Sci. Mon. **45**, 360–365.

Suzuki, H. (**2003**). "Analysis of piano tones with soft and hard touches," in *Proceedings of the Stockholm Music Acoustics Conference (SMAC'03), August 6–9, 2003*, edited by R. Bresin (Department of Speech, Music, and Hearing, Royal Institute of Technology, Stockholm, Sweden), Vol. 1, pp. 179–182.

Traux, B. (**1978**). *Handbook for Acoustic Ecology, World Soundscape Project*, Vol. **5** (A.R.C. Vancouver, BC).

Van den Berghe, G., De Moor, B., and Minten, W. (**1995**). "Modeling a grand piano key action," Comput. Music J. **19**, 15–22.

White, W. B. (**1930**). "The human element in piano tone production," J. Acoust. Soc. Am. **1**, 357–367.

Optical and tomographic imaging of a middle ear malformation in the bullfrog (*Rana catesbeiana*)

Seth S. Horowitz and Andrea Megela Simmons[a)]
Departments of Psychology and Neuroscience, Brown University, Providence, Rhode Island 02912

Darlene R. Ketten
Biology Department, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543
and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02114

(Received 23 October 2004; revised 1 May 2005; accepted 10 May 2005)

Using a combination of *in vivo* computerized tomography and histological staining, a middle ear anomaly in two wild-caught American bullfrogs (*Rana catesbeiana*) is characterized. In these animals, the tympanic membrane, extrastapes, and pars media (shaft) of the stapes are absent on one side of the head, with the other side exhibiting normal morphology. The pars interna (footplate) of the stapes and the operculum are present in their normal positions at the entrance of the otic capsule on both the affected and unaffected sides. The pattern of deformity suggests a partial failure of development of tympanic pathway tissues, but with a preservation of the opercularis pathway. While a definitive proximate cause of the condition could not be determined, the anomalies show similarities to developmental defects in mammalian middle ear formation. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1944627]

PACS number(s): 43.80.Lb [WA] Pages: 1166–1171

I. INTRODUCTION

The unique absorptive characteristics of amphibian skin and egg membranes, combined with their necessary restriction to wetland environments, has made amphibians unique biomarkers for effects of pollutants and parasites on species survival. Occurrences of malformations and mutations in wild amphibian populations include those affecting the limbs (missing, reduced, or multiple), the head (abnormalities in head or jaw shape, missing or reduced eyes), the spine, and the skin (Meteyer, 2000). The appearance of some of these skeletal and morphological abnormalities, particularly of the limbs, is correlated with exposure to a variety of potential mutagens, including pesticides, pathogens, and increased ultraviolet light, particularly in the larval (tadpole) period (Sessions and Ruth, 1990; Blaustein *et al.*, 1997; Johnson *et al.*, 1999; Sessions *et al.*, 1999). Other abnormalities may be genetic in origin, with no known environmental link.

The auditory system of adult anuran amphibians shares some structural and functional similarities with that of other vertebrates. Two peripheral transduction pathways have been identified in anurans. In the tympanic pathway, vibrations of the external tympanic membrane are transmitted to the inner ear via a cartilaginous extrastapes and bony stapes [also termed the extracolumella and columella (Wever, 1985)]. The stapes is believed to be homologous to that of other vertebrates (Lombard and Bolt, 1988). Although the homology of the extrastapes is unknown (Jaslow *et al.*, 1988), the operation of the stapes and extrastapes is functionally analogous to that of the mammalian ossicular chain (Møller, 1963; Mason and Narins, 2002a). The anuran's middle ear apparatus also includes a cartilaginous operculum attached to the caudal end of the oval window and to the shoulder girdle via the opercularis muscle (Wever, 1985). The operculum and opercularis muscle together make up the opercularis system, the function of which has been debated (Lombard and Straughan, 1974; Wever, 1985; Hetherington, 1994; Mason and Narins, 2002b), but which may be involved in the detection of low-frequency seismic and acoustic signals. In the bullfrog (*Rana catesbeiana*), the opercularis pathway and the tympanic pathway mature at different time points during larval and early postmetamorphic development (Hetherington, 1987; Boatright-Horowitz and Simmons, 1995; Horowitz *et al.*, 2001). The different developmental trajectories of these two pathways may allow specific genetic or environmentally produced abnormalities of one system at specific points in time, leaving the other intact.

Here, we describe an unusual cranial deformity in one adult male and one subadult female American bullfrog collected from the same site in which the left tympanic membrane, tympanic annulus, extrastapes, and pars media of the stapes are missing, but with the corresponding structures on the right appearing normal. Abnormalities appear to be confined to structural elements of the middle ear, with seemingly normal gross features of the skull and otic capsule. The nature of the abnormalities suggests the viability of an amphibian model of developmental hearing disorders, including what the clinical literature refers to in humans as "congenital aural atresia" (Schuknecht, 1989).

II. METHODS

A. Specimen collection

The first specimen was an adult male bullfrog captured from a small pond in Rhode Island that was host to a modest but healthy bullfrog chorus (typically 6–8 calling males over

[a)]Electronic mail: andrea_simmons@brown.edu

 0001-4966/2005/118(2)/1166/6/$22.50

the course of a season). This particular site was visited for approximately 10 years prior to the capture of the animal as part of a series of field studies on bullfrog chorusing behavior. No other external abnormalities were observed in any other vocalizing males during this period of time, but the pond was not explicitly censused for this purpose. The animal was observed first in early June during playback experiments. He approached a loudspeaker broadcasting male bullfrog advertisement calls, but remained silent although stationed near the loudspeaker for the duration of the playback (approximately 30 min), a behavior not uncommon in the field. The animal was captured by hand and transported to the laboratory where he was measured, photographed, and housed in a 75 l plastic terrarium filled with sterile soil and an internal 4 l water pool. He was fed live adult crickets *ad libitum*. The second animal was collected from the same site 2 years later as a tadpole in metamorphic climax [stage 43 (Gosner, 1960)]. No external abnormalities were noted at the time of collection. The animal was placed in a tadpole colony tank consisting of an aerated 75 l plastic aquarium and fed salt-free cooked spinach *ad libitum*. Upon entering the terminal stages of metamorphic climax, it was transferred to a terrarium similar to that described for the earlier adult. It was not until after completion of climax and some subsequent development to the size of a young subadult (Boatright-Horowitz and Simmons, 1995) that the abnormality of the left tympanic region was noted. Capture and subsequent handling of the animals conformed with relevant state and federal regulations, and research protocols were approved by the Brown University Institutional Animal Care and Use Committee.

B. Computerized tomographic scanning

In order to characterize the extent of the internal deformities *in vivo*, the animals underwent computerized tomographic (CT scan) imaging at the Radiology Department of Massachusetts Eye and Ear Hospital (male adult) and at the Imaging Center of the Woods Hole Oceanographic Institution (both animals). A series of registered CT images was obtained of the entire animal to determine if any abnormalities extended beyond the visible external abnormalities of the head region. The frogs were lightly anesthetized by submersion in 0.6% tricaine methanesulfonate (MS-222; Sigma) for 20 min and scanned with a Siemens Plus 4 CT unit. Scan images of the adult male were obtained originally using a 1 mm spiral protocol with 1 mm table increments. These images were formatted in 0.3 and 1 mm slice thicknesses in the transaxial plane in both bone and soft tissue kernels. Postscan reformats in coronal and sagittal planes at 0.1 mm were also produced on a Siemens Volume Zoom at the Imaging Center of the Woods Hole Oceanographic Institution. Scan images of the female were obtained in 0.5 and 2 mm thicknesses. Three-dimensional reconstructions of soft and bony elements were obtained using Siemens proprietary software on both of the earlier machines.

C. Histological procedures

Postscanning, the animals were maintained in the laboratory for several weeks. Although the male's behavior seemed otherwise normal, he never vocalized, either spontaneously, in response to natural vocalizations of other males housed in the laboratory, or in response to playbacks of conspecific advertisement calls. The female's behavior also seemed normal, although she never showed orienting responses to playbacks of natural or recorded advertisement calls. Neither of these behaviors is unusual in captive frogs outside of the breeding season. The male was sacrificed by intraperitoneal injection of sodium pentobarbital (100 mg/kg; Abbott), and transcardially perfused with heparinized 0.9%(w/v) saline and 4% formaldehyde in 0.1 M phosphate-buffered saline. The head was removed, embedded in paraffin, sectioned transaxially at 10 μm and stained with Gomori's trichrome to allow better visualization of internal structures. The female was anesthetized with 0.6% MS-222 and the eighth cranial nerve was exposed bilaterally through the roof of the mouth. A fluorescent lipophilic dye (DiI: 1,1′,di-octadecyl-3,3,3′3′-tetramethylindocarbocyanine perchlorate, Molecular Probes) was pressure injected into both the right and left eighth nerve medial to the otic capsule. The animal was allowed to survive for 5 days, then euthanized and perfused as described earlier. The brain and medial portions of the eighth nerves were removed, meninges were cleared, and the brain was sectioned at 50 μm on a vibratome. Both transmitted light and fluorescent sections were viewed using an Olympus BX60 microscope equipped with a fluorescence attachment and images collected on a Pentium 4 computer running MagnaFire software (Optronics).

III. RESULTS

At capture, the male measured 10.8 cm snout-vent length. A dorsal view of his head is shown in Fig. 1(a). The tympanic membrane is absent on the left side [Fig. 1(b)]. The right tympanic membrane [Fig. 1(c)] measured 1.9 cm diameter, which is in the range of tympanic membrane diameters observed in normal adult males (Boatright-Horowitz and Simmons, 1995). The dorsal aspect of the normal (right) tympanic membrane is bounded by a distinct, semilunate supratympanic ridge of connective tissue, extending from the retro-orbital region to the caudal edge of the tympanic membrane itself. The tympanic membrane is attached superiorly to this supratympanic ridge and is normal in appearance with the typical large central patch (the thickened region overlying the attachment region for the extrastapes). A ridge is also present on the abnormal (left) side [Fig. 1(b)], but is reduced in extent and extends laterally rather than dorsally [Figs. 2(a)–2(c)]. The outermost region on the left side of the head is occupied by conventional, smooth epidermis. Neither the central patch of the tympanic membrane nor a tympanic annulus is present.

Figure 1(d) shows the CT-derived skeletal structure of the frog, which largely displays normal form and structure. There is no apparent postcranial skeletal abnormality anywhere in the body; individual skull bones are all present and

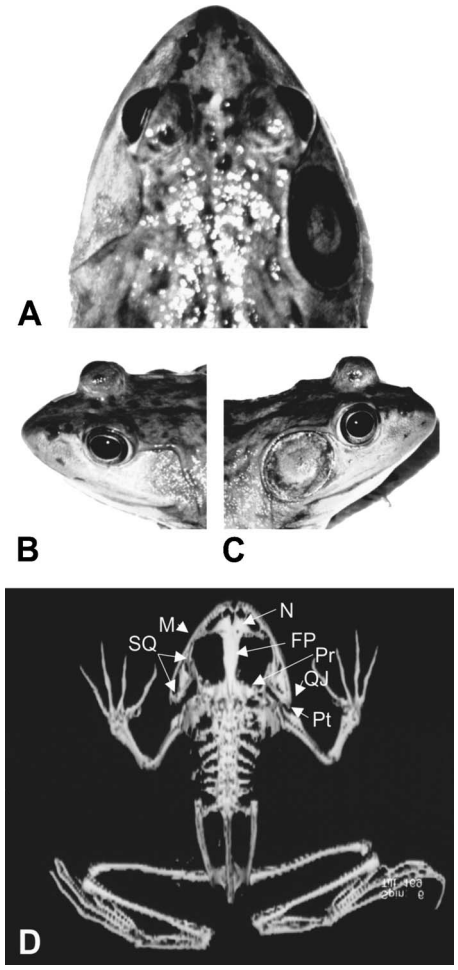


FIG. 1. (a) Dorsal view of adult male frog. The large, normal looking tympanic membrane is on the animal's right. (b) The abnormal side, lacking any tympanic membrane and with a reduced supratympanic ridge, whose caudal aspect passes dorsal-ventrally through what would be the caudal extent of a normal tympanic membrane's central patch. (c) The normal tympanic membrane, showing a robust medial thickening of the central patch. (d) CT scan of skeleton, showing normal axial skeleton, and normal, symmetric skull bones. Abbreviations: FP: frontoparietal; M: maxillary; N: nasal; Pr: Prootic; Pt: pterygoid; SQ: squamosal; QJ: quadrojugatal bone.

symmetric. The only anomalies visible in this section are the lack of the tympanic annulus on the left (overlying the squamosal and pterygoid bones, which are present bilaterally).

Figure 2 shows histological sections through the head region, cut in the transaxial (coronal) plane, to show internal morphology [Figs. 2(a)–2(c)], and corresponding tomographic sections through the head in the living animal [Figs. 2(d)–2(f)]. The absence of the middle ear structures is quite obvious on the left side of the images in both sets of figures. The large white ovals in the otic capsule in the CT sections are the otoliths, which are very dense [Figs. 2(e) and 2(f)]. From these figures, it is apparent that the basic skull anatomy is largely symmetrical, although there may be a slight depression of the dorsal process of the prootic bone on the deformed side. The internal structure of the otic capsule and bony labyrinth also appear to be normal. Figures 2(b) and 2(e) show the presence of a properly formed stapes on the right, and the lack of most of the stapedial shaft (pars media) on the left. The pars interna (footplate) cartilage is present on both sides [Fig. 2(b) right side, Fig. 2(c) left side due to slight bias in section slicing]. The operculum cartilage was observed bilaterally overlying the oval window, visible in more caudal sections (data not shown).

Figure 3 shows a close-up of the area of the tympanic membrane on both the deformed [Fig. 3(a), left] and normal [Fig. 3(b), right] sides, with the thick, filled-in connective tissue on the left and laminated structure of a normal tympanic membrane clear on the right. The lack of the tympanic annulus is also clear in this figure. Figures 3(c) and 3(d) shows sections of the oval window area showing the anterior margin of the pars interna, similar in size and position on both sides of the head. Note that the pars interna is free medially, and fused to bone laterally. It is visible bilaterally in the CT scans as well [Fig. 2(e)].

At the time of examination, the female's snout-vent length was 7.8 cm, with a 0.75 cm tympanic membrane clearly visible on the right side of her head [Figs. 4(a) and 4(c)], consistent with normal measurements for females at this developmental stage (Boatright-Horowitz and Simmons, 1995). As with the male, on the normal (right) side, there was a clear supratympanic ridge dorsal and caudal to the tympanic membrane, and the tympanic membrane itself had a large central patch. The abnormal side showed striking similarities to the malformation observed in the male, with a complete lack of tympanic membrane, reduced supratympanic ridge, and smooth epidermal covering overlying the region the tympanic membrane should occupy [Fig. 4(b)]. Transaxial CT scans of the animal's otic region [Fig. 4(d)] also showed great similarity to that observed in the male, with normally placed otoliths, and normal skull and postcranial skeletal anatomy, but with no stapedial shaft, extrastapes or tympanic membrane structures on the abnormal side. The footplate of the stapes and the operculum were present on both normal and abnormal sides. Transport of DiI from the eighth nerve to the medulla (dorsal lateral nucleus and vestibular nucleus complex) was qualitatively similar, showing extensive fiber and soma label, on both the normal and deformed sides.

IV. DISCUSSION

The proximate cause for the lack of tympanic membrane, tympanic annulus, extrastapes and pars media of the stapes combined with otherwise normal skeletal anatomy is unknown. Our inability to obtain reliable laboratory data on the animals' hearing sensitivity did not allow us to quantify the extent of any hearing loss. Even though the male animal was not observed to vocalize either in the field or in the laboratory, his behavior in the field, approaching playbacks of conspecific advertisement calls, indicated that he probably had some auditory function and at least some rudimentary ability to localize sound sources. The extent of the female's auditory function is unknown, but any auditory function in each animal could arise from several sources. First, unilateral tympanic input was clearly possible given the normal anatomy on the right side of the animals' heads. Second, the observation that the operculum cartilages were present on both the affected and unaffected sides of the head implies

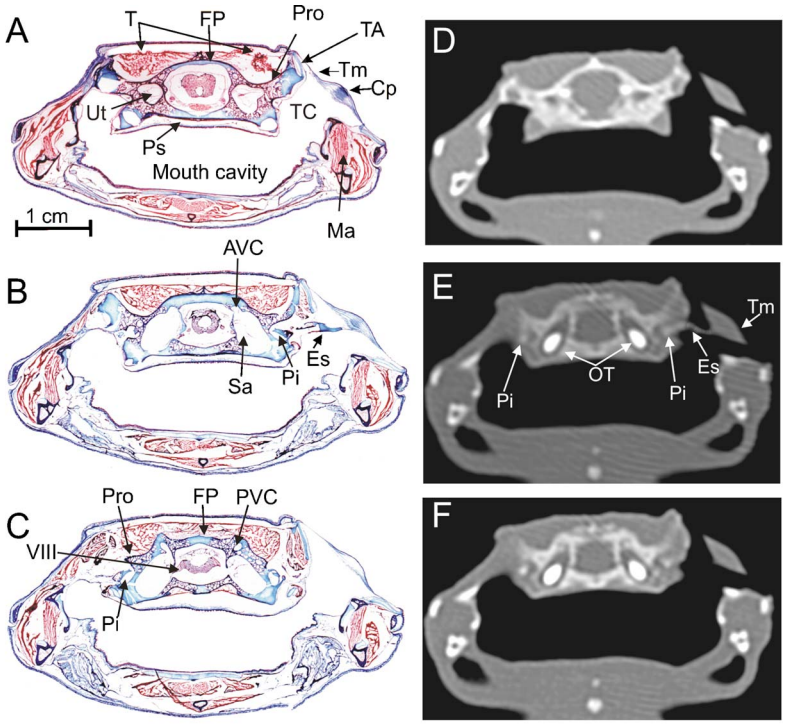


FIG. 2. (Color online). (a–c, left). 10 micron Gomori trichrome stained transaxial (coronal) section at levels of (a) auditory midbrain, (b) cerebellum, and (c) auditory medulla/VIII nerve from adult male frog. (d–f, right): 1.0 mm CT transaxial scans from living frog at levels corresponding to histological sections on left. Saccular (Sa) and Utricular (Ut) spaces are labeled, although at higher resolution, the organs themselves are plainly visible and appear normal in both orientation and extent. Abbreviations: AVC: anterior vertical canal; CP: central patch; Es: extrastapes; Ma: masseter muscle; OT: otoliths; Pi: tympanic columella pars interna; Ps: parasphenoid; PVC: posterior vertical canal; Sa: saccular space; T: temporalis muscle; TA: tympanic annulus; TC: tympanic cavity; Tm: tympanic membrane; Ut: utricular space; VIII: eighth cranial nerve.

that the animals possessed a functional opercularis system, which could convey sensitivity to low-frequency signals (Lombard and Straughan, 1974). Third, in some, typically smaller, anurans, the lungs and body wall serve as an input pathway to the inner ear (Narins *et al.*, 1988). Several extant species of anurans have reduced or missing tympani along with well-developed inner ears and intact operculum cartilages (Jaslow *et al.*, 1988). Vibrations of the body wall transmitted to the inner ear are important in mediating low frequency sensitivity in these "earless" anurans (Hetherington and Lindquist, 1999). Comparisons of auditory sensitivity in harlequin frogs (*Atelopus*) with and without a tympanic ear indicate that the tympanic pathway conveys greater sensitivities to high frequencies, but that the absence of a tympanic ear is associated with greater sensitivities to low frequencies (Lindquist *et al.*, 1998). It is unclear if a lung/body wall input pathway mediates hearing in larger bullfrogs with a normal tympanic system, but might be important in our specimen animals.

Because the male animal was captured as an adult, it was not possible to ascertain with certainty whether he was hatched in the same pond from which he was collected, or whether he migrated there as an adult from neighboring bodies of water. The presence of a second animal, captured as a

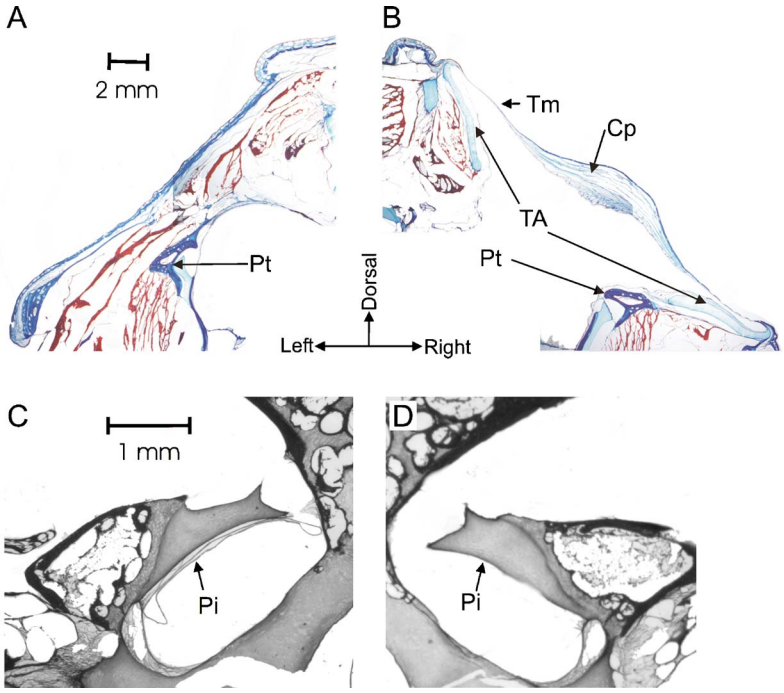


FIG. 3. (Color online). (a,b) 10 micron Gomori trichrome stained sections showing detail of tympanic region in malformed (a, left) and normal (b, right) sides of adult male frog. (c,d) Grayscale image of 10 micron Gomori trichrome stained sections showing detail of the pars interna of the stapes in oval window region on left (c) and right (d) sides, demonstrating presence of normal cartilage bilaterally. Abbreviations as in Fig. 2.

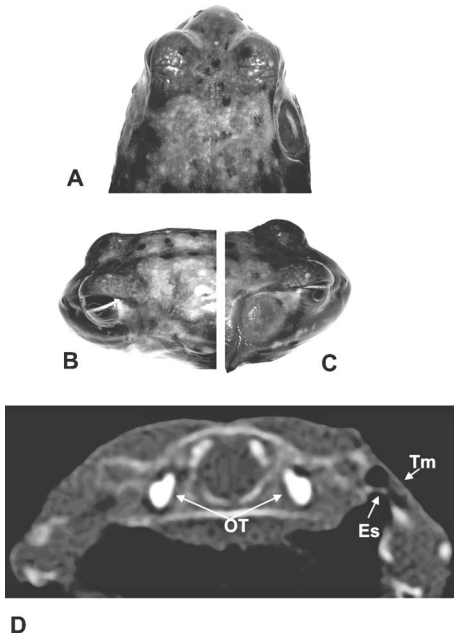


FIG. 4. (a) Dorsal view of female subadult frog. The normal looking tympanic membrane is on the animal's right. (b) The abnormal side, lacking any tympanic membrane and with a reduced supratympanic ridge. (c) The normal tympanic membrane, showing a robust central patch and supratympanic ridge. (d) CT scan of otic region showing normal middle ear structures on the right and loss of tympanic membrane, extrastapes, and parts of the stapes on the left. Abbreviations: OT: Otoliths; St: stapes; Tm: tympanic membrane.

tadpole and so bred in the same site, suggests that the presence of some agent within the aquatic environment may be the causative factor in expression of the deformity. It is interesting to note that the animals were collected several years after the onset of treatment of sewage catchbasins in the area with methoprene as a mosquito control measure. Methoprene, a commonly used domestic and agricultural pesticide, breaks down in the presence of ultraviolet light into compounds with the same structure and receptor binding properties as retinoic acids, which induce middle and external ear anomalies in mammals (Lammer, 1991; Mallo, 1997; Moerike *et al.*, 2002), and that have been implicated in the formation of craniofacial deformities in leopard frogs (*Rana pipiens*) (Ankley *et al.*, 1998). Given the homology between the mammalian and anuran stapes (Lombard and Bolt, 1988) and the functional analogy between the mammalian ossicular chain and the frog stapes/extrastapes system (Møller, 1963; Mason and Narins, 2002a), one possibility is that exposure to retinoic acid or its derivatives may have led to the frog's unilateral deformity. The presence of the female, captured as a tadpole, rules out the possibility that direct injury produced the deformity.

It is possible that some genetic anomaly produced the observed deformities. In mammals, components of the tympanic pathway arise via differentiation of neural-crest derived mesenchymal tissue, from the first (malleus, incus) and second (stapes) branchial arches, respectively, with the tympanic membrane derived from interactions between tissue from the first branchial cleft and branchial pouch. These developmental pathways are under the control of a wide variety of genes which are expressed at different time points in development (Mallo, 2001). These differential time courses controlling development of elements of the auditory periphery can lead to various types of anomalies, each of which can seriously impact auditory function. In many instances, malformations are restricted to developmentally related structures, which allows estimates of the time of damage. For example, in mice, exposure to retinoic acid yields differential malformations of middle ear elements depending on gestational time of exposure and hence the temporal sequence of migration of their precursors from the neural crest (Mallo, 1997). Furthermore, in many cases, both sides of the same embryo reached different stages of development, suggesting that the program regulating differentiation of the parts of the middle ear can act independently on each side (Mallo, 1997).

The observed pattern of deformities in the two frogs probably reflects the common embryological derivatives of the affected structures. Early induction studies in ranids demonstrate that transplanted portions of the tympanic annulus could induce formation of a tympanic membrane in anomalous locations (Helff, 1928). In bullfrogs, the tympanic membrane forms at the conclusion of metamorphic climax, and is often first evident by a thinning and darkening of the epidermis overlying the tympanic cavity at approximately Gosner stage 44 (Boatright-Horowitz and Simmons, 1995). The stapes forms at approximately this time as well, with the pars interna at the rostral end of the oval window (Horowitz *et al.*, 2001). The tympanic annulus appears to condense slightly before this, at about stage 43 (unpublished data). The operecularis system develops and reaches maturity earlier than the tympanic system, with the operculum cartilage forming over the oval window by about larval stage 40, and the opercularis muscle connection completed by stage 42 (Hetherington, 1987), coincident with the onset of metamorphic climax. This normal pattern of development suggests that the deformity described here most likely arose during metamorphic climax stages.

Certain human otolaryngological pathologies, such as congenital aural atresia, are often presented in pediatric patients with gradations ranging from minor external ear malformation to severe disruption of the anatomy of the entire tympanic pathway and inner ear. Aural atresia (absence or incomplete formation of the ear canal and external pinna) is a congenital syndrome seen in humans (Schuknecht, 1989; Shah and Shah, 2002) of poorly understood etiology. Usually, it involves abnormalities of the outer and middle ears, but with no involvement of inner ear structures. The incidence of abnormalities in the tympanic membrane varies; in some cases, it is abnormally small, while in more severe cases, it is totally absent. The syndrome can be unilateral or bilateral, is more common in males than females, and may occur in conjunction with other craniofacial syndromes (Llano-Rivas *et al.*, 1995; Kountakis *et al.*, 1995). Further study of the mechanism producing the absence of the tympanic membrane in anurans, with its qualitative similarity to aural atresia, may shed light on the etiology of this otolaryngological syndrome.

ACKNOWLEDGMENTS

Research protocols were approved by the Brown University Animal Care and Use Committee. The second author holds a scientific collection permit from the State of Rhode Island. The authors thank Paul Monfils, Rhode Island Hospital, for help with the histological preparations, and the reviewers for their helpful comments on the manuscript. Julie Arruda and Scott Cramer assisted with scanning and image reformatting of the CT data. This research was supported by a grant from the National Institutes of Health to A.M.S. (DC05257).

Ankley, G. T., Tietge, J. E., DeFoe, D. L., Hjensen, K. M., Holcombe, G., Durhan, E. J., and Diamond, A. (**1998**). "Effects of ultraviolet light and methoprene on survival and development of *Rana pipiens*," Envir. Toxicol. Chem. **17**, 2530–2542.

Blaustein, A. R., Kiesecker, J. M., Chivers, D. P., and Anthony, R. G. (**1997**). "Ambient UV-B radiation causes deformities in amphibian embryos," Proc. Natl. Acad. Sci. U.S.A. **94**, 13735–13737.

Boatright-Horowitz, S. S., and Simmons, A. M. (**1995**). "Postmetamorphic changes in auditory sensitivity of the bullfrog midbrain," J. Comp. Physiol., A **177**, 577–590.

Gosner, K. L. (**1960**). "A simplified table for staging anuran embryos and larvae with notes on identification," Herpetologica **16**,183–190.

Helff, O. M. (**1928**). "Studies on amphibian metamorphosis. III. The influence of the annular tympanic cartilage on the formation of the tympanic membrane," Physiol. Zool. **1**, 463–495.

Hetherington, T. E. (**1987**). "Timing of development of the middle ear of Anura (Amphibia)," Zoomorph. **106**, 289–300.

Hetherington, T. E. (**1994**). "The middle ear muscle of frogs does not modulate tympanic responses to sound," J. Acoust. Soc. Am. **95**, 2122–2125.

Hetherington, T. E., and Lindquist, E. G. (**1999**). "Lung-based hearing in an "earless" anuran amphibian," J. Comp. Physiol., A **184**, 395–401.

Horowitz, S. S., Chapman, J. A., Kaya, U., and Simmons, A. M. (**2001**). "Metamorphic development of the bronchial columella of the larval bullfrog (*Rana catesbeiana*)," Hear. Res. **154**, 12–25.

Jaslow, A. P., Hetherington, T. E., and Lombard, R. E. (**1988**). "Structure and function of the amphibian middle ear," in *The Evolution of the Amphibian Auditory System*, edited by B. Fritzsch, M. J. Ryan, W. Wilczynski, T. E. Hetherington, and W. Walkowiak (Wiley Interscience, New York), pp. 69–92.

Johnson, P. T. J., Lunde, K. B., Ritchie, E. G., and Launer, A. E. (**1999**). "The effect of trematode infection on amphibian limb development and survivorship," Science **284**, 802–804.

Kountakis, S. E., Helidonis, E., and Jahrsdoerfer, R. A. (**1995**). "Microtia grade as an indicator of middle ear development in aural atresia," Arch. Otolaryngol. Head Neck Surg. **121**, 885–886.

Lammer, E. (**1991**). "Preliminary observations on isoretinoin-induced ear malformations and pattern formation of the external ear," J. Craniofac Genet. Dev. Biol. **11**, 292–295.

Lindquist, E. D., Hetherington, T. E., and Volman, S. F. (**1998**). "Biomechanical and neurophysiological studies on auditon in eared and earless harlequin frogs (*Atelopus*)," J. Comp. Physiol., A **183**, 265–271.

Llano-Rivas, I., González-del Angel, A., del Castillo, V., Reyes, R., and Carnevale, A. (**1999**). "Microtia: A clinical and genetic study at the National Institute of Pediatrics in Mexico City," Arch. Med. Res. **30**, 120–124.

Lombard, R. E., and Bolt, J. R. (**1988**). "Evolution of the stapes in paleozoic tetrapods: Conservative and radical hypotheses," in *The Evolution of the Amphibian Auditory System*, edited by B. Fritzsch, M. J. Ryan, W. Wilczynski, T. E. Hetherington, and W. Walkowiak (Wiley Interscience, New York), pp. 37–68.

Lombard, R. E., and Straughan, I. R. (**1974**). "Functional aspects of anuran middle ear structures," J. Exp. Biol. **61**, 71–93.

Mallo, M. (**1997**). "Retinoic acid disturbs mouse middle ear development in a stage-dependent fashion," Dev. Biol. **184**, 175–186.

Mallo, M. (**2001**). "Formation of the middle ear: Recent progress on the developmental and molecular mechanisms," Dev. Biol. **231**, 410–419.

Mason, M. J., and Narins, P. M. (**2002a**). "Vibrometric studies of the middle ear of the bullfrog *Rana catesbeiana* I. The extrastapes," J. Exp. Biol. **205**, 3153–3165.

Mason, M. J., and Narins, P. M. (**2002b**). "Vibrometric studies of the middle ear of the bullfrog *Rana catesbeiana* II. The operculum," J. Exp. Biol. **205**, 3167–3176.

Meteyer, C. U. (2000). Field Guide to Malformations of Frogs and Toads, with radiographic interpretations. Biological Science Report USGS/BRD/BSR-2000-005.

Moerike, S., Pantzar, D. T., and De Sa, D. (**2002**). "Temporal bone pathology in fetuses exposed to isoretinoin," Pediatr. Dev. Pathol. **5**, 405–409.

Møller, A. R. (**1963**). "Transfer function of the middle ear," J. Acoust. Soc. Am. **35**, 1526-1534.

Narins, P. M., Ehret, G., and Tautz, J. (**1988**). "Accessory pathway for sound transfer in a neotropical frog," Proc. Natl. Acad. Sci. U.S.A. **85**, 1508–1512.

Schuknecht, H. F. (**1989**). "Congenital aural atresia," Laryngoscope **99**, 908–917.

Sessions, S. K., and Ruth, S. B. (**1990**). "Explanation for naturally occurring supernumerary limbs in amphibians," J. Exp. Zool. **254**, 38–47.

Sessions, S. K., Franssen, R. A., and Horner, V. L. (**1999**). "Morphological clues from multilegged frogs: are retinoids to blame?," Science **284**, 800–802.

Shah, R. K., and Shah, U. K. (**2002**). "External auditory canal atresia," eMedicine Journal **3**, 1–15.

Wever, E. G. (**1985**). *The Amphibian Ear* (Princeton University Press, Princeton, NJ).

Receiving beam patterns in the horizontal plane of a harbor porpoise (*Phocoena phocoena*)

Ronald A. Kastelein[a] and Mirjam Janssen
Sea Mammal Research Company (SEAMARCO), Julianalaan 46, 3843 CC Harderwijk, The Netherlands.

Willem C. Verboom
TNO Observation Systems, Department of Underwater Technology, P.O. Box 96864, 2509 JG Den Haag, The Netherlands.

Dick de Haan
Netherlands Institute for Fisheries Research (RIVO), P.O. Box 68, 1970 AB IJmuiden, The Netherlands.

(Received 20 December 2004; revised 9 May 2005; accepted 11 May 2005)

Receiving beam patterns of a harbor porpoise were measured in the horizontal plane, using narrow-band frequency modulated signals with center frequencies of 16, 64, and 100 kHz. Total signal duration was 1000 ms, including a 200 ms rise time and 300 ms fall time. The harbor porpoise was trained to participate in a psychophysical test and stationed itself horizontally in a specific direction in the center of a 16-m-diameter circle consisting of 16 equally-spaced underwater transducers. The animal's head and the transducers were in the same horizontal plane, 1.5 m below the water surface. The go/no-go response paradigm was used; the animal left the listening station when it heard a sound signal. The method of constants was applied. For each transducer the 50% detection threshold amplitude was determined in 16 trials per amplitude, for each of the three frequencies. The beam patterns were not symmetrical with respect to the midline of the animal's body, but had a deflection of 3–7° to the right. The receiving beam pattern narrowed with increasing frequency. Assuming that the pattern is rotation-symmetrical according to an average of the horizontal beam pattern halves, the receiving directivity indices are 4.3 at 16 kHz, 6.0 at 64 kHz, and 11.7 dB at 100 kHz. The receiving directivity indices of the porpoise were lower than those measured for bottlenose dolphins. This means that harbor porpoises have wider receiving beam patterns than bottlenose dolphins for the same frequencies. Directivity of hearing improves the signal-to-noise ratio and thus is a tool for a better detection of certain signals in a given ambient noise condition. [DOI: 10.1121/1.1945565]

PACS number(s): 43.80.Lb, 43.66.Gf [WA] Pages: 1172–1179

I. INTRODUCTION

The harbor porpoise (*Phocoena phocoena*) has well-developed echolocation (Busnell and Dziedzic, 1976; Kastelein *et al.*, 1997a, b; 1999). The species produces narrow-band (~16 kHz), low-amplitude [source level (SL) about 157 dB re 1 μPa, p/p] echolocation signals with a peak frequency between 120 and 130 kHz (Schevill *et al.*, 1969; Møhl and Andersen, 1973; Kamminga and Wiersma, 1981; Verboom and Kastelein, 1995, 1997, 2003; Au *et al.*, 1999). The transmitting beam patterns in the horizontal and vertical plane have been studied by Au *et al.* (1999), and are wider than in the bottlenose dolphin (*Tursiops truncatus*), beluga whale (*Delphinapterus leucas*) and false killer whale (*Pseudorca crassidens*; Au, 1993; Au *et al.*, 1995).

The underwater hearing sensitivity of the harbor porpoise has been studied in two electrophysiological experiments in very small tanks (Popov *et al.*, 1986; Bibikov, 1992) and in two psychophysical experiments (Andersen, 1970; Kastelein *et al.*, 2002). Based on the latter, more elaborate study, the range of best hearing of the harbor porpoise (defined as 10 dB within the lowest threshold) is very large (16–140 kHz). The hearing between 64 and 180 kHz is more sensitive than that of most other odontocetes tested so far (Johnson, 1967; Hall and Johnson, 1971; Jacobs and Hall, 1971; White *et al.*, 1978; Ljungblad *et al.*, 1982; Awbrey *et al.*, 1988; Thomas *et al.*, 1988; Wang *et al.*, 1992; Nachtigall *et al.*, 1995; Szymanski *et al.*, 1999; Sauerland and Dehnhardt, 1998; Tremel *et al.*, 1998; Kastelein *et al.*, 2003). Only killer whales (*Orcinus orca*) have a similarly low threshold (Szymanski *et al.*, 1999).

Four studies on bottlenose dolphins showed that in this species, as in other mammals, hearing sensitivity has directional deviations, and noise from different directions causes different levels of masking of sounds produced in front of the animal. In a study by Au and Moore (1984) sounds were produced up to 100° to either side of the animal's frontal midline, and relatively high frequencies (30, 60, and 120 kHz) were tested because the study was geared towards echolocation. Zaytseva *et al.* (1975) studied the effect of a broadband masking sound (50–100 kHz) on the detection ability of an 80 kHz, 0.6 s, signal which was produced directly in front of the animal. The noise transducer was offered at 0°, 7°, 15°, 30°, 50°, 90°, and 180° from the longitudinal body axis. Schlundt *et al.* (1998) tested the hearing

[a]Electronic mail: researchteam@zonnet.nl

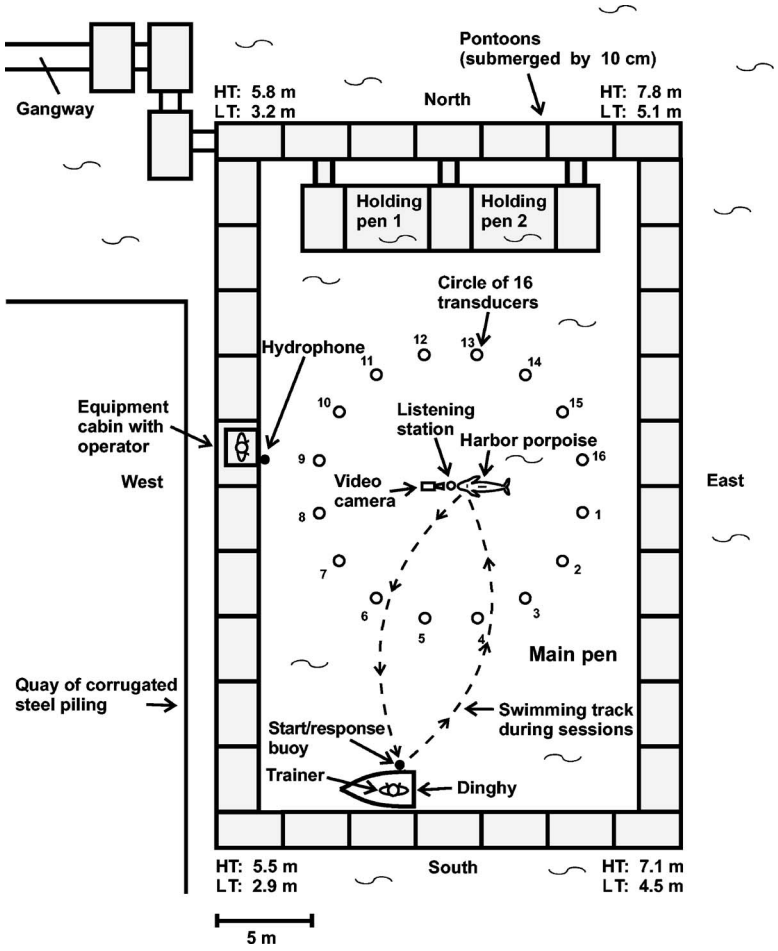


FIG. 1. Top view of the study area, showing the 16 underwater transducers in a 16-m-diameter circle around the central listening station and the underwater camera (filming the porpoise from above). Also shown are the monitoring hydrophone, the equipment cabin with the operator and the trainer in the dinghy. The water depths near the corners of the pen at high tide (HT) and low tide (LT) are also shown.

sensitivity of a bottlenose dolphin for three frequencies (2, 8, and 12 kHz) with the transducer in two positions: in front of the animal and below its head. Popov and Supin (1990) showed, in evoked potential studies with bottlenose dolphins and two river dolphin species, that their hearing is most sensitive when sounds are produced in front of the animals. The receiving beam becomes narrower with higher frequencies (Popov *et al.*, 1992; Supin and Popov, 1993).

Evoked potential data for the harbor porpoise show strong directional variation in hearing thresholds for signals between 30 and 160 kHz. Harbor porpoises are most sensitive to sounds arriving from angles within 15°–30° of straight ahead (Voronov and Stosman, 1983).

Directional hearing is an important property of a hearing system, either just for listening to sounds from the environment or as a property of a biosonar system. Directivity improves the signal-to-noise ratio and thus is a tool for a better detection of certain signals in a given ambient noise condition. Frontally-oriented narrow transmitting and receiving beams allow odontocetes to localize objects and separate them in a multiobject field and to minimize the amount of acoustic interference received, and therefore improve the animals' detection ability.

The aims of the present study were to determine the receiving beam patterns of a harbor porpoise psychophysically for 16, 64, and 100 kHz narrow-band frequency modulated (FM) signals in a 360° circle in the horizontal plane, and to compare the results with directional hearing of the bottlenose dolphin. The present study is part of a series of acoustic and behavioral studies with the ultimate goal of reducing accidental bycatch of harbor porpoises in gillnets either via deterring sounds, or via enhanced net detectability by echolocation (Kastelein *et al.*, 1995, 1997a, 2000, 2001, 2002).

II. MATERIALS AND METHODS

A. Study animal

The study animal was a male harbor porpoise (PpSH052) that had undergone veterinary treatment. The animal stranded in November 1998 on the Dutch Island of Ameland at the approximate age of 10 months. At the time of the study the animal was healthy, around three years old, weighed 29 kg and had a body length (tip rostrum to notch in tail fluke) of 119 cm. The animal was fed six times a day at 0900, 1030 (during a research session), 1200 (during a research session), 1330 (during a research session), 1500 (during a research session) and 1630 h.

B. Study area

The porpoise was housed in a large floating pen (34 m $\times$ 20 m; 3.5 m deep at the sides and 4–6 m deep in the center depending on the tide; Fig. 1). The surrounding pontoons (plywood boxes filled with Styrofoam and coated with fiberglass) were submerged by about 10 cm. The net at the

bottom of the pen was made of nylon and the sides were made of polypropylene. Both types of net had a stretched mesh size of 9 cm and a twine diameter of 3 mm. Seawater flowed freely through the net, and the twine was covered with algae. The floating pen was in a harbor that was horseshoe-shaped (500 m × 280 m) with the entrance to the northeast. The harbor was in the southwest of The Netherlands, at Neeltje Jans (51°37′N,03°40′E). The entrance was near the inside of the Oosterschelde surge barrier, which is only closed during exceptionally high tides and storms. The barrier was open at the time of the study. Therefore the tidal range inside the harbor was similar to that in the nearby North Sea. During the sessions no shipping occurred within 2 km of the study area. The sea floor below the pen was flat and covered with sandy silt. During the study the salinity in the pen was measured weekly and varied between 3.1% and 3.6%. The mean monthly water temperature decreased from 13 °C in October to 11 °C in November 2000. The following environmental parameters were recorded at the start of each session: water depth at a specific location of the pen, underwater visibility, estimated wind speed, and wind direction. The underwater visibility (as determined by using a Secchi disk) varied between 1 and 2 m. During the experiments the study animal was kept in the main pen and the two pool mates (porpoises) were kept in holding pens (Fig. 1).

C. Stimuli

Directional hearing was tested using narrow-band FM signals with ISO R266 standardized frequencies of 16, 64, and 100 kHz. The modulation range was ±1% of the center frequency and the modulation frequency was 100 Hz. The rationale behind the selected signal frequencies was that frequencies around 16 kHz commonly occur in the lowest part of the frequency spectrum of some commercially-available acoustic alarms to prevent accidental bycatch of porpoises in gillnets, while 100 kHz was the maximum frequency of the spectrum that could be produced with a sufficient SL by these transducers (120–130 kHz would have been preferred in relation to the peak frequency of the porpoise's echolocation clicks; Møhl and Andersen, 1973; Verboom and Kastelein, 1995, 1997, 2003), and 64 kHz was selected as a frequency between the two extremes.

FM test signals were used to reduce the chance of occurrence of standing waves. FM signals were generated by a waveform generator (Hewlett Packard, model 33120A) and a custom-built selector box, consisting of a signal shaper, a driver and a switch to select the desired transducer. Total signal duration was 1000 ms including a 200 ms rise time and 300 ms fall time. The rationale behind the selected signal duration was that in hearing studies usually 500–1000 ms signals are used, and the selected short (1000 ms) signal duration prevented the animal in the present study from orienting itself towards an active transducer during signal presentation. The stimulus was transmitted by transducers (LabForce 1 BV, model 90.02.01), which were omnidirectional in the horizontal plane. Correct setting of the voltage levels (amplitude) and the frequency of the stimulus was checked each session with an oscilloscope (Philips, model PM 3233).

Sixteen transducers (equally spaced at 360/16=22.5° intervals) were suspended from ropes, strung across the pen. The rationale behind the locations of the transducers was that this was the setup of a previous study on sound source localization. The setup allowed the porpoise's hearing sensitivity to be tested in a horizontal plane from 16 fixed directions around its body. During sessions the transducers were at a depth of 1.5 m in a 16-m-diameter circle around a central listening station which consisted of a water-filled stainless steel tube with a 5-cm-long and 3-cm-wide rubber knob near the end, pointing exactly towards the east side of the pen (Fig. 1). The knob was also 1.5 m below the water surface. For the safety of the porpoises, the listening station and all 16 transducers were lifted to 85 cm above the water surface with pulley systems, after each session.

D. Acoustic calibration and monitoring

The sound pressure levels (SPLs, dB re 1 μPa, rms) of all transducers were measured with a hydrophone (Brüel & Kjaer, model 8101; the calibration curve of this hydrophone was flat up to 100 kHz) positioned at the location of the porpoise's head when stationed in the listening position (1.5 m below the water surface), a conditioning amplifier (Brüel & Kjaer, Nexus 2690), and a computer with an analog data acquisition card (National Instruments, PCI-MIO-16E-1, 12-bit resolution, sample rate 512 kHz). The system was calibrated with a pistonphone (Brüel & Kjaer, 4223). The analysis of SPLs was carried out using a RIVO-designed analysis module based on Labview 4.1 software (National Instruments). This way the signal generating system was calibrated weekly during the course of the one month study period.

Before each session, the equipment and transducers were checked by using a hydrophone (LabForce 1 BV, model 90.02.01) connected to a heterodyne bat detector (Batbox III, Stag Electronics, Steyning, UK). The hydrophone was 1.5 m below the water surface, next to the equipment cabin (Fig. 1). To monitor the audible part of the underwater background noise during sessions, the same hydrophone was connected to a charge amplifier (Brüel & Kjaer, model 2635), the output of which was connected to an amplified loudspeaker.

E. Methodology

Operant conditioning, using positive reinforcement, was used to train the porpoise. Preliminary tests showed that the animal needed no warm-up trials before a session began.

A session began when the trainer called the animal to the start/response buoy (which was attached to the dinghy), by tapping on it. The signal operator had set the frequency for the session and selected the transducer to be activated during the first trial, as well as the signal amplitude. The operator and trainer communicated via a radio connection and headsets.

In a preliminary test, the rough hearing threshold levels for the three signal frequencies of each of the 16 transducers were determined. Based on these results, three amplitudes (6 dB steps) were selected around the roughly established thresholds (a level the animal never, or almost never heard, a level he often heard, and a level he usually heard). Based on the detection rate at these three exposure levels psychometric functions could be drawn from which the 50% detection threshold levels could be derived. During the actual experiment, one of the three amplitudes was offered in each session (method of constants), and the animal's reaction to this amplitude was recorded.

At a hand signal of the trainer, the animal swam towards the listening station and positioned itself (with its rostrum pointing to the west) in the horizontal plane (Fig. 1). This way, his lower jaw was below the rubber knob. He was filmed from above, and the accuracy of the animals' positioning could be observed on a monitor by the operator in the equipment cabin. To enhance the contrast of the dark dorsal side of the porpoise in the dark background, a stripe of zinc ointment was put on the porpoise's back between the blowhole and the dorsal fin. A maximum difference of only 2° (indicated by lines drawn on the monitor screen) between the animal's body axis and the stationing knob's axis was accepted in both horizontal directions. Trials were cancelled when the animal was not in the correct position at the listening station. The operator informed the trainer when the animal was not in the correct position and the trainer then signaled to the animal (by tapping on the side of the dinghy) that the trial had ended, thus calling the animal back to the start/response buoy. In that case no reward was given.

After the animal stationed correctly, a trial could be classified as a signal-present or a signal-absent trial. Signal trials were conducted by waiting a random period between 2 and 8 s after the porpoise had stationed before one of the 16 transducers was activated. The go/no go response paradigm was used. As soon as the animal detected the sound, it left the listening station (go response), and returned to the start/response buoy (Fig. 1). The signal operator signaled to the trainer that the response was correct, after which the trainer blew a whistle and the porpoise received a fish. If the animal did not respond to the sound (no-go response) the signal operator informed the trainer that the trial had ended. The trainer then signaled to the animal (by tapping on the side of the dinghy) that the trial had ended, thus calling the animal back to the start/response buoy. In that case no reward was given. If the animal responded before a signal was produced (prestimulus response), the signal operator told the trainer to end the trial and not provide a reward.

For signal-absent trials, the signal operator told the trainer, after a random time period between 3 and 8 s after the porpoise had stationed, to end the trial by signaling to the animal (by blowing a whistle) to return to the start/response buoy and receive a fish reward. If the porpoise left the listening station before the whistle was blown (prestimulus response), the signal operator notified the trainer to end the trial and not provide a reward. The amount of fish given as a reward for correct go and no-go trials was the same. After a correct response trial, the next trial would start as soon the porpoise had swallowed the reward.

TABLE I. The relative 50% detection threshold levels of the harbor porpoise (in dB), in 16 directions for 16, 64, and 100 kHz narrow-band FM signals (for the transducer locations see Fig. 1).

Transducer No.	Direction (° relative to body length axis)	Signal frequency (kHz) 16	64	100
1	−168.7	−8	−12	−11
2	−146.2	−5	−11	−18
3	−123.7	−7	−10	−18
4	−101.3	−7	−7	−14
5	−78.8	−6	−9	−14
6	−56.3	−1	−7	−9
7	−33.8	−3	−4	−9
8	−11.3	0	0	−2
Body axis	0	⋯	⋯	⋯
9	11.3	−1	0	0
10	33.8	−3	−1	−12
11	56.3	−3	−8	−14
12	78.8	−3	−4	−17
13	101.3	−5	−7	−17
14	123.7	−3	−4	−17
15	146.2	−4	−10	−13
16	168.7	−11	−17	−15

Only one signal frequency was tested per session. A session consisted of 25 trials and each transducer was activated once. The activation occurred in a random order. Each session consisted therefore of 16 (64%) signal trials and 9 (36%) signal-absent ("control" or "catch") trials. The signal-absent trials were randomly distributed between the signal-present trials. The operator observed the animal on the monitor in the cabin (Fig. 1), and recorded its responses. Occasional behavioral maintenance sessions (no data collected) were conducted in order to maintain the porpoise's correct position at the listening station.

Sessions were conducted between 12 October and 14 November 2000. Sessions were not carried out in winds over 4 Beaufort, or in rain. Generally, four sessions were conducted daily. The order in which the three frequencies were tested was random during the study period. A total of 2304 signal trials (3 frequencies $\times$ 16 transducers $\times$ 16 trials per transducer per frequency per amplitude $\times$3 amplitudes) were used in the analysis.

The detection thresholds were calculated by drawing a psychometric function (amplitude versus percent signal detection) per transducer for each frequency. The 50% detection amplitudes were derived from these graphs (accuracy about + or −1 dB).

III. RESULTS

The relative hearing sensitivity of the porpoise for the three test frequencies in the 16 transducer directions is shown in Table I. The prestimulus response rate was 4%. The signal to noise ratio was >10 dB.

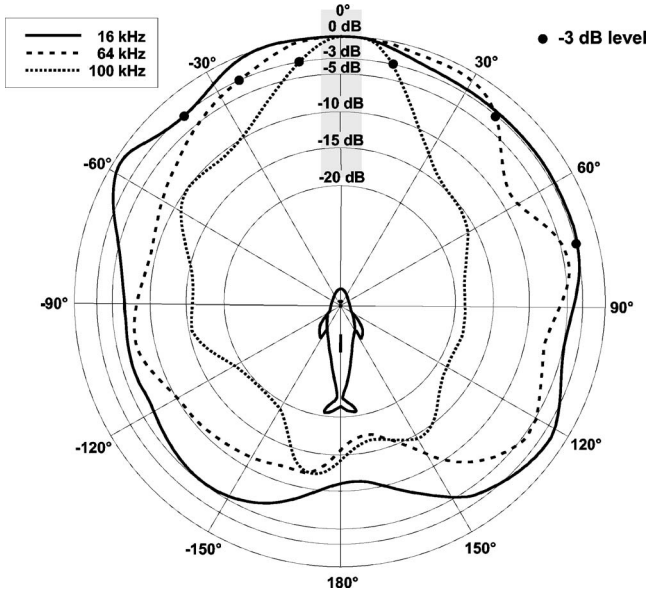


FIG. 2. The receiving beam patterns of the harbor porpoise in the horizontal plane for narrow-band FM signals with center frequencies of 16, 64, and 100 kHz. The dots represent the "−3 dB levels." Note that the beam patterns are not symmetrical with respect to the midline of the animal's body, but have a deflection of 3–7° to the right.

The hearing thresholds at the acoustic axis (0° azimuth) were not measured, but were calculated by extrapolating the data points mathematically. This was done by "smoothing" the hearing threshold patterns (custom-built software in Matlab 6, The Math Works Inc.) followed by setting the hearing threshold at the 0° azimuth at 0 dB for each of the three test frequencies. Thereafter the relative thresholds in the 16 directions were normalized (the correction factor was <1 dB) relative to calculated thresholds at the 0° azimuth (Fig. 2). This graph represents the spatial hearing sensitivity of the porpoise in the horizontal plane. The beam pattern is frequency dependent, becoming narrower (i.e., more directional) with increasing frequency. A skew occurred between the hearing directivity and the animal's body length axis. This skew was between 3 and 7° depending on the test frequency.

Directionality is often expressed as the beam width between −3 dB points (down 3 dB on each side of the 0° azimuth). The significance of these points is that the −3 dB beam corresponds to 50% intensity level. The −3 dB beam width for 100 kHz signals was 22° (−10° and +12°) and for 64 kHz signals 64° (−25° and +39°). At 16 kHz the −3 dB beam width was 115° (−40° and +75°).

From the receiving beam patterns the directivity index can be calculated. The directivity index is a measure of the effectiveness of an acoustic receiver in limiting the effects of omnidirectional background noise. This study only provides information about the beam patterns in the horizontal plane. To allow calculation of the directivity index, one needs to measure the hearing thresholds in all directions. If the animal's anatomy and physiology were symmetrical, the total spatial pattern would be rotation-symmetrical, but in reality there will always be deflections. The skew found in this study, is an example of such a phenomenon. However, in terms of directivity index (in dB) this skew is negligible. Assuming that the beam pattern is rotation-symmetrical according to an average of the horizontal beam pattern halves, then the receiving directivity indices can be calculated (Urick, 1983). The directivity indices thus found are 4.3 dB at 16 kHz, 6.0 dB at 64 kHz, and 11.7 dB at 100 kHz (Fig. 3). For comparison, the directivity indices of the bottlenose dolphin (based on the receiving beam patterns reported by Au and Moore, 1984) were calculated in the same way as those of the harbor porpoise in the present study (Fig. 3). The directivity indices of the harbor porpoise are lower than those of the bottlenose dolphin for the same frequencies.

IV. DISCUSSION AND CONCLUSIONS

A. Evaluation

In general the porpoise positioned itself incorrectly about once per session. By carrying out training sessions between research sessions, the correct positioning behavior was maintained.

The present study differed in several aspects from the study of Au and Moore (1984) in which the receiving beams of a bottlenose dolphin were determined. Au and Moore used a 1–2 dB up-down staircase method to determine the hearing thresholds. In the present study the method of constants was used to derive the hearing thresholds, leading to a less accurate threshold determination due to the larger (6 dB) steps. In the study by Au and Moore (1984) sounds were

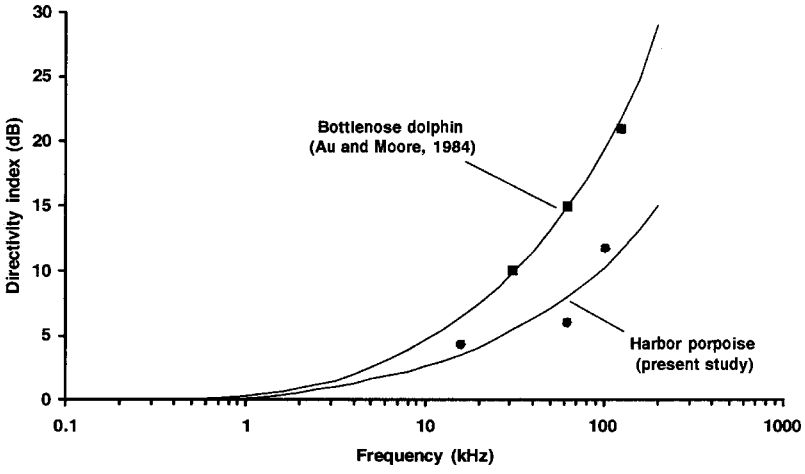


FIG. 3. The calculated directivity indices for the harbor porpoise based on data of the present study and those for the bottlenose dolphin based on data by Au and Moore (1984). Also shown are the lines of best fit to the data points of each species.

produced up to 100° to either side of the animal's frontal body axis, and the animal was visually aware of the location of the single signal transducer (which was moved only between sessions). In the present study, the porpoise was in the centre of 16 transducers in the horizontal plane, and could not acquire additional information on the direction of the sound. This may have lead to higher thresholds, but possibly more natural thresholds, as in the wild a sound can come from any direction. In the study by Au and Moore (1984) relatively high frequencies (30, 60, and 120 kHz) were tested because the study was geared towards dolphin echolocation. In the present study, lower frequencies were used partly due to equipment limitations, but also because the audibility of, usually low-frequency, man-made noise was of interest. Au and Moore (1984) used a bite-plate to keep the animal in position when testing the directivity of the hearing of bottlenose dolphins. In the present study no bite-plate was used and the porpoise had to focus more on maintaining the correct listening position, but its lower and upper jaw were in a natural occlusion. Whether the different methods affected the obtained results is not clear.

One possible explanation for the skew in the response pattern is that the two ears (or whatever the left and right sensors of sound are in the harbor porpoise) have different sensitivities, thereby biasing the animal's response to directional sources. Another explanation could be the asymmetry of the porpoise skull (Yurick and Gaskin, 1987).

The lack of a transducer right in front of the porpoise (on 0° azimuth) in the present study is not as problematic as its seems, as the lowest detection thresholds are probably not for sounds coming from right in front of an animal. Harbor porpoises are most sensitive to sounds arriving from angles within 15°–30° of straight ahead (Voronov and Stosman, 1983). Norris and Harvey (1974) found the maximum hydrophone response in the region of the auditory bulla of a bottlenose dolphin when sound came from 20° off the midline. Renaud and Popper (1975) showed that the best pure tone sound source localization occurred when the azimuth was 15°, an angle, which closely approximates the best angle of reception in the bottlenose dolphin. Au and Moore (1984) also found that the angle of best hearing deviated from the 0° azimuth, and was frequency dependent. Renaud and Popper (1975) suggest that the angle at which sound enters the two pan bone areas of the lower jaw (also called acoustic windows) determines the detection threshold for each ear.

B. Comparison of directivity indices of receiving and transmitting beams

Au *et al.* (1999) measured the transmitting beam pattern of a harbor porpoise in the horizontal, as well as in the vertical plane. These measurements were carried out in the same floating pen as the one in the present study and were limited to a sector of +/− 40° in front of the animal. The transmitting directivity index in this 40° sector, calculated in the same ways as in the present study, was 19.8 dB at 127.5 kHz (the average frequency of the echolocation signals). According to the best-fit curve of Fig. 3, the receiving directivity index for the harbor porpoise at 127.5 kHz is approximately 11 dB. Thus, the harbor porpoise's receiving beam is wider than its transmitting beam. This is also the case for the bottlenose dolphin (Au and Moore, 1984). Beam patterns are formed because of the geometry of the head and the physics of sound propagation within the head of an odontocete for sound propagating outwards, as well as inwards (Cranford *et al.*, 1996). Since different parts of the head take part in the transmission (among others the melon) and reception (among others the lower jaw) process, there are no good reasons why the beams would be the same or similar. Whether there is any evolutionary advantage of having a broader reception beam than a transmission beam is questionable.

C. Ecological significance

The receiving beam of the harbor porpoise becomes narrower as the frequency of signals increases. The narrower the beam, the less noise from its environment an animal will perceive. Thus, the echoes of harbor porpoise's outgoing echolocation signals are mainly masked by sound coming from directly in front of the animal. The effects of discrete noise sources on echolocation can be minimized by directing the beam away from the noise sources. The receiving beam (present study) and transmission beam (Au *et al.*, 1999) of the harbor porpoise are both wider than those of the bottlenose dolphin (Au and Moore, 1984; Au, 1993). The beam differences between species are probably linked to the differences in body dimensions; the larger the transducer or distance between receivers [in odontocetes this is either the distance between the (middle+inner) ears, lower jaws, or fat tissues lateral of the ears], the more directive the beams. The harbor porpoise uses narrow-band high-frequency (120–130 kHz) echolocation signals, while echolocation signals of bottlenose dolphins are more broadband and of lower frequency (between 50 and 120 kHz; Au, 1993). Wide bandwidths support the transmission of much information, so being a narrow-band echolocator like the harbor porpoise is not necessarily an advantage; target resolution in space and time may be enhanced by bandwidth. Although the receiving beams become narrower with increasing frequency, increasing detection ability, sound absorption losses become markedly greater thus limiting detection distances. Also, above about 30 kHz, the background noise may be of thermal origin (depending on whether it is near the coast, or in the middle of the ocean) and increases with increasing frequency, making detection more difficult.

The receiving directivity index of a harbor porpoise for its echolocation signals (120–130 kHz) is roughly 11 dB. The receiving directivity index of bottlenose dolphins for their broadband echolocation signals varies between 13 and 19 dB for signals between 50 and 100 kHz (Fig. 3).

Why is there a need for low frequency (omnidirectional) hearing sensitivity, when porpoises are known to transmit only narrow-band, very high frequency, pulses? The hearing of porpoises is probably not only used for echolocation. It possibly also serves to detect predators like killer whales, which produce lower frequency echolocation signals and even lower frequency social calls (Miller *et al.*, 2004). In addition, most odontocetes seem to use passive sonar to derive information about their environment.

Another reason to know the directivity of harbor porpoise hearing for frequencies lower than its echolocation signals, is to determine the audibility of acoustic alarms (pingers) to harbor porpoises. These pingers are designed to deter porpoises away from gillnets (Kastelein *et al.*, 2001). Also to determine the audibility of other man-made sounds, such as those made by offshore windmills (Koschinski *et al.*, 2003) and underwater data communication sounds (Kastelein *et al.*, 2005) to porpoises, the directivity index of porpoises for low frequency sounds is of importance.

ACKNOWLEDGMENTS

The authors thank Helmi ter Horst, Huub Vlemmix, Harrie Janssen, Lia Janssen, and Bas Koreman for their help with the data collection. They thank Niek Straver (Labforce company) for supplying the transducers, Rob Triesscheijn for drawing the graphs and Teun van den Dool (TNO-TPD) for his software development assistance. They also thank Peter van der Sman (Shell, The Netherlands), Alexander Supin (Institute of Ecology and Evolution, Moscow, Russia), Lee Miller (Odense University, Denmark), Nancy Vaughan Jennings (University of Bristol, UK), and two anonymous reviewers for their valuable comments on this manuscript. Funding for this project was obtained from The North Sea Directorate (DNZ, through Wanda Zevenboom, Contract No. 76/318701, IBO 4.2, 2000) of the Directorate-General of the Netherlands Ministry of Transport, Public Works and Water Management (RWS), and SEAMARCO, Harderwijk, The Netherlands. The porpoise's training and testing were authorized by the Netherlands Ministry of Agriculture, Nature Management and Fisheries, Department of Nature Management. Endangered Species Permit FEF27 06/2/98/0184.

Andersen, S. (**1970**). "Auditory sensitivity of the Harbour Porpoise *Phocoena phocoena*," in *Investigations on Cetacea*, edited by G. Pilleri (Institute for Brain Research, Bern), Vol. **3**, pp. 255–259.

Au, W. W. L. (**1993**). *The Sonar of Dolphins* (Springer-Verlag, New York).

Au, W. W. L., Kastelein, R. A., Rippe, T., and Schooneman, N. M. (**1999**). "Transmission beam pattern and echolocation signals of a harbor porpoise (*Phocoena phocoena*)," J. Acoust. Soc. Am. **106**, 3699–3705.

Au, W. W. L., Pawloski, J. L., Nachtigall, P. E., Blonz, M., and Gisiner, R. C. (**1995**). "Echolocation signals and transmission beam pattern of a false killer whale (*Pseudorca crassidens*)," J. Acoust. Soc. Am. **98**, 51–59.

Au, W. L., and Moore, P. W. B. (**1984**). "Receiving beam patterns and directivity indices of the Atlantic bottlenose dolphin *Tursiops truncatus*," J. Acoust. Soc. Am. **75**, 255–262.

Awbrey, F. T., Thomas, J. A., and Kastelein, R. A. (**1988**). "Low-frequency underwater hearing sensitivity in belugas, *Delphinapterus leucas*," J. Acoust. Soc. Am. **84**, 2273–2275.

Bibikov, N. G. (**1992**). "Auditory brainstem responses in the Harbor Porpoise (*Phocoena phocoena*)," in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 197–211.

Busnell, R. G., and Dziedzic, A. (**1967**). "Resultats metrologiques experimentaux de l'echolocation chez le *Phocaena phocaena* et leur comparaison avec ceux de certaines chauves-souris," in *Animal Sonar System, Biology and Bionics*, edited by R. G. Busnell (Lab. Physiol. Acoust. Jouey-en-Josas, France), pp. 307–356.

Cranford, T. W., Amundin, M., and Norris, K. S. (**1996**). "Functional morphology and homology in the odontocete nasal complex: Implications for sound generation," J. Morphol. **228**, 223–285.

Hall, J. D., and Johnson, C. S. (**1971**). "Auditory thresholds of a killer whale," J. Acoust. Soc. Am. **51**, 515–517.

Jacobs, D. W., and Hall, J. D. (**1972**). "Auditory thresholds of a fresh water dolphin, *Inia geoffrensis*, Blainville," J. Acoust. Soc. Am. **51**, 530–533.

Johnson, S. C. (**1967**). "Sound detection thresholds in marine mammals," in *Marine Bio-acoustics*, edited by W. N. Tavolga (Pergamon, New York), Vol. **2**, pp. 247–260.

Kamminga, C., and Wiersma, H. (**1981**). "Investigations of Cetacean Sonar II. Acoustical similarities and differences in odontocete sonar signals," Aquat. Mam. **8**, 41–62.

Kastelein, R. A., Bunskoek, P., Hagedoorn, M., Au, W. W. L., and de Haan, D. (**2002**). "Audiogram of a harbor porpoise (*Phocoena phocoena*) measured with narrow-band frequency-modulated signals," J. Acoust. Soc. Am. **112**, 334–344.

Kastelein, R. A., Goodson, A. D., Lien, J., and de Haan, D. (**1995**). "The effects of acoustic alarms on Harbour porpoise (*Phocoena phocoena*) behaviour," in *Harbour Porpoises, Laboratory Studies to Reduce Bycatch*, edited by P. E. Nachtigall, J. Lien, W. W. L. Au, and A. J. Read (De Spil, Woerden, The Netherlands), pp. 157–167.

Kastelein, R. A., de Haan, D., Goodson, A. D., Staal, C., and Vaughan, N. (**1997a**). "The effects of various sounds on a harbour porpoise (*Phocoena phocoena*)," in *The Biology of the Harbour Porpoise*, edited by A. J. Read, P. R. Wiepkema, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 367–383.

Kastelein, R. A., Au, W. W. L., Rippe, H. T., and Schooneman, N. M. (**1999**). "Target detection by an echolocating harbor porpoise (*Phocoena phocoena*)," J. Acoust. Soc. Am. **105**, 2493–2498.

Kastelein, R. A., de Haan, D., Vaughan, N., Staal, C., and Schooneman, N. M. (**2001**). "The influence of three acoustic alarms on the behaviour of harbour porpoises (*Phocoena phocoena*) in a floating pen," Mar. Env. Res. **52**, 351–371.

Kastelein, R. A., Rippe, H. T., Vaughan, N., Schooneman, N. M., Verboom, W. C., and de Haan, D. (**2000**). "The effect of acoustic alarms on the behavior of harbor porpoises (*Phocoena phocoena*) in a floating pen," Marine Mammal Sci. **16**, 46–64.

Kastelein, R. A., Hagedoorn, M., Au, W. W. L., and de Haan, D. (**2003**). "Audiogram of a striped dolphin (*Stenella coeruleoalba*)," J. Acoust. Soc. Am. **113**, 1130–1137.

Kastelein, R. A., Schooneman, N. M., Verboom, W. C., and Vaughan, N. (**1997b**). "The ability of a harbour porpoise (*Phocoena phocoena*) to discriminate between objects buried in sand," in *The Biology of the Harbour Porpoise*, edited by A. J. Read, P. R. Wiepkema, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 329–342.

Kastelein, R. A., Verboom, W. C., Muijsers, M., Jennings, N. V., and van der Heul, S. (**2005**). "The influence of acoustic emissions for underwater data transmission on the behaviour of harbour porpoises (*Phocoena phocoena*) in a floating pen," Mar. Env. Res. **59**, 287–307.

Koschinski, S., Culik, B. M., Damsgaard Hendriksen, O., Tregenza, N., Ellis, G., Jansen, C., and Kathe, G. (**2003**). "Behavioural reactions of free-ranging porpoises and seals to noise of a simulated 2 MW windpower generator," Mar. Ecol.: Prog. Ser. **265**, 263–273.

Ljungblad, D. K., Scoggins, P. D., and Gilmartin, W. G. (**1982**). "Auditory thresholds of a captive eastern Pacific bottlenosed dolphin, *Tursiops spp*.," J. Acoust. Soc. Am. **72**, 1726–1729.

Miller, P. J. O., Shapiro, A. D., Tyack, P. L., and Solow, A. R. (**2004**). "Call-type matching in vocal exchanges of free-ranging resident killer whales, *Orcinus orca*," Anim. Behav. **67**, 1099–1107.

Møhl, B., and Andersen, S. (**1973**). "Echolocation: High-frequency component in the click of the Harbour Porpoise (*Phocoena ph.* L.)," J. Acoust. Soc. Am. **53**, 1368–1372.

Nachtigall, P. E., Au, W. W. L., Pawloski, J. L., and Moore, P. W. B. (**1995**). "Risso's dolphin (*Grampus griseus*) hearing thresholds in Kaneohe Bay, Hawaii," in *Sensory Systems of Aquatic Mammals*, edited by R. A. Kastelein, J. A. Thomas, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 49–53.

Norris, K. S., and Harvey, G. W. (**1974**). "Sound transmission in the porpoise head," J. Acoust. Soc. Am. **56**, 659–664.

Popov, V. V., Ladygina, T. F., and Supin, A. Ya. (**1986**). "Evoked potentials of the auditory cortex of the porpoise, *Phocoena phocoena*," J. Comp. Physiol., A **158**, 705–711.

Popov, V., and Supin, A. (**1990**). "Electrophysiological studies of hearing in some cetaceans and a manatee," in *Sensory Abilities of Cetaceans, Laboratory and Field Evidence*, edited by J. A. Thomas and R. A. Kastelein (Plenum, New York), pp. 405–415.

Popov, V. V., Supin, A. Ya, and Klishin, V. O. (**1992**). "Electrophysiological study of sound conduction in dolphins," in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 269–276.

Renaud, D. L., and Popper, A. N. (**1975**). "Sound localization by the bottlenose porpoise, *Tursiops truncates*," J. Exp. Biol. **63**, 569–585.
Sauerland, M., and Dehnhardt, D. (**1998**). "Underwater audiogram of a Tucuxi (*Sotalia fluviatilis guianensis*)," J. Acoust. Soc. Am. **103**, 1199–1204.
Schevill, W. E., Watkins, W. A., and Ray, C. (**1969**). "Click structure in the porpoise, *Phocoena phocoena*," J. Mammal. **50**, 721–728.
Schlundt, C. E., Carder, D. A., and Ridgway, S. H. (1998). "The effect of projector position on the hearing thresholds of dolphins (*Tursiops truncatus*) at 2, 8 and 12 kHz," Poster at the Biological Sonar Conference, Carvoeiro, Portugal, March 1998.
Supin, A. Ya., and Popov, V. V. (**1993**). "Direction-dependent spectral sensitivity and interaural spectral difference in a dolphin: Evoked potential study," J. Acoust. Soc. Am. **93**, 3490–3495.
Szymanski, M. D., Bain, D. E., Kiehl, K., Pennington, S., Wong, S., and Henry, K. R. (**1999**) "Killer whale (*Orcinus orca*) hearing: Auditory brainstem response and behavioral audiograms," J. Acoust. Soc. Am. **84**, 936–940.
Thomas, J. A., Chun, N., Au, W. W. L., and Pugh, K. (**1988**). "Underwater audiogram of a false killer whale (*Pseudorca crassidens*)," J. Acoust. Soc. Am. **84**, 936–940.
Tremel, D. P., Thomas, J. A., Ramirez, K. T., Dye, G. S., Bachman, W. A., Orban, A. N., and Grimm, K. K. (**1998**). "Underwater hearing sensitivity of a Pacific white-sided dolphin, *Lagenorhynchus obliquidens*," Aquat. Mam. **24**, 63–69.
Urick, R. J. (**1983**). *Principles of Underwater Sound* (McGraw-Hill, New York) p. 42.
Verboom, W. C., and Kastelein, R. A. (**1995**). "Acoustic signals by Harbour porpoises (*Phocoena phocoena*)," in *Harbour Porpoises, Laboratory Studies to Reduce Bycatch*, edited by P. E. Nachtigall, J. Lien, W. W. L. Au, and A. J. Read (De Spil, Woerden, The Netherlands), pp. 1–39.
Verboom, W. C., and Kastelein, R. A. (**1997**). "Structure of harbour porpoise (*Phocoena phocoena*) click train signals," in *The biology of the harbour porpoise*, edited by A. J. Read, P. R. Wiepkema, and P. E. Nachtigall (De Spil, Woerden, The Netherlands), pp. 343–362.
Verboom, W. C., and Kastelein, R. A. (**2003**). "Structure of harbour porpoise (*Phocoena phocoena*) echolocation signals with high repetition rates," in *Echolocation in Bats and Dolphins*, edited by J. A. Thomas, C. Moss, and M. Vater (University of Chicago Press, Chicago), pp. 40–43.
Voronov, V. A., and Stosman, I. M. (**1983**). "On sound perception in the dolphin *Phocoena phocoena*," J. Evol. Biochem. Physiol. **18**, 352–357 (Zh. Evol. Biokhim Fiziol **18**, 499–506, **1982**).
Wang, D., Wang, K., Xiao, Y., and Sheng, G. (**1992**). "Auditory sensitivity of a Chinese river dolphin, *Lipotes vexillifer*," in *Marine Mammal Sensory Systems*, edited by J. A. Thomas, R. A. Kastelein, and A. Ya. Supin (Plenum, New York), pp. 213–221.
White, M. J., Jr., Norris, J., Ljungblad, D. K., Baron, K., and di Sciara, G. (1978). "Auditory thresholds of two beluga whales (*Delphinapterus leucas*)," HSWRI Techn. Rep. No. 78-109, Hubbs Marine Research Institute, San Diego, California.
Yurick, D. B., and Gaskin, D. E. (**1987**). "Morphometric and meristic comparisons of skulls of harbour porpoise *Phocoena phocoena* (L.) from the North Atlantic and North Pacific," Ophelia **27**, 53–75.
Zaytseva, K. A., Akopian, A. I., and Morozov, V. P. (**1975**). "Noise resistance of the dolphin auditory analyzer as a function of noise direction," Biofizika **20**, 519–521.

A passive acoustic monitoring method applied to observation and group size estimation of finless porpoises

Kexiong Wang
Institute of Hydrobiology, the Chinese Academy of Sciences, Wuhan 430072, People's Republic of China and Graduate School of the Chinese Academy of Sciences, Beijing, 100039, People's Republic of China

Ding Wang[a)]
Institute of Hydrobiology, the Chinese Academy of Sciences, Wuhan 430072, People's Republic of China

Tomonari Akamatsu
National Research Institute of Fisheries Engineering, Ebidai, Hasaki, Kashima, Ibaraki 314-0421, Japan

Songhai Li and Jianqiang Xiao
Institute of Hydrobiology, the Chinese Academy of Sciences, Wuhan 430072, People's Republic of China and Graduate School of the Chinese Academy of Sciences, Beijing, 100039, People's Republic of China

(Received 21 December 2004; revised 25 April 2005; accepted 11 May 2005)

The present study aimed at determining the detection capabilities of an acoustic observation system to recognize porpoises under local riverine conditions and compare the results with sighting observations. Arrays of three to five acoustic data loggers were stationed across the main channel of the Tian-e-zhou Oxbow of China's Yangtze River at intervals of 100–150 m to record sonar signals of free-ranging finless porpoises (*Neophocaena phocaenoides*). Acoustic observations, concurrent with visual observations, were conducted at two occasions on 20–22 October 2003 and 17–19 October 2004. During a total of 42 h of observation, 316 finless porpoises were sighted and 7041 sonar signals were recorded by loggers. The acoustic data loggers recorded ultrasonic signals of porpoises clearly, and detected the presence of porpoises with a correct detection level of 77.6% and a false alarm level of 5.8% within an effective distance of 150 m. Results indicated that the stationed passive acoustic observation method was effective in detecting the presence of porpoises and showed potential in estimating the group size. A positive linear correlation between the number of recorded signals and the group size of sighted porpoises was indicated, although it is faced with some uncertainty and requires further investigation.
 [DOI: 10.1121/1.1945487]

PACS number(s): 43.80.Ka, 43.80.Jz, 43.66.Gf [WA] Pages: 1180–1185

I. INTRODUCTION

The finless porpoises (*Neophocaena phocaenoides*) in the Yangtze River exhibit up and downstream movements and may also move between the mainstream and adjacent lakes (Wei *et al.*, 2002b). It is important to develop a practical way to monitor such movement patterns over the long term to understand more about the ecology of the species and develop conservation measures. Satellite-relayed tracking (Zhang *et al.*, personal communication) was conducted on finless porpoises in the Yangtze River. The porpoises were fitted with cotton cloth vests designed to fit around the front part of the animals' body, and tags were sewn onto the vests. As this species has no dorsal fin for tags to be securely attached to, the vests with tags came off very quickly. Detection of underwater sounds produced by individuals and groups has been identified as an effective method to monitor cetaceans in the wild (Akamatsu *et al.*, 1998; van Parijs *et al.*, 1998; Jefferson *et al.*, 2002). It was suggested that in some cases sound emission rates could even be used to estimate population abundance (van Parijs *et al.*, 2002). The finless porpoise produces sonar signals frequently (Akamatsu *et al.*, 2000; Akamatsu *et al.*, 2005a), and was once observed effectively by a vessel-based acoustic survey in the Yangtze River (Akamatsu *et al.*, 2001). With reliable correct detection and small false alarm levels, the acoustic survey system detected the presence of porpoises within 300 m of the vessel. A porpoise click detector (POD) was once used to make an estimate of the track line detection probability for visual ship surveys of finless porpoises in Hong Kong and adjacent waters of China (Jefferson *et al.*, 2002). The POD was able to detect most sightings that were visually detected within a 250-m perpendicular distance. These studies suggest that it might be promising to monitor the presence of porpoises using an on-site acoustic data logger array and to estimate group size based on the number of detected sonar signals. In addition, a fixed system that can automatically record signals of porpoises for long periods of time has the added advantage that it does not disturb the animals. So far, the reliability of such stationed acoustical detection systems for porpoises

[a)]Electronic mail: wangd@ihb.ac.cn

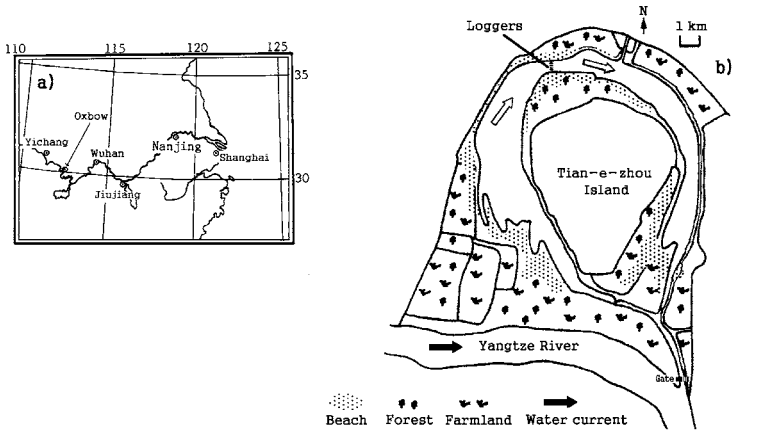


FIG. 1. The geographic location of the study site and shape of the Tian-e-zhou Oxbow which periodically connects to the Yangtze River through installed gate. (a) Map showing Oxbow location at middle reach of the Yangtze River; (b) Oxbow map (bar=1 km) showing position of loggers and dimensions of different types of habitat as indicated by symbols. Open arrow =0–0.5 m/s in oxbow; Solid arrow=Yangtze River flow (higher speed).

has not yet been evaluated quantitatively with concurrent visual observations.

In the present study, passive acoustic observation using a stationed acoustic data logger array, concurrent with visual observations, were recorded in Tian-e-zhou Oxbow of China's Yangtze River, to determine the detection probability of the acoustic observation system on the presence of porpoises, and the possible relationship between the number of sonar signals detected acoustically and number of porpoises sighted.

II. MATERIALS AND METHODS

A. Study area and species

The Tian-e-zhou Oxbow is an old course of the Yangtze River, and is located in Shishou of the Hubei Province, central China [Fig. 1(a)]. It was cut off naturally from the mainstream in 1972, and was approved as a seminatural reserve for *ex situ* protection of baiji (*Lipotes vexillifer*) and finless porpoises by the Chinese government in 1992 (Zhang *et al.*, 1995; Wang *et al.*, 2000; Wei *et al.*, 2002a). It is 21 km long and 1–2.0 km wide, and has a maximum depth of 20 m. The water velocity ranges from 0 to 0.5 m/s and this current speed changes with the operational mode of the constructed gate and the water level of the Yangtze River [Fig. 1(b)]. When this study was carried out, the porpoise group in the oxbow consisted of 23 individuals (Wei *et al.*, 2002a). Some porpoises were introduced to the reserve from the mainstream of the Yangtze River during the 1990s, and others were born in the oxbow (Zhang *et al.*, 1995; Wang *et al.*, 2000; Wei *et al.*, 2002a).

B. Acoustic data loggers

Seven sets of miniature pulse event data loggers were deployed. The stereo data logger (W20-AS, Little Leonardo, Tokyo, Japan, diameter: 22 mm; length: 122 mm; weight: 77 g) was equipped with two hydrophones (System Giken Co. Ltd., Japan, −210 dB/1 V peak-to-peak *re* 1 μPa sensitivity) and a bandpass filter of 70–300 kHz to eliminate background noise. It was equipped with an analog–digital converter, a 256-MB flash memory and a CPU (PIC18F6620, Microchip, USA). The chips were integrated inside a pressure-resistant aluminum cylinder. The detection threshold of the loggers at the on-axis direction was 129 dB whereas it was 140 dB when the sound came from the 90° off-axis direction. The acoustic data logger continuously detected underwater sounds, and as soon as the received amplitude exceeded the threshold, pulse events were recorded. It recorded the intensity, timing and the source direction of the ultrasonic pulses for up to 60 h with a sampling rate of 2000 events per second.

C. Logger array and visual observations

Arrays of loggers were stationed across the oxbow channel [Fig. 1(b)]. One array of five data loggers was deployed from October 20 to 22, 2003, and two arrays one with three and one with four loggers were deployed between October 17 and 19, 2004. The distance between neighboring loggers was 100 m in 2003, and 150 m in 2004. The loggers were placed from the north shore of the oxbow to the opposite south shore. Water depth profile at the logger locations are shown in Fig. 2. Each logger was fixed to a float and a weight, and was stationed at a depth of 0.5 m. Two visual observers on the north bank (7.5 m above the water surface) searched the up and downstream sides of the array. The observers recorded the time, the minimum estimated group size and the estimated distance of the group from the north shore as the porpoises moved through the array. The range to the passing porpoise was estimated visually, but the distance of each logger float to the north shore was measured prior to observations using a portable GPS and therefore were used as a basis for estimating distance.

The group size was recorded as the minimum estimated number of animals that swam through the acoustic observation array in each 1-min bin. A correct detection by the acoustic observation system was defined as the detection of more than a cutoff number of pulses within a 1-min time window either before or after the moment of the visual sighting of porpoises. A false alarm was defined as the detection of more than a cutoff number of pulses without any visual sighting within a 1-min time window both before and after the acoustic detections.

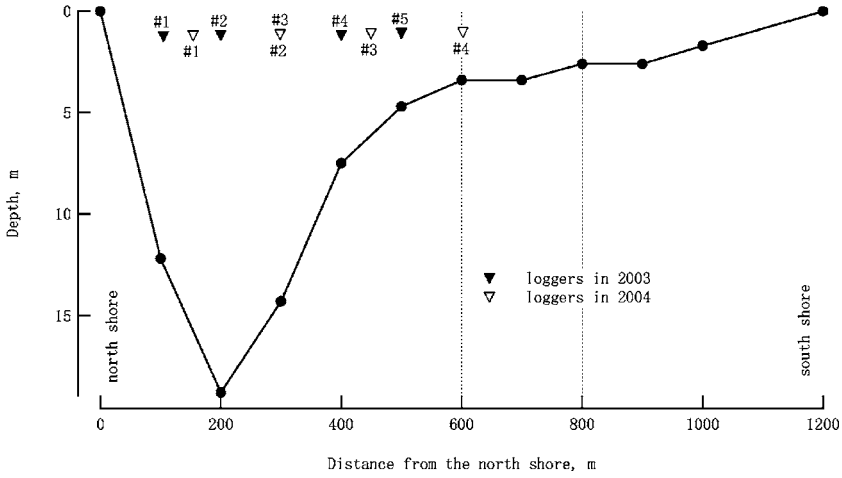


FIG. 2. Cross section of the bottom profile at the transect with loggers positioned in 2003 and 2004.

D. Data analysis

We developed a data analysis program based on Igor (Wavematrics, Lake Oswego, Oregon 97035, USA) for the study to calculate interpulse intervals of the recorded pulse events and to count the number of pulses per minute. The pulses with irregular changes of successive interpulse intervals (two times or more and half or less than the previous one) were considered as noise (Akamatsu *et al.*, 1998; Akamatsu *et al.*, 2001), and were eliminated. The remaining pulses, whether they were within a pulse train or not, were considered as individual sonar pulses, and were used for calculation of the number of pulses per minute. Additionally, we correlated pulse-number with animal numbers, rather than pulse-trains with animal numbers, and therefore, we did not calculate the number of pulse-trains per minute in this study.

III. RESULTS

A. Presence detection of porpoises

Porpoises were generally observed moving up or downstream between the north shore and the logger at the south end of the array. Fewer porpoises were observed moving between the logger at the south end of the array and the south shore where the water was shallow.

During 42 h of observation, 316 porpoises were sighted at 99 sighting events, and 7041 individual sonar pulses were recorded by acoustic data loggers. The interpulse intervals, whether they were within a pulse train or not, were mostly between 22–24 ms and the changes of the typical successive interpulse intervals were mostly less than 20 ms (Fig. 3). Comparisons between the number of finless porpoises sighted and the pulse signals recorded every minute are presented in Fig. 4, showing results of the 3-day observations in 2003. A clear coincidence between visual observation and acoustic data logger detection is shown.

The receiver operating characteristic (ROC) curve based on the cutoff number of pulses in a minute is shown in Fig. 5. At the cutoff number of pulses (9), correct detection (77.6%) and small false alarm levels (5.8%) were indicated.

The number of pulses detected by the loggers and the distance between porpoises and loggers are shown in Fig. 6. In total, 68.8% of the correct acoustic detections (more than 9 pulses in a minute) were within 100 m of the logger, 79.7% were within 150 m, and the maximum distance of correct detection of porpoises was 250 m. On average, the maximum sound pressure level of pulses corresponding with visual detection of porpoises within 50 m was 140.4 dB $p-p$ *re* 1 μPa.

B. Group sizes and ultrasonic events

There were 23 individuals in the oxbow, but smaller groups, especially groups consisted of 1–3 individuals were

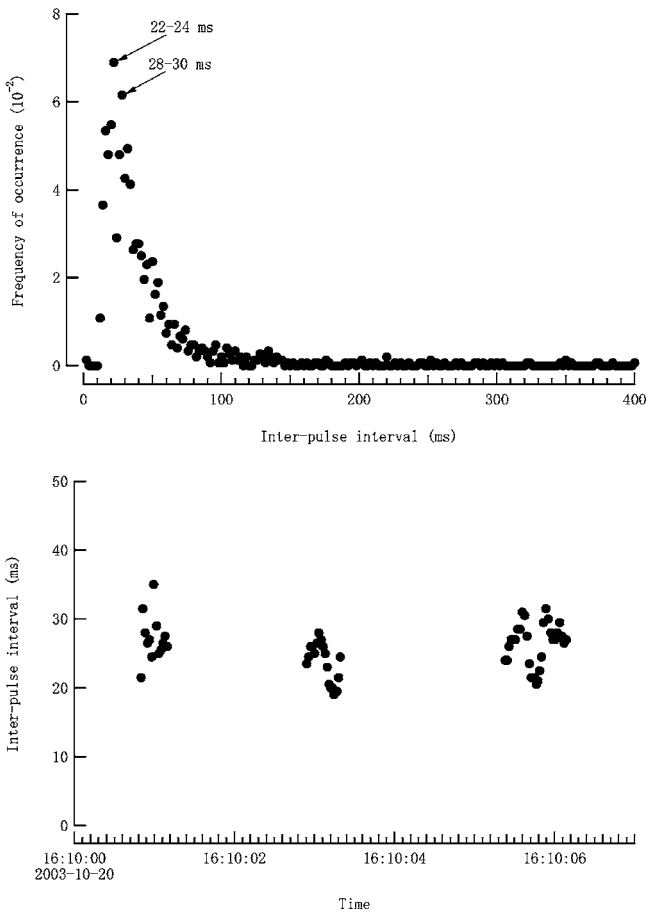


FIG. 3. Frequency of occurrence in interpulse intervals of the porpoises' sonar signals (upper). The animals frequently produced 22–30 ms pulse intervals. Change of pulse intervals in three pulse trains (lower). The pulse intervals were fluctuated in a train and the difference of successive pulse intervals was mostly less than 20 ms.

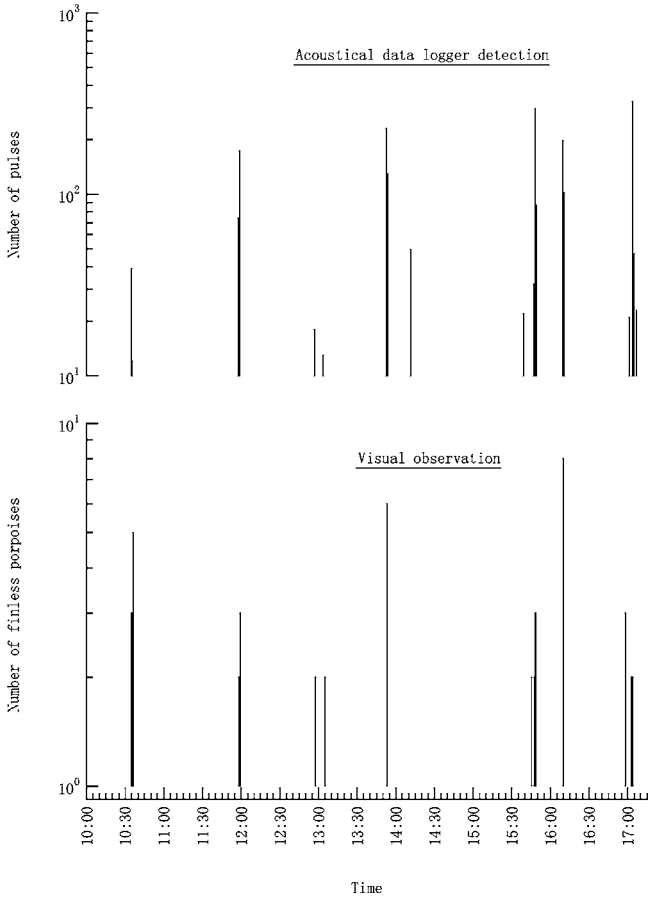


FIG. 4. Comparisons between numbers of porpoises and pulse signals every minute during a 3-day observation period in 2003. A clear coincidence between visual observation and acoustic data logger detection was indicated.

usually observed moving through the array. The relationship between the number of pulses recorded by loggers (more than 9 pulses in a minute) and porpoises sighted is shown in Fig. 7. A very limited positive linear correlation between the number of pulses and porpoises exists ($n=34, R^2=0.1609, p=0.019$), showing the potential of the methodology, however the variability is presently too high to draw firm conclusions.

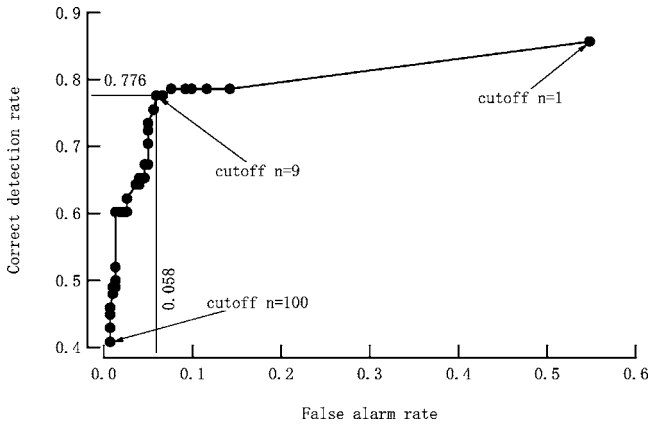


FIG. 5. ROC curve of the present acoustic array system. The correct detection rate decreased rapidly for a cut-off number >9, while the false alarm rate increased rapidly for a cut-off number <9. Therefore, 9 was chosen as the best compromise between maximizing correct detections and minimizing false alarms.

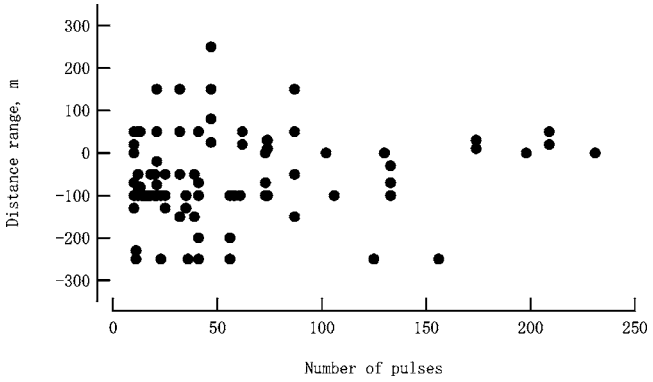


FIG. 6. Numbers of pulses detected by loggers and distance variances between porpoises and loggers. All the correct acoustical detections (more than 9 pulses in a minute) were within 250 m, and mostly within 150 m from the porpoises.

IV. DISCUSSION

The finless porpoise clicks were high frequency narrowband ultrasonic pulses, and the peak frequencies of typical clicks ranged from 87 to 145 kHz with an average of 125±6.92 kHz, and the durations ranged from 30 to 122 μs with an average of 68±14.12 μs (Akamatsu *et al.*, 1998; Li *et al.*, 2005). Therefore, the bandpass filter of 70–300 kHz built-in the logger was effective in both passing signals and eliminating noises. Porpoise pulses had a wide variety of intervals from 8 to 400 ms, and the difference of successive interpulse intervals was normally less than 20 ms (Akamatsu *et al.*, 1998), indicating that the event-sampling rate of the loggers in this study was adequate to record porpoise pulse events and the pulses recorded by the loggers in this study were typical sonar signals of porpoises (Fig. 3).

According to field observations, most porpoise groups traversed the array in 30–60 s. Thus, the 1-min time window used in the definition of group size was reasonable to partition different groups. On average, the porpoise produced 14.5–19.1 click trains in a minute, and a click train consists of several tens of ultrasonic pulses (Akamatsu *et al.*, 2001). Even though the acoustic logger might fail to detect some sonar signals with intensities below the detection threshold of the equipment, the 1-min time unit was adequate for analysis of pulse counts of the porpoise. A porpoise might

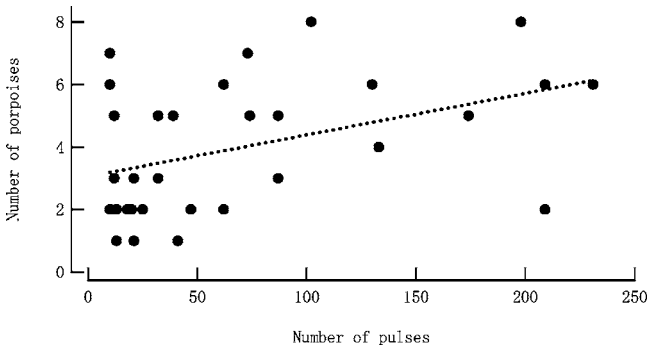


FIG. 7. Correlation between number of pulses and number of porpoises. A very limited positive linear correlation was indicated that the number of porpoises (Y) was about 0.0133 times the number of pulses (X) ($n=34, R^2=0.1609, p=0.019$).

produce sound before or after visual identification, and then the two 1-min time windows used in definition of correct detection were adequate to detect the sound produced by the porpoise before or after visual identification. On the other hand, the two 1-min time windows used in the definition of false alarm were necessary to exclude any possibility of porpoise existence around the array before and after acoustic detection, since the average dive time of deep dives of the porpoise was 32.6–70.9 s (Akamatsu *et al.*, 2002). Additionally, some periods of silence between sonar signal productions were observed in the porpoise, sometimes lasting over 10 s (Akamatsu *et al.*, 2005a). Even if porpoises were silent while passing through the array, they would still produce some sounds before and after the silence, and the sounds would be detected and recorded by the loggers. Therefore, the existence of the period of silence might have no obvious effect on acoustic detection of the presence of the porpoise.

In brief, this study indicated that high frequency signals of finless porpoises in the oxbow can be detected by the logger array, and the system is useful for acoustic observation of porpoises in the wild. A previous acoustic survey with vessel-based hydrophones had a correct detection rate of 82%, a false alarm level of 0.9%, and a cutoff number of 15 pulses (Akamatsu *et al.*, 2001). In comparison to this survey, our acoustic data logger array had a somewhat lower correct detection rate (77.6%), a higher false alarm rate (5.8%), and a smaller cut-off number (9). The smaller cutoff pulse number in the oxbow might have been caused by the closer distance of porpoises to the loggers. In the Yangtze River, the distance of porpoises to hydrophones usually was more than 100 m, but much less in our study in the Oxbow.

A ROC curve is based on the performance of a detector against a known number of stimuli. In this study, both the visual and the acoustic detections lacked controls, so there was no way of testing either method against a control. However, the aim of this study was to evaluate the feasibility of acoustic detection of porpoises evaluated against concurrent visual observations, and to compare both correct detection and false alarm of the logger method with those of a previous hydrophone method (Akamatsu *et al.*, 2001). Therefore, the comprehensible ROC curve was needed and was relatively useful in this study.

The detected average maximum sound pressure level of pulses within 50 m was 140.4 dB in the present study. According to the calculation of attenuations (Akamatsu *et al.*, 2001), the source sound level of the porpoise was estimated to be less than 167.4 dB. The detection threshold of the loggers was 129 dB at the on-axis direction and 140 dB at the 90° off-axis direction. The theoretical maximum detection distance of the logger array was therefore 55 m at 140 dB detection threshold level and 692 m at 129 dB detection threshold level. The actual maximum detection distance (250 m) of the logger array was within the theoretical detection range.

It is likely that an acoustic array stationed near the surface would miss some animals in deep water because of higher sound attenuation, although the porpoise used sonar almost continuously with relatively constant sensing effort at each swimming depth (2, 4, and 8 m) (Akamatsu *et al.*, 2005b). The maximum depth of the oxbow is 20 m, and the average dive depth of porpoises in the oxbow was 1.7 m and the maximum dive depth was 11.5 m (Akamatsu *et al.*, 2000). Theoretically, the loggers could detect acoustically the porpoises even if they pass in deep waters beneath loggers because the maximum dive depth was far less than 55 m of the theoretical maximum detection distance of the logger array even the signals came from a 90° off-axis direction, and therefore the effect of depth on acoustic detection was not considered in this study.

Porpoise sonar emission might be more frequent at nighttime than in the daytime (Wang, 1996). The interpulse interval of sonar was longer in open waters than in small closed waters, and the change of interpulse intervals in a tank ($8 \times 5 \times 2$ m^3) was monotonous decrement, and was fluctuated in open waters (Akamatsu *et al.*, 1998). Porpoises might use sonar with long interpulse intervals for navigation and use sonar with short interpulse intervals in approaching prey. Therefore, both environmental condition and porpoise behavior may influence acoustic detections. The present study was carried out only in day time and in a fixed location, and the possible effects of environmental condition and porpoise behavior on acoustic detections of loggers were not considered. However, if waters, working time, and porpoise behavior vary greatly, the acoustic detection of loggers might be fluctuated greatly, and the effects on pulse reception rate of loggers should be considered.

Calls of Pacific humpback dolphin have been used to estimate the abundance of an inshore population (van Parijs *et al.*, 2002). The frequency range of the recordings was limited (5–22 kHz) and the estimation of group size was based on visible spectrograms of all types of sounds (van Parijs *et al.*, 2002). Low frequency vocalizations of the finless porpoise were not observed in the present study because of the low cut frequency filter of the data loggers, although the species might produce low frequency sounds (Wang, 1996). Moreover, the low frequency underwater noises in the Yangtze River would interfere with recordings of low frequency sounds of the porpoise. Therefore, using low frequency sounds to estimate abundance of porpoises in the river is probably not appropriate. The array of loggers employed in this study was resistant to low frequency background noises and recorded only higher frequency signals of porpoises, and it could be applied effectively to acoustic detection of presence of porpoise in the Yangtze River.

The clear coincidence between visual observation and acoustic data logger detection (Fig. 4) suggest that the passive acoustical method might be potential in group size estimation of the porpoise in the wild, and the very limited correlation between numbers of pulses and number of porpoises indicated in Fig. 7 seemed to support this. However, further research is especially needed to establish a reliable correlation between number of pulses and animal sightings to improve the regression formula for estimation of the number of porpoises. Additionally, the intercept (about three individuals sighted, but no acoustic detection) indicated in Fig. 7 seemed to show that acoustic detection might be biased towards larger groups and smaller groups would be missed more frequently. In other words, the array would underestimate the

number of porpoises. Although the correlation was imperfect, the positive linear tendency between the number of detected pulses and group size existed in acoustic detections both of finless porpoise and of Pacific humpback dolphin. The tendency seemed to suggest that all the group members, rather than one or some members produced sounds while the members navigated together.

Further application of the acoustic array system in the Yangtze River is anticipated. To do that, a large battery capacity for longer operation is needed. A potential application would be the monitoring of possible movement of porpoises between the mainstream and adjacent lakes. For the purpose, two parallel arrays of loggers can be stationed across the mouth of the lake to record the signals of porpoises moving through two arrays. The difference of signals detection time between two arrays would provide the movement direction of the animals to be determined. Simultaneously the number of pulses helps to estimate the group size of the moving porpoises. This study has shown the potential for stationed passive acoustic monitoring in studies of cetacean movement, ecology, and conservation.

ACKNOWLEDGMENTS

We wish to thank Q. Zhao, Z. Wei, X. Wang, X. Zhang, J. Zheng, B. Yu, X. Q. Zhang, X. Y. Wang, W. Dong, X. Zhao, M. Wu, J. Peng, and T. Morisaka for their cooperation in the study. We thank Dr. Harald Rosentbal and Ms. Gill Braulik for their modification and editorial assistance on an early version of this manuscript. We would also like to express our appreciation to two anonymous reviewers for their constructive comments and inputs on this manuscript. The Field Station of Tian-e-zhou Baiji Nature Reserve, National Institute of Polar Research of Japan, and Little Leonardo Co. greatly supported our experiments. This research was supported by the Chinese Academy of Sciences (CAS) and Institute of Hydrobiology, CAS (No. KSCX2-SW-118 and 220103) and Program for Promotion of Basic Research Activities for Innovative Biosciences of Japan.

Akamatsu, T., Wang, D., Nakamura, K., and Wang, K. (**1998**). "Echolocation range of captive and free-ranging baiji (*Lipotes vexillifer*), finless porpoise (*Neophocaena phocaenoides*), and bottlenose dolphin (*Tursiops truncatus*)," J. Acoust. Soc. Am. **104**, 2511–2516.

Akamatsu, T., Wang, D., Wang, K., and Naito, Y. (**2000**). "A method for individual identification of echolocation signals in free-ranging finless porpoises carrying data loggers," J. Acoust. Soc. Am. **108**, 1353–1356.

Akamatsu, T., Wang, D., Wang, K., and Wei, Z. (**2001**). "Comparison between visual and passive acoustic detection of finless porpoises in the Yangtze River, China," J. Acoust. Soc. Am. **109**, 1723–1727.

Akamatsu, T., Wang, D., Wang, K., Wei, Z., Zhao, Q., and Naito, Y. (**2002**). "Diving behavior of freshwater finless porpoises (*Neophocaena phocaenoides*) in an oxbow of the Yangtze River, China," ICES J. Mar. Sci. **59**, 438–443.

Akamatsu, T., Wang, D., Wang, K., and Naito, Y. (**2005a**) "Biosonar behavior of free-ranging porpoises," Proc. R. Soc. London, Ser. B **272**, 797–801.

Akamatsu, T., Wang, D., and Wang, K. (**2005b**) "Off-axis sonar beam pattern of free-ranging finless porpoises measured by a stereo pulse event data logger," J. Acoust. Soc. Am. **117**, 3325–3330.

Jefferson, T. A., Hung, S. K., Law, L., Torey, M., and Tregenza, N. (**2002**). "Distribution and abundance of finless porpoises in Hong Kong and adjacent waters of China," Raffles Bull. Zool., Suppl. **10**, 43–55.

Li, S., Wang, K., Wang, D., and Akamatsu, T. (**2005**). "Echolocation signals of the free-ranging Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientialis*)," J. Acoust. Soc. Am. **117**, 3288–3296.

van Parijs, S. M., Smith, J., and Corkeron, P. J. (**2002**). "Using calls to estimate the abundance of inshore dolphins: A case study with Pacific humpback dolphins Sousa chinensis," J. Appl. Ecol. **39**, 853–864.

van Parijs, S. M., Thompson, P. M., Hastie, F. D., and Bartels, B. A. (**1998**). "Modification and deployment of a sonobuoy for recording underwater vocalizations from marine mammals," Marine Mammal Sci. **14**, 310–315.

Wang, D. (**1996**). "A preliminary study on sound and acoustic behavior of the Yangtze River finless porpoise, *Neophocaena phocaenoides*," Acta Hydrobiol. Sinica **20**, 127–133.

Wang, D., Liu, R., Zhang, X., Yang, J., Wei, Z., Zhao, Q., and Wang, X. (**2000**). "Status and conservation of the Yangtze finless porpoises," in *Biology and Conservation of Freshwater Cetaceans in Asia*, edited by R. R. Reeves, B. D. Smith, and T. Kasuya (Occas. Pap. IUCN Spec. Surv. Commn.), IUCN Spec. Surv. Commn., Gland, Switzerland, Vol. 23, pp. 81–85.

Wei, Z., Wang, D., Kuang, X., Wang, K., Wang, X., Xiao, J., Zhao, Q., and Zhang, X. (**2002a**). "Observations on behavior and ecology of the Yangtze finless porpoise (*Neophocaena phocaenoides asiaeorientalis*) group at Tian-e-Zhou Oxbow of the Yangtze River," Raffles Bull. Zool., Suppl. **10**, 97–103.

Wei, Z., Wang, D., Zhang, X., Zhao, Q., Wang, K., and Kuang, X. (**2002b**). "Population size, behavior, movement pattern and protection of Yangtze finless porpoise at Balijiang section of the Yangtze River," Resour. Environ. Yangtze Basin **11**, 427–432.

Zhang, X., Wei, Z., Wang, X., Yang, J., and Chen, P. (**1995**). "Studies on the feasibility of establishment of a semi-natural reserve at Tian-e-zhou (swan) oxbow for baiji, *Lipotes vexillifer*," Acta Hydrobiol. Sinica **19**, 110–123.

The dependencies of phase velocity and dispersion on trabecular thickness and spacing in trabecular bone-mimicking phantoms

Keith A. Wear[a)]
U.S. Food and Drug Administration, Center for Devices and Radiological Health, HFZ-142, 12720 Twinbrook Parkway, Rockville, Maryland 20852

(Received 25 February 2005; revised 24 April 2005; accepted 2 May 2005)

Frequency-dependent phase velocity was measured in trabecular-bone-mimicking phantoms consisting of two-dimensional arrays of parallel nylon wires (simulating trabeculae) with thicknesses ranging from 152 to 305 μm and spacings ranging from 700 to 1000 μm. Phase velocity varied approximately linearly with frequency over the range from 400 to 750 kHz. Dispersion was characterized by the slope of a linear least-squares regression fit to phase velocity versus frequency data. The increase in phase velocity (compared with that in water) at 500 kHz was approximately proportional to the (1) square of trabecular thickness, (2) inverse square of trabecular spacing, and (3) volume fraction occupied by nylon wires. The first derivative of phase velocity with respect to frequency was negative and exhibited nonlinear, monotonically decreasing dependencies on trabecular thickness and volume fraction. The dependencies of phase velocity and its first derivative on volume fraction in the phantoms were consistent with those reported in trabecular bone. [DOI: 10.1121/1.1940448]

PACS number(s): 43.80.Qf [FD] Pages: 1186–1192

I. INTRODUCTION

Bone sonometry is now an accepted method for diagnosis of osteoporosis (Laugier, 2004). Speed of sound (SOS) in trabecular bone is highly correlated with bone mineral density (Rossman *et al.*, 1989; Tavakoli and Evans, 1991; Zagzebski *et al.*, 1991; Njeh *et al.*, 1996; Laugier *et al.*, 1997; Nicholson *et al.*, 1998; Hans *et al.*, 1999; Trebacz and Natali, 1999), which is an indicator of systemic osteoporotic fracture risk (Cummings *et al.*, 1993). Calcaneal ultrasonic measurements (SOS combined with broadband ultrasonic attenuation or BUA) have been shown to be predictive of hip fractures in women in prospective (Hans *et al.*, 1996; Bauer *et al.*, 1997; Huopio *et al.*, 2004) and retrospective (Schott *et al.*, 1995; Turner *et al.*, 1995; Glüer *et al.*, 1996; and Thompson *et al.*, 1998) studies. SOS and BUA have also been shown to be as effective as central dual energy x-ray absorptiometry in identification of women at high risk for prevalent osteoporotic vertebral fractures (Glüer *et al.*, 2004).

Despite the clinical utility of SOS, the mechanisms responsible for variations of SOS in trabecular bone are not well understood yet. This paper describes a phantom study designed to provide insight into the relationship between SOS and trabecular microarchitecture. This includes an investigation of the role of microarchitecture in determining dispersion. Unlike soft tissues, which typically exhibit positive dispersion (phase velocity increasing with ultrasonic frequency) (O'Donnell *et al.*, 1981; Waters *et al.*, 1999), trabecular bone exhibits negative dispersion (Nicholson *et al.*, 1996; Strelitzki and Evans, 1996; Droin *et al.*, 1998; Wear, 2000a, 2001a).

[a)]Electronic mail: kaw@cdrh.fda.gov

II. METHODS

A. Phantoms

Seven phantoms consisting of parallel nylon wires (simulating trabeculae) in two-dimensional rectangular grid arrays (custom-built by Computerized Imaging Reference Systems, Norfolk, VA) were interrogated. See Fig. 1. The nylon wire diameter corresponded to trabecular thickness, which in the standard nomenclature for bone histomorphometry is denoted by Tb.Th (Parfitt *et al.*, 1987). Four values for Tb.Th were used: 152, 203, 254, and 305 μm. (These values correspond to 0.006, 0.008, 0.010, and 0.012 in., which are readily available nylon wire thicknesses.) See Table I. The mean value for Tb.Th for human calcaneus is 127 μm (Ulrich *et al.*, 1999).

Trabecular spacing, s, is given by

$$s = \mathrm{Tb}\,.\,\mathrm{Sp} + \mathrm{Tb.Th}, \tag{1}$$

where Tb.Sp is trabecular separation. Four values for s were used: 700, 800, 900, and 1000 μm. The mean value for Tb.Sp in human calcaneus is 684 μm (Ulrich *et al.*, 1999), which corresponds to a mean value for s equal to 684 μm+127 μm=811 μm.

The volume fraction, VF, occupied by wire (trabeculae) is given by

$$\mathrm{VF} = \frac{\pi(\mathrm{Tb.Th}/2)^2}{s^2}. \tag{2}$$

VF in bone is often denoted by BV/TV, the ratio of bone volume to tissue volume (Parfitt *et al.*, 1987). Porosity, β, is given by $\beta = 1 - \mathrm{VF}$. The range of VF spanned by the seven phantoms (1.8%–11.4%) roughly corresponds to the

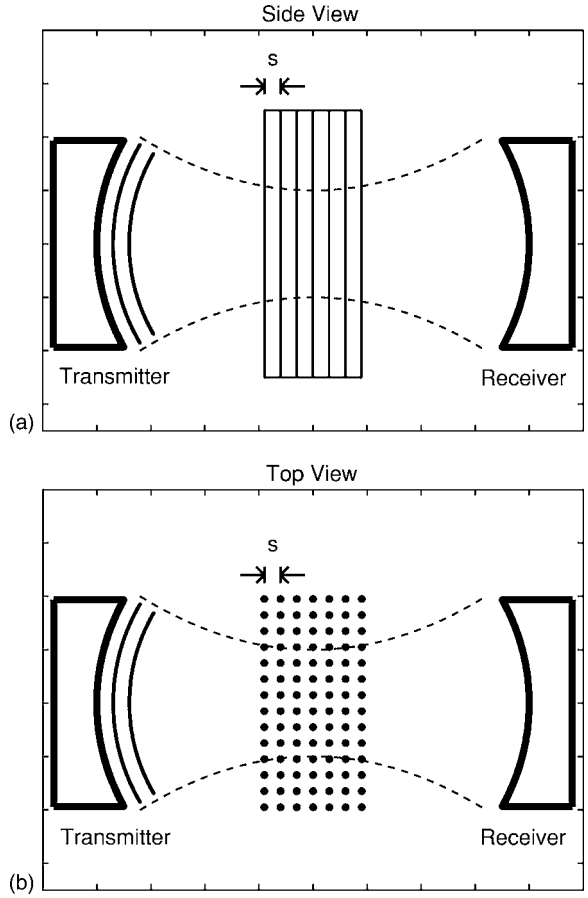


FIG. 1. (a) Side view and (b) top view of experimental setup.

range reported for human calcaneus, 2%–14% (Wear, 2005).

The grid arrays were immersed in a water tank so that water filled the spaces between the wires. This phantom design was somewhat simplistic in that it (1) substituted nylon for mineralized bone and water for marrow, (2) contained only rod-like structures and not plate-like structures that are also known to exist in trabecular bone, and (3) was perfectly periodic unlike trabecular bone, which is far less regular in structure. Its relevance, as discussed in the following, depended on its ability to reproduce frequency-dependent phase velocity properties similar to those observed in trabecular bone.

Some justification for the substitution of water for marrow is provided by the fact that the longitudinal sound speed in water (1480 m/s) is probably commensurate with that in marrow. Measurements of sound speed in isolated marrow are difficult to come by, but sound speeds in most soft tissues fall in the range from 1400 to 1600 m/s (Duck, 1990). Many *in vitro* experiments in bone are performed with water instead of marrow and yield results consistent with *in vivo* measurements. Nicholson and Bouxsein (2002) compared phase velocities in marrow-filled and water-filled human calcaneus *in vitro*. They found a good correlation ($r^2=0.77$) between the two but somewhat higher values in water (1563 ± 25 m/s vs 1520 ± 36 m/s). Hoffmeister *et al.* (2002a) found no significant difference between the two in bovine trabecular tibia.

The longitudinal sound speed in nylon (2600 m/s) is somewhat lower than that for mineralized bone material (2800–4000 m/s, near 500 kHz) (Duck, 1990) but still far greater than that for water or marrow. In addition, nylon wires exhibit frequency-dependent scattering similar to that exhibited by trabecular bone (Wear, 2004).

A previously reported phantom design, consisting of cubic granules of gelatin immersed in epoxy, has been shown to be useful for the prediction of the dependencies of phase velocity, dispersion, and attenuation on porosity of trabecular bone (Clarke *et al.*, 1994; Strelitzki *et al.*, 1997). One advantage of the parallel-nylon-wire-in-water design is that it allows straightforward investigation of the effects of Tb.Th and s on phase velocity and dispersion.

TABLE I. Phantom properties. Tb.Th is trabecular (nylon wire) thickness. The variable s is the interwire spacing, which is equal to the sum of Tb.Th and Tb.Sp (trabecular separation). The volume fraction (VF) is the fraction of volume occupied by nylon wire. Porosity=1−VF.

Tb.Th(μm)	s=Tb.Th+Tb.Sp(μm)	Volume fraction	Porosity
152	1000	0.018	0.982
152	900	0.022	0.978
152	800	0.028	0.972
152	700	0.037	0.963
203	800	0.051	0.949
254	800	0.079	0.921
305	800	0.114	0.886

B. Ultrasonic methods

A Panametrics (Waltham, MA) 5800 pulser/receiver was used. Samples were interrogated in through-transmission in a water tank using a pair of coaxially aligned Panametrics 500 kHz, broadband, 0.75 in. diameter, unfocused transducers. The propagation path between transducers was 3 in. (7.62 cm). Received rf signals were digitized (8 bit, 10 MHz) using a LeCroy (Chestnut Ridge, NY) 9310C Dual 400 MHz oscilloscope and stored on computer (via GPIB) for off-line analysis. Seven measurements (of ten rf lines each) were obtained on each phantom. Phantoms were removed from the tank and then repositioned between measurements.

Frequency-dependent phase velocity, $c_p(f)$, was computed using

$$c_p(f) = \frac{c_w}{1 + \frac{c_w \Delta\phi(f)}{2\pi f d}}, \tag{3}$$

where f is frequency, $\Delta\phi(f)$ is the difference in unwrapped phases (see next paragraph) of the received signals with and without the phantom in the water path, d is the phantom thickness (12.7 mm), and c_w is the temperature-dependent speed of sound in distilled water given by (Kaye and Laby, 1973)

$$c_w = 1402.9 + 4.835T - 0.047\,016T^2 + 0.000\,127\,25T^3 \text{ m/s} \tag{4}$$

and T is the temperature in degrees Celsius. Temperature, measured with a digital thermometer, was 19.5° for these

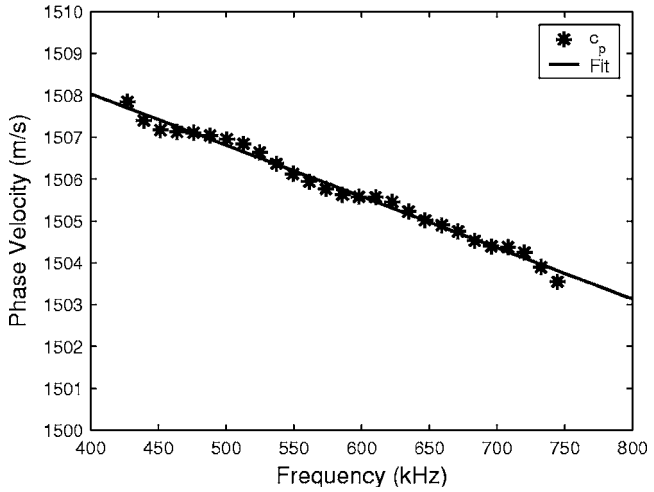


FIG. 2. Measurements (*) of phase velocity (c_p) vs frequency for phantom with Tb.Th=254 μm and s=800 μm. A linear least-squares regression fit to the data is also shown (solid line).

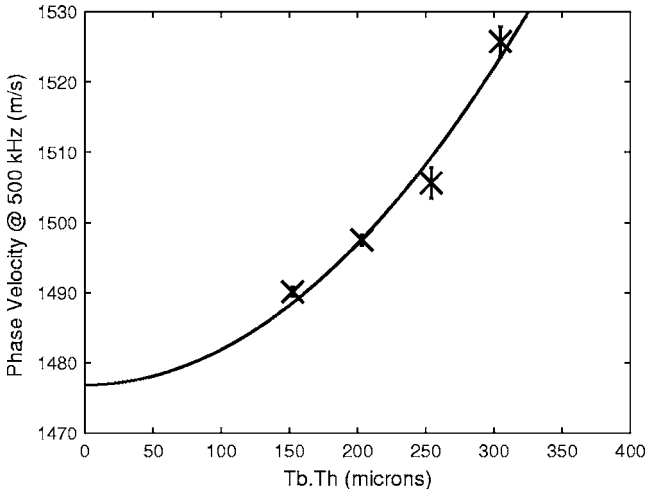


FIG. 3. Phase velocity at 500 kHz vs Tb.Th for four phantoms with s =800 μm. Error bars denote standard deviations. A quadratic fit, $c_p(500\ \text{kHz}) = 1477 + 502[\text{Tb.Th(mm)}]^2$ m/s, is also shown.

measurements, which meant that c_w was 1480 m/s.

The unwrapped phase difference, $\Delta\phi(f)$, was computed as follows. Fast Fourier Transforms (FFTs) of the digitized received signals were taken. The phase of the signal at each frequency was taken to be the inverse tangent of the ratio of the imaginary to real parts of the FFT at that frequency. Since the inverse tangent function yields principal values between $-\pi$ and π, the phase had to be unwrapped by adding an integer multiple of 2π to all frequencies above each frequency where a discontinuity appeared.

Dispersion was characterized by the slope, dc_p/df, of a linear least-squares regression fit of $c_p(f)$ vs f over the range from 400 to 750 kHz, which roughly corresponded to the system −6 dB bandwidth.

III. RESULTS

Figure 2 shows measurements of phase velocity (c_p) versus frequency for one phantom. Phase velocity declined quasilinearly with frequency for all phantoms.

Figure 3 shows measurements of c_p(500 kHz) vs Tb.Th for four phantoms with a constant value of s(800 μm). A quadratic fit, $c_p(500\ \text{kHz}) = 1477 + 502[\text{Tb.Th(mm)}]^2$ m/s, is also shown.

Figure 4 shows measurements of c_p(500 kHz) vs s for four phantoms with a constant value of Tb.Th (152 μm). A curve fit, $c_p(500\ \text{kHz}) = 1482 + 5.5/[s(\text{mm})]^2$ m/s, is also shown. The functional forms of the curve fits in Figs. 3 and 4 suggest that $c_p(500\ \text{kHz}) - c_w$ is approximately proportional to VF. [See Eq. (2)].

Figure 5 shows measurements of c_p(500 kHz) vs VF on all seven phantoms. A linear fit, $c_p(500\ \text{kHz}) = 1479 + 387(\text{VF})$ m/s, is in good agreement with the data.

Figure 6 shows measurements of dc_p/df vs Tb.Th for four phantoms with a constant value of s(800 μm). A power law fit, $dc_p/df = -10\,700[\text{Tb.Th(mm)}]^{4.8}$ is also shown. The fact that the exponent is so far removed from 2 suggests that, unlike change in phase velocity, dc_p/df is not simply proportional to VF.

Figure 7 shows measurements of dc_p/df vs s for four phantoms with a constant value of Tb.Th (152 μm). A meaningful curve fit to this data is precluded by small values of dc_p/df, relatively large error bars, and small variation in dc_p/df, observed over the range of s studied.

Figure 8 shows measurements of dc_p/df vs VF for all seven phantoms. A power law fit, $dc_p/df = -5950\text{VF}^{2.4}$ is also shown. Unlike the case with phase velocity, the relationship between dc_p/df and VF is nonlinear. Values for dc_p/df ranged from −2 to −35 m/sMHz, which is consistent with values reported in human calcaneus *in vitro*. See Table II.

IV. DISCUSSION

Phase velocity in trabecular-bone-mimicking phantoms is a linear, monotonically increasing, function of volume fraction. This seems to be true regardless of whether changes in volume fraction arise from changes in trabecular thickness or changes in trabecular spacing. Phase velocity in human calcaneus is also highly influenced by volume fraction. In

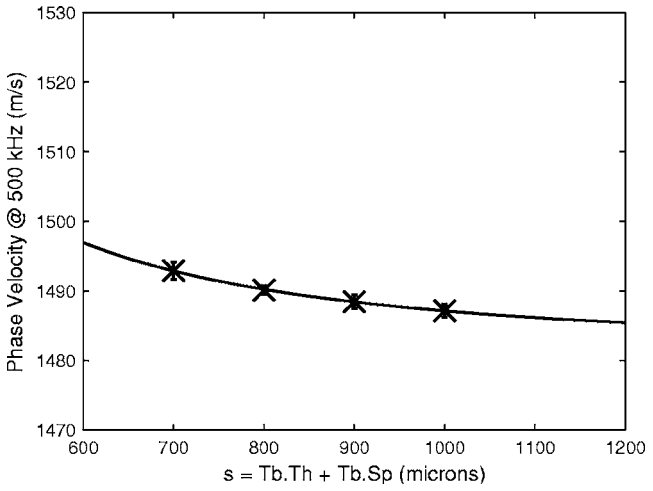


FIG. 4. Phase velocity at 500 kHz vs s for four phantoms with Tb.Th =152 μm. Error bars denote standard deviations. A curve fit, $c_p(500\ \text{kHz}) = 1482 + 5.5/[s(\text{mm})]^2$ m/s, is also shown.

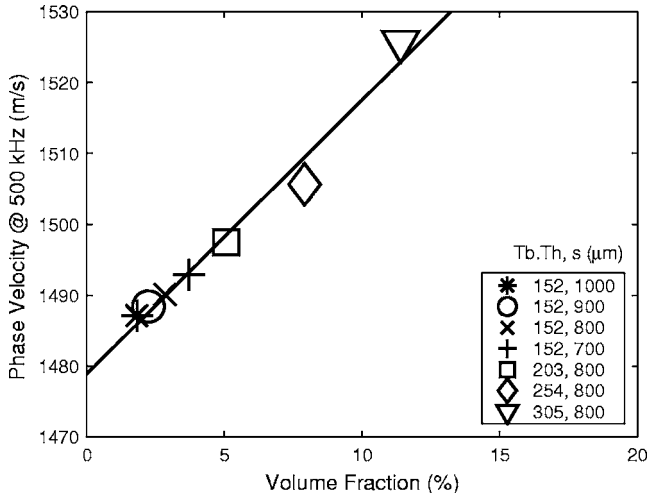


FIG. 5. Phase velocity at 500 kHz vs volume fraction for all seven phantoms. A linear fit, $c_p(500\ \text{kHz}) = 1479 + 387(\text{VF})$ m/s, is also shown.

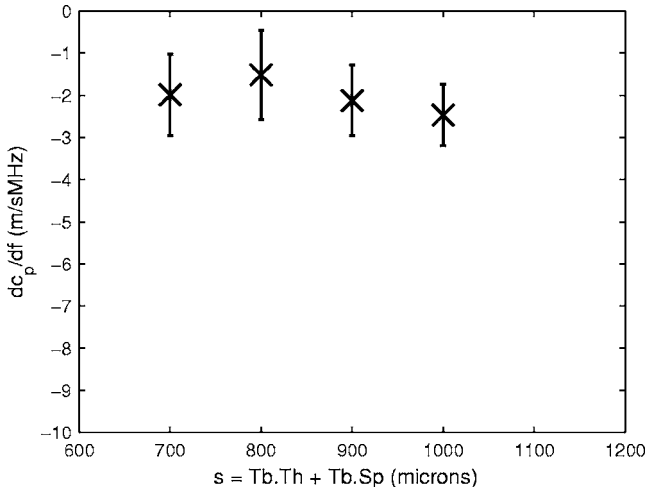


FIG. 7. The first derivative of phase velocity with respect to frequency, dc_p/df, vs s for four phantoms with Tb.Th = 152 μm. Error bars denote standard deviations.

this case, phase velocity varies nonlinearly, but still increases monotonically, with volume fraction. The variation may be predicted accurately using Biot theory (Wear *et al.*, 2005). Similar findings have been reported for bovine trabecular bone (Williams, 1992; Hosokawa and Otani, 1997; Hosokawa and Otani, 1998; Haire and Langton, 1999; Lee *et al.*, 2003; Mohamed *et al.*, 2003). An earlier application of Biot theory to trabecular bone was reported by McKelvie and Palmer (1991).

The first derivative of phase velocity with respect to frequency, dc_p/df, in trabecular-bone-mimicking phantoms is a nonlinear, monotonically decreasing function of volume fraction. Measurements of dc_p/df in phantoms may be compared with previously reported measurements of dc_p/df in human calcaneus (Wear, 2000a) plotted versus estimates of volume fraction based on an assumption of constant bone material density (Wear *et al.*, 2005), shown in Fig. 9. There is too much scatter in the human calcaneus data to allow confident conclusions regarding the functional dependence of dc_p/df on volume fraction. Nevertheless, the human data in Fig. 9 largely fall within the range of the phantom data in Fig. 8. For trabecular bovine tibia interrogated in the mediolateral orientation, dispersion has been reported to be a linear, monotonically decreasing function of density (Waters *et al.*, 2005, Fig. 6). Although dc_p/df may potentially carry important diagnostic information, dc_p/df measurements in bone, even *in vitro*, tend to exhibit high variability. *In vivo* application of this measurement would be very challenging with currently available techniques.

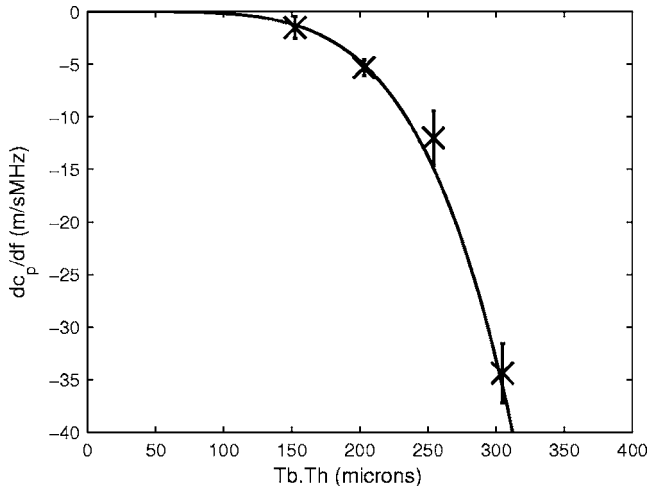


FIG. 6. The first derivative of phase velocity with respect to frequency, dc_p/df, vs Tb.Th for four phantoms with s = 800 μm. Error bars denote standard deviations. A power law fit, $dc_p/df = -10\,700[\text{Tb. Th(mm)}]^{4.8}$ is also shown.

The physical mechanism responsible for negative dispersion in the nylon wire phantoms is unknown. It has been shown, however, that negative dispersion in media consisting of alternating parallel slabs of two components may be predicted with the so-called "stratified model" (Brekhovskikh, 1980; Hughes *et al.*, 1999). The stratified model predicts values of dc_p/df commensurate with those observed in polystyrene/water phantoms and in human calcaneus *in vitro* (Wear, 2001a). The modified Biot–Attenborough theory (Lee *et al.*, 2003) and the Kramers–Kronig relations (Waters *et al.*, 2005) have successfully modeled dispersion in bovine trabecular bone.

Strelitzki *et al.* (1997) reported measurements of $c_p(600\ \text{kHz})$ and dc_p/df in phantoms consisting of cubic gelatin granules suspended in epoxy. Comparison of the

TABLE II. Estimates of the first derivative of phase velocity with respect to frequency, dc_p/df, in human calcaneus from Nicholson *et al.* (1996, Table I), Strelitzki and Evans (1996, Table II), Droin *et al.* (1998, Table I), and Wear (2000a, Table I). N is the number of calcaneus samples upon which measurements were based.

Author(s)	N	Frequency range (kHz)	dc_p/df (mean ± standard deviation) × (m/sMHz)
Nicholson *et al.*	70	200–800	−40
Strelitzki and Evans	10	600–800	−32 ± 27
Droin, Berger, and Laugier	15	200–600	−15 ± 13
Wear	24	200–600	−18 ± 15

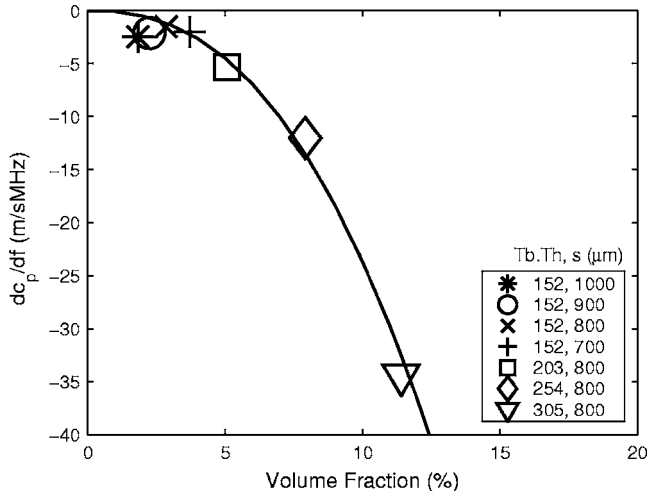


FIG. 8. The first derivative of phase velocity with respect to frequency, dc_p/df, vs volume fraction for all seven phantoms. A power law fit, $dc_p/df = -5950\ \mathrm{VF}^{2.4}$ is also shown.

present study with the work of Strelitzki *et al.* is complicated by the facts that (1) the two phantom designs used different materials, and (2) the phantom sets spanned different, nonoverlapping, ranges of volume fraction. Streliztki *et al.* examined a range of VF from 17% to 54% while the present study examined a range from 1.8% to 11.4%, which was chosen to approximate the range reported in human calcaneus, 2%–14% (Wear, 2005). As in the present study, Strelitzki *et al.* found phase velocity to increase with VF, but they observed a quadratic rather than a linear variation. Contrary to the present study, they found dc_p/df to increase with VF. Extrapolation of the nonlinear trend of Strelitzki *et al.* to VF=0 (where dc_p/df would be expected to be near 0), however, would suggest dc_p/df decreasing with VF at low values for VF ($<10\%$), consistent with the present study.

The data reported in the present paper suggest that for certain simple alterations in microarchitecture (changes in Tb.Th and s), c_p and dc_p/df are primarily determined by volume fraction. For more complicated alterations, this may not hold. For example, many studies have demonstrated a substantial anisotropy of SOS in trabecular bone (Nicholson *et al.*, 1998; Hosokawa and Otani, 1998; Hans *et al.*, 1999; Hughes *et al.*, 1999; Luo *et al.*, 1999; Hoffmeister *et al.*, 2000, 2002a, 2000b), suggesting that the physical arrangement of the trabeculae, and not just the quantity of trabecular material, influences SOS. In clinical calcaneal-based bone sonometry, however, the mediolateral orientation is always used, and the degree of microarchitectural alteration encountered in diagnostic applications can be expected to be more subtle than the comparatively drastic difference measured in anisotropy studies (e.g., medio-lateral versus antero-posterior versus supero-inferior orientations). Consequently, the simple alterations considered in the present phantom study may be relevant to the clinical situation.

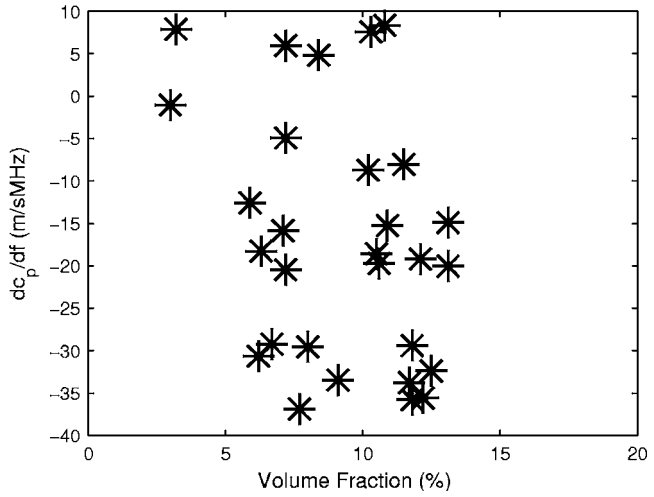


FIG. 9. The first derivative of phase velocity with respect to frequency, dc_p/df, vs volume fraction in 30 human calcaneus samples.

Under conditions when phase velocity (c_p) and dc_p/df are primarily determined by VF, group velocity (c_g) must also be primarily determined by VF. This may be seen by considering the relationship between the two velocity measures [Duck, 1990, Eq. (4.2)],

$$c_g = \frac{c_p}{1 - \frac{f_c}{c_p}\left(\frac{dc_p}{df}\right)}. \tag{5}$$

Many clinical and laboratory measurements of SOS in bone, however, are neither phase nor group velocity. They are based on time-of-flight measurements of broadband pulses through bone in which a marker (e.g., a zero-crossing or a threshold) on the pulse wave form is designated to measure pulse arrival time. Rather than using the pulse envelope maximum (as would be required for c_g), it is common in bone sonometry to choose a marker closer to the leading edge of the pulse. SOS measurements obtained in this way differ from c_g by an amount that increases monotonically with attenuation (Wear, 2000b; 2001b). This discrepancy is negligible for soft tissues but substantial for highly attenuating media such as bone. Therefore, the dependence of SOS on volume fraction may be expected to differ somewhat from that for c_p or c_g. Nevertheless, Luo *et al.* (1999), using a finite-difference simulation based on the two-dimensional elastic wave equation in conjunction with a threshold near the leading edge for time-of-flight estimation, also predicted a close, monotonically increasing, relationship between velocity and volume fraction in trabecular bone. Nicholson *et al.* (2001) found that the correlation between volume fraction and c_p(600 kHz), $r=0.86$, was very close to the correlation between volume fraction and signal velocity (SOS based on the pulse leading edge as an arrival time marker), $r=0.88$, in human calcaneus *in vitro*.

The present study may help explain findings by other researchers investigating relationships between phase velocity and microarchitecture. Nicholson *et al.* (2001), reporting measurements on 69 human calcaneal trabecular bone cubes, found moderate correlations between c_p(600 kHz) (mediolateral orientation) and Tb.Th ($r=0.49$) and between c_p(600 kHz) and Tb.Sp ($r=-0.47$). As expected, the signs of these correlation coefficients are consistent with the phantom measurements of the present study. [Figures 3–5 and Eq. (2) suggest, however, that $\mathrm{Tb.Th}^2$ and s^{-2} may have been more appropriate independent variables in the regression analysis

than Tb.Th and Tb.Sp.] More important, Nicholson *et al.* found a high correlation between c_p(600 kHz) and VF ($r=0.86$) but that multivariate regression models to predict c_p(600 kHz) from VF and Tb.Th or Tb.Sp did not significantly increase that correlation coefficient. In other words, phase velocity contains little or no information regarding Tb.Th or Tb.Sp beyond that already contained in VF. Chaffai *et al.* (2002) reported similar findings in human calcaneus. (Chaffai *et al.* actually used bone mineral density rather than VF as an independent variable. These two parameters, both of which are essentially reflections of quantity of bone, were highly correlated with each other in the study of Chaffai *et al.*, however, with $r=0.92$.) The present phantom study offers an explanation for the results of Nicholson *et al.* and Chaffai *et al.* As can be seen in Fig. 5, c_p(500 kHz) in phantoms is highly correlated with VF but is relatively insensitive to the particular combination of Tb.Th and s that produce VF.

In this investigation, the dependencies of phase velocity and its first derivative (with respect to frequency) on trabecular thickness, trabecular spacing, and volume fraction were measured in phantoms, yielding insight into relationships between frequency-dependent phase velocity and microarchitecture in bone. The phantom design allowed easy separation of effects due to changes in trabecular thickness from those due to changes in trabecular spacing. The dependencies of phase velocity and its first derivative on volume fraction in the phantoms were consistent with those reported in trabecular bone. The measurements in phantoms help explain why previous investigators have found in multiple regression analyses that trabecular thickness and trabecular spacing carry little predictive information regarding phase velocity beyond that carried by volume fraction alone.

ACKNOWLEDGMENTS

The author is grateful to Heather Miller, C.I.R.S., Norfolk, VA, for assistance in phantom design and construction. The mention of commercial products, their sources, or their use in connection with material reported herein is not to be construed as either an actual or implied endorsement of such products by the Food and Drug Administration.

Bauer, D. C., Gluer, C. C., Cauley, J. A., Vogt, T. M., Ensrud, K. E., Genant, H. K., and Black, D. M. (**1997**). "Broadband ultrasound attenuation predicts fractures strongly and independently of densitometry in older women," Arch. Intern Med. **157**, 629–634.

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

Chaffai, S., Peyrin, F., Nuzzo, S., Porcher, R., Berger, G., and Laugier, P. (**2002**), "Ultrasonic characterization of human cancellous bone using transmission and backscatter measurements: Relationships to density and microstructure," Bone (N.Y.) **30**, 229–237.

Clarke, A. J., Evans, J. A., Truscott, J. G., Milner, R., and Smith, M. A. (**1994**). "A phantom for quantitative ultrasound of trabecular bone," Phys. Med. Biol. **39**, 1677–1687.

Cummings, S. R., Black, D. M., Nevitt, M. C., Browner, W., Cauley, J., Ensrud, K. E., Genant, H. K., Palermo, L., Scott, J., and Vogt. T. M. (**1993**). "Bone density at various sites for prediction of hip fractures," Lancet **341**, 72–75.

Droin, P., Berger, G., and Laugier, P. (**1998**). "Velocity dispersion of acoustic waves in cancellous bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 581–592.

Duck, F. A. (**1990**). *Physical Properties of Tissue* (Cambridge University Press, Cambridge, UK).

Glüer, C., Cummings, S. R., Bauer, D. C., Stone, K., Pressman, A., Mathur, A., and Genant, H. K. (**1996**). "Osteoporosis: Association of recent fractures with quantitative US findings," Radiology **199**, 725–732.

Glüer, C. C., Eastell, R., Reid, D. M., Felsenbert, D., Roux, C., Barkmann, R., Timm, W., Blenk, T., Armbrecht, G., Stewart, A., Clowes, J., Thomasius, F. E., and Kolta, S. (**2004**). "Association of five quantitative ultrasound devices and bone densitometry with osteoporotic vertebral fractures in a population-based sample: The OPUS study," J. Bone Miner. Res. **19**, 782–793.

Haire, T. J., and Langton, C. M. (**1999**). "Biot Theory: A review of its application to ultrasound propagation through cancellous bone," Bone (N.Y.) **24**, 291–295.

Hans, D., Dargent-Molina, P., Schott, A. M., Sebert, J. L., Cormier, C., Kotzki, P. O., Delmas, P. D., Pouilles, J. M., Breart, G., and Meunier, P. J. (**1996**). "Ultrasonographic heel measurements to predict hip fracture in elderly women: The EPIDOS prospective study," Lancet **348**, 511–514.

Hans, D., Wu, C., Njeh, C. F., Zhao, S., Augat, P., Newitt, D., Link, T., Lu, Y., Majumdar, S., and Genant, H. K. (**1999**). "Ultrasound velocity of trabecular cubes reflects mainly bone density and elasticity," Calcif. Tissue Int. **64**, 18–23.

Hoffmeister, B. K., Auwarter, J. A., and Rho, J. Y. (**2002a**). "Effect of marrow on the high frequency ultrasonic properties of cancellous bone," Phys. Med. Biol. **47**, 3419–3427.

Hoffmeister, B. K., Whitten, S. A., Kaste, S. C., and Rho, J. Y. (**2002b**). "Effect of collagen and mineral content on the high-frequency ultrasonic properties of human cancellous bone," Osteoporosis Int. **13**, 26–32.

Hoffmeister, B. K., Whitten, S. A., and Rho, J. Y. (**2000**). "Low-megahertz ultrasonic properties of bovine cancellous bone," Bone (N.Y.) **26**, 635–642.

Hosokawa, A., and Otani, T (**1997**). "Ultrasonic wave propagation in bovine cancellous bone," J. Acoust. Soc. Am. **101**, 558–562.

Hosokawa, A., and Otani, T (**1998**). "Acoustic anisotropy in bovine cancellous bone," J. Acoust. Soc. Am. **103**, 2718–2722.

Hughes, E. R., Leighton, T. G., Petley, G. W., and White, P. R. (**1999**). "Ultrasonic propagation in cancellous bone: A new stratified model," Ultrasound Med. Biol. **25**, 811–821.

Huopio, J., Kroger, H., Honkanen, R., Jurvelin, J., Saarikoski, S., and Alhava, E. (**2004**). "Calcaneal ultrasound predicts early postmenopausal fractures as well as axial BMD. A prospective study of 422 women," Osteoporosis Int. **15**, 190–195.

Kaye, G. W. C., and Laby, T. H. (**1973**). *Table of Physical and Chemical Constants* (Longman, London, UK).

Laugier, P., "An overview of bone sonometry," (**2004**). International Congress Series **1274**, 23–32.

Laugier, P., Droin, P., Laval-Jeantet, A. M., and Berger, G. (**1997**). "In vitro assessment of the relationship between acoustic properties and bone mass density of the calcaneus by comparison of ultrasound parametric imaging and quantitative computed tomography," Bone (N.Y.) **20**, 157–165.

Lee, K. I., Roh, H., and Yoon, S. W. (**2003**). "Acoustic wave propagation in bovine cancellous bone: Application of the modified Biot-Attenborough model," J. Acoust. Soc. Am. **114**, 2284–2293.

Luo, G., Kaufman, J. J., Chiabrera, A., Bianco, B., Kinney, J. H., Haupt, D., Ryaby, J. T., and Siffert, R. S. (**1999**). "Computational methods for ultrasonic bone assessment," Ultrasound Med. Biol. **25**, 823–830.

McKelvie, M. L., and Palmer, S. B. (**1991**). "The interaction of ultrasound with cancellous bone," Phys. Med. Biol. **36**, 1331–1340.

Mohamed, M. M., Shaat, L. T., and Mahmoud, A. N. (**2003**). "Propagation of ultrasonic waves through demineralized cancellous bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 279–288.

Nicholson, P. H. F., and Bouxsein, M. L. (**2002**). "Bone marrow influences quantitative ultrasound measurements in human cancellous bone," Ultrasound Med. Biol. **28**, 369–375.

Nicholson, P. H. F., Lowet, G., Langton, C. M., Dequeker, J., and Van der Perre, G. (**1996**). "Comparison of time-domain and frequency-domain approaches to ultrasonic velocity measurements in trabecular bone," Phys. Med. Biol. **41**, 2421–2435.

Nicholson, P. H. F., Muller, R., Cheng, X. G., Ruegsegger, P., Van der Perre, G., Dequeker, J., and Boonen, S. (**2001**). "Quantitative ultrasound and trabecular architecture in the human calcaneus," J. Bone Miner. Res. **16**, 1886–1892.

Nicholson, P. H. F., Muller, R., Lowet, G., Cheng, X. G., Hildebrand, T., Ruegsegger, P., Van Der Perre, G., Dequeker, J., and Boonen, S. (**1998**).

"Do quantitative ultrasound measurements reflect structure independently of density in human vertebral cancellous bone?," Bone (N.Y.) **23**, 425–431.
Njeh, C. F., Hodgskinson, R., Currey, J. D., and Langton, C. M. (**1996**). "Orthogonal relationships between ultrasonic velocity and material properties of bovine cancellous bone," Med. Eng. Phys. **18**, 373–381.
O'Donnell, M., Jaynes, E. T., and Miller, J. G. (**1981**). "Kramers-Kronig relationship between ultrasonic attenuation and phase velocity," J. Acoust. Soc. Am. **69**, 696–701.
Parfitt, A. M., Drezner, M. K., Glorieux, F. H., Kanis, J. A., Malluche, H., Meunier, P. J., Ott, S. M., and Recker, R. R. (**1987**). "Bone histomorphometry: Standardization of nomenclature, symbols, and units," J. Bone Miner. Res. **2**, 595–609.
Rossman, P., Zagzebski, J., Mesina, C., Sorenson, J., and Mazess, R. (**1989**). "Comparison of speed of sound and ultrasound attenuation in the os calcis to bone density of the radius, femur and lumbar spine," Clin. Phys. Physiol. Meas. **10**, 353–360.
Schott, M., Weill-Engerer, S., Hans, D., Duboeuf, F., Delmas, P. D., and Meunier, P. J. (**1995**). "Ultrasound discriminates patients with hip fracture equally well as dual energy x-ray absorptiometry and independently of bone mineral density," J. Bone Miner. Res. **10**, 243–249.
Strelitzki, R., and Evans J. A. (**1996**). "On the measurement of the velocity of ultrasound in the os calcis using short pulses," Eur. J. Ultrasound **4**, 205–213.
Strelitzki, R., Evans, J. A., and Clarke, A. J. (**1997**). "The influence of porosity and pore size on the ultrasonic properties of bone investigated using a phantom material," Osteoporosis Int. **7**, 370–375.
Tavakoli, M. B., and Evans, J. A. (**1991**). "Dependence of the velocity and attenuation of ultrasound in bone on the mineral content," Phys. Med. Biol. **36**, 1529–1537.
Thompson, P., Taylor, J., Fisher, A., and Oliver, R. (**1998**). "Quantitative heel ultrasound in 3180 women between 45 and 75 years of age: Compliance, normal ranges and relationship to fracture history," Osteoporosis Int. **8**, 211–214.
Trebacz, H., and Natali, A. (**1999**). "Ultrasound velocity and attenuation in cancellous bone samples from lumbar vertebra and calcaneus," Osteoporosis Int. **9**, 99–105.
Turner, H., Peacock, M., Timmerman, L., Neal, J. M., and Johnston, Jr., C. C. (**1995**). "Calcaneal ultrasonic measurements discriminate hip fracture independently of bone mass," Osteoporosis Int. **5**, 130–135.
Ulrich, D., van Rietbergen, B., Laib, A., and Rüegsegger, P. (**1999**). "The ability of three-dimensional structural indices to reflect mechanical aspects of trabecular bone," Bone (N.Y.) **25**, 55–60.
Waters, K. R., Hoffmeister, B. K., and Javarone, J. A. (**2005**). "Application of the Kramers-Kronig relations to measurements of attenuation and dispersion in cancellous bone," *Proceedings of the 2004 IEEE Ultrasonics Symposium*.
Waters, K. R., Hughes, M. S., Mobley, J., Brandenburger, G. H., and Miller, J. G. (**1999**). "Kramers-Kronig dispersion relations for ultrasonic attenuation obeying a frequency power law," *Proceedings of the 1999 IEEE Ultrasonics Symposium*, Vol. **1**, pp. 537–541.
Wear, K. A. (**2000a**). "Measurements of phase velocity and group velocity in human calcaneus," Ultrasound Med. Biol. **26**, 641–646.
Wear K. A. (**2000b**), "The effects of frequency-dependent attenuation and dispersion on sound speed measurements: Applications in human trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 265–273.
Wear, K. A. (**2001a**). "A stratified model to predict dispersion in trabecular bone," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 1079–1083.
Wear, K. A. (**2001b**). "A numerical method to predict the effects of frequency dependent attenuation and dispersion on speed of sound estimates in cancellous bone," J. Acoust. Soc. Am. **109**, 1213–1218.
Wear, K. A. (**2004**). "Measurement of dependence of backscatter coefficient from cylinders on frequency and diameter using focused transducers—with applications in trabecular bone," J. Acoust. Soc. Am. **115**, 66–72.
Wear, K. A., Laib, A., Stuber, A. P., and Reynolds, J. C. (2005). "Comparison of measurements of phase velocity in human calcaneus to Biot theory," J. Acoust. Soc. Am. (in press).
Williams, J. L. (**1992**). "Ultrasonic wave propagation in cancellous and cortical bone: Predictions of some experimental results by Biot's theory," J. Acoust. Soc. Am. **92**, 1106–1112.
Zagzebski, J. A., Rossman, P. J., Mesina, C., Mazess, R. B., and Madsen, E. L. (**1991**). "Ultrasound transmission measurements through the os calcis," Calcif. Tissue Int. **49**, 107–111.

Acoustic radiation from a fluid-filled, subsurface vascular tube with internal turbulent flow due to a constriction

Yigit Yazicioglu, Thomas J. Royston,[a] Todd Spohnholtz, Bryn Martin, and Francis Loth
Mechanical Engineering, University of Illinois at Chicago, 842 West Taylor Street, MC 251, Chicago, Illinois 60607-7022

Hisham S. Bassiouny
Surgery, Section of Vascular Surgery, University of Chicago 5841 South Maryland Avenue, MC 5028, Chicago, Illinois, 60637

(Received 18 February 2005; revised 17 May 2005; accepted 24 May 2005)

The vibration of a thin-walled cylindrical, compliant viscoelastic tube with internal turbulent flow due to an axisymmetric constriction is studied theoretically and experimentally. Vibration of the tube is considered with internal fluid coupling only, and with coupling to internal-flowing fluid and external stagnant fluid or external tissue-like viscoelastic material. The theoretical analysis includes the adaptation of a model for turbulence in the internal fluid and its vibratory excitation of and interaction with the tube wall and surrounding viscoelastic medium. Analytical predictions are compared with experimental measurements conducted on a flow model system using laser Doppler vibrometry to measure tube vibration and the vibration of the surrounding viscoelastic medium. Fluid pressure within the tube was measured with miniature hydrophones. Discrepancies between theory and experiment, as well as the coupled nature of the fluid–structure interaction, are described. This study is relevant to and may lead to further insight into the patency and mechanisms of vascular failure, as well as diagnostic techniques utilizing noninvasive acoustic measurements. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1953267]

PACS number(s): 43.80.Qf, 43.40.Rj, 43.40.Ey, 43.20.Mv [FD] Pages: 1193–1209

I. INTRODUCTION

Physician-based stethoscopic auscultation of blood vessels has long been used as a simple noninvasive means of *qualitatively* assessing their patency. Increased audible frequency (sonic) sounds are associated with constrictions or other geometric alterations in the vessel geometry. These can be a result of plaque build-up in arteries, such as the coronary or common carotid, which may lead to a localized loss of circulation. A constriction can also result from intimal thickening, which commonly occurs near arteriovenous (AV) grafts that are used for dialysis patients. In recent years there has been growing interest in the correlation of sonic phenomena with vascular pathology from two perspectives, mechanistic and diagnostic.

First, from a mechanistic point of view, it is believed that identifying the cause and effect relationships between sonic phenomena and other symptoms associated with vascular pathology could lead to improved surgical practices or treatments. For example, consider AV grafts and venous anastomotic intimal hyperplasia (VAIH). Individuals with end-stage renal disease would die within a few weeks or months if not sustained by some form of dialysis therapy or a kidney transplant. An AV graft is constructed by the joining of an artery to a vein to provide an access site for hemodialysis patients. By bypassing the high resistance vessels (arterioles and capillaries), high flow rates can be achieved that are necessary for efficient hemodialysis. A synthetic graft material, polytetrafluoro-ethylene (PTFE), is often used for these grafts. More than half of the AV grafts fail and require surgical reconstruction within 3 years.[1] The majority of these graft failures is caused by occlusive VAIH, which is characterized by a narrowing or stenosis of the vein downstream of the graft junction. While the natural healing response after surgery causes some degree of intimal thickening, the biomechanical environment appears to be responsible for progression of intimal thickening to occlusive VAIH. Biomechanical forces in the AV graft are unique, with generally high wall shear stress (WSS) acting on the vein, flow separation, and significant pressure fluctuations that vibrate the vein wall and surrounding tissue. Studies in a canine animal model have shown that perivascular tissue vibration is correlated with VAIH ($r=0.92, p<0.001$).[2] Vibration likely occurs as a result of the transitional and turbulent flow patterns that exist in the AV graft due to the high flow rate and complex geometry.[3,4] An obvious question is whether this vibration has helped to catalyze or accelerate VAIH, or is it merely a benign symptom. If it is the former, then perhaps a modified graft geometry or construction should be considered in order to minimize this vibration.

Second, from a diagnostic viewpoint, whether or not turbulence-induced tissue vibration is an exacerbating catalyst of pathology or merely a benign by-product, its existence affords its use as a diagnostic indicator. To have quantitative utility and to be independent of individual physician skill and experience, a more rigorous analysis of the sonic phenomena is needed than can be obtained via the human ear and stethoscopic auscultation. In this context, there have

[a]Electronic mail: troyston@uic.edu

been numerous studies reported in the literature for more than three decades that have empirically and stochastically correlated sonic acoustic "signatures" with associated pathology.[5–17] A common observation in these studies is that sound intensity tends to increase with stenosis severity (until occlusion becomes nearly total), especially at audible frequencies on the order of hundreds of hertz (~100 to 1000 Hz), and that this sound is associated with turbulence in the stenosed vessel distal to the constriction. Some have sought to put this into a diagnostic "imaging" format via use of a two-dimensional array of acoustic transducers mounted on the skin surface that can be used to passively beamform on the sonic source.[18,19] Others have shown that the Doppler mode of conventional ultrasound (US) imaging technology can be used to roughly quantify the level of tissue vibration associated with VAIH in AV grafts.[2] Similarly, another study reported the development of an ultrasonic pulse–echo multigate technique using quadrature phase demodulation to obtain simultaneous measures of tissue vibration and blood velocity at multiple depths, again showing correlation in a case study of a severe stenosis in a human infrainguinal vein by-pass graft.[20]

A barrier to further refinement of diagnostic techniques or improved mechanistic comprehension is the limited understanding of the complex and coupled transitional or turbulent fluid and structural dynamics of the constricted blood vessel embedded in viscoelastic soft tissue layers. Numerous empirical and semiempirical models have been proposed to correlate measurements of the turbulent pressure field downstream of the occlusion with the geometry of the occluded vessel and the flow rate.[5,8–12,16] Note that, even for relatively simple geometries, an exact solution using computational fluid dynamic (CFD) simulations is as of yet unavailable for the case of transitional fluid behavior and compliant vessel wall dynamics. Accurate CFD simulations for transitional flow problems with anatomically correct, yet assumed rigid vessel walls require substantial CPU time and are just now being reported.[3,4]

Other studies have attempted to model the relationship between the turbulent field and resulting vessel vibration.[6,12,21] These studies have essentially noted that the tube-wall vibration spectra can differ significantly from the turbulence spectra as a result of the tube's frequency-dependent mobility. Additionally, it still remains somewhat unclear as to which scenario is more prevalent: (1) broad turbulence generates wall vibrations with some resonant spectral content that may subsequently cause coherent oscillations in the blood flow; or (2) coherent vortex shedding of the blood flow causes wall vibrations with distinct spectral content.[20] Still others have then tried to predict measurements by surface sensors, accounting for intervening tissue layers.[16,18] In addition to axisymmetry of the blood vessel and constriction, Ref. 16 assumed that the surrounding soft tissue is also axisymmetrically arranged about the vessel. The surrounding tissue is then treated as another fluid medium, only supporting compression waves, not shear or surface waves. On the other hand, Ref. 18 only considered shear wave radiation in the soft tissue from the turbulent source, assigning general attenuation rates to it based on geometric spreading and material viscosity.

Almost absent from the literature is a closed-form theoretical model of the entire coupled fluid/structure problem, starting from the generation of turbulence, the corresponding vessel wall vibration, and resulting surrounding tissue vibration. As noted above, a few studies have attempted this, but with the indicated limiting assumptions.[16,18] The present article reviews a theoretical and experimental study that endeavors to improve the fundamental understanding of this complex, coupled problem by considering a simple, axisymmetric, constricted vascular phantom model, in terms of its individual components as well as the assembled, coupled system. Laser Doppler vibrometry and miniature catheter hydrophone measurements provide an extensive experimental quantification of the fluid and structural behavior. While the theoretical model for fluid turbulence must be empirical, given the current state of the art, the remainder of the closed-form analytical model is based on first principles. Low-amplitude displacements of the vessel wall and surrounding tissue are assumed, which enables a linear treatment of the solid tissue dynamics. While this assumption should enable one to capture much of the dynamic phenomena present *in vivo* and in the *in vitro* phantom model described in this article, it is acknowledged that some documented phenomena will not be predicted. For example, flexible tube collapse, or buckling, after a stenosis due to the lowered intraluminal pressure has been reported *in vivo* and carefully studied *in vitro*. It can occur under both laminar and turbulent flow conditions. This involves large deformation of the lumen and requires nonlinear analysis, and is not within the scope of this article. See Ref. 22 for a review of this topic and its biological applications.

In the present article, experiments and theoretical developments are reviewed that attempt to quantify the fluid environment, its coupling to the vessel wall, and the resulting sound radiation into media exterior to the vessel wall. It is emphasized that the focus here is on developing a closed-form analytical model that may yield unique insight. Consequently, this article is divided into the following sections:

(1) description of the constricted flow phantom model and experimental measurement methods;
(2) adaptation and experimental evaluation of an empirical model for fluid turbulence;
(3) development and experimental evaluation of a theoretical model for fluid–vessel coupling; and
(4) development and experimental evaluation of a theoretical model for radiation into surrounding fluid or viscoelastic medium, including measurement of the sonic phenomena on the surface of a viscoelastic medium within which the constricted vessel is embedded.

II. THE AXISYMMETRIC CONSTRICTED FLOW MODEL

A. The experimental model

A simple axisymmetric tube geometry and flow constriction is considered. The compliant fluid-filled latex tube (Latex Penrose Tubing, Sherwood Medical, St. Louis, MO) is

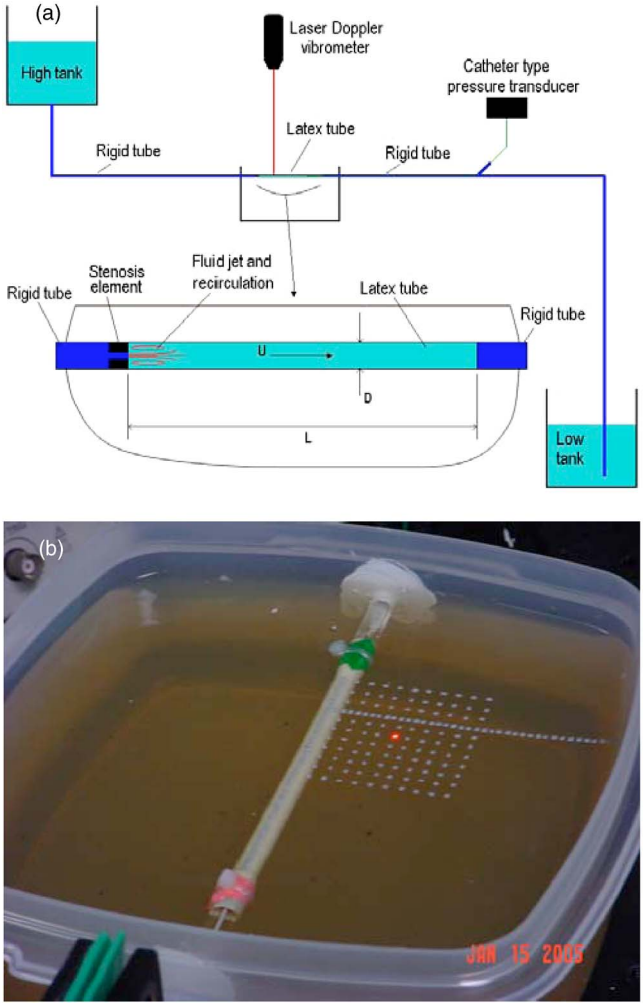


FIG. 1. (Color online) Experimental system. (a) Schematic. (b) Photograph of compliant tube embedded in gel phantom. Retroreflective tape visible on gel surface.

externally unconstrained (except at its ends) or lies just below and parallel to the horizontal free surface of a water-filled container or gel phantom model, as indicated in Fig. 1. Approximately 5% axial strain is imposed on the latex tube in all cases to prevent sagging in the externally unconstrained case. The experimental model geometry and parameter values are provided in Fig. 1 and Table I. Gravity-fed flow rates producing Reynold's (Re) numbers sufficient for turbulent generation downstream of the constriction and within the regime of biological relevance are considered. By adjusting the height of the upper water reservoir and adjusting a valve downstream of the compliant tube section, the flow rate and mean pressure within the compliant tube section are independently controlled. Degassed water is used. Steady flow rate conditions are used, as the frequency band of dynamic response associated with turbulent behavior is above and well separated from low-frequency dynamics associated with the pulsatile nature of blood flow *in vivo*.[11]

B. Experimental measurement methods

Mean and dynamic pressure within the tube, tube radial wall velocity, and gel phantom vertical velocity (when present) were measured for various flow conditions. Pressure measurements were made using a catheter-type pressure transducer (SPR-524, Millar, Houston, TX), with a bandwidth of ~10 kHz according to the manufacturer. The pressure transducer had a 1.08-mm-diameter tip where the active element was located at the end of a 0.71-mm-diameter wire. The tip was fed into the flow system from the downstream side through a sealed connector. The pressure measurements were made in 2.5-mm increments from 2.5 to 100 mm downstream of the constriction. The radial position within the tube of the catheter tip was consistently closer to the tube wall than the tube axis.

A noncontacting laser Doppler vibrometer (LDV) (CLV 800-FF/1000, Polytec, Auburn, MA) with threshold sensitivity of ~1 μm/s, low-pass filtered at 5 kHz, was used to measure tube and phantom surface velocity. Small pieces of adhesive retroreflective tape (3*M*, St. Paul, MN) measuring $\sim 1 \times 1$ mm were placed on the tube and phantom surface to improve the LDV signal. When used on the tube surface, reflective tape pieces were spaced 2.5 mm apart along the flow direction of the 100-mm tube, resulting in 41 data points (coinciding with internal pressure measurement points). On the phantom surface, they were spaced in 2.5 -mm increments along the tube axis and 2.5- or 5-mm increments lateral to it [see Fig. 1(b)]. Tube radial velocity was also measured in some cases when the tube itself was submerged in water or buried in the tissue-mimicking phantom. This measurement was more challenging because of the partial scattering of laser light at the water and phantom surfaces, which were in motion due to the turbulence-generated vibration; measurements through the phantom material during the higher flow rate case were not possible.

A two-channel digital dynamic signal analyzer (35670a, Agilent, Palo Alto, CA) was used to capture the experimental data. The acquired signals were processed using the signal

TABLE I. Experimental system parameter values.

Parameter	Value
Inner diameter of flexible tube (D)	6.4 mm
Diameter of constricted zone (d)	2.3 mm
Length of flexible tube (L)	100 mm
Wall thickness of flexible tube (h)	0.3 mm
Distance from gel or water surface to top of latex tube (h_g)	6.5 mm
Density of latex tube material (ρ_L)	1086 kg/m^3
Young's modulus for latex tube material[a] (E_L)	800 kPa
Linear viscous loss factor for latex tube material[b] (χ)	2×10^{-4} s
Poisson's ratio for latex tube material[a] (υ)	0.495
Water density[c] (ρ_f)	1000 kg/m^3
Speed of sound in water[c] (C_f)	1490 m/s
Phantom gel density (Ref. 31) (ρ_e)	1000 kg/m^3
Phantom gel volume elasticity (Ref. 31) (λ_1)	2.6 GPa
Phantom gel shear elasticity (Ref. 31) (μ_1)	4.5 kPa
Phantom gel shear viscosity (Ref. 31) (μ_2)	4 Pa s

[a]Based on quasistatic measurements in lab.
[b]Based on comparing theoretical predictions to experimental results in Figs. 6 and 7.
[c]Values based on nominal room temperature of 21 °C at atmospheric pressure.

TABLE II. A tabular approximation of $F_{n1}[x/D]$ based on Fig. 12 of Tobin and Chang (Ref. 9).

x/D	0	1	1.5	2	2.5	3	4	6	8	10	15	70
$F_{n1}[x/D]$	0	0.02	0.03	0.0355	0.0355	0.025	0.01	0.004	0.003	0.0025	0.002	0.002

analyzer and postprocessed with MATLAB V7.0. Time data were acquired at a 4096-Hz sample rate with the corresponding 120-dB/decade antialias filter set at 1600 Hz. The auto power spectra of 64 independent time records were averaged for each measurement point for each case.

III. AXISYMMETRIC FLOW CONSTRICTION AND TURBULENT FLOW

A. Theory based on empirical analysis of compliant wall studies

As noted in the Introduction, there have been numerous studies of flow in cylindrical channels with axisymmetric constrictions that are similar to the one depicted in Fig. 1(a) Researchers have theoretically, computationally, and experimentally analyzed the turbulent flow field in both relatively rigid and compliant tubes. Results and derivations reported in several specific articles were found to be most useful for the present analysis.[9,18,23]

Given the axisymmetric tube and constriction, it is assumed that the pressure distribution on the inner tube wall is also axisymmetric. The dynamic (acoustic) pressure $p(x,t)$ is only dependent on one spatial variable, the axial location x along the tube, as well as time, t. Its dependence on time and its dependence on x, are random, but with underlying time-averaged trends.[23] The turbulent coherent structures that produce dynamic pressure variations are assumed to propagate down the tube at a mean speed equal to the steady-state flow speed, but with significant random variation in the speed. Consequently, while $p(x,t)$ is random, a deterministic spectral approximation may be reasonable. A deterministic expression for $p(x,f)$, where f is the cyclic frequency in hertz (Hz), will be developed based on the literature.

Tobin and Chang[9] studied steady flow of water in a straight, compliant (latex rubber) cylindrical tube of diameter $D=7.94$ mm with a constricted region of various diameters d. The latex tube rested horizontally on 100 mm of cotton batting to isolate it from building vibrations. The tube itself

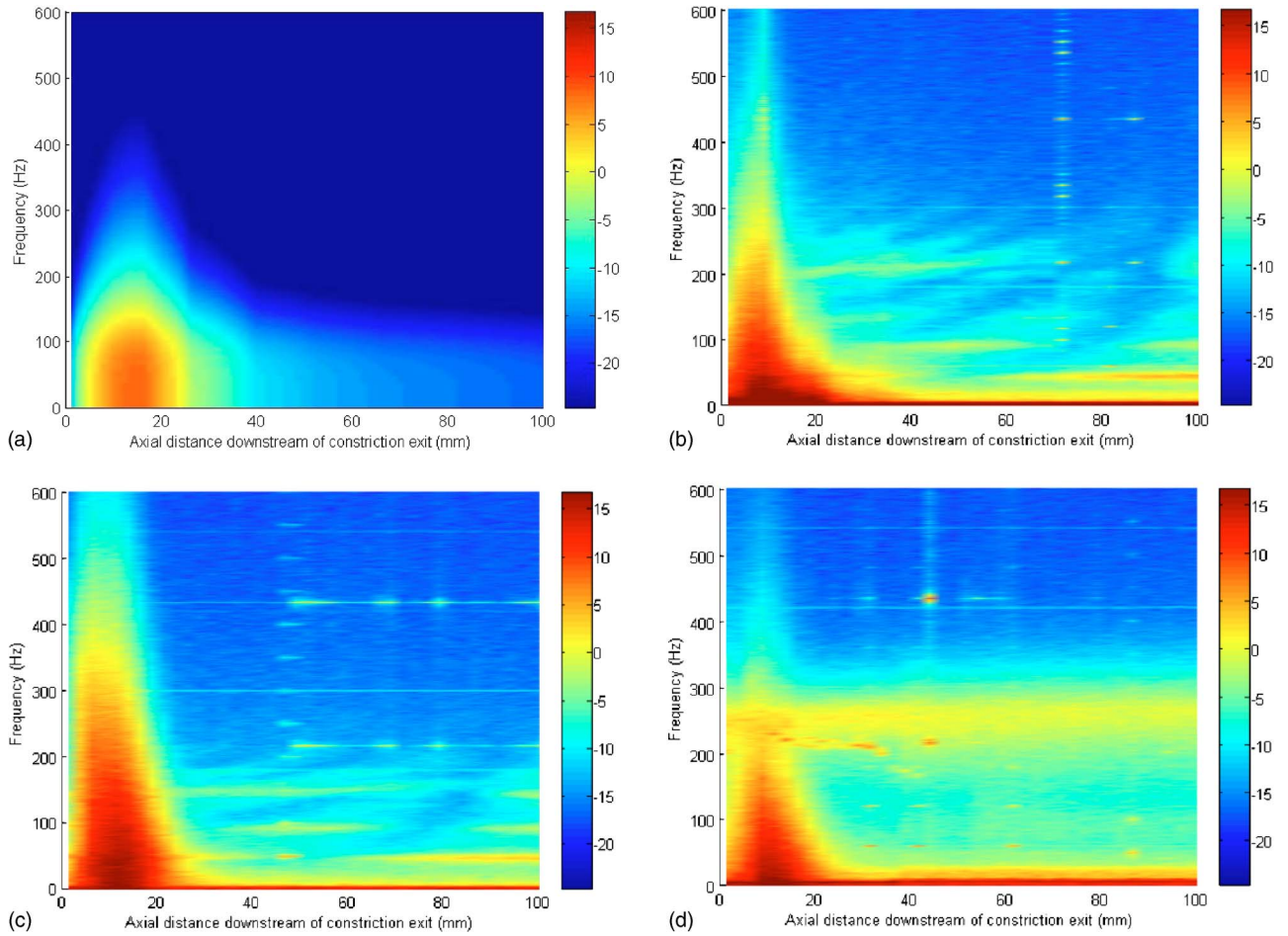


FIG. 2. (Color online) Acoustic pressure (dB *re*: 1 Pa) near wall inner surface as function of axial position and frequency for $\mathrm{Re}_D=1000$. All cases are for a tube with a constriction ending at 0 mm. (a) Theory. (b) Experiment in compliant tube in air. (c) Experiment in compliant tube in water. (d) Experiment in rigid tube. Online version uses color scale for dB.

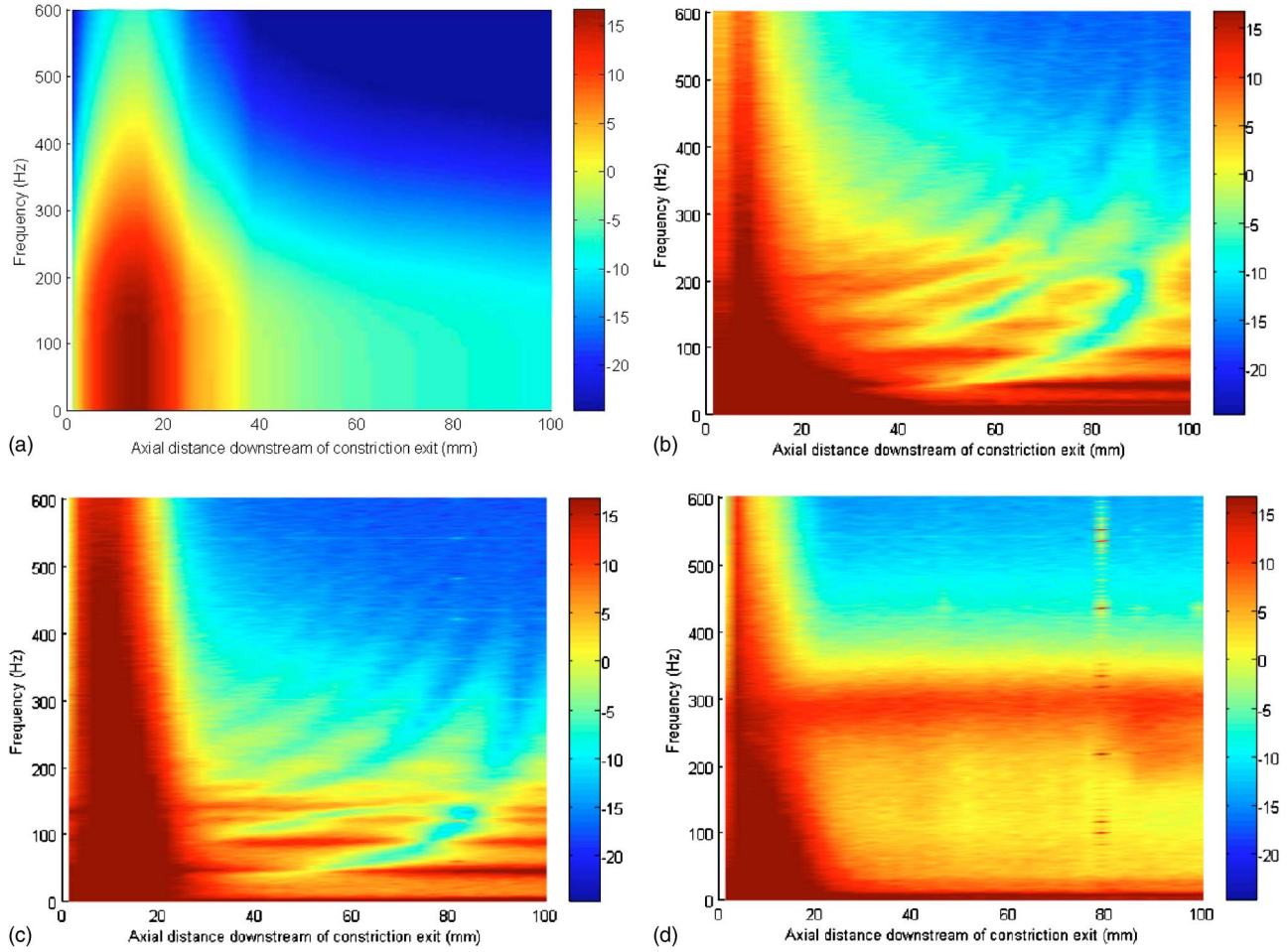


FIG. 3. (Color online) Acoustic pressure (dB *re*: 1 Pa) near wall inner surface as function of axial position and frequency for $\mathrm{Re}_D=2000$. All cases are for a tube with a constriction ending at 0 mm. (a) Theory. (b) Experiment in compliant tube in air. (c) Experiment in compliant tube in water. (d) Experiment in rigid tube. Online version uses color scale for dB.

was estimated to have an elastic modulus at the upper limit of the vascular physiological range, though no specific tube material property values were provided. The blunt and axisymmetric constriction was 12.7 mm long, and five different d/D ratios resulting in 75% to 95% reductions in area were considered. A wall pressure tap of diameter 1.75 mm was located downstream of the stenosis. Flow velocities, U, ranging from 60–500 mm/s and associated Reynolds numbers, based on D and U (Re_D) of 500–4000 were considered. The pressure sensor used at the tap was capable of measuring the dynamic response at least up to 2000 Hz.

For a range of $\mathrm{Re}_D=1500$ to 4000 they found a fairly consistent relationship between the dynamic root-mean-square wall pressure p_{rms} and the distance x downstream from the stenosis such that

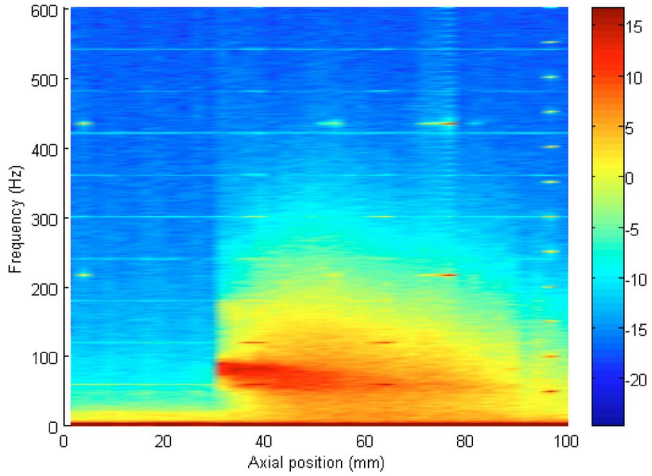


FIG. 4. (Color online) Acoustic pressure (dB *re*: 1 Pa) near wall inner surface as function of axial position and frequency for $\mathrm{Re}_D=2000$ in an *unconstricted* compliant tube (experiment). Online version uses color scale for dB.

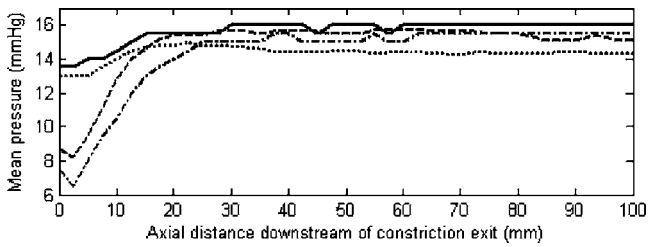


FIG. 5. Constricted tube mean fluid pressure (mm Hg gauge) as a function of axial position. Key: ——— $\mathrm{Re}_D=1000$, compliant tube; --- $\mathrm{Re}_D=1000$, rigid tube; – – – $\mathrm{Re}_D=2000$, compliant tube; –·– $\mathrm{Re}_D=2000$, rigid tube.

$$p_{\mathrm{rms}} = \rho u_j^2 \frac{d}{D} F_{n1}[x/D], \tag{1}$$

where F_{n1} denotes a nonlinear relationship, ρ is the fluid density, u_j is the spatial mean systolic jet velocity in the constriction, and x is the distance downstream from the end of the constriction.

For this same range of Reynolds numbers and constrictions, Tobin and Chang[9] also identified a consistent relationship between the power spectral density $E(f)$ and frequency f at the position x/D, where p_{rms} is maximum such that

$$E(f) \equiv \frac{p(f)^2}{\Delta f} = \rho^2 D u_j^3 \left(\frac{d}{D}\right)^2 F_{n2}[fD/u_j], \tag{2}$$

where F_{n2} denotes a nonlinear monotonic relationship and $p(f)^2/\Delta f$ denotes the power spectral density of the wall pressure variation as a function of frequency f and dependent on the filter bandwidth that was used, such that the units are pascals^2/Hz. As the nondimensional frequency increases (beyond the corner frequency) the log–log spectra approach a negative slope of 5.3. Specifically, a curve fit to the data in Ref. 9 yields

$$F_{n2}[fD/u_j] = \frac{0.002\,08}{1 + 20(fD/u_j)^{5.3}}. \tag{3}$$

Also, note that

$$p_{\mathrm{rms}}^2 = \langle p(t)^2 \rangle = \left(\frac{1}{T}\int_0^T p(t)^2 dt\right) = \int_0^\infty E(f) df, \tag{4}$$

where T is a suitable averaging time such that the value has asymptotically converged. The above formula can be used to approximate, let's say, the root-mean-square pressure value between frequencies $f - \frac{1}{2}$ Hz and $f + \frac{1}{2}$ Hz such that

$$p_{\mathrm{rms}}(f) \approx \rho D^{1/2} u_j^{3/2} \left(\frac{d}{D}\right) (F_{n2}[fD/u_j])^{1/2}. \tag{5}$$

Next, by taking the spectral content in this 1-Hz band to be concentrated at f and in phase, one could express this component of the spectrum in the following form:

$$p(e^{j2\pi ft}, \phi_f) = \sqrt{2}\rho U^{3/2} \frac{D^{5/2}}{d^2} (F_{n2}[fd^2/UD])^{1/2} e^{j(2\pi ft + \phi_f)}. \tag{6}$$

In this expression the term ϕ_f denotes an unknown phase angle. Under this condition, and taking x/D at the point of the maximum value for p_{rms}, then, based on Fig. 12 of Ref. 9

$$p_{\mathrm{rms}} \approx \left(\sum_{f=1}^{\infty} (p[e^{j2\pi ft}, \phi_f])^2\right)^{1/2} \approx \rho U^2 \frac{D^3}{d^3} F_{n1}[x_{\max}/D]$$
$$\approx \rho U^2 \frac{D^3}{d^3}(0.0355). \tag{7}$$

According to Ref. 9, the spectral distribution of the wall pressure variation at different downstream positions maintains the same general form as at the point of maximum p_{rms}, but does vary to some degree. Specifically, as one moves downstream from near the position of maximum pressure, both the amplitude and corner frequency of the spectrum decrease. However, quantitative measurements are not provided. In the present treatment, it will be assumed that the relative spectral distribution will stay the same, but be attenuated in overall amplitude to correctly follow the p_{rms} level as it decreases when one moves away from its maximum location.

Specifying a value for ϕ_f is another matter. This issue is taken up in some detail by Keith and Abraham[23] for the more general case of turbulent pressure flow over a wall. The baseline assumption is that the turbulent wall pressure convection velocity u_c matches that of the stream velocity. Owsley and Hull[18] have assumed that the turbulent wall pressure convection moves axially at the jet velocity u_j of the stenotic constriction. Then, one would have that the phase difference, $\Delta\phi_f$, as a function of frequency f between two points separated by a distance Δx would equal $2\pi f \Delta x/u_j$. However, studies have shown that this simplifying assumption results in an overprediction of cross-spectral properties between two different axial locations of wall pressure measurement. Its effect on predictions of sound radiation to points exterior to the flow tube is unclear. A more accurate model would account for the fact that, near the wall, flow is slower than along the central axis of the flow tube. Larger, coherent structures of lower frequency are located more toward the axial center with smaller structures near the wall. The small structures, however, dissipate more rapidly with distance from the stenosis, and the larger structure's effect on wall pressure variations may become relatively more prominent further downstream. It follows that u_c may be a complex, frequency-dependent quantity. In the present study, the approximation used by Owsley and Hull[18] is followed.

These assumptions lead to

$$p[e^{j2\pi ft}, x, \phi_f] = P_0 e^{j(2\pi f - \phi_f)} = 1.82 F_{n1}[x/D]\rho U^{3/2} \times \frac{D^{5/2}}{d^2}\left(\frac{1}{1 + 20(fd^2/UD)^{5.3}}\right)^{1/2} e^{j(2\pi ft - \phi_f)}, \tag{8a}$$

where

$$\phi_f[f, x] = \frac{2\pi f x d^2}{UD^2}. \tag{8b}$$

Here, $F_{n1}[x/D]$ is the value interpolated from Fig. 12 of Tobin and Chang.[9] Linear interpolation via a tabular approximation was used in simulations reported here and is provided in Table II. Note that the studies in Ref. 9 were conducted with different flow constrictions and upstream pressures to achieve different Reynold's numbers and percent area stenoses. Apparently, no attempt was made to ensure that the mean pressure downstream of the constriction was consistent.

B. Theoretical predictions and experimental measurements

Based on the above theoretical analysis, and given the geometry of the experimental system provided in Fig. 1 and

Table I, predictions of the acoustic pressure on the wall inner surface as a function of axial distance downstream of the constriction exit and as a function of frequency were computed [Figs. 2(a) and 3(a)] for flows with Re_D of ~1000 and ~2000. (This calculation is made neglecting a ~2.8% increase in D in the compliant tube due to the internal mean pressure of ~15 mm Hg.) This corresponds to Reynold's numbers in the rigid constricted zone of Re_d equals 2783 and 5565, respectively. It is apparent that the greatest acoustic pressure amplitude occurs roughly 15 mm downstream of the constriction exit for both cases. Spectral content up through several hundred hertz is significant, but then attenuates at higher frequencies.

As described in Sec. II B, internal mean and dynamic pressure measurements were made using a miniature catheter hydrophone. It is recognized that the presence of the hydrophone likely will alter the flow field. Additionally, the hydrophone is not acquiring a pressure measurement directly at the channel inner wall, but rather at a location within the flow channel near the wall. Measurements were obtained for flow rates of Re_D= ~1000 and ~2000, within the compliant latex tube, as well as within a more rigid, hard plastic tube of the same inner dimensions, in order to assess the effect of tube wall compliance on the flow field. Measurements were also obtained in compliant and "rigid" tubes with the axisymmetric flow constriction removed.

Selected results are presented in Figs. 2–5. In comparing theoretical predictions to experiment, it appears that the theory has captured some of the general trends, both in terms of amplitude and spatial-spectral distribution, though the match is not perfect. In the experiment, the maximum pressure appears to be closer to 10–13 mm downstream of the constriction, as opposed to 15 mm. This discrepancy may be due in part to the hydrophone position. As the turbulent field spreads radially after exiting the constriction, it will reach the hydrophone before it reaches the tube wall. Additionally, the presence of the hydrophone may cause turbulence to occur earlier at a position further upstream.

There are other downstream spatial-spectral features evident in the experimental measurement that are not evident in the theoretical predictions, particularly observable in the Re_D=2000 case. In comparing experimental measurements for the constricted compliant tube versus a constricted rigid tube, the downstream spectral content is very different, which suggests that it is coupled with the tube wall dynamics. Note that the band of increased vibration that occurred along the entire length of the rigid test section at around 250–300 Hz was also present, though to a far less extent, for a rigid tube *without* a flow constriction (measurements not shown). An experimental modal analysis of the test setup indicated that a beamlike bending resonance of the rigid tube was present with a spectral peak in the 250–300 Hz range; it is suspected that this resonance may have been excited by flow dynamics and in turn increased the spectral content of the dynamic pressure measured in the fluid.

For the compliant tube cases, multiple bands of increased vibration along the length of the axis appear; these generally disappear, or at least recede to below the noise floor when the constriction is removed. These bands, which are not predicted by theory, are not due to the presence of the hydrophone in the flow channel, as they also appear to affect the vibration of the compliant tube (shown later) even when the hydrophone was not inserted. (In all constricted compliant studies, vibration of the compliant tube was identical with respect to whether or not the hydrophone was present.) Additionally, axisymmetric resonant modes of tube vibration are evident in the pressure readings, particularly in Figs. 3(b) and 3(c). Note that in one unconstricted study in the compliant tube at Re_D=2000, when the hydrophone was placed at roughly 30 to 100 mm from the compliant tube entrance, turbulence was generated that appeared to be caused by the hydrophone's presence (Fig. 4). For the same unconstricted test in a rigid tube, no turbulence occurred. Likewise, at Re_D=1000, in rigid and compliant tubes without constrictions, no turbulence occurred.

Overall, spectral bands downstream of the constriction do appear to be altered by the flow rate (Re), whether the tube is rigid or compliant, and, in the case of the compliant tube, also depend on the mean fluid pressure and whether external fluid loading is present. Hence, they are associated with coupled fluid and structural dynamics, which are not captured using the simple theory proposed here and likely would not be accurately predicted using a more complex theory or numerical (CFD) simulation that did not account for compliant walls. While the mean pressure did drop immediately downstream of the constriction (Fig. 5), it remained above atmospheric pressure in this experiment, and large deformation buckling or tube collapse was not observed.

IV. FLUID–VESSEL COUPLING

A. Theory

1. Tube of infinite length

Consider a cylindrical tube of radius "a" and thickness "h" that is infinite in length. It will be assumed that the tube material is isotropic, and the tube itself can be considered to be thin (e.g., $h/a<0.1$). Extensions of the analysis to the case of orthotropy[24] and "thick shell" theory are relatively straightforward but introduce additional complexity that is not relevant to the present analysis. The vibrational motion of a thin, isotropic cylindrical elastic shell can be described by the Donnell–Mushtari (DM) equations.[25] In addition to the assumption that h/a is small, this theory also assumes that resulting shell dynamic displacements are small, transverse normal stress acting on planes parallel to the shell middle surface are negligible, and fibers of the shell normal to the middle surface remain so after deformation and are themselves not subject to elongation. Presence of a compressible fluid within and/or exterior to the shell and the resulting fluid–structure interaction has been considered.[26–29] Alternative derivations for thicker shells have been applied to this problem with interior fluid, including use of the Kennard shell equations, which add a few additional terms to account for curvature but do not increase the degrees of freedom.[28] Use of first-order shear deformation theory to augment the DM theory, which adds two rotational degrees of freedom to account for shear deformation through the

shell thickness, has also been considered in conjunction with compressible fluid interior and exterior to the shell.[29]

Consider harmonic line forces applied around the circumference of the tube at axial position x_0 and specified by

$$p_t(x_0,\theta,t) = F_t \cos(n\theta)\delta(x-x_0)e^{j\omega t}, \tag{9}$$

where F_t is force per unit length, δ denotes the Dirac delta function, θ is the azimuthal angle in the tube with $\theta=0$ denoting vertically up, $\omega=2\pi f$, and $n=0,1,2,\ldots$. The excitation of the tube wall due to interior turbulent flow is approximated by resolving it into a finite number of such line forces. Given the case of harmonic excitation, it is convenient to express the shell displacements and applied forces in terms of inverse Fourier transforms in the axial wave number k_{ns}, here taking the case that $x_0=0$.[27]

$$u = \frac{1}{\sqrt{2\pi}}\sum_{n=0}^{\infty}\sum_{s=0}^{\infty}\int_{-\infty}^{\infty}\bar{U}_{ns}\cos(n\theta)e^{j(-k_{ns}x+\omega t+\pi/2)}\partial k_{ns}, \tag{10}$$

$$v = \frac{1}{\sqrt{2\pi}}\sum_{n=0}^{\infty}\sum_{s=0}^{\infty}\int_{-\infty}^{\infty}\bar{V}_{ns}\sin(n\theta)e^{-j(k_{ns}x-\omega t)}\partial k_{ns}, \tag{11}$$

$$w = \frac{1}{\sqrt{2\pi}}\sum_{n=0}^{\infty}\sum_{s=0}^{\infty}\int_{-\infty}^{\infty}\bar{W}_{ns}\cos(n\theta)e^{-j(k_{ns}x-\omega t)}\partial k_{ns}, \tag{12}$$

$$\bar{p}_t = (1/\sqrt{2\pi})F_t\cos(n\theta)e^{j\omega t}. \tag{13}$$

Here, u, v, and w denote vibratory displacements in the axial, azimuthal, and radial directions, respectively. One obtains the following by inserting these expressions into the DM equations and utilizing orthogonality of the n and s components in the summations in Eqs. (10)–(12):

$$\begin{bmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{bmatrix}\begin{bmatrix} \bar{U}_{ns} \\ \bar{V}_{ns} \\ \bar{W}_{ns} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \Omega^2 F_t/\sqrt{2\pi}\rho_s h\omega^2 \end{bmatrix}, \tag{14}$$

with

$$L_{11} = -\Omega^2 + (k_{ns}a)^2 + n^2\left(\frac{1-\nu}{2}\right), \tag{15a}$$

$$L_{12} = L_{21} = n(k_{ns}a)\left(\frac{1+\nu}{2}\right), \tag{15b}$$

$$L_{13} = L_{31} = \nu(k_{ns}a), \tag{15c}$$

$$L_{22} = -\Omega^2 + \frac{1-\nu}{2}(k_{ns}a)^2 + n^2, \tag{15d}$$

$$L_{23} = L_{32} = n, \tag{15e}$$

and

$$L_{33} = 1 + \beta^2[(k_{ns}a)^2 + n^2]^2 - \Omega^2 - FL_i - FL_e, \tag{15f}$$

where FL_i and FL_e, respectively, account for internal and external fluid loading in the coupled problem. It is assumed that these are compressible fluids that satisfy the acoustic wave equation in cylindrical coordinates. To ensure the fluid remains in contact with the tube wall, the fluid radial motion and the tube radial motion must be equal at the interface of the tube and fluid. If the shell is submerged in a compressible fluid of infinite extent, this coupling condition results in the following expressions:

$$FL_i = \Omega^2\frac{\rho_i}{\rho_s h k_{ri}}\left[\frac{J_n[k_{ri}a]}{J_n'[k_{ri}a]}\right], \tag{16}$$

$$FL_e = \Omega^2\frac{\rho_e}{\rho_s h k_{re}}\left[\frac{H_n^{(2)}[k_{re}a]}{H_n'^{(2)}[k_{re}a]}\right], \tag{17}$$

where

$$\Omega^2\left(\frac{c_L}{c_i}\right)^2 = (k_{ns}a)^2 + (k_{ri}a)^2, \tag{18a}$$

$$\Omega^2\left(\frac{c_L}{c_e}\right)^2 = (k_{ns}a)^2 + (k_{re}a)^2, \tag{18b}$$

$$\Omega = \omega a/c_L, \tag{18c}$$

and

$$c_L = \sqrt{E(1+j\omega\xi)/\rho_s(1-\nu^2)}. \tag{18d}$$

Here, J_n and $H_n^{(2)}$, respectively, refer to an nth-order Bessel function of the first kind and an nth-order second Hankel function, which denotes outgoing wave propagation consistent with the convention $e^{j\omega t}$ denoting harmonic motion. Also, the primes on J_n and $H_n^{(2)}$ denote differentiation with respect to the arguments $k_{ri}a$ and $k_{re}a$, respectively. If the external fluid does have finite boundaries, then the expression in Eq. (17) would need to be modified. Additionally, ρ_s, ρ_i, and ρ_e refer to the density of the shell (tube) material, internal fluid, and external fluid, respectively. And, c_L refers to the complex extensional phase speed of the tube material, dependent on its Young's modulus, E, linear viscous loss factor, ξ, density, ρ_s, and Poisson's ratio ν.

Then, the spectral radial displacement amplitude [as a function of $(k_{ns}a)$] is

$$\bar{W}_{ns} = \left(\frac{\Omega^2 F_t}{\sqrt{2\pi}\rho_s h\omega^2}\right)I_{33}, \tag{19a}$$

where

$$I_{33} = (L_{11}L_{22} - L_{12}L_{21})/|\mathbf{L}|. \tag{19b}$$

Application of the inverse transform gives the radial displacement as

$$w(x/a,n,s) = \frac{\Omega^2 F_t}{2\pi\rho_s h a\omega^2}\int_{-\infty}^{\infty} I_{33}e^{-j(k_{ns}a)(x/a)}d(k_{ns}a) = Y_{ns}(x/a)F_0/j\omega, \tag{20}$$

where $Y_{ns}(x/a)$ is the transfer mobility for a particular branch s and circumferential mode of vibration n such that

$$\frac{\dot{w}(x/a,n,s)}{F_t} = Y_{ns}(x/a)$$

$$= \frac{j\Omega^2}{2\pi\rho_s ha\omega}\int_{-\infty}^{\infty} I_{33}e^{-j(k_{ns}a)(x/a)}d(k_{ns}a). \quad (21)$$

By utilizing the theorem of residues, the transfer mobility (i.e., response at x due to excitation at $x_0=0$) can be written as the sum of the residues evaluated at the poles: i.e.,

$$Y_{ns}(x/a) = \frac{\Omega^2}{\rho_s ha\omega}\sum_{s=1}^{\infty} \text{Re } s_s, \quad (22a)$$

$$\text{where Re } s_s = \frac{(L_{11}L_{22} - L_{12}L_{21})e^{-j(k_{ns}a)(x/a)}}{(\det|L|)'}, \quad (22b)$$

where the prime denotes the derivative with respect to $(k_{ns}a)$.

To couple this analysis with the analysis of the turbulent field in the previous section, the line force F_t is obtained by taking the axisymmetric ($n=0$) pressure calculation from the turbulent analysis and approximating it as a circumferential line force in terms of short segments along the cylinder. So, $F_t=P_0^{\star}\Delta L$, where ΔL is made small enough that the results asymptotically converge. Note that, if pressure data were available from, say a numerical simulation that involved a nonaxisymmetric constriction and a resulting nonaxisymmetric pressure distribution, higher azimuthal order ($n>0$) components could be used to predict the resulting nonaxisymmetric tube vibration.

2. Tube of finite length

The compliant section of tube of length L in the experimental setup (Fig. 1) is approximated as pinned at its ends such that simply supported boundary conditions exist. Adapting the above analysis for the harmonic line force and finite length shell yields

$$u = \sum_{n=0}^{\infty}\sum_{m=1}^{\infty} U_{nm}\cos(m\pi x/L)\cos(n\theta)e^{j\omega t}, \quad (23)$$

$$v = \sum_{n=0}^{\infty}\sum_{m=1}^{\infty} V_{nm}\sin(m\pi x/L)\sin(n\theta)e^{j\omega t} \quad (24)$$

$$w = \sum_{n=0}^{\infty}\sum_{m=1}^{\infty} W_{nm}\sin(m\pi x/L)\cos(n\theta)e^{j\omega t}. \quad (25)$$

Again, by insertion of these expressions into the DM equations and utilizing the orthogonality of the mode shapes, delineated by m and n, this leads to

$$\begin{bmatrix} L_{11} & L_{12} & L_{13} \\ L_{21} & L_{22} & L_{23} \\ L_{31} & L_{32} & L_{33} \end{bmatrix}\begin{bmatrix} U_{nm} \\ V_{nm} \\ W_{nm} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2\Omega^2 F_t \sin(k_m x_0)/L\rho_s h\omega^2 \end{bmatrix}, \quad (26)$$

with

$$L_{11} = -\Omega^2 + (k_m a)^2 + n^2\left(\frac{1-\nu}{2}\right), \quad (27a)$$

$$L_{12} = L_{21} = -n(k_m a)\left(\frac{1+\nu}{2}\right), \quad (27b)$$

$$L_{13} = L_{31} = -\nu(k_m a), \quad (27c)$$

$$L_{22} = -\Omega^2 + \frac{1-\nu}{2}(k_m a)^2 + n^2, \quad (27d)$$

$$L_{23} = L_{32} = n, \quad (27e)$$

and

$$L_{33} = 1 + \beta^2[(k_m a)^2 + n^2]^2 - \Omega^2 - FL_i - FL_e. \quad (27f)$$

The internal and external fluid loading terms, FL_i and FL_e, would be the same as in the infinite case if rigid diaphragms existed at $x=0$ and $x=L$ extending from $r=0$ to ∞ for $\theta=0$ to 2π; this is not the case in the experimental setup. Consider the internal loading term first, FL_i. At $x=0$ the axisymmetric constriction exists at the connection of the rigid tube to the compliant tube, resulting in an impedance mismatch. At $x=L$, while an axisymmetric constriction does not exist, there is a change from a compliant to a rigid wall, which results in an impedance mismatch. These end impedances are not infinite, which would be the case if rigid diaphragms existed. This difference is expected to have a more significant effect at lower axial modal orders, with reducing significance as axial modal order increases, since boundary conditions become less important as modal order increases. Nonetheless, in the present theoretical study rigid diaphragms in the internal fluid at $x=0$ and L are assumed, and Eq. (16) is used with the following in place of Eq. (18a):

$$\Omega^2\left(\frac{c_L}{c_i}\right)^2 = (k_{nm}a)^2 + (k_{ri}a)^2. \quad (28)$$

With regard to the external fluid loading term, FL_e, for the experimental case the external fluid extends much less than the wavelength of sound in it; thus, its effect on tube vibrations is approximated simply as a mass load. This leads to

$$FL_E = \Omega^2\frac{\rho_e}{\rho_s}\frac{h_{eq}(h_{eq}+2a)}{2ha}, \quad h_{eq} = a + h_g + h/2, \quad (29)$$

where h_g denotes the depth from the top of the tube to the free surface of the external fluid. This expression is equivalent to adding to the tube a mass per unit length representing an external fluid layer of thickness h_{eq} completely surrounding the tube, where h_{eq} is the average distance of a point on the tube wall to the nearest free surface.

Radial velocity of the tube at axial location x and angular position θ due to the circumferential line force per unit length F_t at $x=x_0$ can then be expressed in terms of a modal superposition

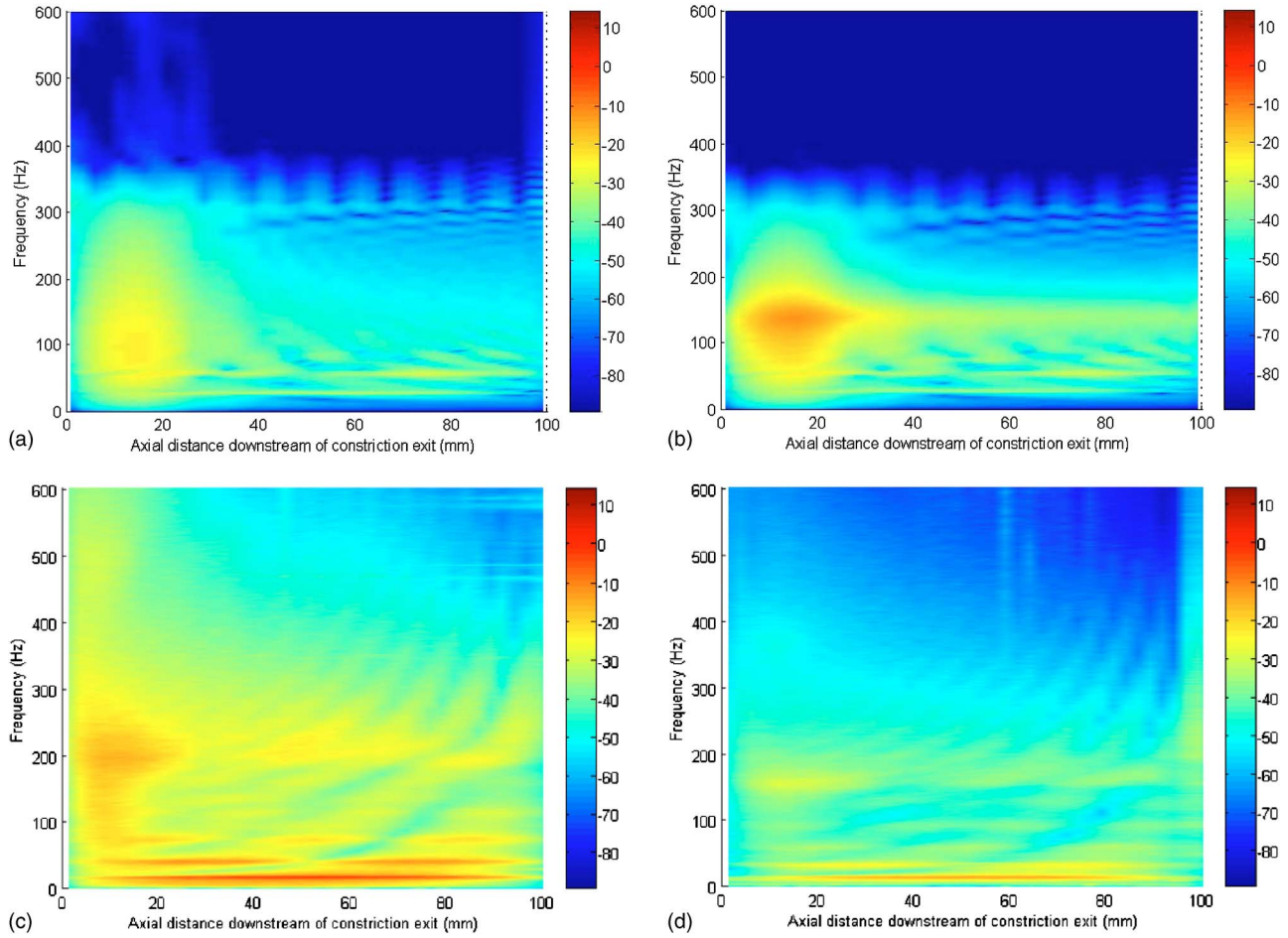


FIG. 6. (Color online) Constricted compliant tube radial wall velocity (dB *re*: 1 mm/s) as a function of axial position and frequency for $\mathrm{Re}_D=1000$. (a) Theory, in air. (b) Theory, submerged in water. (c) Experiment, in air. (d) Experiment, submerged in water. Online version uses color scale for dB.

$$\dot{w}[x,\theta,t]=\sum_{n=0}^{\infty}\sum_{m=1}^{\infty} I_{33} j\omega \frac{2\Omega^2 F_t \sin[k_m x_0]}{L\rho_s h\omega^2} \times \sin(m\pi x/L)\cos(n\theta)e^{j\omega t}, \tag{30}$$

where I_{33} is given by Eq. (19b) with values of $\mathbf{L}$ defined in Eqs. (26) and (27).

B. Theoretical predictions and experimental measurements

The predicted axisymmetric pressure distribution from Sec. III was used as an input to the finite-length tube model to predict the resulting vessel wall axisymmetric radial velocity $\dot{w}$ as a function of axial position and frequency. Circumferential line forces per unit length $F_t(x,f)$ were used every 1.2 mm axially to approximate $p_t(x,f)$, whose magnitude was depicted in Figs. 2(a) and 3(a). Theoretical predictions are shown in Figs. 6(a), 6(b), 7(a), and 7(b) for the flow cases of $\mathrm{Re}_D=1000$ and 2000 for the case of only internal fluid (fluid-filled tube suspended in air), and for the case of internal and external fluid. Also shown in these same figures are the corresponding experimental measurements of the tube wall vibration, using laser Doppler vibrometry; see Figs. 6(c), 6(d), 7(c), and 7(d). In comparing experimental cases without and with external fluid loading, it appears that the external fluid loading is primarily acting as a mass load, decreasing the frequency of, but not altering, the spatial-spectral character of the system. Likewise, in comparing similar cases with $\mathrm{Re}_D=1000$ vs $\mathrm{Re}_D=2000$, the change in flow rate appears to alter the intensity of vibration significantly, but it alters the spatial-spectral distribution of the vibration less significantly. This emphasizes the importance of the surrounding structural dynamic properties, which are consistent and unchanged when flow rate changes.

In comparing the theoretical cases to the experimental cases, again it appears that some but not all phenomena are captured. Even if one takes the experimental pressure measurements shown in Figs. 2(b), 2(c), 3(b), and 3(c) and uses these as an input to the theoretical tube model of Sec. IV, a closer, but still not exact match to experiment can be achieved. (Results of this "hybrid" calculation are not shown.) In the theoretical simulation, all geometric and material parameters (Table I) were measured or identified independently of this experiment, except the linear viscous damping term used for latex, ξ; this value was adjusted based on a rough comparison of theory and experiment, but was kept the same in all simulations. Use of a material damping term with a nonlinear dependence on frequency would have improved the match between theory and experiment. Also, during the course of the experimental study, three different latex tubes were used and some variability of elastic modulus, up to 10%, was observed between different tubes and as

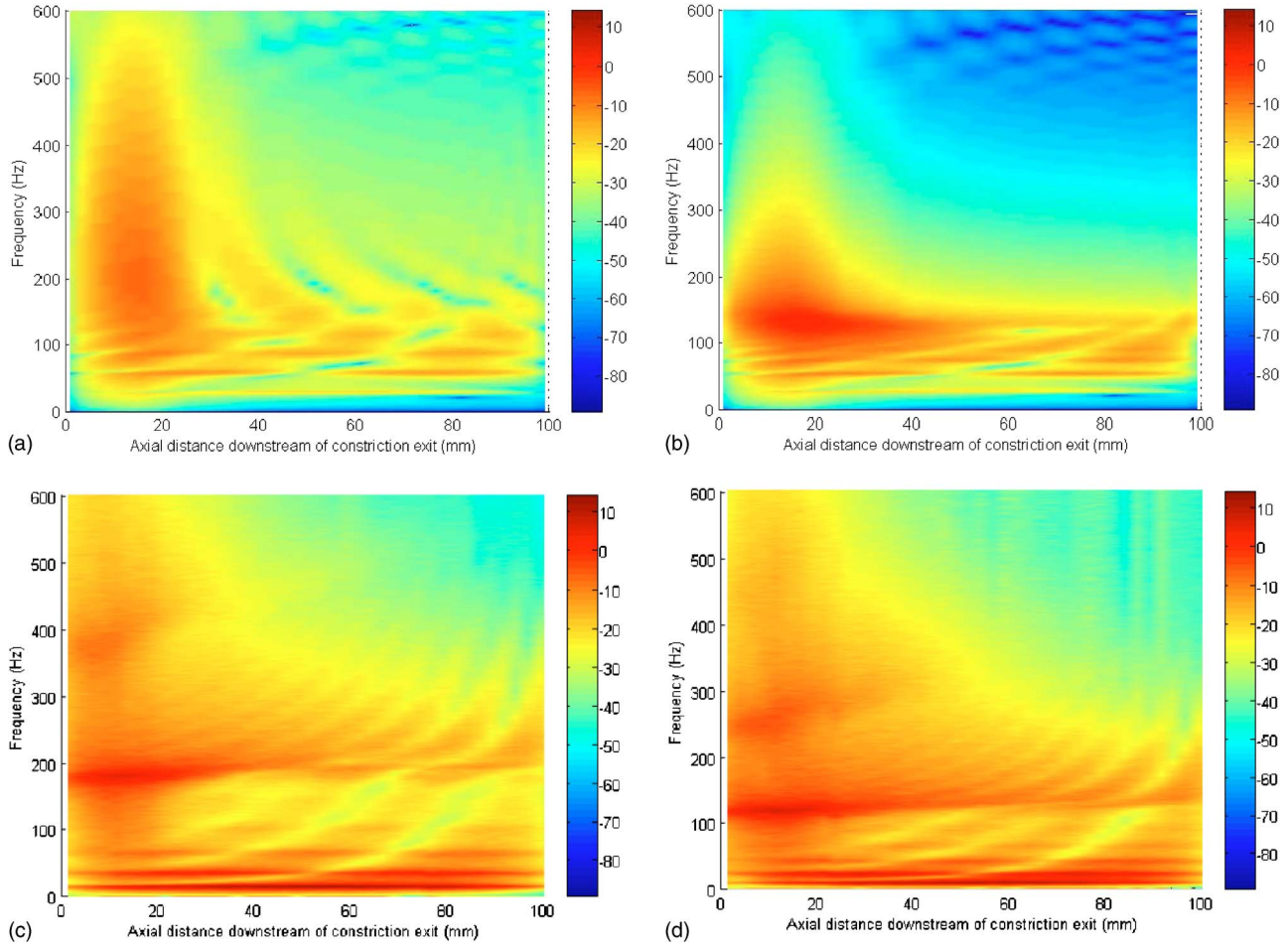


FIG. 7. (Color online) Constricted compliant tube radial wall velocity (dB *re*: 1 mm/s) as a function of axial position and frequency for $\mathrm{Re}_D=2000$. (a) Theory, in air. (b) Theory, submerged in water. (c) Experiment, in air. (d) Experiment, submerged in water. Online version uses color scale for dB.

a function of time in the same tube due to usage and aging; Young's modulus would decrease slightly after repeated use.

Some experimentally observed dynamics are simply not captured in the theoretical treatment. It is suspected that the mean pressure drop immediately after the constriction (see Fig. 5) does alter the system somewhat in this region, relative to the theoretical model, which assumed a uniform mean pressure throughout the length of the compliant tube in determining tube dimensions. As predicted above, due to the implicit assumption of rigid diaphragms in the fluid at either end of the flexible tube, the lowest tube structural resonant modes for the experiment tend to shift down in frequency, relative to theory, though for modes above the first three, agreement between theory and experiment is good.

For the experimental cases shown in Figs. 6 and 7, the mean (head) pressure in the compliant tube was adjusted to ~15 mm Hg gauge by adjusting the height of the upstream reservoir. It was found that increasing the mean pressure in the fluid channel substantially increased some, but not all, of the spectral features of the tube vibration. Figure 8 shows a case for $\mathrm{Re}_D=1000$ with the compliant constricted tube in air (no external water or gel loading), but with an internal mean pressure of ~60 mm Hg gauge. The conditions of this case are identical to that shown in Fig. 6(c) except for the increase in mean pressure of ~45 mm Hg. Such a change in pressure increases the compliant tube diameter by ~8.4% and decreases its thickness by ~7.7%; this should lower axisymmetric tube structural resonances by about 7% assuming linear elasticity. Static tests of tube diameter as a function of

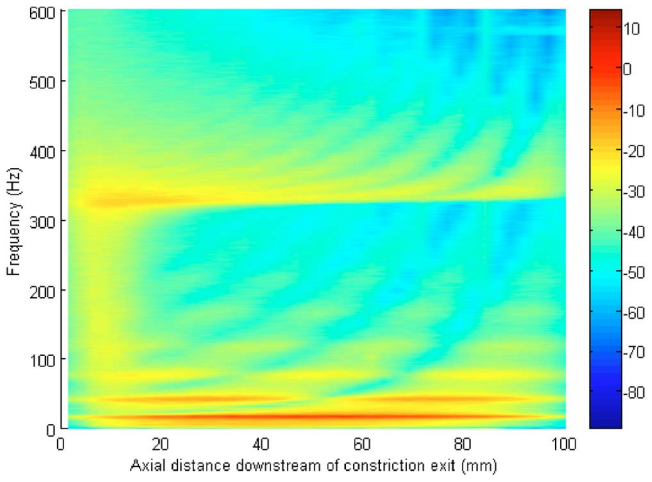


FIG. 8. (Color online) Measured constricted compliant tube radial wall velocity (dB *re*: 1 mm/s) as a function of axial position and frequency for $\mathrm{Re}_D=1000$, in air. Effect of head pressure. High tank (Fig. 1) at same height as in $\mathrm{Re}_D=2000$ case resulting in downstream mean pressure of ~60 mm Hg gauge. Online version uses color scale for dB.

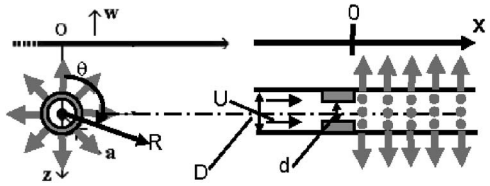


FIG. 9. Approximating the radiation of vibro-acoustic energy from the flexible tube vibration with finite dipoles.

internal static pressure confirmed that the tube elasticity was slightly nonlinear, with a ~10% increase in elastic modulus when internal pressure increased from 15 to 60 mm Hg. This increase actually would counter the effect of the geometric changes and result in less of a predicted decrease in resonant frequency. Significant changes are not evident in the frequencies of the first few standing wave patterns for the tube, below 100 Hz, supporting this analysis.

However, at higher frequencies, there is a significant change in the spectral content between the two cases [Figs. 6(c) and 8]. This same trend was observed in a number of experiments at various Re_D and mean pressures. While increasing the mean pressure did not significantly alter the lower frequency standing wave response of the tube, it did increase the frequency and sometimes sharpen the spectral content of the higher frequency vibration that was strongest just downstream of the constriction (from ~200 to ~325 Hz for $\mathrm{Re}_D=1000$ at 15 and 60 mm Hg mean pressure, respectively). Theoretical simulations, which used different values for diameter and thickness of the tube dependent on the mean pressure, showed no significant changes in their prediction of the internal dynamic pressure and tube vibration when the internal mean pressure was varied from 15 to 60 mm Hg. It is hypothesized that the observed phenomenon is due to nonlinear behavior just downstream of the constriction that is caused in part by the larger dynamic forces at this point and the substantially reduced mean pressure immediately distal to the constriction, shown in Fig. 5. Such behavior in a collapsible tube has been previously demonstrated.[22]

V. RADIATION INTO THE SURROUNDING VISCOELASTIC MEDIUM

A. Theory

Now, consider the axisymmetrically vibrating tube to be embedded in a viscoelastic phantom material with properties comparable to soft biological tissue. The term for external fluid loading, given above in Eq. (29), may be suitable when, in fact, the external medium is a fluid. However, when the tube is embedded parallel to and near the surface of a viscoelastic medium, a different approach may be considered. The elastic and viscous forces of the viscoelastic material may be non-negligible. Consequently, a composite tube wall is defined as being composed of the latex material and a thickness of the gel material taken to be the depth that the latex is below the free surface, h_g. The composite tube thickness h_c is thus the combined thickness of the latex material h and its depth below the surface, h_g. The composite Young's

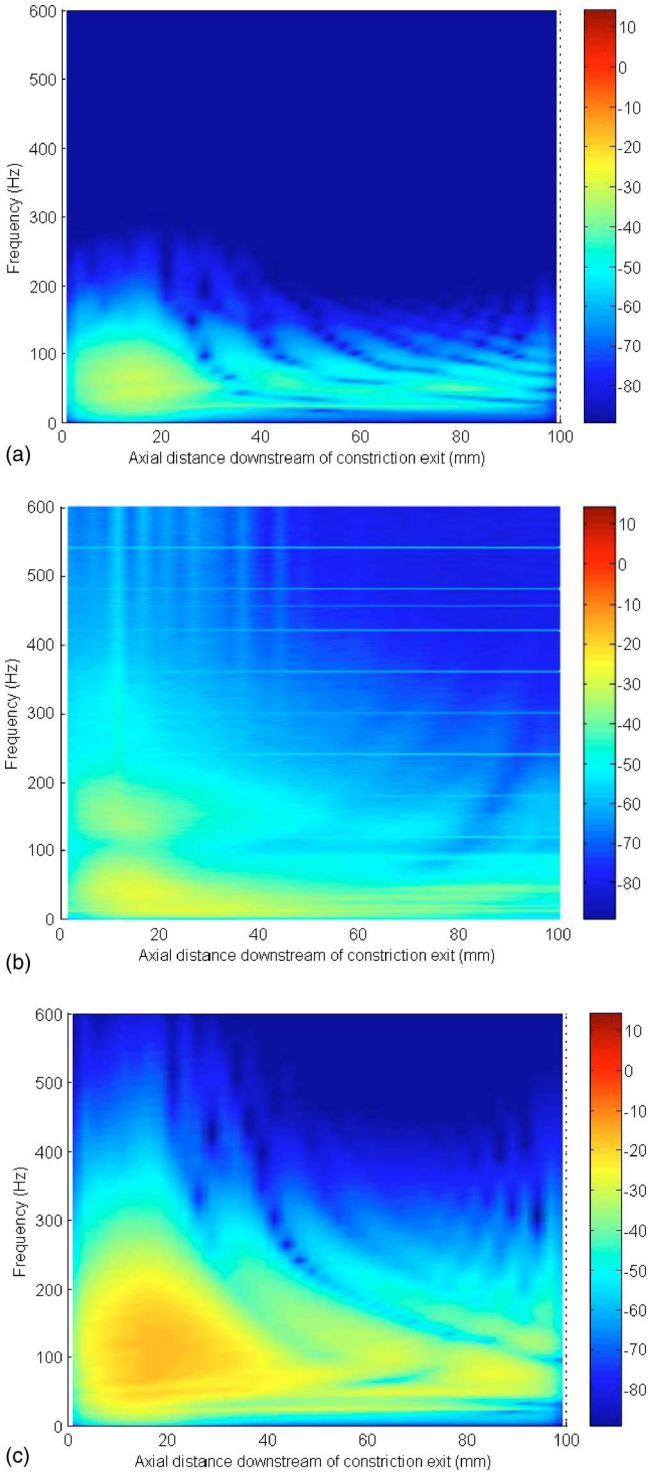


FIG. 10. (Color online) Constricted compliant tube radial wall velocity (dB *re*: 1 mm/s) while embedded in soft tissue gel phantom as a function of axial position and frequency. (a) Theory, $\mathrm{Re}_D=1000$. (b) Experiment, $\mathrm{Re}_D=1000$. (c) Theory, $\mathrm{Re}_D=2000$. Experimental measurement at $\mathrm{Re}_D=2000$ not possible. Online version uses color scale for dB.

modulus E_c is taken to be that of the thickness-weighted sum of the individual elastic moduli of the Latex and gel material

$$E_c = \frac{h}{h_c}E_L + \frac{h_g}{h_c}E_g, \tag{31}$$

with $E_g \approx 3\mu$, where μ is the complex shear modulus of the gel phantom material (Table I). This is then used in place of

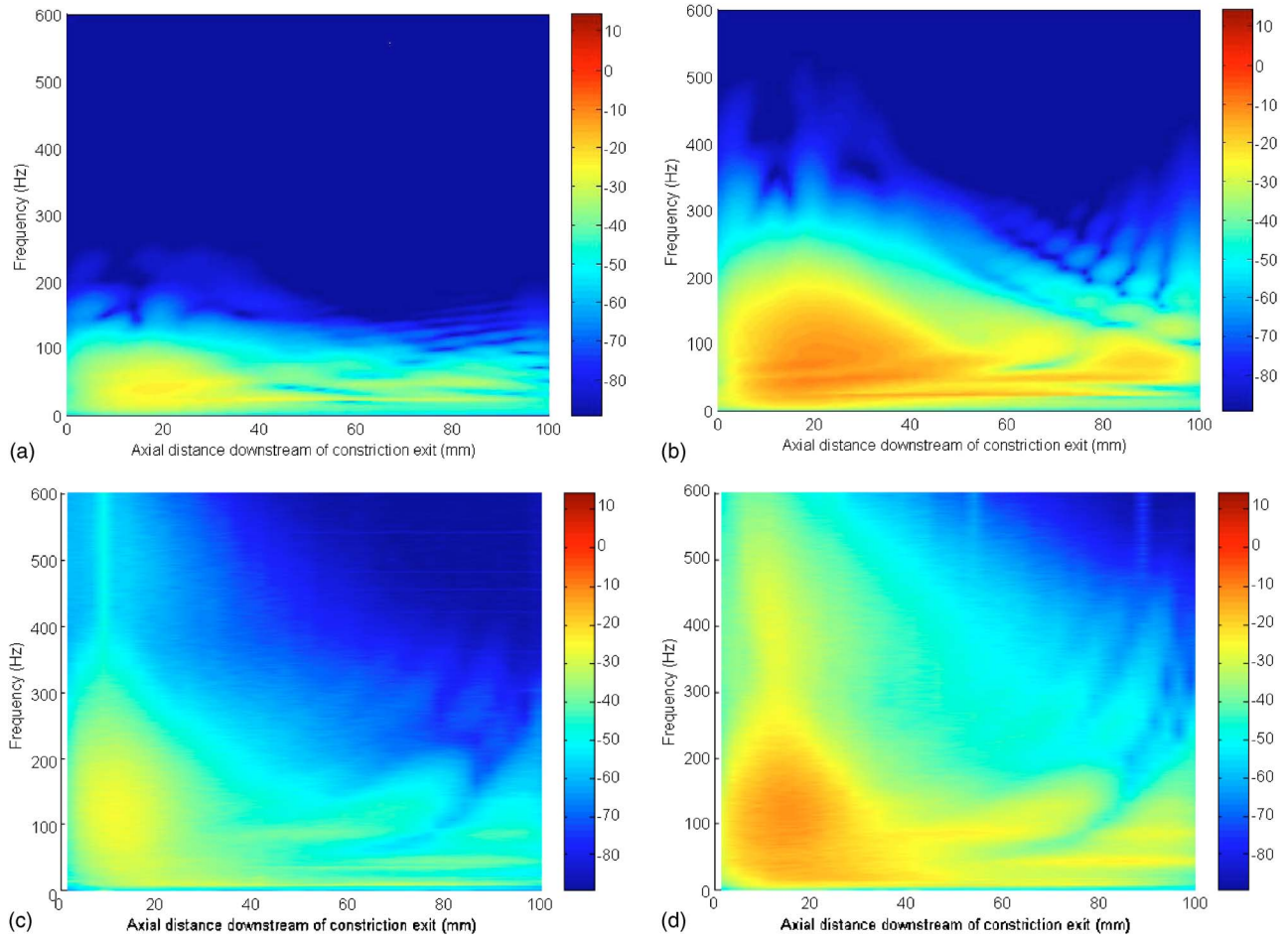


FIG. 11. (Color online) Vertical velocity (dB *re*: 1 mm/s) at phantom surface for constricted compliant tube embedded in soft tissue gel phantom as a function of axial position and frequency directly above tube. (a) Dipole theory, $\mathrm{Re}_D=1000$. (b) Dipole theory, $\mathrm{Re}_D=2000$. (c) Experiment, $\mathrm{Re}_D=1000$. (d) Experiment, $\mathrm{Re}_D=2000$. Online version uses color scale for dB.

E in Eq. (18d) and the term $FL_E=0$. With these assumptions, the radial vibration of the tube wall can be predicted with Eq. (30) and used as the input to a model that will predict the resulting motion at the gel surface. Note that the motion at the phantom surface just above the tube should closely match that of the tube itself for small values of h_g.

Alternatively, to predict vibratory motion at any location on the phantom surface or within the phantom, the axisymmetrically vibrating tube can be approximated via a finite number of elementary acoustic sources, finite monopoles or dipoles, spaced sufficiently close together. Given that the theoretically modeled radial motion of the tube is axisymmetric, a row of monopoles may seem like a logical approach. However, monopoles do not generate shear waves, only compression waves. In the actual experiment, wall motion will be nonaxisymmetric given the random nature of turbulence, which would result in the generation of shear waves in addition to compression waves. Dipoles do radiate both compression and shear waves, and may result in a more realistic simulation in some cases for some frequency regimes. In a previous study, it was shown that, for elementary acoustic sources comparable in dimension to the compliant tube considered here, and just below the surface of a soft-tissue viscoelastic phantom material, the resulting phantom surface motion is composed of contributions from both shear and compression waves, with shear wave contributions being dominant below ~100 Hz and near the source and compression waves becoming dominant above ~150 Hz and as one moves further from the source.[30]

Two simulation approaches, using monopoles or dipoles, are developed here. In either case, scattering of the field of one elementary source caused by another or by itself after reflection from a boundary is neglected. That is, the monopoles and dipoles will be treated as infinitesimal. Additionally, the viscoelastic medium will be considered to be a semi-infinite half-space. The effect of these assumptions on the predicted field for single monopole or dipole sources is discussed in Royston *et al.*[30] Also in Ref. 30, it was shown that, generally as one moves away from being directly over the source, the surface response to an infinitesimal monopole or dipole just below the surface is reasonably approximated by simply doubling the theoretical response at the location of the half-space surface that is calculated based on the assumption of an infinite medium. This approximation, and neglecting multiple reflections from the source, is expected to worsen the match of theory to experiment when measurements are taken directly over the source.

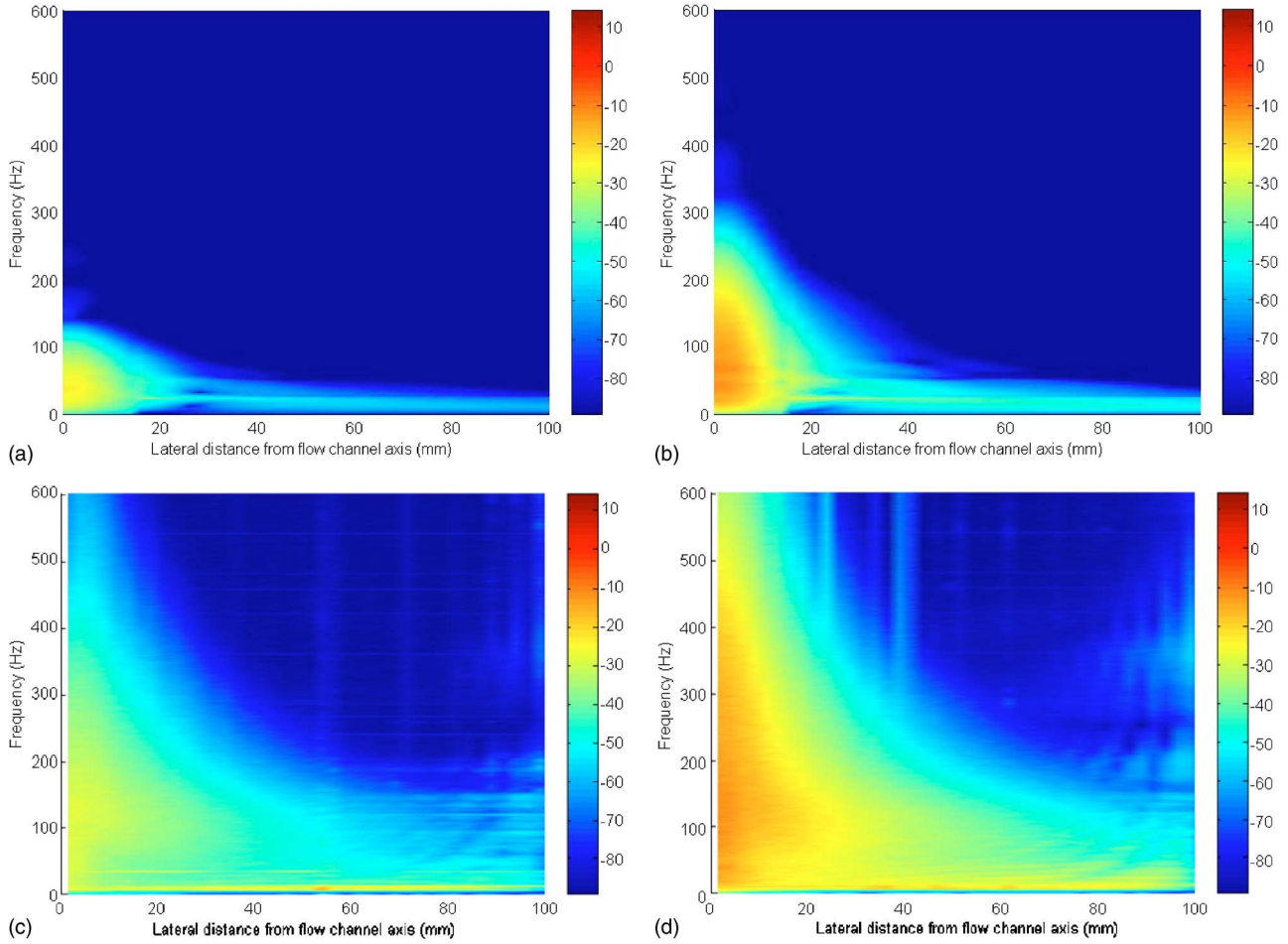


FIG. 12. (Color online) Vertical velocity (dB *re*: 1 mm/s) at phantom surface for constricted compliant tube embedded in soft tissue gel phantom downstream of constriction as a function of lateral position moving away from the tube. (a) Dipole theory at 15 mm downstream, $\mathrm{Re}_D=1000$. (b) Dipole theory at 15 mm downstream, $\mathrm{Re}_D=2000$. (c) Experiment at 10 mm downstream, $\mathrm{Re}_D=1000$. (d) Experiment at 10 mm downstream, $\mathrm{Re}_D=2000$. Online version uses color scale for dB.

1. Monopole approach

For the tube, radial wall velocity is given by Eq. (30). It can be calculated at any axial location x and azimuthal angle θ. Suppose its value is calculated at a finite number of points along the tube with axial resolution ΔL. The calculated radial wall displacement at one of these points w_x is assumed to represent the axisymmetric radial motion of that segment of the tube from $x-\Delta L/2$ to $x+\Delta L/2$. The volume displacement of this segment of the tube will be equated to the volume displacement of a finite monopole of radius a_m undergoing radial displacement of amplitude u_0 and located at the geometric center of the tube segment it represents. This leads to

$$\Delta L a 2\pi w_x = 4\pi a_m^2 u_0. \tag{32}$$

Set $\sqrt{2}a_m$ equal to $\sqrt{\Delta L a}$; then, $w_x=u_0$. The corresponding spherically symmetric radiated field of the monopole is

$$u_R = u_0 \frac{h_1^{(2)}[k_\alpha R]}{h_1^{(2)}[k_\alpha a_m]} e^{j\omega t}, \tag{33}$$

where $h_1^{(2)}$ denotes a spherical Hankel function for outgoing waves, R denotes the distance from the monopole location to the point of measurement, and k_α denotes the complex wave number for compression wave motion in the viscoelastic phantom material.[30] The vertical response at the free surface of the gel phantom can then be approximated by summing the response to all of the monopoles used in the discretization, accounting for the angle that the outgoing spherical wave makes with the vertical direction.

2. Dipole approach

For this case, the tube response is discretized with axial resolution ΔL and, this time, azimuthal resolution $a\Delta\theta$. The radial displacement calculated at the center of a segment of the tube extending ΔL by $a\Delta\theta$ is denoted as $w_{x\theta}$, and that entire segment of the tube is approximated as having radial displacement $w_{x\theta}$. See Fig. 9. A finite dipole of radius a_d vibrating with amplitude u_0 displaces a volume $\pi a_d^2 u_0$ directly in front of it. This is equated to the volume displacement of the portion of the tube the dipole represents, which is given by $\Delta L a \Delta\theta w_{x\theta}$. If one takes $\sqrt{\pi a_d}=\sqrt{\Delta L a \Delta\theta}$, then $w_{x\theta}=u_0$. The dipole is positioned at the geometric center of the tube wall segment it represents, and its axis is oriented normal outward to that segment. The corresponding radiated field is

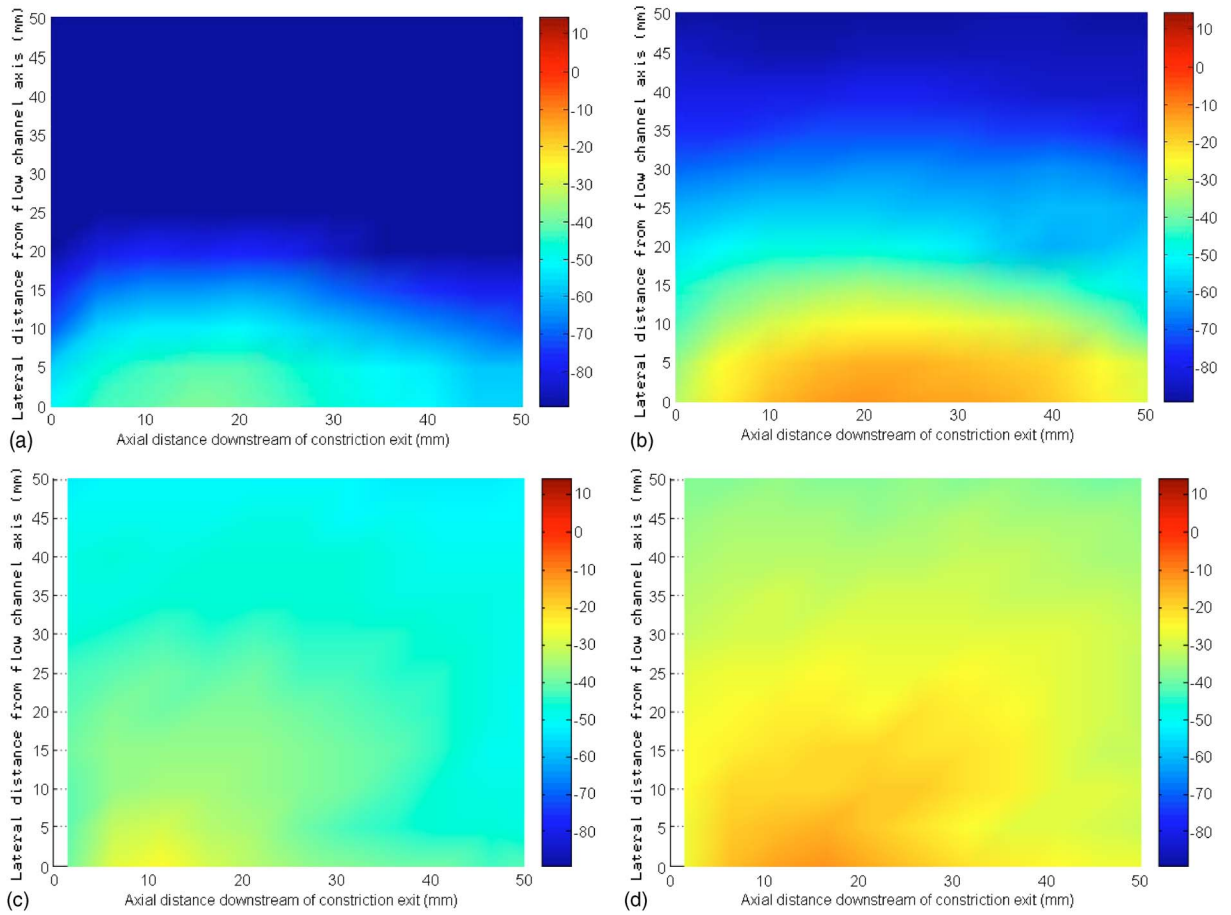


FIG. 13. (Color online) Vertical velocity (dB *re*: 1 mm/s) at phantom surface for constricted compliant tube embedded in soft tissue gel phantom as a function of axial and lateral position at 100 Hz. (a) Dipole theory, $\mathrm{Re}_D=1000$. (b) Dipole theory, $\mathrm{Re}_D=2000$. (c) Experiment, $\mathrm{Re}_D=1000$. (d) Experiment, $\mathrm{Re}_D=2000$. Online version uses color scale for dB.

$$u_R = N_1 \cos(\psi)\{[2 + 2jk_\alpha R - (k_\alpha R)^2]e^{-jk_\alpha R} + 2N_2(-jk_\beta R - 1)e^{-jk_\beta R}\}e^{j\omega t}/R^3, \tag{34}$$

$$u_\psi = -N_1 \sin(\psi)\{(-jk_\alpha R - 1)e^{-jk_\alpha R} + N_2[1 + jk_\beta R - (k_\beta R)^2]e^{-jk_\beta R}\}e^{j\omega t}/R^3, \tag{35}$$

where in this expression ψ denotes the angle made between the principle axis of that particular dipole and a vector drawn from the dipole to the point of measurement.[30] Here, R denotes the distance from the dipole location to the point of measurement, u_R and u_ψ denote the radial and tangential components of the generated vibratory field relative to the dipole, and k_α and k_β denote the complex wave number of the radiated compression and shear waves, respectively. Note that only dipole and measurement point combinations for which $-\pi/2 < \psi < \pi/2$ are used in the summation; i.e., only the field radiating radially outward from the vessel wall is used. The coefficients N_1 and N_2 can be specified based on the boundary conditions for "welded" contact or "lossless slip" contact. The welded contact case results in

$$N_1 = -u_0 a^3 e^{jk_\alpha a} \times \frac{3 + 3jk_\beta a - (k_\beta a)^2}{(1 + jk_\beta a)(k_\alpha a)^2 + 2(1 + jk_\alpha a)(k_\beta a)^2 - (k_\alpha a)^2(k_\beta a)^2}, \tag{36}$$

$$N_2 = e^{-j(k_\alpha - k_\beta)a}\frac{3 + 3jk_\alpha a - (k_\alpha a)^2}{3 + 3jk_\beta a - (k_\beta a)^2}. \tag{37}$$

The vertical response at the free surface is then approximated by summing the response to all of the dipoles used in the discretization, taking into account the angle that u_R and u_ψ make with the vertical direction.

B. Theoretical predictions and experimental measurements

Based on the approaches described above, the radial velocity of the latex tube and the vertical velocity at the gel surface for the constricted vessel and geometry of Fig. 1 were calculated at points directly above the embedded vessel and at points lateral to the tube axis downstream of the constriction. These results are shown in Figs. 10–13. The theoretical predictions in Figs. 11–13 were based on using sets of eight dipoles equally spaced circumferentially and axially

spaced at 2.5 mm to approximate the tube of length $L = 100$ mm. Increasing the dipole resolution further did not noticeably alter the results. Predictions using the dipole approach matched experimental measurements to a greater degree than those using the monopole approach; results based on the monopole approach are not shown. It is believed that this is due to the inability of the monopole approach to generate shear waves. Generally, it was observed experimentally that the overall vibration levels were greater than predicted via the dipole approach. This is believed to be partially due to the finite volume of the viscoelastic phantom resulting in multiple reflections, which the half-space theory does not capture.

In comparing Figs. 10(a)–10(c), 11(c), and 11(d), note that the prediction of the latex radial wall velocity closely matches the measured radial wall velocity and the measured gel vertical surface velocity just above the vessel, as expected given the shallow depth of the vessel (~6.5 mm from the surface to the top of the latex tube). The dipole approximation provides a less accurate prediction [Figs. 11(a) and 11(b)], but still captures the trends in frequency and location. The dipole approximation also enables a prediction of surface motion not directly over the vessel, capturing experimental trends as shown in Figs. 12 and 13. Experimentally, and in the theoretical simulation, it is observed that the surrounding viscoelastic gel material substantially, though not completely, dampens the resonant properties of the water-filled latex tube, resulting in a more uniform spectral content epicentered just downstream of the constriction, yet still possessing substantial frequency content well through several hundred hertz.

VI. CONCLUSION

The vibration of a thin-walled cylindrical, compliant viscoelastic tube with internal flow and an axisymmetric constriction that results in turbulent fluid flow was studied theoretically and experimentally. The developed closed-form analytical model of the coupled fluid and structural system may provide a baseline for future, more comprehensive analyses, analytical and computational, that may improve upon some of the indicated shortcomings or extend the approach to more complex and realistic geometries and material properties. Additionally, the reported experimental study provides a unique and comprehensive set of measurements and associated discussion relevant to vascular dynamics and diagnostics.

Vibration of the tube was considered with internal fluid coupling only, and with coupling to internal flowing fluid and external stagnant fluid or external tissuelike viscoelastic material. The theoretical analysis included the adaptation of a model for turbulence in the internal fluid and its vibratory excitation of and interaction with the tube wall and surrounding fluid or viscoelastic medium. Analytical predictions compared favorably with experimental measurements. Reasons for identified discrepancies between theory and experiment were provided. It was hypothesized that the primary reasons for these discrepancies were that (1) the empirical turbulence model did not fully account for compliant tube vibration and how it alters the turbulent field, and (2) the linear structural model of the tube did not capture some of the apparent nonlinear phenomena that were highly dependent upon the mean pressure within the tube. Note, it did appear that, when the tube was embedded in the viscoelastic material, the added inertial and dissipative loading on the tube attenuated these sources of discrepancy.

Extension of the theoretical modeling approach to more realistic and complex geometries may require numerical adaptations to predict the resulting transitional or turbulent flow field and surrounding solid tissue vibration in this highly coupled system. However, the analytical model provided here could be used to validate numerical techniques. Additionally, the developed theory could be used to provide estimates of the dynamic forces *in vivo* that relate to pathology and may be difficult or impossible to measure directly, such as the dynamic stress levels created in a vascular wall due to turbulent flow. Also, it is envisioned that the developed theoretical and experimental techniques may be helpful in evaluating and improving medical diagnostic technologies that utilize audible frequency phenomena generated below the skin surface to provide diagnostic information as an acoustic image or in another form.

ACKNOWLEDGMENTS

The financial support of the National Institutes of Health (EB002511 and HL55296) and the Whitaker Foundation (BME RG 01-0198) are acknowledged.

[1] R. Y. Kanterman, T. M. Vesely, T. K. Pilgram, B. W. Guy, D. W. Windus, and D. Picus, "Dialysis access grafts: Anatomic location of venous stenosis and results of angioplasty," Radiology **195**, 135–139 (1995).
[2] M. F. Fillinger, E. R. Reinitz, R. A. Schwartz, D. E. Resetarits, A. M. Paskanik, D. Bruch, and C. E. Bredenberg, "Graft geometry and venous intimal-medial hyperplasia in arteriovenous loop grafts," J. Vasc. Surg. **11**, 556–566 (1990).
[3] F. Loth, P. F. Fischer, N. Arslan, C. D. Bertram, S. E. Lee, T. J. Royston, W. E. Shaalan, and H. S. Bassiouny, "Transitional flow at the venous anastomosis of an arteriovenous graft: Potential activation of the ERK1/2 mechanotransduction pathway," ASME J. Biomech. Eng. **125**, 49–61 (2003).
[4] S. W. Lee, F. Loth, T. J. Royston, P. F. Fischer, H. S. Bassiouny, and J. K. Grogan, "Flow induced vein wall vibration in an arteriovenous graft," J. Fluids Struct. (in press).
[5] R. S. Lees and C. F. Dewey, "Phonoangiography: A new noninvasive diagnostic method for studying arterial disease," Proc. Natl. Acad. Sci. U.S.A. **67**, 935–942 (1970).
[6] B. Kim and W. K. Corcoran, "Experimental measurement of turbulence spectra distal to stenosis," J. Biomech. **7**, 335–342 (1974).
[7] G. W. Duncan, J. O. Gruber, C. F. Dewey, Jr., G. S. Meyers, and R. S. Lees, "Evaluation of carotid stenosis by phonoangiography," N. Engl. J. Med. **293**, 1124–1128 (1975).
[8] R. Gupta, J. W. Miller, A. P. Yoganathan, F. E. Udwadia, and W. H. Corcoran, "Spectral analysis of arterial sounds: A noninvasive method of studying arterial disease," Med. Biol. Eng. **13**, 700–705 (1975).
[9] R. J. Tobin and I.-D. Chang, "Wall pressure spectra scaling downstream of stenosis in steady tube flow," J. Biomech. **9**, 633–640 (1976).
[10] J. J. Fredberg, "Origin and character of vascular murmurs: Model studies," J. Acoust. Soc. Am. **61**, 1077–1085 (1977).
[11] W. H. Pitts III and C. F. Dewey, Jr., "Spectral and temporal characteristics of post-stenotic turbulent wall pressure fluctuations," ASME J. Biomech. Eng. **101**, 89–95 (1979).
[12] J. Wang, B. Tie, W. Welkowitz, J. Semmlow, and J. Kostis, "Modeling sound generation in stenosed coronary arteries," IEEE Trans. Biomed. Eng. **37**, 1087–1094 (1990).
[13] Y. Kurokawa, S. Abiko, and K. Watanabe, "Noninvasive detection of in-

tracranial vascular lesions by recording blood flow sounds," Stroke **25**, 397–402 (1994).
[14]M. Akay, Y. Akay, W. Welowitz, J. L. Semmlow, and J. B. Kostis, "Application of adaptive filters to noninvasive acoustical detection of coronary occlusions before and after angioplasty," IEEE Trans. Biomed. Eng. **39**, 176–183 (1992).
[15]A. Akay, M. Akay, W. Welkowitz, S. Lewkowicz, and Y. Palti, "Dynamics of sounds caused by partially occluded femoral arteries in dogs," Ann. Biomed. Eng. **22**, 493–500 (1994).
[16]A. O. Borisyuk, "Noise field in the human chest due to turbulent flow in a larger blood vessel," Flow, Turbul. Combust. **61**, 269–284 (1999).
[17]H. A. Mansy, S. J. Hoxie, N. H. Patel, and R. H. Sandler, "Computerized analysis of auscultory sounds associated with vascular patency of hemodialysis access," Med. Biol. Eng. Comput. **43**, 56–62 (2005).
[18]N. L. Owsley and A. J. Hull, "Beamformed nearfield imaging of a simulated coronary artery containing a stenosis," IEEE Trans. Med. Imaging **17**, 900–909 (1998).
[19]C. E. Chassaing, S. D. Stearns, M. H. van Horn, and C. A. Ryden, "Noninvasive turbulent blood flow imaging system," United States Patent No. 6,278,890 issued 8/21/2001.
[20]M. I. Plett, K. W. Beach, B. Dunmire, K. G. Brown, J. F. Primozich, and E. Strandness, Jr., "*In vivo* ultrasonic measurement of tissue vibration at a stenosis: A case study," Ultrasound Med. Biol. **27**, 1049–1058 (2001).
[21]R. L. Kirkeeide, D. F. Young, and N. R. Cholvin, "Wall vibrations induced by flow through simulated stenoses in models and arteries," J. Biomech. **10**, 431–441 (1977).
[22]J. B. Grotberg and O. E. Jensen, "Biofluid mechanics in flexible tubes," Annu. Rev. Fluid Mech. **36**, 121–147 (2004).
[23]W. L. Keith and B. M. Abraham, "Effects of convection and decay of turbulence on the wall pressure wavenumber-frequency spectrum," J. Fluids Eng. **119**, 50–55 (1997).
[24]R. K. Jain, "Vibration of fluid-filled, orthotropic cylindrical shells," J. Sound Vib. **37**, 379–388 (1974).
[25]A. Leissa, *Vibration of Shells* (Acous. Soc. Amer., 1993).
[26]M. C. Junger and D. Feit, *Sound, Structures and Their Interaction* (Acous. Soc. Amer., 1993).
[27]C. R. Fuller, "The input mobility of an infinite circular cylindrical elastic shell filled with fluid," J. Sound Vib. **87**, 409–427 (1983).
[28]B. J. Brevart and C. R. Fuller, "Energy exchange between the coupled media of impulsively excited, fluid-filled, elastic cylinders," J. Sound Vib. **190**, 763–774 (1996).
[29]X. M. Zhang, M. Fatemi, R. R. Kinnick, and J. F. Greenleaf, "Noncontact ultrasound stimulated optical vibrometry study of coupled vibration of arterial tubes in fluids," J. Acoust. Soc. Am. **113**, 1249–1257 (2003).
[30]T. J. Royston, Y. Yazicioglu, and F. Loth, "Surface response of a viscoelastic medium to subsurface acoustic sources with application to medical diagnosis," J. Acoust. Soc. Am. **113**, 1109–1121 (2003).
[31]T. J. Royston, H. A. Mansy, and R. H. Sandler, "Excitation and propagation of surface waves on a viscoelastic half-space with application to medical diagnosis," J. Acoust. Soc. Am. **106**, 3678–3686 (1999).

A model for estimating ultrasound attenuation along the propagation path to the fetus from backscattered waveforms

Timothy A. Bigelow[a] and William D. O'Brien, Jr.[b]
Bioacoustics Research Laboratory, Department of Electrical and Computer Engineering, University of Illinois, 405 North Mathews, Urbana, Illinois 61801

(Received 26 October 2004; revised 28 April 2005; accepted 11 May 2005)

Accurate estimates of the ultrasound pressure and/or intensity incident on the developing fetus on a patient-specific basis could improve the diagnostic potential of medical ultrasound by allowing the clinician to increase the transmit power while still avoiding the potential for harmful bioeffects. Neglecting nonlinear effects, the pressure/intensity can be estimated if an accurate estimate of the attenuation along the propagation path (i.e., total attenuation) can be obtained. Herein, a method for determining the total attenuation from the backscattered power spectrum from the developing fetus is proposed. The boundaries between amnion and either the fetus' skull or soft tissue are each modeled as planar impedance boundaries at an unknown orientation with respect to the sound beam. A mathematical analysis demonstrates that the normalized returned voltage spectrum from this model is independent of the planes orientation. Hence, the total attenuation can be estimated by comparing the location of the spectral peak in the reflection from the fetus to the location of the spectral peak in a reflection obtained from a rigid plane in a water bath. The independence of the attenuation estimate and plane orientation is then demonstrated experimentally using a Plexiglas plate, a rat's skull, and a tissue-mimicking phantom. © *2005 Acoustical Society of America.*
[DOI: 10.1121/1.1945564]

PACS number(s): 43.80.Ev, 43.80.Qf, 43.20.El [FD] Pages: 1210–1220

LIST OF SYMBOLS

f = frequency
f_{peak} = frequency corresponding to the spectral peak at each inclination angle (i.e., $|V_{plane}(\omega)|/\max_{\forall\omega}|V_{plane}(\omega)| \cong \exp[-(f-f_{peak})^2/2\sigma_{\omega p}^2]$).
$g(\vec{r}_d,\vec{r}')$ = effective Green's function valid from the scattering region to the detector
G_o = geometric gain value for velocity potential field at focus when W_{source} is approximated by a Gaussian (m)
G_T = dimensionless aperture gain function that accounts for the focusing of the ultrasound source
H = dimensionless filtering characteristics for the ultrasound source
k = effective wave number along the propagation path
$\tilde{k}$ = effective complex wave number along the propagation path (i.e., $\tilde{k}=k+i\alpha$)
K_{uV} = conversion constant relating voltage to particle velocity for ultrasound source (m/s V^{-1})
$\mathbf{M}_{\mathbf{image}}$ = matrix used to generate image point
$\vec{n}_f$ = the outward normal for the plane at arbitrary angle to beam axis
p_{plane} = pressure field from rigid plane placed near focal plane
$\vec{r}_f$ = locations on inclined plane
$\vec{r}_1$ = points on aperture plane of image source
$\vec{r}_T,\vec{r}_d$ = locations on aperture plane of transmitter/detector
S_f = surface of inclined plane near focal plane
S_I = aperture area of image source
S_T = aperture area of ultrasound transmitter
V_{inc} = voltage applied to the ultrasound source during transmit
V_{plane} = voltage from ultrasound source due to the backscatter from rigid plane near focus
w_x,w_y,w_z = equivalent Gaussian dimensions on receive of pressure field in focal region
x_I,y_I,z_I = coordinate location of image point
z_f = distance of rigid plane to the focal plane
z_p = distance along the beam axis from the focus to the intersection of the plane with the beam axis
z_T,z_d = distance of the aperture plane of the ultrasound transmitter/detector to the focal plane
α = effective attenuation along the propagation path
α_{error} = error in attenuation associated with inclination angle of plane
α_o = slope of attenuation assuming strict linear frequency dependence (i.e., $\alpha=\alpha_o f$)
Γ_{plane} = reflection coefficient of plane
θ_f,ϕ_f = angles describing orientation of plane with beam axis
ξ_x,ξ_y,ξ_z = coordinate system for the image source

[a]Electronic mail: bigelow@uiuc.edu
[b]Author to whom correspondence should be addressed. Electronic mail: wdo@uiuc.edu

$\hat{\xi}_x, \hat{\xi}_y, \hat{\xi}_z$ = unit normal vectors defining coordinate system for the image source
ρ = density
$\sigma_{\omega p}$ = Gaussian bandwidth for reflected voltage from inclined plane (i.e., $|V_{plane}(\omega)|/\max_{\forall\omega}|V_{plane}(\omega)| \cong \exp[-((f-f_{peak})^2/2\sigma_{\omega p}^2]]$).
ω = radian frequency
$\omega_{\max}$ = radian frequency corresponding to spectral peak
$\cdot$ = dot product

I. INTRODUCTION

Over the years, ultrasound as a medical diagnostic tool has become increasingly valuable. This can be attributed, in part, to its ability to display gray-scale images of anatomy and color-coded blood flow in real time. A specific benefit that is provided by diagnostic ultrasound is its ability to "safely" image embryos and fetuses *in utero*. This allows the medical professionals to accurately diagnose and monitor the different embryo/fetal developmental stages during pregnancy. Although high-quality images can be obtained for many patients, some patients are challenging to image due to increased attenuation along the propagation path, thus degrading the backscattered signals. A simple solution would be to increase the transmit power of the ultrasound source, but increased output power translates to increased tissue heating as well as the possibility of other mechanical bioeffects (AIUM, 1988, 1993, 2000; NCRP, 2002). Of particular importance is the heating near the developing cranial bone because heating of the developing brain tissue has the potential to result in long-term neurological disorders (Barnett, 2000). Heating is of greater concern when the ultrasound beam impinges on the developing cranial bone due to the large absorption of the ossified bone tissue (Myers, 2004; Barnett, 2001; Fuji *et al.* 1999; Drewniak and Dunn, 1996; O'Neil *et al.*, 1994; Haken *et al.*, 1992; Wu and Du, 1990; Carstensen *et al.*, 1990; Drewniak *et al.*, 1989). Currently, these heating concerns are avoided by requiring that the output levels be kept artificially much less than those anticipated to produce biologically significant temperature increases.

The risk for tissue heating is currently assessed by assuming that the attenuation along the propagation path is 0.3 dB/cm-MHz, a conservative lower limit for the attenuation of biological tissue (AIUM/NEMA, 1998, Abbott, 1999). However, in a study of 23 nonpregnant female volunteers (Siddiqi *et al.*, 1992), the measured attenuation coefficient for the abdominal wall varied from 0.40 to 4.0 dB/cm-MHz (mean of 1.39 dB/cm-MHz with a standard deviation of ±0.89 dB/cm-MHz). Hence, based only on attenuation of the abdominal wall, the potential exists for patient variability as high as 3.4 dB/cm-MHz, or as high as 1.8 dB/cm-MHz using the two standard deviation criterion. But, for transabdominal ultrasound scanning, the entire propagation path needs to be considered, including abdominal wall, bladder, myometrium, and vaginal wall. In another *in vivo* study (Siddiqi *et al.*, 1999) that also combined the results from two other separate *in vivo* studies, the total attenuation along the propagation path from the abdominal wall through the vagina was from 0.8±0.4 dB/cm-MHz for 57 subjects with empty bladders and 0.6±0.3 dB/cm-MHz for 64 subjects with full bladders. Based on total attenuation, the potential exists for patient variability as high as 0.8 dB/cm-MHz for empty bladders and 0.6 dB/cm-MHz for full bladders (two standard deviation criterion). These examples demonstrate a large patient-specific variability.

As the attenuation along the propagation path is increased, more of the transmitted ultrasound energy is absorbed before it can reach the developing skull, thus reducing the *in vivo* exposure pressure/intensity. Hence, the resulting temperature increase is reduced as the total attenuation along the propagation path is increased. As a result, the transmit power could be increased to improve the image quality on the challenging patients without endangering the fetus provided that the *in vivo* and *in situ* pressure and/or intensity could be estimated on a patient-specific basis. Neglecting nonlinear effects, the *in vivo* pressure and/or intensity can be obtained by measuring the pressure and/or intensity in a water bath using a calibrated hydrophone and then derating the measured waveform using the attenuation coefficient as measured for a specific patient.

In this study, one possibility of estimating the attenuation along the propagation path leading to the fetus on a patient-specific basis has been investigated. The hypothesis is that if an accurate model of the scattering structure can be developed, then the attenuation along the propagation path can be estimated by comparing the *in vivo* backscattered spectrum to a reference spectrum obtained in a water bath. The simplest scattering structure would be a planar boundary between two impedances where the sound beam is incident at some arbitrary angle with respect to the planar boundary. One possible example of a planar boundary *in utero* might be the tissue/skull boundary of the fetus. Hence, the theoretical voltage spectrum from an inclined planar boundary is first derived. When the backscattered radio frequency (rf) spectrum from the planar boundary is then compared to a reference spectrum, an estimate for the attenuation along the propagation path is obtained by measuring the frequency location of the main spectral peak. The theoretical analysis is then validated using both a Plexiglas plate and a rat's skull in water baths at different inclination angles. Unfortunately, the simple plane model may not be valid for the skull in the early developmental stages due to the ossification process of the skull. Hence, the simple plane model is also evaluated for a soft-tissue/water boundary, similar to the aminiotic fluid/soft tissue boundary *in utero*, using a soft-tissue-mimicking phantom in a water bath. Lastly, the applicability of the results to estimating the exposure conditions of the fetus is discussed.

II. THEORETICAL CALCULATIONS

In this section, the voltage received by the ultrasound source when operating in pulse echo mode when a planar boundary is placed in the focal region will be derived. It is assumed that the velocity potential field in the focal region follows a three-dimensional Gaussian distribution as was verified by Bigelow and O'Brien (2004a). The coordinate

system for the problem is shown in Fig. 1. In this figure, $\vec{n}_f$ is the outward normal for the plane that intersects the z axis (beam axis) at a distance of z_p from the focus (positive for $z_p > 0$) with n_{fx}, n_{fy}, n_{fz} given by

$$\begin{aligned} n_{fx}\hat{x} &= \sin(\theta_f)\cos(\phi_f)\hat{x}, \\ n_{fy}\hat{y} &= \sin(\theta_f)\sin(\phi_f)\hat{y}, \\ n_{fz}\hat{z} &= \cos(\theta_f)(-\hat{z}). \end{aligned} \tag{1}$$

Hence, the equation describing the plane is given by

$$x\sin(\theta_f)\cos(\phi_f) + y\sin(\theta_f)\sin(\phi_f) + (z_p - z)\cos(\theta_f) = 0. \tag{2}$$

The goal is to solve for the reflected longitudinal waves in the region of the ultrasound source. Shear waves were not included in the initial developments in order to simplify the analysis. The impact of neglecting shear waves is beyond the scope of the current work, but, in general, the influence of shear waves on the reflected longitudinal waves is small when the incident angle of the wave with respect to the plane normal is small (Mayer, 1964). Hence, the derived equations will be limited to small incident angles for the ultrasound beam.

After applying the Kirchoff-Helmholtz integral theorem, the pressure field reflected by a planar specular reflector in the focal region is given by (Pierce, 1991)

$$\begin{aligned} p_{plane}(\vec{r}_d, \omega) = \frac{1}{4\pi}\iint_{S_f} d\vec{r}_f \vec{n}_f \cdot (g(\vec{r}_d, \vec{r}_f)\nabla_f p_{plane}(\vec{r}_f, \omega) \\ - p_{plane}(\vec{r}_f, \omega)\nabla_f g(\vec{r}_d, \vec{r}_f)), \end{aligned} \tag{3}$$

which, after applying the appropriate Green's function g, becomes (Pierce, 1991)

$$p_{plane}(\vec{r}_d, \omega) = \frac{1}{2\pi}\iint_{S_f} d\vec{r}_f \vec{n}_f \cdot \left(\frac{e^{i\tilde{k}|\vec{r}_d - \vec{r}_f|}}{|\vec{r}_d - \vec{r}_f|}\nabla_f p_{plane}(\vec{r}_f, \omega)\right). \tag{4}$$

Once the reflected pressure field at the planar surface [i.e., $p_{plane}(\vec{r}_f, \omega)$] is known, the reflected pressure field everywhere [i.e., $p_{plane}(\vec{r}_d, \omega)$] can be determined.

In order to find the pressure field immediately after reflection, an appropriate image source was generated corresponding to the original transmitting source similar to the work done by Pedersen and Orofino (1996). The field from the image source was then determined at the plane. Image sources are normally used to quickly satisfy the boundary conditions of either pressure release or rigid surfaces (i.e., magnitude of reflection coefficient equal to one). When the magnitude of the reflection coefficient is different from one, the use of image sources must be modified to include the angle of incidence of the fields decomposed into a set of plane waves (Morse and Ingard, 1968). However, if the incident angles are small, the angular dependence of the reflection coefficient can be ignored in a first-order calculation.

The location of the image source can be found by reflecting the original source across the plane. The reflection of an arbitrary point across the plane was performed by transforming the coordinate system until the plane was aligned with a new x-y plane with the origin at the intersection between the beam axis of the ultrasound source and the plane using a translation transformation and two rotation transformations. Then, the point was reflected to the other side of the plane along the z axis, and the coordinate system was returned to its original orientation with two additional rotation transformations and a translation transformation (Hearn and Baker, 1997). The transformations are illustrated in Eq. (5):

$$\begin{bmatrix} x_I \\ y_I \\ z_I \\ 1 \end{bmatrix} = \overbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & z_p \\ 0 & 0 & 0 & 1 \end{bmatrix}}^{Translation} \overbrace{\begin{bmatrix} -\cos(\phi_f) & \sin(\phi_f) & 0 & 0 \\ -\sin(\phi_f) & -\cos(\phi_f) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}^{\phi-Rotation} \overbrace{\begin{bmatrix} -\cos(\theta_f) & 0 & -\sin(\theta_f) & 0 \\ 0 & 1 & 0 & 0 \\ \sin(\theta_f) & 0 & -\cos(\theta_f) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}^{\theta-Rotation}$$

$$\times \overbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}^{Reflection} \overbrace{\begin{bmatrix} -\cos(\theta_f) & 0 & \sin(\theta_f) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\theta_f) & 0 & -\cos(\theta_f) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}^{\theta-Rotation} \overbrace{\begin{bmatrix} -\cos(\phi_f) & -\sin(\phi_f) & 0 & 0 \\ \sin(\phi_f) & -\cos(\phi_f) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}^{\phi-Rotation} \overbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -z_p \\ 0 & 0 & 0 & 1 \end{bmatrix}}^{Translation} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{M}_{\mathbf{image}}^{1,1} & \mathbf{M}_{\mathbf{image}}^{1,2} & \mathbf{M}_{\mathbf{image}}^{1,3} & \mathbf{M}_{\mathbf{image}}^{1,4} \\ \mathbf{M}_{\mathbf{image}}^{2,1} & \mathbf{M}_{\mathbf{image}}^{2,2} & \mathbf{M}_{\mathbf{image}}^{2,3} & \mathbf{M}_{\mathbf{image}}^{2,4} \\ \mathbf{M}_{\mathbf{image}}^{3,1} & \mathbf{M}_{\mathbf{image}}^{3,2} & \mathbf{M}_{\mathbf{image}}^{3,3} & \mathbf{M}_{\mathbf{image}}^{3,4} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \mathbf{M}_{\mathbf{image}} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \tag{5}$$

where

$$\mathbf{M}^{:,1}_{\mathbf{image}} = \begin{bmatrix} 1-2\sin^2(\theta_f)\cos^2(\phi_f) \\ -2\sin^2(\theta_f)\sin(\phi_f)\cos(\phi_f) \\ 2\sin(\theta_f)\cos(\phi_f)\cos(\theta_f) \\ 0 \end{bmatrix},$$

$$\mathbf{M}^{:,2}_{\mathbf{image}} = \begin{bmatrix} -2\sin^2(\theta_f)\sin(\phi_f)\cos(\phi_f) \\ 1-2\sin^2(\theta_f)\sin^2(\phi_f) \\ 2\sin(\theta_f)\sin(\phi_f)\cos(\theta_f) \\ 0 \end{bmatrix},$$

$$\mathbf{M}^{:,3}_{\mathbf{image}} = \begin{bmatrix} 2\sin(\theta_f)\cos(\phi_f)\cos(\theta_f) \\ 2\sin(\theta_f)\sin(\phi_f)\cos(\theta_f) \\ 1-2\cos^2(\theta_f) \\ 0 \end{bmatrix}, \quad (6)$$

$$\mathbf{M}^{:,4}_{\mathbf{image}} = \begin{bmatrix} -2z_p\sin(\theta_f)\cos(\phi_f)\cos(\theta_f) \\ -2z_p\sin(\theta_f)\sin(\phi_f)\cos(\theta_f) \\ 2z_p\cos^2(\theta_f) \\ 1 \end{bmatrix},$$

and (x_I, y_I, z_I) is the image of the original point. Therefore, the field from the image source would be given by

$$p_{plane}(\vec{r}_f,\omega) = \frac{-i\omega\rho\Gamma_{plane}K_{uV}(\omega)V_{inc}(\omega)H(\omega)}{2\pi} \times \iint_{S_I} d\vec{r}_I G_T(\vec{r}_I,\omega)\frac{e^{i\tilde{k}|\vec{r}_I-\vec{r}_f|}}{|\vec{r}_I-\vec{r}_f|}, \quad (7)$$

where Γ_{plane} accounts for the fact that the plane may not be a perfect reflector. Also, assuming that the planar surface is near the focus (i.e., $|\vec{r}_f|$ is small compared to $|\vec{r}_I|$), Eq. (7) becomes

$$p_{plane}(\vec{r}_f,\omega) \cong \frac{-i\omega\rho\Gamma_{plane}K_{uV}(\omega)V_{inc}(\omega)H(\omega)}{2\pi} \times \iint_{S_I} d\vec{r}_I G_T(\vec{r}_I,\omega)\frac{e^{i\tilde{k}r_I}}{r_I}e^{i\tilde{k}\left(\frac{\vec{r}_I}{r_I}\right)\vec{r}_f}, \quad (8)$$

which can be written in closed form if the velocity potential field near the focus can be approximated as a three-dimensional Gaussian distribution, an assumption that was shown to be valid for scattering problems by Bigelow and O'Brien (2004a). Hence, Eq. (8) becomes

$$p_{plane}(\vec{r}_f,\omega) \cong \frac{-i\omega\rho\Gamma_{plane}K_{uV}(\omega)V_{inc}(\omega)H(\omega)G_o}{2\pi} \times e^{-((\xi_x(\vec{r}_f)/w_x)^2+(\xi_y(\vec{r}_f)/w_y)^2+(\xi_z(\vec{r}_f)/w_z)^2)}e^{i\tilde{k}(z_T-\xi_z)}, \quad (9)$$

where (ξ_x, ξ_y, ξ_z) refer to the coordinate system for the image source and z_T is the distance of the aperture plane of the image source from its focus (same as for the original source).

The locations of points (ξ_x, ξ_y, ξ_z) now need to be determined in the original coordinate system. Because the ξ_x, ξ_y, and ξ_z axes are the image of the x, y, and z axes, Eq. (5) yields

$$\begin{aligned} &\hat{\xi}_x\cdot\hat{x} = \mathbf{M}^{1,1}_{\mathbf{image}}, \quad \hat{\xi}_y\cdot\hat{x} = \mathbf{M}^{1,2}_{\mathbf{image}}, \quad \hat{\xi}_z\cdot\hat{x} = \mathbf{M}^{1,3}_{\mathbf{image}}, \\ &\hat{\xi}_x\cdot\hat{y} = \mathbf{M}^{2,1}_{\mathbf{image}}, \quad \hat{\xi}_y\cdot\hat{y} = \mathbf{M}^{2,2}_{\mathbf{image}}, \quad \hat{\xi}_z\cdot\hat{y} = \mathbf{M}^{2,3}_{\mathbf{image}}, \\ &\hat{\xi}_x\cdot\hat{z} = \mathbf{M}^{3,1}_{\mathbf{image}}, \quad \hat{\xi}_y\cdot\hat{z} = \mathbf{M}^{2,3}_{\mathbf{image}}, \quad \hat{\xi}_z\cdot\hat{z} = \mathbf{M}^{3,3}_{\mathbf{image}}, \end{aligned} \quad (10)$$

and the origin of the (ξ_x, ξ_y, ξ_z) coordinate system is located at $(\mathbf{M}^{1,4}_{\mathbf{image}}, \mathbf{M}^{2,4}_{\mathbf{image}}, \mathbf{M}^{3,4}_{\mathbf{image}})$. Therefore, (ξ_x, ξ_y, ξ_z) in (x, y, z) coordinates are given by

$$\begin{aligned} \xi_x &= x\mathbf{M}^{1,1}_{\mathbf{image}} + y\mathbf{M}^{2,1}_{\mathbf{image}} + z\mathbf{M}^{3,1}_{\mathbf{image}} + \mathbf{M}^{1,4}_{\mathbf{image}}, \\ \xi_y &= x\mathbf{M}^{1,2}_{\mathbf{image}} + y\mathbf{M}^{2,2}_{\mathbf{image}} + z\mathbf{M}^{3,2}_{\mathbf{image}} + \mathbf{M}^{2,4}_{\mathbf{image}}, \\ \xi_z &= x\mathbf{M}^{1,3}_{\mathbf{image}} + y\mathbf{M}^{2,3}_{\mathbf{image}} + z\mathbf{M}^{3,3}_{\mathbf{image}} + \mathbf{M}^{3,4}_{\mathbf{image}}, \end{aligned} \quad (11)$$

thus providing an expression for $p_{plane}(\vec{r}_f, \omega)$ in closed form.

Now $\vec{n}_f\cdot\nabla_f p_{plane}(\vec{r}_f,\omega)$ can be calculated and substituted into Eq. (4). After simplify the results by assuming that θ_f is small, Eq. (4) becomes

$$p_{plane}(\vec{r}_d,\omega) = \frac{-\tilde{k}\omega\rho\Gamma_{plane}K_{uV}(\omega)V_{inc}(\omega)H(\omega)G_o\cos(\theta_f)}{(2\pi)^2} \cdot \iint_{S_f} d\vec{r}_f \frac{e^{i\tilde{k}|\vec{r}_d-\vec{r}_f|}}{|\vec{r}_d-\vec{r}_f|} \times e^{-((\xi_x(\vec{r}_f)/w_x)^2+(\xi_y(\vec{r}_f)/w_y)^2+(\xi_z(\vec{r}_f/w_z))^2}e^{i\tilde{k}(z_T-\xi_z)}, \quad (12)$$

where ξ_x, ξ_y, and ξ_z are given by Eq. (11).

After determining the reflected fields in the region containing the original ultrasound source, these fields need to be translated to a voltage output by the transducer. The voltage output by the source can be found from (Bigelow and O'Brien, 2004a)

$$V_{plane}(\omega) = \frac{H(\omega)}{i\omega\rho S_T K_{uV}(\omega)}\int\int_{S_T} d\vec{r}_d G_T(\vec{r}_d,\omega)\frac{\partial p_{plane}}{\partial z_d}, \quad (13)$$

where

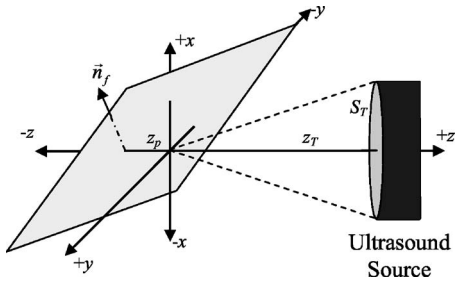


FIG. 1. Coordinate system for reflection from plane placed in the focal region.

$$\frac{\partial}{\partial z_d}p_{plane}(\vec{r}_d,\omega) \cong \frac{-i\tilde{k}^2\omega\rho\Gamma_{plane}K_{uV}(\omega)V_{inc}(\omega)H(\omega)G_o\cos(\theta_f)}{(2\pi)^2}\cdot\iint_{S_f} d\vec{r}_f\left(e^{-((\xi_x(\vec{r}_f)/w_x)^2+(\xi_y(\vec{r}_f)/w_y)^2+(\xi_z(\vec{r}_f)/w_z)^2)}e^{i\tilde{k}(z_T-\xi_z)}\right)$$
$$\times\frac{e^{i\tilde{k}r_d-i\tilde{k}\vec{r}_f\cdot(\vec{r}_f/r_d)}}{r_d}. \tag{14}$$

Hence, the reflected voltage from the inclined plane is given by

$$V_{plane}(\omega) \cong \frac{-\tilde{k}^2\Gamma_{plane}V_{inc}(\omega)H^2(\omega)G_o\cos(\theta_f)}{S_T(2\pi)^2}\cdot\iint_{S_T} d\vec{r}_d G_T(\vec{r}_d,\omega)\iint_{S_f} d\vec{r}_f$$
$$\times\left(e^{-((\xi_x(\vec{r}_f)/w_x)^2+(\xi_y(\vec{r}_f)/w_y)^2+(\xi_z(\vec{r}_f)/w_z)^2)}e^{i\tilde{k}(z_T-\xi_z)}\right)\frac{e^{i\tilde{k}r_d-i\tilde{k}\vec{r}_f\cdot(\vec{r}_d/r_d)}}{r_d}$$
$$=\frac{-\tilde{k}^2\Gamma_{plane}V_{inc}(\omega)H^2(\omega)G_o^2\cos(\theta_f)}{S_T(2\pi)^2}e^{i2\tilde{k}z_T}\iint_{S_f} d\vec{r}_f\left(e^{-\left(\frac{\xi_x^2+x_f^2}{w_x^2}+\frac{\xi_y^2+y_f^2}{w_y^2}+\frac{\xi_z^2+z_f^2}{w_z^2}+i\tilde{k}(\xi_x+z_f)\right)}\right), \tag{15}$$

where the velocity potential fields have been assumed to have a three-dimensional Gaussian distribution once again. The integral in Eq. (15) can be integrated by expressing all of the spatial coordinates in terms of x_f and y_f. Performing the integration while assuming that θ_f is small yields

$$V_{plane}(\omega) \cong \frac{-\pi w_x w_y\tilde{k}^2\Gamma_{plane}V_{inc}(\omega)H^2(\omega)G_o^2\cos(\theta_f)}{2S_T(2\pi)^2}\exp\left(-\frac{2z_p^2}{w_z^2}-2i\tilde{k}z_p+2i\tilde{k}z_T\right)$$
$$\times\left(1-\frac{w_x^2\tan^2(\theta_f)\cos^2(\phi_f)}{2w_z^2}-\frac{w_y^2\tan^2(\theta_f)\sin^2(\phi_f)}{2w_z^2}\right)\exp\left(\frac{w_y^2\tan^2(\theta_f)\sin^2(\phi_f)}{2}\left(\frac{4z_p^2}{w_z^4}+i\tilde{k}\frac{4z_p}{w_z^2}-\tilde{k}^2\right)\right)$$
$$\times\exp\left(\frac{w_x^2\tan^2(\theta_f)\cos^2(\phi_f)}{2}\left(\frac{4z_p^2}{w_z^4}+i\tilde{k}\frac{4z_p}{w_z^2}-\tilde{k}^2\right)\right), \tag{16}$$

where z_p is the distance along the beam axis from the focus to the intersection of the plane with the beam axis. Equation (16) can be further simplified by noting that typically $4z_p^2/w_z^4 \ll \mathrm{Re}\{\tilde{k}^2\}$ and $w_x, w_y \ll w_z$, thus yielding

$$V_{plane}(\omega) \cong \frac{-w_x w_y\tilde{k}^2\Gamma_{plane}V_{inc}(\omega)H^2(\omega)G_o^2\cos(\theta_f)}{8\pi S_T}$$
$$\times e^{i2\tilde{k}(z_T-z_p)}e^{-2z_p^2/w_z^2}\exp\left(-\tilde{k}^2\frac{\tan^2(\theta_f)}{2}\right.$$
$$\left.\times(w_x^2\cos^2(\phi_f)+w_y^2\sin^2(\phi_f))\right). \tag{17}$$

Now that an expression for the voltage returned from the inclined plane has been derived, the equation can be checked by considering a special case. First, if $w_x = w_y$, as would be the case for circularly symmetric source, then Eq. (17) becomes

$$V_{plane}(\omega) \cong \frac{-w_x^2\tilde{k}^2\Gamma_{plane}V_{inc}(\omega)H^2(\omega)G_o^2\cos(\theta_f)}{8\pi S_T}$$
$$\times e^{i2\tilde{k}(z_T-z_p)}\exp\left(-\frac{2z_p^2}{w_z^2}-w_x^2\tilde{k}^2\frac{\tan^2(\theta_f)}{2}\right), \tag{18}$$

which is no longer dependent on ϕ_f as would be expected. Second, if θ_f goes to zero, then Eq. (18) becomes

$$V_{plane}(\omega) \cong \frac{-w_x w_y\tilde{k}^2\Gamma_{plane}V_{inc}(\omega)H^2(\omega)G_o^2}{8\pi S_T}$$
$$\times e^{i2\tilde{k}(z_T-z_p)}e^{-2z_p^2/w_z^2}, \tag{19}$$

which is the same as was derived previously for a plane placed parallel to the focal plane (Bigelow and O'Brien, 2004a).

Equation (17) is a complete description of the backscattered voltage spectrum, but the equation contains more infor-

mation than is required to determine the total attenuation along the propagation path. Hence, $|V_{plane}(\omega)|$ in Eq. (17) is normalized by its peak value at the specific angle for the plane yielding

$$\frac{|V_{plane}(\omega)|}{\max_{\forall\omega}|V_{plane}(\omega)|} \cong \frac{V_{inc}(\omega)H^2(\omega)e^{-2\alpha_o f(z_T-z_p)}}{V_{inc}(\omega_{\max})H^2(\omega_{\max})e^{-2\alpha_o f_{\max}(z_T-z_p)}}, \tag{20}$$

where it is assumed that w_x and w_y are strictly proportional to wavelength (i.e., kw_x and kw_y are independent of frequency) and that the attenuation, assumed to be much smaller than the wave number, is strictly proportional to frequency (i.e., $\alpha=\alpha_o f$). Equation (20) is independent of the inclination angle of the plane because all of the angle dependences in Eq. (17) are removed by the normalization. Although the inclination angle of the plane impacts the absolute magnitude of the reflected spectrum, the shape of the reflected spectrum is not affected by θ_f. Hence, if we compare the normalized spectrum from a plane positioned at an unknown orientation in an unknown lossy medium to the normalized reference spectrum from a plane placed at the focal plane in a water bath, then the only difference between the two spectra should be the attenuation along the propagation path, $\alpha_o(z_T-z_p)$, in the unknown lossy medium.

There are many ways to compare the reference and measured spectrum. The most robust method for the work reported herein was to fit the normalized voltage spectrum from Eq. (20) with a Gaussian function of the form $|V_{plane}(\omega)|/\max_{\forall\omega}|V_{plane}(\omega)| \cong \exp((f-f_{peak})^2/2\sigma_{\omega p}^2)$. Although this is a crude assumption, it is reasonably accurate over a limited range of frequencies near the spectral peak for backscattered waveforms. The attenuation along the propagation path leading to the planar boundary in the unknown medium would then be given by the downshift in the center frequency expressed as (Narayana and Ophir, 1983)

$$\alpha_o(z_T-z_p) = \frac{f_{peak}|_{reference} - f_{peak}|_{unknown}}{2\sigma_{\omega p}^2}. \tag{21}$$

III. EXPERIMENTAL VERIFICATION

Before Eq. (21) can be used to estimate *in vivo* signal intensities, we need to verify that the normalized backscattered voltage spectrum is independent of the orientation of the plane [i.e., demonstrate the validity of Eq. (20)]. In this section, the backscattered voltage spectrum from a Plexiglas plate, a rat's skull, and a tissue mimicking phantom (Model 539; ATS Laboratories, Inc., Bridgport, CT) are compared for various orientation angels in a water bath (i.e., α_o assumed lossless). Water was selected due to its ease of use, but the conclusions would be valid with any attenuating fluid because when estimating the attenuation from backscattered waveforms the accuracy and precision of the estimate are not dependent on the attenuation along the propagation path (Huisman and Thijssen, 1996; Bigelow and O'Brien, 2005). The experiments were performed using a spherically focused $f/2$ transducer (Valpey Fisher Instruments, Inc., Hopkinton, MA). The transducer had a diameter of 2.1 cm, a center frequency of 8.7 MHz, and a −3-dB bandwidth of 1.6 MHz as measured from a wire reflection (Raum and O'Brien, 1997). Also, the equivalent Gaussian dimensions for the transducer were measured as is discussed by Bigelow and O'Brien (2004b) and were found to be $w_z=17.1\lambda+924\times10^{-6}$ m and $w_x=1.57\lambda+27.0\times10^{-6}$ m. The transducer was placed in a water bath and shock excited using a Panametrics 5900 pulser/receiver (Waltham, MA) operating in pulse-echo mode, and the returned waveforms were recorded using a digital oscilloscope at a sampling frequency of 100 MHz (Lecroy 9354 TM; Chestnut Ridge, NY) and averaged at least 100 times to remove electronic noise. The impact of noise on the estimation technique was beyond the scope of the current investigation.

A. Plexiglas experiment

The initial experiments were performed using a piece of smooth Plexiglas to act as the plane placed near the focal plane of the transducer. The Plexiglas plate was placed in the focal plane by positioning the plate so that the pulse-echo peak-to-peak voltage was a maximum. The Plexiglas was then rotated so that θ_f swept out angles from −10° to 10°, while maintaining a z_p of approximately zero; the pulse-echo waveform was recorded in steps of 1° change of θ_f. The scan was repeated four times with the focus moved along the axis of rotation for the plate, perpendicular to the beam axis, in steps of 0.1 mm (i.e., 0, 0.1, 0.2, and 0.3 mm). The results at each angle for all four scans were then averaged together. Also, the value of z_p was then estimated from the location of the time-domain waveforms and the corresponding spectra were multiplied by $\exp(2(z_p/w_z)^2)$. However, the correction term for z_p did not significantly affect the final results (max value for $|z_p|$ of 395 μm).

Recall that the purpose of the derivation was to use the inclined plane to approximate the skull interface of the fetus. The backscattered signals from the skull could then be analyzed to assess properties such as attenuation of the intervening tissue layers. Hence, the experimental results were evaluated in terms of the errors of the attenuation estimate (i.e., deviations from known α_o of 0 dB/cm-MHz) that would have resulted from the backscattered signals at each inclination angle of the plane. Because the total attenuation was to be measured from the downshift in the location of the main spectral peak, errors in the attenuation estimate would result from the voltage spectrum not peaking at the same frequency for all incident angles. Hence, the error in the attenuation estimate in Np/MHz for different inclination angles would be given by

$$\alpha_{error} = \frac{f_{peak}(\theta_f=0) - f_{peak}(\theta_f)}{2\sigma_{\omega p}^2}, \tag{22}$$

where f_{peak} is the frequency corresponding to the spectral peak at each inclination angle.

Figure 2(a) shows the error in the attenuation calculated using Eq. (22) for θ_f swept out angles from −10° to 10°. The frequencies selected for the Gaussian fit given by $|V_{plane}(\omega)|/\max_{\forall\omega}|V_{plane}(\omega)| \cong \exp(-(f-f_{peak})^2/2\sigma_{\omega p}^2)$ were all the frequencies for which the backscattered voltage spec-

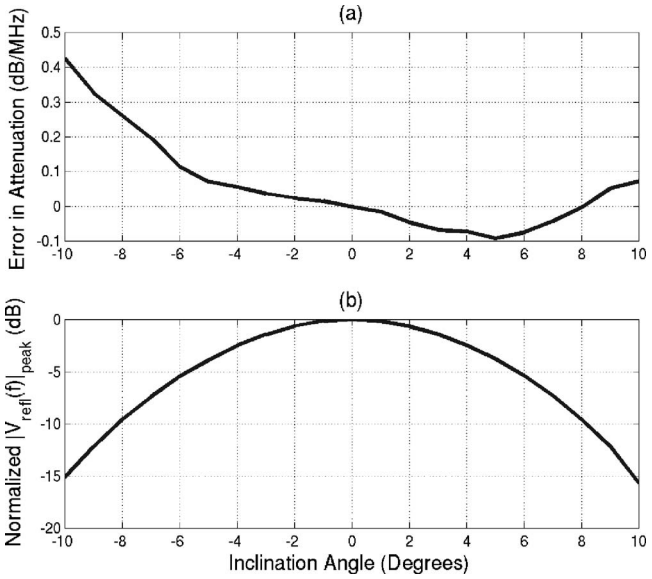


FIG. 2. (a) Error in attenuation estimates resulting from inclination angle and (b) normalized magnitude of spectrum peak of backscattered signal for various inclination angles of the Plexiglas plate.

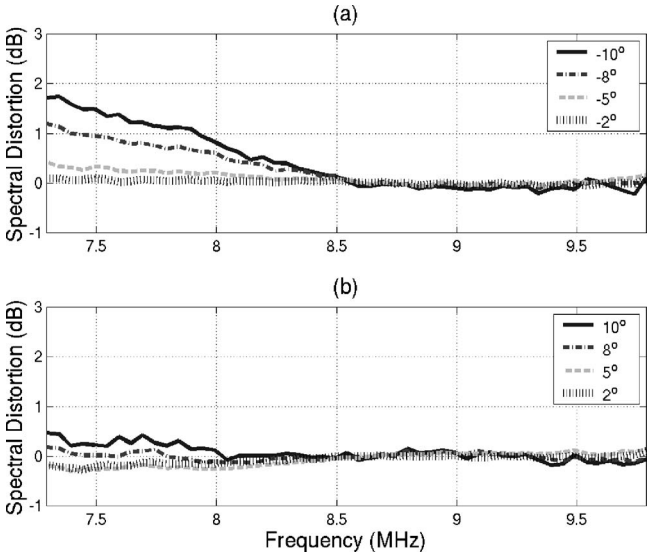


FIG. 3. Amount of spectral distortion introduced for inclination angles of (a) −10°, −8°, −5°, −2° and (b) 10°, 8°, 5°, and 2° for frequencies in the −6 -dB bandwidth of the transducer.

trum was greater than $\frac{1}{2}$ its maximum value for a θ_f of 0° (i.e., 7.29 to 9.79 MHz). The error introduced by θ_f is small (less than 0.1 dB/MHz) for angles from −6° to at least 10°. The asymmetry in the attenuation errors indicates that the Plexiglas plate was not perfectly aligned at the beginning of the scan (i.e., an inclination angle of 0° does not correspond to a θ_f of 0°). The magnitude of the voltage at the spectral peak (i.e., f_{peak}) for each value of θ_f relative to the value for a θ_f of 0° is also provided [Figure 2(b)]. The magnitude of the backscatter decreases quickly with increasing inclination angle of the Plexiglas plate. This decrease results from the tangent squared of θ_f in the exponent of Eq. (18).

In order to further validate Eq. (20) using the Plexiglas plate, the amount of spectral distortion introduced at each inclination angle was also assessed for frequencies in the −6 dB bandwidth of the transducer (i.e., 7.29 to 9.79 MHz). The assessment was done by comparing the normalized spectrum given by Eq. (20) at each inclination angle to the normalized spectrum at normal incidence as described by Eq. (23).

$$\text{Distortion}(\omega) = 20\log_{10}\left[\left(\frac{|V_{plane}(\omega)|}{\max_{\forall\omega}|V_{plane}(\omega)|}\right)\Bigg|_{\theta_f}\right] - 20\log_{10}\left[\left(\frac{|V_{plane}(\omega)|}{\max_{\forall\omega}|V_{plane}(\omega)|}\right)\Bigg|_{\theta_f=0^\circ}\right]. \quad (23)$$

The results for θ_f of −10°, −8°, −5°, and −2° are shown in Fig. 3(a) while the results for θ_f of 10°, 8°, 5°, and 2° are shown in Fig. 3(b). The distortion is always less than 2 dB, indicating that the spectral shape is not affected by the inclination angle of the plane as was predicted by Eq. (20). Also, the distortion is slightly larger for the negative values of θ_f [Fig. 3(a)] as compared to the positive values of θ_f [Figure 3(b)], indicating once again that the Plexiglas plate was not perfectly aligned at the beginning of the scan.

B. Bone experiment

After performing the Plexiglas experiments, the procedure was repeated using a rat's skull from an adult Sprague-Dawley rat (Harlan Sprague Dawley Laboratories, Indianapolis, IN). The experimental protocol was approved by the Laboratory Animal Care Advisory Committee at the University of Illinois at Urbana-Champaign and satisfied all campus and National Institutes of Health (NIH) rules for the human use of laboratory animals. The rat was housed in an Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC)-approved animal facility. The rat was placed in a polycarbonate cage with β-chip bedding and a wire bar lid and provided with food and water *ad lib*. The rat was euthanized by cervical dislocation while under anesthesia in a humane fashion prior to extracting the skull. AAALAC (Rockville, MD) is a private nonprofit organization that promotes the humane treatment of animals in science through a voluntary accreditation program.

The skull was obtained by first trimming away the skin on the head. Then, the top portion of the skull was cut using scissors from the eyes to the ears. The cut skull was then pulled gently from the brain and the connective tissue was trimmed away from both sides of the removed skull plate. Finally, the skull was rinsed and placed in a sterile saline solution. The skull/saline solution was refrigerated for several days before the ultrasound experiment was performed.

The skull was positioned so that the focus would be on the outer surface of the skull with the beam at approximately normal incidence by adjusting the position of the skull and transducer to achieve the maximum peak-to-peak voltage of the backscattered time-domain waveform. The skull was rotated so that the inclination angle of the beam would change from −10° to +10° and the backscattered waveform from the

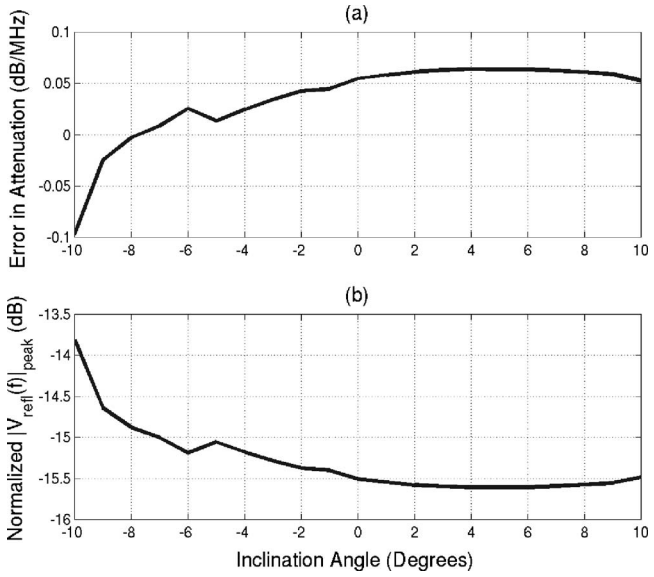


FIG. 4. (a) Error in attenuation estimates resulting from inclination angle and (b) normalized magnitude of spectrum peak of backscattered signal for various inclination angles of the rat skull.

skull was recorded in steps of 1°. However, the axis of rotation did not coincide with the focus so the value of z_p changed as the skull was rotated (max value for $|z_p|$ of 1.15 mm). After acquiring the waveforms, the impact of z_p was removed by analyzing the time-domain signals. The error in the attenuation for each value of inclination angle was then found from

$$\alpha_{error} = \frac{f_{peak}|_{reference} - f_{peak}(\theta_f)}{2\sigma^2_{\omega p}|_{reference}}, \tag{24}$$

where the reference spectrum was the reflected spectrum from the Plexiglas at normal incidence.

The error in the attenuation is shown in Fig. 4(a). In addition, the magnitude of the voltage at the spectral peak (i.e., f_{peak}) for the signal returned from the skull for each inclination angle relative to the magnitude of the voltage at the spectral peak for the signal returned from the Plexiglas for a θ_f of 0° is shown in Fig. 4(b). The magnitude signal from the skull [Fig. 4(b)] is much less than the magnitude of the signal from the Plexiglas plate. The voltage magnitude does not decay with increasing incident angle like the Plexiglas plate that dropped by 15 dB when θ_f was changed by 10°. The error in attenuation due to inclination angle [Fig. 4(a)] for the skull was always smaller than 0.1 dB/MHz. The constant magnitude of the voltage returned from the skull and the low error values in the attenuation are probably due to the surface roughness reducing the impact of the incident angle. In Fig. 2(b), the received voltage decayed quickly with increasing incident angle. Hence, only portions of the skull oriented close to normal incidence are likely to influence the backscattered voltage spectrum. Therefore, the surface roughness of the skull reduces the importance of the inclination angle on the measurement of attenuation from the backscattered waveforms.

C. Tissue-mimicking phantom experiment

Although the rat's skull and associated planar model may be similar in structure to the fetal skull near the end of the gestational period, the planar model would not be appropriate at the beginning of the ossification process. In the developing skull of humans, as well as many other animals, the parietal bones, frontal bone, a small part of the occipital bones, and much of the temporal bones are ossified by the direct formation of bone without a cartilage intermediate step (Carlson, 1994). Rather, the ossification begins near the center of the bone and proceeds along branching tubules until the bone plate has filled in (Nishimura *et al.*, 1977). It is doubtful that the branching tubules could be adequately modeled by the simple planar boundary model developed in this paper. However, a clear impedance boundary exists between the amniotic fluid and soft tissue of the developing fetus. Hence, it may be possible to apply the planar model to the amniotic fluid/soft-tissue boundary.

In order to determine if the reflection of the ultrasound off soft tissue could also be used to estimate the attenuation along the propagation path, the experiment was repeated using a soft-tissue mimicking phantom (Model 539; ATS Laboratories, Inc., Bridgport, CT). The front surface of the phantom was placed in the focal plane of the ultrasound transducer by maximizing the pulse-echo peak-to-peak voltage. Then, the ultrasound transducer was rotated from −10° to +10° in steps of 1°. The transducer was rotated instead of the phantom due to the large size of the phantom. At each angle, the distance between the transducer and the phantom was adjusted until the location of the first pressure peak corresponded in time to the location of the first pressure peak at normal incidence (i.e., z_p approximately zero for all inclination angles) before the backscattered waveform was stored by the oscilloscope.

Before finding the spectrum for each backscattered waveform at each inclination angle, care was taken so that the speckle from within the phantom/soft-tissue did not affect the location of the spectral peak. The problem is avoided when reflecting the ultrasound signal from the skull because the magnitude of the speckle is much smaller than the magnitude of the reflection from the bone due to the larger impedance difference. However, when dealing with the soft tissue, the speckle is on the same order as the reflection from the water/phantom boundary, especially at the larger inclination angles. Therefore, in order to reduce the impact of the speckle, the time-domain waveform for each inclination angle was windowed using a Hamming window that spanned the reflections from the water/phantom boundary while minimizing the distortion introduced by the window. The starting location for the Hamming window was given by the first location in time for which absolute value of the backscattered voltage waveform was greater than 10% its maximum value. In order to find the optimal Hamming window, the length of the window was varied from 1 to 6 periods in steps of 0.05 periods corresponding to the period of the center frequency of the transducer (i.e., 0.1149 to 0.6897 μs in steps of 5.747 ns corresponding to a center frequency of 8.7 MHz). The reference signal (i.e., the Plexiglas at normal

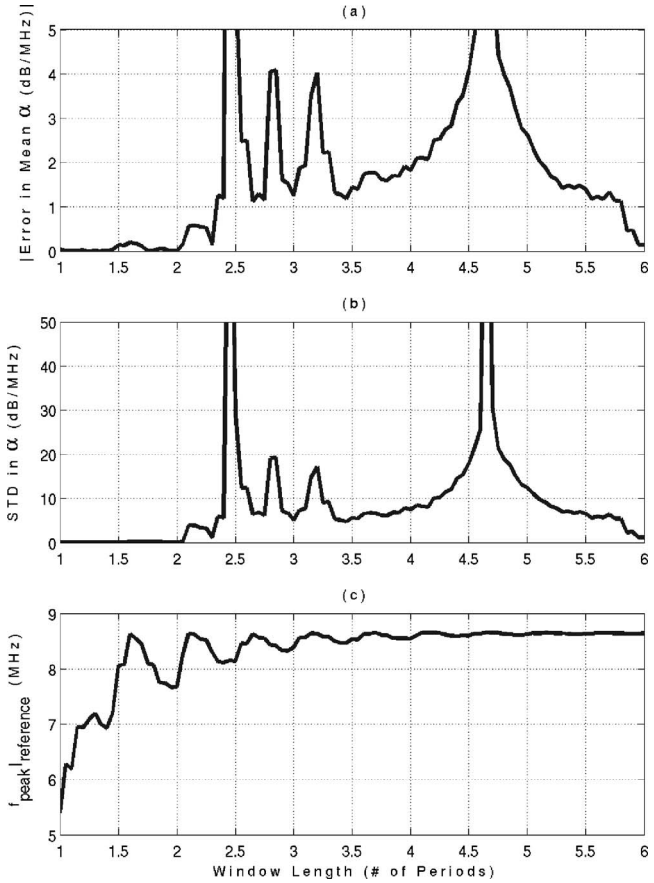


FIG. 5. Mean value of attenuation error (a), standard deviation of attenuation error (b), and frequency location of spectral peak from the reference waveform (c) as a function of Hamming window length for the phantom experiment.

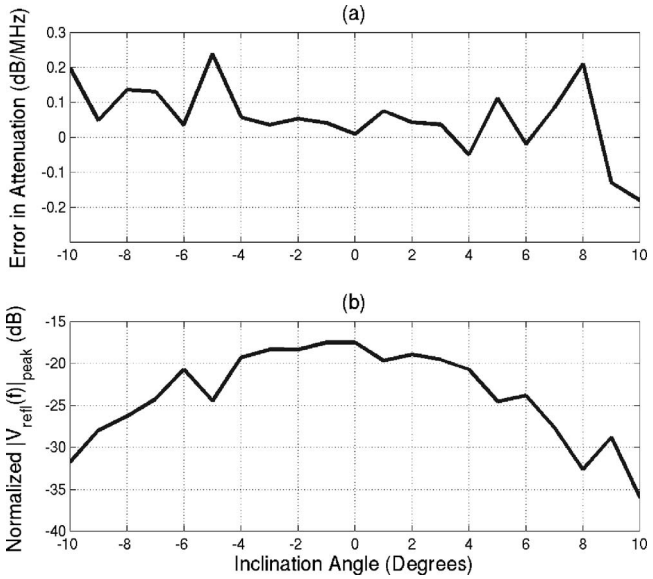


FIG. 6. (a) Error in attenuation estimates resulting from inclination angle and (b) normalized magnitude of spectrum peak of backscattered signal for various inclination angles of the tissue mimicking phantom.

incidence) was also windowed with the same Hamming window. The reference pulse amplitude exceeded 10% of its maximum value for a duration of 0.66 μs. The reference spectrum and phantom spectra were then found by taking the Fourier transform of the windowed time domain signals. Then, the error in attenuation for each of the phantom spectra for each window was calculated using Eq. (24).

The mean value and standard deviation for the attenuation error over all of the inclination angles for each of the window lengths are shown in Figs. 5(a) and 5(b), respectively. In addition, the location of the spectral peak from the reference spectrum for each window length is shown in Fig. 5(c). The accuracy [Fig. 5(a)] and the precision [Fig. 5(b)] of the attenuation estimate degrade quickly when the Hamming window length exceeds 2 periods (window length of 0.23 μs). The frequency location of the reference spectral peak [Fig. 5(c)] asymptotically approaches ~8.7 MHz as the window length is increased and is within 10% of its final value for a window length of 2 periods. The downshift of the reference spectral peak for smaller Hamming windows is a measure of the amount of distortion introduced by the windowing. Hence, a Hamming window length of 2 periods appears to be the best choice for reducing the effect of the speckle while still minimizing the distortion introduced by the window.

The error in the attenuation estimate for each inclination angle using a Hamming window with a length of 2 periods is shown in Fig. 6(a). The error in the attenuation estimate is always less than 0.25 dB/MHz. Therefore, it may be possible to estimate the attenuation along the propagation path from the backscatter returned by the soft fetal tissue. The error in the attenuation estimates shown in Fig. 6(a) was obtained for a single backscattered waveform. The accuracy of the estimate might be improved if the spectra or waveforms from several locations were averaged together. Combining the information from several waveforms is beyond the scope of the current investigation.

In addition to the attenuation estimate, the magnitude of the voltage at the spectral peak (i.e., f_{peak}) for the signal returned from phantom for each inclination angle relative to the magnitude of the voltage at the spectral peak for the signal returned from the Plexiglas for a θ_f of 0° is also shown in Fig. 6(b). The magnitude of the voltage peak returned from the phantom is always less than the magnitude of the voltage peak returned from the skull. Also, the magnitude decays quickly with increasing inclination angle just as was observed with the experiment with the Plexiglas plate. Hence, surface roughness does not play a dominate role in the reflection from the tissue-mimicking phantom. However, surface roughness may have a larger impact on the reflection from the fetus *in utero*.

IV. DISCUSSION AND CONCLUSIONS

A method for determining the attenuation along the propagation path leading to a developing fetus has been proposed. Initially, the fetal skull was modeled as a planar boundary at some unknown orientation with respect to the incident ultrasound beam. Theoretical calculations demonstrated that the shape of the backscattered spectrum would be independent of the inclination angle of the impinging sound beam. Hence, the spectrum from the skull could be compared to a reference spectrum obtained in a water bath to yield an

estimate for the total attenuation along the propagation path. The theoretical calculations were validated using a Plexiglas plate and a rat's skull positioned at various angles with respect to the sound beam. The error in the attenuation estimate due to the inclination angle of the rat's skull was always less than 0.1 dB/MHz for all of the tested inclination angles. The surface roughness of the skull when compared to the smooth Plexiglas appeared to improve the attenuation estimate at the larger incidence angles. Based on the skull results, estimating the attenuation using reflections from the fetal skull appears to be very promising.

After performing the skull experiments, it was noted that the simple planar boundary model would probably not be valid for the skull in the early stages of fetal development. Ossification of most of the skull bones begins with branching tubules that cannot be approximated as a simple planar boundary. Hence, the estimation technique was extended to the amniotic fluid/soft fetal tissue boundary and tested using a water/tissue mimicking phantom boundary. The phantom waveforms were windowed with Hamming windows in order to limit the impact of the speckle from within the phantom on the attenuation estimate. The best window length was found to be 2 periods of the fundamental frequency of the ultrasound transducer. Using this window length, the error in the attenuation estimate was always less than 0.25 dB/MHz. This error might be reduced further if the information from more than one waveform were averaged together.

The phantom results show great promise for extending the estimation technique to any gestational age due to the prevalence of soft fetal tissue very early in development. However, the results should still be validated using fetal tissue from an appropriate animal model. In addition to working with an animal model, the attenuation estimate could probably be improved by using an autoregressive (AR) estimation scheme rather than windowing to reduce the impact of the speckle (Girault *et al.*, 1998). AR estimation may also improve the computation time of the attenuation estimate.

Although the results are promising, estimating the exposure pressure and/or intensity using the attenuation along the propagation path using a simple derating procedure requires that nonlinear propagation be neglected. Neglecting nonlinear effects may not be reasonable especially when estimating the exposure intensity/pressure at higher transmit powers (Christopher and Carstensen, 1996). Hence, before being clinically implemented, the impact of nonlinear propagation needs to be addressed. It may still be possible to "quickly" estimate the *in vivo* exposure pressure and/or intensity to a reasonable level of accuracy on a patient specific basis from the attenuation along the propagation path by assuming that all of the attenuation occurs in an infinitesimally thin sheet at the surface of the transducer. Then, potentially, a look-up table could be designed to estimate the pressure/intensity at the focus.

ACKNOWLEDGMENTS

This work was supported by the University of Illinois Research Board and by a Beckman Institute Graduate Fellowship awarded to T. A. Bigelow. J. P. Blue, Jr. is also acknowledged for his assistance with dissecting the skull.

Abbott, J. G., (**1999**). "Rational and derivation of MI and TI—A review," Ultrasound Med. Biol. **25**, 431–441.

AIUM (**1988**). *Bioeffects and Safety of Diagnostic Ultrasound* (American Institute of Ultrasound in Medicine, Laurel, MD).

AIUM (**1993**). "Bioeffects Considerations for the Safety of Diagnostic Ultrasound (American Institute of Ultrasound in Medicine, Laurel, MD)," J. Ultrasound Med. **7**, S1–S38.

AIUM (**2000**). "Mechanical Bioeffects from Diagnostic Ultrasound: AIUM Consensus Statements (American Institute of Ultrasound in Medicine, Laurel, MD)," J. Ultrasound Med. **19**, 68–168.

AIUM/NEMA (**1998**). *Standard for real time display of thermal and mechanical acoustic output indices on diagnostic ultrasound equipment*, Rev 1 (American Institute of Ultrasound in Medicine, Laurel, MD; National Electrical Manufacturers Association, Rosslyn, VA).

Barnett, S. B. (**2000**). "Biophysical aspects of diagnostic ultrasound," Ultrasound Med. Biol. **26**(Suppl. 1), S68–S70.

Barnett, S. B. (**2001**). "Intracranial temperature elevation from diagnostic ultrasound," Ultrasound Med. Biol. **27**, 883–888.

Bigelow, T. A., and O'Brien, W. D., Jr. (**2004a**). "Scatterer size estimation in pulse-echo ultrasound using focused sources: Theoretical approximations and simulation analysis," J. Acoust. Soc. Am. **116**, 578–593.

Bigelow, T. A., and O'Brien, W. D., Jr., (**2004b**). "Scatterer size estimation in pulse-echo ultrasound using focused sources: Calibration measurements and phantom experiments," J. Acoust. Soc. Am. **116**, 594–602.

Bigelow, T. A., and O'Brien, W. D., Jr. (**2005**). "Evaluation of the spectral fit algorithm as functions of frequency range and Δka_{eff}," IEEE Trans. Ultrason. Ferroelectr. Freq. Control (in press).

Carlson, B. M. (**1994**). *Human Embryology and Developmental Biology* (Mosby, St. Louis, MO).

Carstensen, E. L., Child, S. Z., Norton, S., and Nyborg, W. (**1990**). "Ultrasonic heating of the skull," J. Acoust. Soc. Am. **87**, 1310–1317.

Christopher, T., and Carstensen, E. L. (**1996**). "Finite amplitude distortion and its relationship to linear derating formulae for diagnostic ultrasound systems," Ultrasound Med. Biol. **22**, 1103–1116.

Drewniak, J. L., and Dunn, F. (**1996**). "An experimentally obtainable heat source due to absorption of ultrasound in biological media," J. Acoust. Soc. Am. **100**, 1250–1253.

Drewniak, J. L., Carnes, K. I., and Dunn, F. (**1989**). "*In vitro* ultrasonic heating of fetal bone," J. Acoust. Soc. Am. **86**, 1254–1258.

Fuji, M., Sakamoto, K., Toda, Y., Negishi, A., and Kanai, H. (**1999**). "Study of the cause of the temperature rise at the muscle-bone interface during ultrasound hyperthermia," IEEE Trans. Biomed. Eng. **46**, 494–504.

Girault, J., Ossant, F., Ouahabi, A., Kouame, D., and Patat, F. (**1998**). "Time-varying autoregressive spectral estimation for ultrasound attenuation in tissue characterization," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 650–659.

Haken, B. A., Frizzell, L. A., and Carstensen, E. L. (**1992**). "Effect of mode conversion on ultrasonic heating at tissue interfaces," J. Ultrasound Med. **11**, 393–405.

Hearn, D., and Baker, M. P. (**1997**). *Computer Graphics: C Version*, 2nd ed. (Prentice Hall, Upper Saddle River, NJ), Chap. 11, pp. 408–419.

Huisman, H. J., and Thijssen, J. M. (**1996**). "Precision and accuracy of acoustospectrographic parameters," Ultrasound Med. Biol. **22**, 855–871.

Mayer, W. G. (**1964**). "Mode conversion of ultrasonic waves at flat boundaries," IEEE Trans. Sonics Ultrason. **11**, 1–3.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (McGraw-Hill, New York), Chap. 7, pp. 370–371.

Myers, M. R. (**2004**). "Transient temperature rise due to ultrasound absorption at a bone/soft-tissue interface," J. Acoust. Soc. Am. **115**, 2887–2891.

Narayana, P. A., and Ophir, J. (**1983**). "A closed form method for the measurement of attenuation in nonlinearly dispersive media," Ultrason. Imaging **5**, 17–21.

NCRP (**2002**). *Exposure Criteria for Medical Diagnostic Ultrasound: II. Criteria Based on all Known Mechanisms* (National Council on Radiation Protection and Measurements, Bethesda, MD), NCRP Report No. 140.

Nishimura, H., Nagai, T., Corp. Author(s): National Institute of Dental Research (U.S.; Jikken Dobutsu Chuo Kenkyujo (Kawasaki-shi, Japan) (**1977**). *Prenatal Development of the Human with Special Reference to Craniofacial Structure: An Atlas* (US Dept. of Health, Education, and Welfare, Bethesda, MD).

O'Neil, T. P., Winkler, A. J., and Wu, J. (**1994**). "Ultrasound heating in a tissue-bone phantom," Ultrasound Med. Biol. **20**, 579–588.

Pedersen, P. C., and Orofino, D. P., (**1996**). "Modeling of received ultrasound signals from finite planar targets," J. Acoust. Soc. Am. **43**, 303–311.

Pierce, A. D. (**1991**). *Acoustics: An Introduction to Its Physical Principles and Applications* (Acoustical Society of America, Woodbury, NY), Chap. 4, pp. 180–182.

Raum, K., and O'Brien, W. D., Jr. (**1997**). "Pulse-echo field distribution measurement technique for high-frequency ultrasound sources," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 810–815.

Siddiqi, T. A., Miodovnik, M., Meyer, R. A., and O'Brien, W. D., Jr. (**1999**). "*In vivo* ultrasonographic exposimetry: Human tissue-specific attenuation coefficients in the gynecologic examination," Am. J. Obstet. Gynecol. **180**, 866–874.

Siddiqi, T. A., O'Brien, W. D., Jr., Meyer, R. A., Sullivan, J. M., and Miodovnik, M. (**1992**). "Human *in situ* dosimetery: Differential insertion loss during passage through abdominal wall and myometrium," Ultrasound Med. Biol. **18**, 681–689.

Wu, J., and Du, G. (**1990**). "Temperature elevation generated by a focused Gaussian ultrasonic beam at a tissue-bone interface," J. Acoust. Soc. Am. **87**, 2748–2755.