CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 111, No. 3 March 2002

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 111, NO. 3, MARCH 2002

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 111, NO. 3, MARCH 2002

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.*

Report of the Auditor

Published herewith is a condensed version of our auditors' report for calendar year ended 31 December 2000.

Independent Auditors' Report

To the Executive Council
Acoustical Society of America

We have audited the accompanying statements of financial position of the Acoustical Society of America as of December 31, 2000, and the related statements of activity and cash flows for the year then ended. These financial statements are the responsibility of the Society's management. Our responsibility is to express an opinion on the financial statements based on our audit.

We conducted our audit in accordance with generally accepted auditing standards. Those standards require that we plan and perform the audit to obtain reasonable assurance about whether the financial statements are free of material misstatement. An audit includes examining, on a test basis, evidence supporting the amounts and disclosures in the financial statements. An audit also includes assessing the accounting principles used and significant estimates made by management, as well as evaluating the overall financial statement presentation. We believe that our audit provides a reasonable basis for our opinion.

In our opinion, the financial statements referred to above present fairly, in all material respects, the financial position of the Acoustical Society of America as of December 31, 2000, and the changes in its net assets and its cash flows for the year then ended in conformity with generally accepted accounting principles.

CONROY, SMITH & CO.
16 May 2001
New York, NY

ACOUSTICAL SOCIETY OF AMERICA
STATEMENT OF FINANCIAL POSITION
AS OF 31 DECEMBER 2000
(With Comparative Totals For 1999)

	2000	1999
Assets:		
Cash and cash equivalents	$ 942,470	$ 489,533
Accounts receivable	324,410	372,836
Marketable securities	6,881,069	7,398,853
Furniture, fixtures and equipment—net	107,191	118,078
Other assets	349,219	292,934
Total assets	$8,604,359	$8,672,234
Liabilities:		
Accounts payable and accrued expenses	$ 453,544	$ 274,942
Deferred revenue	1,449,583	1,524,755
Deferred rent liability	41,989	44,192
Total liabilities	$1,945,116	$1,843,889
Net assets:		
Unrestricted	$ 5,518,982	$5,601,878
Temporarily restricted	565,655	598,120
Permanently restricted	574,606	628,347
Total net assets	$6,659,243	$6,828,345
Total liabilities and net assets	$8,604,359	$8,672,234

 0001-4966/2002/111(3)/1131/7/$19.00 © 2002 Acoustical Society of America

ACOUSTICAL SOCIETY OF AMERICA
STATEMENT OF ACTIVITY
FOR THE YEAR ENDED 31 DECEMBER 2000
(With Comparative Totals For 1999)

	Unrestricted net assets	Temporarily restricted net assets	Permanently restricted net assets	Total	1999 Total
Support and revenue:					
Contributions	-	-	$ 623	$ 623	$ 1,310
Dues	$ 658,172	-	-	658,172	628,110
Publishing	2,172,282	-	-	2,172,282	2,088,076
Standards	418,779	-	-	418,779	366,878
Meetings	396,038	-	-	396,038	239,591
Interest and dividends	280,351	$ 29,330	30,874	340,555	327,615
Unrealized gain (loss)	(503,670)	(53,643)	(56,469)	(613,782)	401,546
Other	48,235	-	-	48,235	60,639
Realized gain (loss)	17,668	1,881	1,981	21,530	236,085
Net assets released from restrictions	10,033	(10,033)	-	-	-
Total support and revenue	$3,497,888	($ 32,465)	($ 22,991)	$ 3,442,432	$4,349,850
Expenses:					
Publishing	$1,630,098	-	-	$1,630,098	$1,783,477
Standards	536,893	-	-	536,893	461,999
Administrative and general	559,803	-	-	559,803	499,455
Meetings	449,117	-	-	449,117	250,613
Other	404,873	-	$ 30,750	435,623	409,938
Total expenses	$3,580,784	-	$ 30,750	$ 3,611,534	$3,405,482
(Decrease) in net assets	($ 82,896)	($ 32,465)	($ 53,741)	($ 169,102)	$ 944,368
Net assets, beginning of year	5,601,878	598,120	628,347	6,828,345	5,883,977
Net assets, end of year	$5,518,982	$565,655	$ 574,606	$ 6,659,243	$6,828,345

ACOUSTICAL SOCIETY OF AMERICA
STATEMENT OF CASH FLOWS
FOR THE YEAR ENDED 31 DECEMBER 2000
(With Comparative Totals For 1999)

	Total All Funds 2000	1999
Operating Activities		
(Decrease) in net assets	($ 169,102)	$ 944,368
Adjustments to reconcile net income to net cash provided by operating activities:		
Depreciation and amortization	32,218	31,355
Unrealized (gain) loss on marketable securities	613,782	(401,545)
Changes in operating assets and liabilities:		
(Increase) decrease in accounts receivable	48,426	5,531
Decrease (increase) in other assets	(56,285)	(46,388)
Increase (decrease) in accounts payable and accrued expenses	178,602	69,104
(Decrease) increase in deferred rent liability	(2,203)	(2,202)
Increase (decrease) in deferred revenue	(75,172)	378,031
Net cash flows provided by operating activities	$ 570,266	$ 978,254
Investing Activities		
Purchase of furniture, fixtures, equipment, and leasehold improvements	($ 21,331)	($ 47,336)
Proceeds from sale of securities	2,315,330	1,713,644
Purchase of securities	(2,411,328)	(4,796,676)
Net cash (used in) provided by investing activities	($ 117,329)	($3,130,368)
Increase (decrease) in cash and cash equivalents	$ 452,937	($2,152,114)
Cash and cash equivalents, beginning of year	489,533	2,641,647
Cash and cash equivalents, end of year	$ 942,470	$ 489,533

Preliminary Notice: 143rd Meeting of the Acoustical Society of America

The 143rd Meeting of the Acoustical Society of America will be held Monday through Friday, 3–7 June 2002, at the Pittsburgh Hilton and Towers Hotel in Pittsburgh, Pennsylvania, USA. A block of rooms has been reserved at the Hilton hotel.

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

Technical Program

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

List of Special Sessions

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

Acoustical Oceanography (AO)
Acoustical studies of bubbles in the ocean
(Joint with Underwater Acoustics)
Characterization of bubbly fluids, including measurement techniques and results
Ed Hamilton memorial session on sediment properties
(Joint with Underwater Acoustics)
Acoustic propagation through ocean sediments, dealing in particular with the issues of attenuation and dispersion, in honor of Ed Hamilton's contributions to our understanding of sediment properties.

Animal Bioacoustics (AB)
Acoustics of terrestrial mammals and birds
Sound production and reception, and effects of noise on terrestrial mammals and birds
Session on marine mammals in honor of Bill Evans
Acoustics of whales and pinnipeds

Architectural Acoustics (AA)
Acoustics of meeting facilities
Detailing the design, shortcomings, and successes in meeting facilities
Adaptive reuse of vaudeville and movie palaces for modern music and theater performance
Architectural and acoustic issues in leveraging the location and value in existing theater spaces
Computer modeling in acoustical consulting (Joint with Noise)
Overview of computer modeling techniques as applied to acoustical consulting
Halls for music performance: Another two decades of experience: 1982–2002
Continuation of poster session held at 141st meeting in Chicago in June 2001
Perceptual aspects of real and virtual rooms
(Joint with Psychological and Physiological Acoustics)
Session to initiate interdisciplinary discussion on perception and room acoustics

Biomedical Ultrasound/Bioresponse to Vibration (BB)
Characterization and monitoring of biological flows
Ultrasonic systems and techniques to measure or quantify flow parameters
Funding biomedical engineering research
Panel discussion of funding mechanisms in the area of biomedical ultrasound
Scattering theory applications for biological media
Applications of scattering theory in biomedicine, including multiphase flows
Ultrasonic field characterization (Joint with Signal Processing in Acoustics)
Techniques of ultrasonic field characterization with emphasis on effects of nonlinear propagation and high intensity applications
Ultrasound propagation in and through bone
Models and techniques for quantifying ultrasound propagation in bone or through bone to other tissues

Education in Acoustics (ED)
Hands-on demonstrations in acoustics for high school students
About 15 demos to introduce high school students to the excitement of acoustics

Engineering Acoustics (EA)
Electroacoustic systems for 3-D audio (Joint with Architectural Acoustics)
Electroacoustic system concepts and designs used to produce desired 3-D audio fields
Projectors for high frequency sonars (Joint with Underwater Acoustics)
Survey of underwater projectors in the 1 kHz to 20 kHz band

Musical Acoustics (MU)
Free reed instruments: Historical perspectives and recent research
Historical studies of the development of free reed instruments together with results of recent acoustical research
Interactive computer music systems
Virtual instruments, computer music performance systems, gestural control of sound synthesis, mapping strategies, and new techniques for real-time sound control
Music recognition systems
(Joint with Psychological and Physiological Acoustics)
Automatic methods for performing tasks related to information content of music signals such as pitch detection, tempo and rhythm detection, instrument detection and timbre categorization, voice separation, and music matching
Naturalness in synthesized speech and music
(Joint with Speech Communication)
Problems and techniques in achieving naturalness in synthesized sound, especially in speech and singing

Noise (NS)
Acoustics of glazing (Joint with Architectural Acoustics and Structural Acoustics and Vibration)
Applications of sound isolation concepts in building glazing
Manufacturing process noise control
Advances in noise control methods for manufacturing
Public policy for noise (Joint with Speech Communication and Animal Bioacoustics)
The objectives, issues, and approaches to creating effective means for noise control
Quiet roadways/quiet tires (Joint with Structural Acoustics and Vibration)
Mechanisms for reducing the noise generated by the tire/road interface

Physical Acoustics (PA)
Demonstration devices in physical acoustics (Joint with Education in Acoustics)
Recent developments in wave propagation in random and complex media (Joint with Underwater Acoustics)
A broad look at progress in theory and applications involving wave propagation through random and complex media such as turbulence, internal wavefields, and porous structures

Psychological and Physiological Acoustics (PP)
Hearing aid design: Psychophysics and signal processing
Overview of current status of hearing aid research and discussion of unresolved hearing aid design issues from a variety of perspectives
Perceptual effects in real and virtual rooms
(Joint with Architectural Acoustics)
Focus on how complex, realistic patterns of echoes and reverberation affect perception
Session honoring Ira Hirsh (Joint with Speech Communication)

Signal Processing in Acoustics (SP)
Advances in sonar imaging techniques including synthetic aperture sonar and computed tomography (Joint with Underwater Acoustics)
Advanced techniques in sonar imaging will be discussed evolving from new approaches in synthetic aperture sonar and computed tomography

Speech Communication (SC)
Future of infant speech perception research: Session in memory of Peter Jusczyk (Joint with Psychological and Physiological Acoustics)
Session to honor Peter Jusczyk's many contributions to science
Models of phonetic formation and structure
Behavioral and neurophysiological data related to the acquisition and function of phonetic categories

Structural Acoustics and Vibration (SA)
Active control of ducted fan noise (Joint with Noise)
Current progress and developments in the active control of ducted fan noise
Flow-induced vibration and noise (Joint with Noise)
Topics of interest include experimental techniques and results, theoretical and numerical models for aerospace, surface and submerged vehicles, and piping systems
Pyroshock
Measurement and analysis of pyroshock will be covered

Underwater Acoustics (UW)
Celebration of the work of Morris Schulkin (Joint with Acoustical Oceanography)

Other Technical Events

History Lectures and Hot Topics Session

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

Student Design Competition in Architectural Acoustics

The Technical Committee on Architectural Acoustics of the Acoustical Society of America and the National Council of Acoustical Consultants are sponsoring a student design competition which will be professionally judged at the meeting. The purpose of this design competition is to encourage students enrolled in Architecture, Architectural Engineering, and other University curriculums that involve building design and/or acoustics to express their knowledge of architectural acoustics and building noise control in the schematic design of portions of a building where acoustical considerations are of primary importance. The submitted designs, which will be displayed at the Pittsburgh ASA Meeting, will be judged by a panel of professional architects and acoustical consultants. An award of $1000 will be made to the submitter(s) of the design judged ''first honors.'' Four awards of $500 each will be made to the submitters of four entries judged ''commendation.'' Full details can be found on the ASA web site: asa.aip.org/design.html

Meeting Program

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

Paper Copying Service

Authors are requested to provide one paper copy of their projection material and/or paper(s) to the Paper Copies Desk upon arrival. The copy should contain material on one side only on $8\frac{1}{2}\times11$ inch or A4 paper suitable for photocopy reproduction. Copies of available papers will be made for a nominal charge.

Tutorial Lecture

A tutorial presentation Keep your Ear on the Ball: Display of Targets in the Bat's Sonar Receiver, will be given by James A. Simmons on Monday, 3 June at 7:00 p.m. Lecture notes will be available at the meeting in limited supply. Those who register by 6 May 2002 are guaranteed receipt of a set of notes.

To partially defray the cost of the lecture a registration fee is charged. The fee is $15 for registration received by 6 May and $25 thereafter including on-site registration at the meeting. The fee for students with current ID cards is $7.00 for registration received by 6 May and $12.00 thereafter, including on-site registration at the meeting.

Short Course on Conversational Systems

This short course will be held on Sunday afternoon, 2 June, and Monday morning, 3 June, at Carnegie Mellon University.

Conversational systems combine the technologies of speech recognition, speech synthesis, and language and domain modeling to develop systems in which users and machines interact using voice to access information or perform useful tasks. This short course will describe, discuss, and demonstrate the aspects of speech and language technology used to develop conversational systems. The course will be directed toward technically literate researchers and students who are not necessarily experienced in speech science and language technologies. The course will include overviews and technical commentary on both the fundamentals and contemporary issues underlying automatic speech recognition, speech synthesis, language modeling and semantic inference, and the design of spoken-language systems, including user interface issues. Technical discussions will be supplemented with demonstrations and discussions of the development of spoken language systems using tools and resources that are available to the public in open source form. The course will be conducted at Carnegie Mellon University in Pittsburgh.

The objective of the course is to provide the participant with a technical overview of a suite of techniques that are used for the development of systems that enable users to interact with computers using voice input and output, along with an introduction to publicly available tools that can be used to develop practical systems.

The course instructors are Alan W. Black, Ronald Rosenfeld, Alexander I. Rudnicky, and Richard M. Stern, who are all faculty members of the speech group at Carnegie Mellon University (CMU). Alan Black is well known for his work in speech synthesis, including the development of the Festival Speech Synthesis system while at the University of Edinburgh. Alexander Rudnicky's research has focused on speech recognition, interfaces, and language, and he has led the effort to develop the CMU Communicator system, a dialog management architecture for complex spoken language tasks. Roni Rosenfeld has extensive experience in language modeling for speech recognition and accessible and natural flexible interfaces for spoken language systems, and he is leading the development of the Universal Speech Interface. Richard Stern has extensive experience in signal processing, speech recognition, and auditory perception, and he directs the activities of the CMU Robust Speech Group.

The registration fee is $250.00 and covers attendance, instructional materials, coffee breaks, and transportation between the Hilton Hotel and Carnegie Mellon University. **The number of attendees will be limited so please register early to avoid disappointment**. Only those who have registered by 6 May will be guaranteed receipt of instructional materials. There will be a $50 discount for registration made prior to 6 May. Full refunds will be made for cancellations prior to 6 May. Any cancellation after 6 May will be charged a $50 processing fee. Use the form in the printed call for papers to register for the course or register online at ⟨http://asa.aip.org⟩.

Short Course on Signal Processing for Machinery and Structural Fatigue Prognostics

This short course will be held on Friday afternoon, 7 June, and Saturday morning, 8 June, at the Hilton Pittsburgh Hotel and Towers.

The course is designed for acousticians and scientists with a basic knowledge of digital signal processing and an interest in its application to machinery and structural health monitoring. Applications of digital data acquisition and calibration are stressed for the detection machinery component signatures and structural modal responses. The signature responses are broken down into the feature information using physics-based feature detection algorithms. These features are then tracked for the prediction of remaining useful life under varying load conditions. Signal models are presented for gears, bearings, shaft balance and alignment, and structural modal responses. These physical models are used to convert raw sensor signals to

information features. The features change over time with changing loads and/or growing damage in the machine or structure. Simple examples of features are material stiffness, rms vibration, vibration kurtosis, nonlinearity, modulation, and so on. A damage feature ''state model'' is composed of the ''position, velocity, acceleration, jerk, joust, etc.'' of the damage feature dynamics, thus allowing future prediction of the damage for a given load condition. Many simple examples using simulated and/or actual sensor data will be demonstrated using straightforward Matlab scripts.

The objective is to present a concise overview for physicists and engineers of the essential signal processing tools to detect and predict future fatigue in machinery elements and vibrating structures.

Professor David C. Swanson is the Division Chief Scientist of The Systems and Operations Automation Division of The Applied Research Laboratory and an Associate Professor of Acoustics at The Pennsylvania State University. He is a member of ASA, the IEEE, and the board certified member of Institute of Noise Control Engineers (INCE). His research interests cover a broad range of intelligent acoustic sensor networks including surveillance systems, predictive maintenance, passive sonar, wireless sensor networks, active noise control, and environmental sensing for outdoor sound propagation. Included as part of the course, his recent textbook *Signal Processing for Intelligent Sensors* (Marcel-Dekker, 2000) will be used along with numerous Matlab examples of the applied processing techniques.

The registration fee is $300.00 and covers attendance, instructional materials and coffee breaks. **The number of attendees will be limited so please register early to avoid disappointment.** Only those who have registered by 6 May will be guaranteed receipt of instructional materials. There will be a $50 discount for registration made prior to 6 May. Full refunds will be made for cancellations prior to 6 May. Any cancellation after 6 May will be charged a $50 processing fee. Use the form in the call for papers to register 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 15 April 2002 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 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 Pittsburgh 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 Pittsburgh meeting. Each award will be of the order of $200. It is anticipated that the Committee will grant a maximum of five awards. Applicants should submit a request for support, a copy of the abstract they have submitted for the meeting, and a current resume/vita to Peggy Nelson, Department of Communicative Disorders, University of Minnesota, 164 SE Pillsbury Drive, Minneapolis, MN 55455; Tel: 612-625-4569; nelso477@tc.umn.edu. Deadline for receipt of applications is 15 April 2002.

Students Meet Members for Lunch

The Education Committee has established a program for students to meet with members of the ASA over lunch. Students are strongly encouraged to contact David Blackstock, University of Texas at Austin, Mechanical Engineering Dept., Austin, TX 78712-1063; Tel.: (512) 471-3145, Fax: (512) 471-1045, dtb@mail.utexas.edu prior to the meeting. There will also be a sign-up sheet available at the registration desk for those students who have not responded prior to the meeting. Members who wish to participate are also encouraged to contact David Blackstock. Participants are responsible for the cost of their own meal.

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

Two socials will be held; one on Tuesday night (4 June) in the Hilton and one on a riverboat on Thursday night (6 June). Both socials will begin at 6:00 p.m. with complimentary buffets with cash bars. On Thursday, the riverboat will leave at 6:00 p.m. from the Point State Park located directly in front of the Hilton (there will be a short walk from the Hilton to the boat). The cruise of the three rivers (Allegheny, Monongahela, and Ohio) will provide an excellent view of Pittsburgh and surroundings, such as Mt. Washington. The boat will return to Point State Park at 7:30 p.m. in time for you to return to the Hilton to attend the Technical Committee meetings.

The Plenary session will be held on Wednesday afternoon (5 June) at the Hilton Hotel where Society awards will be presented and recognition of Fellows will be announced. A Fellows' Luncheon will be held on Thursday (6 June) at 12:00 noon. Cost is $25. Use the registration form in the call for papers to purchase tickets or use the online registration form. Please note that Fellows may bring one guest to the luncheon.

Women in Acoustics Luncheon

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

Transportation and Hotel Accommodations

Air Transportation

Pittsburgh is served by the Pittsburgh International Airport (PIT) which was recently rated by Conde Nast Traveler as the best airport in the country and the third best in the world. There are 23 airlines using PIT with 581 nonstop flights to 119 cities. PIT is a hub for US Airways.

Ground Transportation

The Pittsburgh International Airport (PIT) is about 30 minutes by auto from the Pittsburgh Hilton. A taxi will cost about $28 USD. Express Shuttle, which costs $14 USD, runs once an hour from PIT with the Hilton as one of its stops. There is also a Port Authority Transit (PAT) public bus which runs every 20 minutes from PIT with three stops before the Hilton. This bus is a real bargain, costing less than $2 USD. For those who wish to drive, Pittsburgh is six hours or less by auto from Baltimore, Buffalo, Cincinnati, Cleveland, Columbus, Detroit, Erie, Philadelphia, Washington, DC, and State College.

Hotel Accommodations and Reservations

The meeting and all functions, except the Thursday night social, will be held at the Hilton in downtown Pittsburgh, directly across from the Point State Park where the Monongahela and Allegheny Rivers join to form the Ohio River. Please make your reservations directly with the Hilton and ask for one of the rooms being held for the Acoustical Society of America. The reservation cutoff date for the special discounted ASA rates is 5 May 2002; 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 Pittsburgh Hilton and Towers. Early reservations are strongly recommended. Note that the special ASA meeting rates are not guaranteed after 5 May 2002. You must mention the Acoustical Society of America when making your reservations to obtain the special ASA meeting rates.

Pittsburgh Hilton and Towers
Gateway Center
600 Commonwealth Place
Pittsburgh PA 15222
Tel: (412) 391-4600
Toll-free reservations: 1-800-HILTONS
FAX: (412) 594-2277

Rates: Single:	$128	Quad:	$197
Double:	$151	Govt Rate S/D:	$99/$125 (limited quantity)
Triple:	$174	Additional person:	$25

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, by 15 April to the Acoustical Society of America, preferably by e-mail: asa@aip.org or by postal mail to Attn.: Room Sharing, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502. The responsibility for completing any arrangements for room sharing rests solely with the participating individuals.

Weather

Pittsburgh enjoys four distinct seasons, with early June falling between the spring and summer seasons. The expected temperatures should average around 68 °F with a high of 75 °F.

Accompanying Persons Program

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

In addition to being surrounded by the Three Rivers Art Festival which will be held immediately outside the Hilton during this meeting, you may walk, join a day bus tour, or drive on your own to many attractions in Pittsburgh and the surrounding area.

Some attractions which are within a short walking distance from the hotel include: Department stores, such as Kaufmann's, Lazarus, Lord & Taylor and Saks Fifth Avenue, along with many smaller specialty shops. PNC Park—home of the Pittsburgh Pirates. Station Square—home of 40 specialty shops, 18 restaurants, Gateway Clipper fleet of riverboats and two inclines that take you to the top of Mt. Washington. Strip District—a mile long ''strip'' of land along the Allegheny River of former warehouses that have been converted to shops, produce and flower markets, restaurants, and nightclubs. Heinz Hall—this elegantly restored concert hall is home of the Pittsburgh Symphony Orchestra. Tours of Heinz Hall are being planned. Carnegie Science Museum—this museum features 250 hands-on exhibits (including several on acoustics), a surround theater, interactive planetarium, large miniature railroad and village, and the *Requim*, a submarine from World War II which is open for tours. Sports Works—this interactive museum on the science of sports features over 40 exhibits. Andy Warhol Museum—this museum features a permanent collection of Andy Warhol art and archives, as well as temporary exhibitions from other artists. Mt. Washington—this ridge overlooks downtown Pittsburgh from the south across the Monongahela River. In addition to a spectacular view of downtown Pittsburgh, you can enjoy fine dining at one of the many restaurants located along Grandview Avenue.

Some attractions within a short drive of downtown Pittsburgh include the Pittsburgh Zoo, home of over 4000 animals and a large aquarium. Old Economy Village, which contains 17 of the original buildings from the Harmony Society. Johnstown Flood Museum about 70 miles from downtown Pittsburgh, which features permanent exhibits of one of America's most infamous natural disasters. Allegheny Portage Railroad National Park, about 80 miles from downtown Pittsburgh, shows traces of the first railroad crossing of the Allegheny mountains, including the incline plane railroad which provided a critical link, and the Horseshoe Curve.

Two day trip tours are planned, both of which will include bus transportation, lunch, and tours.

Oakland Area—Only a few miles east of downtown Pittsburgh, this area is home of Carnegie Mellon University, University of Pittsburgh (with the 42-story Cathedral of Learning), Carlow College, Chatham College, the Carnegie Museum of Art, the Carnegie Museum of Natural History, and Phipps Conservatory & Botanical Gardens. The Carnegie Art Museum contains one of the finest collections of impression, post-impression, and contemporary works. The Carnegie Museum of Nature History has one of the world's best dinosaur collections. Tropical plants, a butterfly garden, miniature orchids, and seasonal flower collections are just a few of the exhibits in the Phipps Gardens.

Frank Lloyd Wright architecture—A day tour is planned of Frank Lloyd Wright's masterpieces Fallingwater and Kentuck Knob, which are located about an hour and a half from Pittsburgh. Fallingwater is one of Frank Lloyd Wright's best known homes. Kentuck Knob includes contemporary sculptures, historical artifacts, and a garden shop that enhance its beautiful woodland setting.

Registration Information

The registration desk at the meeting will open on Monday, 3 June at the Hilton 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 17 May you must register on-site**.

Registration fees are as follows:

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

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 (2002) of membership. (Full price for dues: $100). Invited speakers who are members of the Acoustical Society of America are expected to pay the registration fee, but **nonmember invited speakers** who participate in the meeting for one day only may register without charge. Nonmember invited speakers who wish to participate in the meeting for more than one day will be charged the member registration fee, which will include a one-year membership in the ASA upon completion of an application form.

NOTE: A $25.00 PROCESSING FEE WILL BE CHARGED TO THOSE WHO WISH TO CANCEL THEIR REGISTRATION AFTER 6 MAY.

USA MEETINGS CALENDAR

Listed below is a summary of meetings related to acoustics to be held in the U.S. in the near future.

2002

10–13 March	Annual Meeting of American Institute for Ultrasound in Medicine, Nashville, TN [American Institute for Ultrasound in Medicine, 14750 Sweitzer Lane, Suite 100, Laurel, MD 20707-5906; Tel.: 301-498-4100 or 800-638-5352; Fax: 301-498-4450; E-mail: conv_edu@aium.org; WWW: www.aium.org].
3–7 June	143rd Meeting of the Acoustical Society of America, Pittsburgh, PA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].
19–21 Aug.	INTER-NOISE 2002, Dearborn, MI [INTER-NOISE 02 Secretariat, The Ohio State University, Department of Mechanical Engineering, 206 West 18th Ave., Columbus, OH 43210-1107 or E-mail: hp@internoise2002.org].
2–6 Dec.	Joint Meeting: 144th Meeting of the Acoustical Society of America, 3rd Iberoamerican Congress on Acoustics, and 9th Mexican Congress on Acoustics, Cancun, Mexico [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502;

Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org/cancun.html].

2003

28 April–2 May 145th Meeting of the Acoustical Society of America, Nashville, TN [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

10–14 Nov. 146th Meeting of the Acoustical Society of America, Austin, TX [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

2004

24–28 May 75th Anniversary Meeting (147th Meeting) of the Acoustical Society of America, New York, NY [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

29 Nov.–3 Dec. 148th Meeting of the Acoustical Society of America, San Diego, CA [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW: asa.aip.org].

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

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

BOOK REVIEWS

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

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

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

Foundations of Engineering Acoustics

Frank Fahy

Academic Press, San Diego, 2000.
784 pp. Price: $84.95 hc ISBN: 0122476654.

The author, a well-known acoustician, describes engineering acoustics as concerning ''all acoustical and related vibrational aspects of engineering design, manufacture, products, systems and operation.'' Engineering acoustics involves the manipulation of sound, the principal aim being to control it as an undesirable by-product that is potentially harmful to human health and well-being, according to the author. Thus, the book presents the fundamentals of acoustics with emphasis on those relating to noise generation by machinery, vehicles and industrial plants, and to the practice of noise control engineering.

Following a brief introductory chapter on ''Sound Engineering,'' the author presents well-written chapters on ''The Nature of Sound and Some Sound Wave Phenomena,'' ''Sound in Fluids,'' and ''Impedance.'' The next chapter, ''Sound Energy and Intensity,'' is also the subject of a very useful book by the author, published in 1989 with a second edition in 1995. This is followed by chapters on ''Sources of Sound'' and ''Sound Absorption and Sound Absorbers.''

The next chapters on ''Sound in Waveguides,'' Sound in Enclosures,'' ''Structure-borne Sound,'' and ''Transmission of Sound through Partitions'' strike a nice balance between the theoretical and the practical aspects of engineering acoustics. The chapter on ''Sound in Waveguides,'' the longest in the book, begins with a semiquantitative description, in terms of multiple reflection of plane pulse wavefronts, of the generation of undamped and damped plane-wave modes and resonances in uniform ducts terminated by reactive and resistive elements. Mathematical analysis of propagation in ducts having both rigid and finite-impedance walls leads to a brief presentation of performance data for lined ducts and splitter attenuators.

The chapter on ''The Behavior of Sound in Enclosures,'' which relates to auditoria, recording studios, vehicle interiors, and noise control enclosures, opens with an illustration of the temporal evolution of an impulsively excited sound field in a reverberant enclosure using an image model. A conventional wave model analysis demonstrates that the rapid increase of reflection arrivals revealed by the impulse model is complemented by a rapid increase with frequency of the density of natural frequencies, to a point where deterministic modal modeling is of little value. A simple model based upon the enclosed-space Green's functions reveals the factors that govern vibroacoustic coupling between structures and enclosed fluids. The chapter on ''Transmission of Sound through Partitions'' is largely extracted from his earlier book, *Sound and Structural Vibration* (Academic, New York, 1987).

In the final chapter on ''Reflection, Scattering, Diffraction and Refraction'' the reader is introduced to a graphical technique that broadly elucidates the origin and characteristics of diffraction by edges, together with practical data relating to screens. The process of scattering by solid objects is explained and illustrated by an analysis of scattering by a rigid sphere and a thin disk. Finally, a simple example of the application of ray-tracing analysis to refraction of sound is given, along with a brief account of refraction by wind and temperature gradients near the surface of the Earth.

Seven appendices treat ''Complex Exponential Representation of Harmonic Functions,'' ''Frequency Analysis,'' ''Spatial Fourier Analysis of Space-dependent Variables,'' ''Coherence and Cross-Correlation,'' ''The Simple Oscillator,'' ''Measures of Sound, Frequency Weighting and Noise Rating Indicators,'' and ''Demonstrations and Experiments.'' A number of teachers may find the latter appendix, which passes on some of the pedagogical know-how the author has accumulated during his 35 years of teaching, to be worth the price of the book alone.

Professor Fahy writes well, as readers of his previous books will testify. Difficult concepts are explained clearly, and at an appropriate level of mathematical rigor. Although some might question the introduction of the Dirac delta function and Green's function in a text for undergraduates, the author argues that they are essential to the understanding of the equation that relates sound fields to the boundary conditions satisfied by the fluid, for example.

Each chapter has questions (problems), and the answers are given in the back of the book, along with brief explanations, in many cases, of how the solution might proceed, a feature that readers who are using this text as a means of self-instruction will greatly appreciate.

If the book has a weakness, it may be the dearth of discussion of computational methods in engineering acoustics, such as finite- and boundary-element methods of acoustic modeling. This omission is partly due to space limitations and partly due to the author's belief that the variational approach to dynamics upon which they rest will be unfamiliar to most readers.

This book is a welcome addition to the acoustics literature, and it should be in the library of anyone teaching courses in engineering acoustics or noise control. It will no doubt receive a considerable number of adoptions as a textbook for undergraduate and graduate courses in these areas.

THOMAS D. ROSSING
Physics Department
Northern Illinois University
DeKalb, Illinois 60115

Introduction to Wave Propagation in Nonlinear Fluids and Solids

D. S. Drumheller

Cambridge University Press, 1998.
Price: $140.00 (hardcover) $52.95 (paperback) ISBN: 0521583136 (hardcover) 0521587468 (paperback).

This book introduces the reader to the mechanics and physics that describe shock waves in compressible materials. It gives equal emphasis to propagation in both gases and solid materials. It does so at an introductory to intermediate level. The background assumed of the reader is limited to no more than advanced calculus and introductory linear algebra, though some background in the mechanics of a continuous medium would be helpful. The title has been carefully chosen: it does not indicate that the book is about nonlinear waves in fluids and solids, so much as it is about strong waves in fluids and solids that respond nonlinearly. The book describes only one-dimensional, nonlinear, compressional waves; it does not describe other varieties of one-dimensional nonlinear waves such as dispersive water waves. Moreover, the book is not predominantly concerned with wave propagation: at least half the book describes the thermodynamics of materials and the constitutive relations of specific material systems.

The book is comprised of four long chapters and two appendices. The

 0001-4966/2002/111(3)/1142/3/$19.00 © 2002 Acoustical Society of America

second appendix contains useful tables of linear and nonlinear material parameters. The book has 31 references listed in the back matter, all to books written for use by research students or researchers, and none published after 1994. References to research papers are not extensive and are given as footnotes; none are to papers published after 1994. There is a very helpful list of most of the symbols used in the text placed in the front matter, making it possible to weave back and forth among the chapters. The index is too brief, approximating one page of index per 100 pages of text, so that, for example, *stress, deviatoric* is listed, but *deviatoric stress, see stress* is not. The pages are well designed with adequate space between sections, sentences, and words, and with a restrained use of bold and italic type to set off titles and important phrases. The typeface is Times Roman. The diagrams are large and easy to read, and are uniformly well thought out. The equations are nicely set off from the text; only equations referred to in the text are numbered, as is common practice. Type, or type and a diagram, cover 13 by 20 cm, the outside and bottom margins are 3 cm wide, and the inner and top margins are 1.5 cm, so that, at 500 pages of text, this is not a short book. The binding of the paperback edition is adequate if the book is used by a single reader, but not if used by multiple readers.

Chapter 1 describes the continuum mechanics needed to understand the remainder of the book. It treats fluids and solids equally, and uses spatial (Eulerian) and material (Lagrangian) descriptions for both. All readers, regardless of their background, should read this chapter so that they know where to find various definitions as they proceed through the book. Readers who come to the book with only a background in the acoustics of fluids will find much here that is new. The reviewer found the treatment very good, especially the way in which spatial and material descriptions were interwoven, but he did find the discussion of rotations a bit awkward. By defining $\{\mathbf{u},\mathbf{v}\}$ as the signed angle from $\mathbf{u}$ to $\mathbf{v}$, the author eventually gets himself into an awkward situation wherein he writes, for an element of a matrix representing rotation, $\alpha_{k(m)}=\alpha_{(m)k}$, while $\alpha_{k(m)}\neq\alpha_{m(k)}$. This chapter is 78 pages long, has 39 figures, and 62 exercises. The reviewer did not attempt to solve any of these exercises, nor any of the exercises in the other three chapters.

Chapter 2 is the most important chapter from a student's perspective because it describes nonlinear waves and how they form shocks. Much of the analysis is based on Riemann's solution to one-dimensional hyperbolic equations. The author begins his discussion by solving the linear wave equation with Riemann's method, which is a very innovative way to introduce the technique. Further, the combining of space-time diagrams, which indicate how the waves propagate, with stress-particle velocity diagrams, which explain how the material changes from one state to another, provides an especially clear explanation of the mechanics of what is happening. In other books often only the Riemann invariants—rather abstract objects—are used to explain propagation. However, in part by confining himself to considering only a nonlinear-elastic material, the author gives no discussion of the simple kinematic wave equation and of how shocks are geometrically fitted into its solution. Thus, an opportunity to indicate how first-order partial differential equations model nonlinear waves and how they can be solved is missed. The interactions of a nonlinear wave with a barrier or another wave are very complex; the author ends this chapter with clear, well-written qualitative descriptions of these interactions, but the quantitative discussion is very limited. Numerical solutions are usually needed to sort out many of the quantitative details of propagation and interaction, so that the first of the two appendices continues this discussion as it tries to give the reader a sense of which algorithms to choose for these calculations and why. This chapter is 146 pages long, has 125 figures, and 67 exercises.

Chapter 3 describes the thermodynamics of continuous materials when their state is often far from an equilibrium one. The materials described are elastic, thermoelastic, viscous, and thermoviscous. The word thermomechanics, used by the author throughout, is intended to convey that this is a presentation of thermodynamics written to indicate the symbiotic relation between the mechanical and thermal processes induced in a material as strong compressive waves propagate in it. The reviewer learned much from this chapter. Of particular interest was learning that the second law could be used as a constraint on the constitutive relations. In fact, this chapter and the following one describing constitutive relations provide the main reason for a researcher in acoustics to read this book. Explanations of this sort are not easy to find outside of the research literature, and that literature can be contradictory and polemical. Nevertheless, some comment as to why relations that one was told as an undergraduate only apply to quasistatic processes continue to describe accurately processes that are very dynamic would have been helpful. This chapter is 87 pages long, has 33 figures, most very well thought out, and 42 exercises.

Chapter 4 describes constitutive relations for four specific material systems followed by a brief description of combustion and phase transformations. The four systems are (1) an ideal gas; (2) a Mie–Grüneisen solid; (3) an elastic–plastic solid; and (4) a saturated porous solid. A microscopic physical model is sketched in each case to help to motivate the particular construction of a constitutive relation. At 153 pages this is the longest chapter and, as suggested previously, indicates that the book is directed as much at understanding material behavior as at understanding wave propagation. It has 82 figures and 44 exercises.

The author indicates in his preface that the first two chapters of the book can be used as a textbook for a one-semester course discussing continuum mechanics and wave propagation. The instructor should assume that each student will read approximately 20 pages and solve 4–5 homework problems each week, for 12 of the 15 weeks in a semester. The topics will hold the interest of mechanics students and acoustics students with a distinct interest in physical acoustics.

There has been, for some time, a need for an introduction to nonlinear compressional waves that form shocks, and the related mechanics and physics; this excellent book fulfills that need.

JOHN G. HARRIS
Center for QEFP
Northwestern University
2137 North Sheridan Road
Evanston, Illinois 60208-3020

Classic Papers in Shock Compression Science

James N. Johnson and Roger Cheret, Editors

Springer-Verlag, New York, 1998.
524 pp. Price: $99.95 (hard cover) ISBN: 0387984100.

This book contains a selection of the classical papers, eleven papers by ten authors, in shock compression science. A photo and a brief biography of the authors precede each paper. One of the editors, Cheret, wrote a more in-depth biography of Hugoniot. The editors state their goal as the documentation of the process of incremental discovery and the advancement of the fundamental understanding of shock compression science. The editors point out in the preface that no attempt was made to reach a uniform use of symbols. Each paper appears with its original notations. For the sake of completeness they also decided to publish each paper in its entirety, even when only a fraction of the paper deals with shock compression science.

The first paper (63 pp.) is ''A paper on the theory of sound'' by S.-D. Poisson (1808). Poisson first derives the lossless equations of motion in an isothermal fluid. Then, he discusses solutions of the linearized equations, reflection of sound using the image method, and reflection off ellipsoidal and paraboloidal surfaces. Poisson also discusses the parameters that affect the sound propagation speed. He notes that an isothermal sound speed calculation yields an underprediction of about 16% compared with the reported measured value. By including a second linear term (with factor k) in the pressure density relation to account for the elasticity of the air and adjusting the value of this coefficient he was able to obtain good agreement between theory and measurement for a value of $1+k=1.4254$. Poisson then studies finite amplitude motion. He finds the exact solution for progressive waves as a function of the velocity potential. It is interesting to note that Poisson failed to recognize the importance of his solution with regard to waveform distortion. He limits his discussion to a pulse that begins and ends with zero particle velocity and observes that the length of such a pulse will never change. He then concludes that ''irrespective of whether the sound is loud or soft, it is always transmitted at the same velocity.'' Poisson did not appreciate that the propagation speed varies within the pulse.

The second paper (9 pp.) is ''On a difficulty in the theory of sound'' by G. G. Stokes. The paper is a combination of the original 1848 paper and the revised version of 1883. Stokes addresses ''a very remarkable difficulty'' that Chablis had discovered, namely, that the points of zero velocity of a wave coincide with points of maximum velocity. Stokes shows for the first time graphs of waveform distortion caused by the dependence of the propa-

gation speed on particle velocity. Stokes goes on to calculate the minimum time for a wave to develop a vertical slope. He then proposes a surface of discontinuity across which abrupt changes of density and velocity exist. He also concludes that shock formation destroys the progressive wave nature and that reflection must occur. Stokes derives conservation laws of mass and momentum that must hold across the shock front. These relations are two of the three familiar Rankine–Hugoniot relations, but apparently Stokes received no credit for developing these laws. Stokes did not recognize that shock propagation is accompanied by a loss of energy as was pointed out to him later by Kelvin and Rayleigh.

The third paper (22 pp.) is ''On the mathematical theory of sound'' by S. Earnshaw (1860). Earnshaw first rederives the lossless progressive wave equations and then develops the solution for wave propagation in a tube generated by an arbitrary piston motion. Earnshaw therefore solves the forced problem whereas Poisson treated the free problem. Earnshaw first develops a solution for an isothermal gas, then for an adiabatic gas, and finally for an arbitrary pressure–density relationship. He discusses in detail the formation of shocks, which he calls bores, and rarefaction and condensation waves. Earnshaw points out a possible consequence of finite amplitude sound propagation, namely that, ''the command to fire a gun, if instantly obeyed, and the report of the gun, might be heard at a long distance in reverse order.'' Earnshaw further points out the generation of two waves by the piston, a forward and backward traveling wave. He then proceeds to derive an expression for the maximum mass flow that can travel through the tube. He also points to the existence of rarefaction shocks. Finally Earnshaw explains the increase in propagation speed for an isothermal gas as the result of the addition of the particle velocity (he calls it the wind velocity) to the sound speed. For an adiabatic gas Earnshaw failed to see the extra component that occurs from the nonlinearity of the pressure–density relation.

The fourth paper (20 pp.) is ''The propagation of planar air waves of finite amplitude'' by B. Riemann (1860). Here Riemann introduces the method of characteristics. He is also the first to consider compound wave fields. He considers two quantities now known as the Riemann invariants and shows that they are propagated with the relative speeds of propagation. Riemann then proceeds to invert the equations and obtains an ordinary differential equation of the space and time variables as a function of the Riemann invariants. Riemann rediscovers the theory of shocks but made the incorrect hypothesis of a lossless pressure–density relation.

The fifth paper (15 pp.) is ''On the thermodynamic theory of waves of finite longitudinal disturbance'' by W. J. M. Rankine (1870). Rankine was the first to recognize that heat transfer within the gas is necessary for a finite amplitude wave to maintain its shape. He was then able to derive the conservation laws for a steady shock in a perfect gas (now called the Rankine–Hugoniot shock relations) that connect the flow field behind the shock with that ahead of it. In a footnote Rankine mentions that Kelvin pointed out that a rarefaction shock, although mathematically possible, is an unstable motion. Rankine then goes on to find an analytical solution for the profile of a steady shock in a heat conducting, inviscid, and perfect gas.

The sixth and seventh papers are by P. H. Hugoniot. They are preceded by an extended biography (11 pp.) of Hugoniot by Cheret. The first paper by Hugoniot is ''On the propagation of motion in bodies and in perfect gases in particular—I'' (83 pp., 1887). It contains three chapters that deal mostly with the mathematics and physics of waves. The first chapter is a review of the theory of characteristics. In the second chapter Hugoniot derives the equations of motion for a solid, liquid, and gas. Both plane and spherical waves are treated. In the third chapter solutions of the wave equations are discussed. Hugoniot goes on to define the sound speed in fluids. The second paper is a continuation of the first (114 pp., 1889) and has two chapters. Chapter four is on plane fluid motion neglecting viscosity and heat conduction. Hugoniot considers excitation by a piston at the end of a cylindrical pipe. First he treats finite amplitude propagation for a general fluid. Then he applies the results to a perfect gas. Both isothermal and adiabatic cases are treated. Two examples are presented. The first is that of a piston with uniform acceleration and the second is that of a periodic vibrational motion. Here Hugoniot shows two graphs that demonstrate the cumulative nature of waveform distortion. The biggest distortion occurs at the head of the wave and the least near the piston face. Hugoniot also calculates the shock formation distance. Hugoniot states ''when discontinuities are introduced, the original relation between pressure and volume is suddenly changed.'' The last chapter is dedicated to the development of the shock relations that are now famous and referred to as the Rankine–Hugoniot relations. Although Hugoniot's work appeared after that of Riemann and Rankine, Cheret points out in his biography of Hugoniot that Hugoniot was probably not aware of the work of either one. Hugoniot assumed an inviscid and thermally nonconducting gas and his derivations are for an unsteady shock. Hugoniot shows that a shock wave that is reversible and adiabatic would violate the principle of conservation of energy.

The eighth paper is ''Aerial plane waves of finite amplitude'' by Lord Rayleigh (44 pp., 1910). He gives an excellent review of finite amplitude sound. First he discusses finite amplitude waves without dissipation and reviews the work of Poisson, Earnshaw, Riemann, and Stokes. Upon discussing the distortion of finite amplitude waves into a shock, Rayleigh notes ''tendency to discontinuity may be held in check by forces of a dissipative nature.'' The second and third paragraphs are on waves of permanent regime under the influence of dissipative forces. Rayleigh refers to Rankine's work and points out the deficiencies in Rankine's assumption that heat conduction is responsible for the permanency of the wave. Rayleigh also points out that the shock profile calculation of Rankine is limited for shocks with pressure ratio less than a critical value determined by the ratio of specific heats. Next Rayleigh reviews the work of Hugoniot and concludes that Hugoniot's third shock relation is exactly the same as that developed by Rankine 15 years earlier, remarkable since both started from very different assumptions. Then he investigates the effect of viscosity in addition to heat conduction on shock profile. He arrives at a third-order differential equation and mentions that ''a complete analytical solution of our equation is not to be expected.'' Rayleigh continues to find a numerical solution. He shows that the shock thickness of a shock with a density ratio of 3 is about 30 μm.

The ninth paper is ''The conditions necessary for discontinuous motion in gases'' by G. I. Taylor (8 pp., 1910). Taylor shows that heat conduction and viscosity are the causes of dissipative forces in a real gas. He then goes on to calculate the shock profile of a heat conducting and viscous gas. Taylor obtains an analytical solution for weak shocks and derives the formula for the thickness of a weak shock, now called the Taylor shock thickness.

The tenth paper is ''On the theory of shock waves for an arbitrary equation of state'' by H. A. Bethe (72 pp., 1942). Bethe mentions in his introduction that shock wave theory had been mostly developed for ideal gases and that the question of shock wave stability had received little attention. First he develops the theory of shock waves for fluids with an arbitrary equation of state. Then he looks at the theory of small shocks and shows that the entropy increase is a cubic function of shock strength. He also calculates the velocity of the shock wave and shows that it is always supersonic with respect to the medium in front of the shock and subsonic with respect to that behind the shock. Finally Bethe derives a set of three conditions that ensure stability of the shock wave.

The eleventh paper is ''Shock waves in arbitrary fluids'' by H. Weyl (23 pp., 1949). Because of World War II Weyl had no knowledge of the report written by Bethe. The first part of the paper is on thermodynamics and the shock phenomenon and parallels Bethe's work. In the second part, the problem of the shock layer, Weyl develops the concept of the dissipation layer similar to the Prandtl boundary layer. In doing so he uses the asymptotic expansion method, a method now in widespread use in physics.

For anyone working in nonlinear acoustics or shock waves this book is excellent reading. Although not complete in its collection of important papers, e.g., Airy's and Challis' papers are missing, it contains the majority of important papers. A drawback of the book is the multitude of typesetting errors, misspelled words, and errors in formulas. In addition to the errors, some of the translations from French seem to have been done in a rather ''word by word'' fashion. At times, this makes the English version cumbersome to read. In summary this book contains a selection of the important historical papers in the development of nonlinear acoustics and shock waves and thus provides an excellent overview of the history of nonlinear acoustics and shock waves.

BART LIPKENS
Mechanical Engineering
Virginia Commonwealth University
Richmond, Virginia 23284-3015

REVIEWS OF ACOUSTICAL PATENTS

Lloyd Rice
11222 Flatiron Drive, Lafayette, Colorado 80026

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

Reviewers for this issue:

GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
ALIREZA DIBAZAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
IBRAHIM M. HALLAJ, *Wolf Greenfield & Sacks P.C., 600 Atlantic Avenue, Boston, Massachusetts 02210*
HASSAN NAMARVAR, *Department of BioMed Engineering, University of Southern California, Los Angeles, California 90089*
DAVID PREVES, *4 Deerfield Drive, Princeton Junction, New Jersey 08550*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ERIC E. UNGAR, *Acentech, Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*

5,596,550

43.30.Yj LOW COST SHADING FOR WIDE SONAR BEAMS

Linwood M. Rowe, Jr. *et al.* , assignors to Northrop Grumman Corporation
21 January 1997 (Class 367/103); filed 20 June 1995

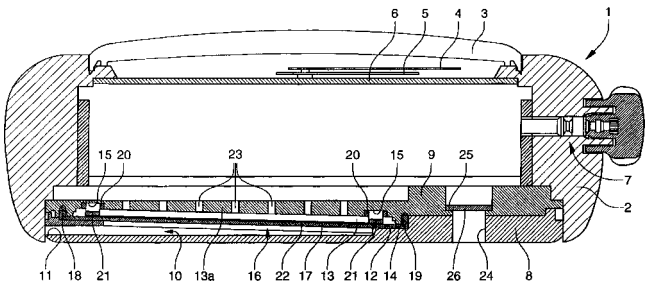
By combining the outputs of just two sensor elements (their spacing relative to the wavelength is not specified) of which one sensor has a weighting which is only 23% of the other but is in phase with it, the authors suggest that a beam pattern as narrow as that obtainable from an array containing many more sensors can be realized.—WT

6,014,347

43.30.Yj DEVICE CAPABLE OF BEING SUBMERGED AND INCLUDING AN ACOUSTIC TRANSDUCER

Jean-Pierre Mignot and Paul Dinnissen, assignors to Asulab S.A.
11 January 2000 (Class 368/88); filed in Switzerland 28 August 1997

A wrist watch case is shown in cross section. Item **2** is the case proper, **3** the crystal, **4** and **5** the hands, and **6** the dial. The back cover **8** creates an inlet cavity **10** that communicates with the exterior space, which may be water, via a channel **11**. A watertight interior cavity, labeled recess **13**, is created when circular membrane **17** is attached to support surfaces **12** and **19** that surround membrane **17**. A bender-type piezoelectric transducer assembly **16** is bonded onto membrane **17**. This combination is free to vibrate within water cavity **10** and air cavity **13**. The associated electronics for the transducer are presumably housed elsewhere in the watch case. The membrane **17** is edge-supported by a set of resilient lugs **21**. Static pressure compensation is provided via membrane **26**, a water-resistant but air-permeable membrane of sintered Teflon.—WT

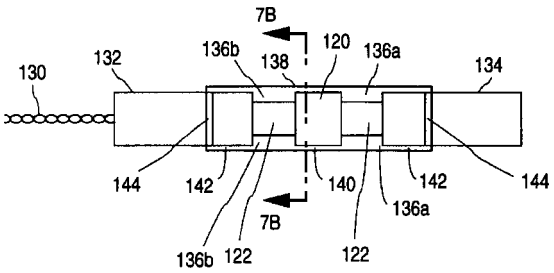


6,277,077

43.35.Wa CATHETER INCLUDING ULTRASOUND TRANSDUCER WITH EMISSIONS ATTENUATION

Axel F. Brisken and N. Parker Willis, assignors to Cardiac Pathways Corporation
21 August 2001 (Class 600/459); filed 16 November 1998

The catheter covered by this patent is said to improve performance by providing ultrasound damping regions. The catheter consists of an elongated body member, an ultrasound transducer located on the body member, and an ultrasound damping region adjacent to the transducer. The transducer is



configured to emit or receive an ultrasound signal. The damping region is configured to improve uniformity of an ultrasound signal in three-dimensional space. The preferred methods of damping include air or a material containing air.—DRR

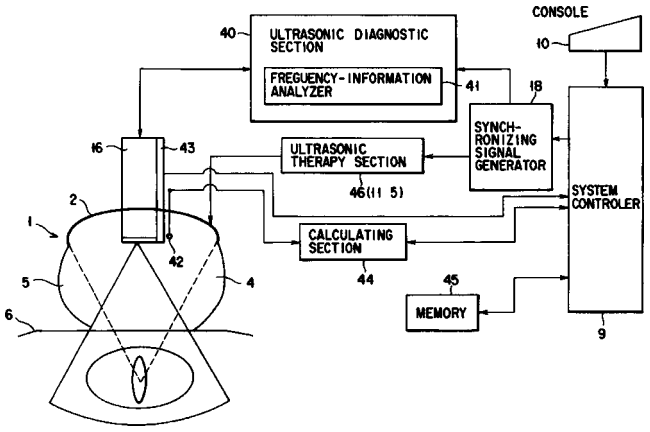
6,280,402

43.35.Wa ULTRASOUND THERAPEUTIC APPARATUS

Yoshiharu Ishibashi *et al.* , assignors to Kabushiki Kaisha Toshiba
28 August 2001 (Class 601/2); filed in Japan 31 March 1995

The patent covers a therapeutic apparatus consisting of a therapeutic ultrasonic source and driver circuit and an *in vivo* imaging probe for the purpose of obtaining a tissue tomographic image in the vicinity of the focus of the ultrasonic waves. The imaging probe receives echoes of the ultrasonic

 0001-4966/2002/111(3)/1145/12/$19.00 © 2002 Acoustical Society of America



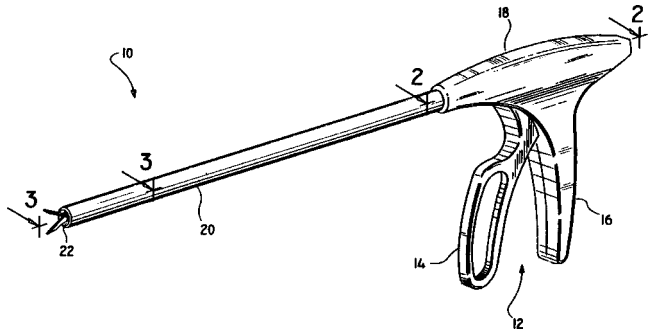
pulses. The ultrasonic wave-driving conditions are adjusted on the basis of the received echo signals, which contain information about the actual intensity of the ultrasonic waves within the body, resulting in a claimed improvement in the safety and reliability of therapy.—DRR

6,280,407

43.35.Wa ULTRASONIC DISSECTION AND COAGULATION SYSTEM

Ronald Manna *et al.* , assignors to United States Surgical Corporation
28 August 2001 (Class 604/22); filed 14 August 1997

This ultrasonic surgical instrument includes an angled blade and a clamp member particularly suited for performing dissection and coagulation of tissue. The system consists of three parts: an ultrasonic instrument, a control module, and a pedal actuator. The instrument has a housing and an elongated portion extending from the housing. An ultrasonic transducer inside the housing is coupled to a cutting blade. A cutting surface is angled



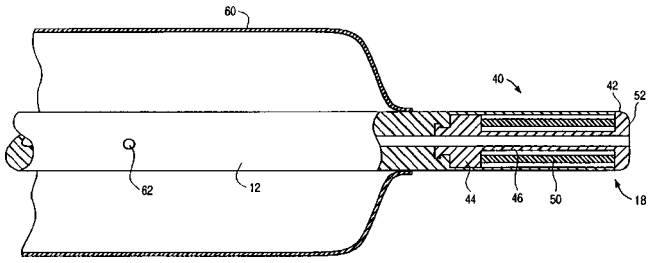
with respect to the longitudinal axis of the elongated body and, thus, with respect to the axis of vibration. A tissue clamp positioned next to the blade is movable from an open position which holds the cutting surface away from the tissue to a clamped position which aligns the blade with the tissue contact surface. The clamp member and the angled blade combine to improve contact between tissue and the blade cutting surface.—DRR

6,287,272

43.35.Wa BALLOON CATHETERS HAVING ULTRASONICALLY DRIVEN INTERFACE SURFACES AND METHODS FOR THEIR USE

Axel F. Brisken and Vartan E. Ghazarossian, assignors to Pharmasonics, Incorporated
11 September 2001 (Class 604/22); filed 5 September 1996

The catheter consists of a catheter body containing an oscillating driver, an interface surface mechanically coupled to the driver, and an inflatable balloon deployed near the interface surface. The balloon may be an



angioplasty balloon, in which case the interface surface will deliver ultrasonic or other vibratory energy into a blood vessel as part of an angioplasty or similar procedure. Alternatively, the catheter may consist of a pair of axially spaced isolation balloons, in which case the interface surface can deliver ultrasonic or other vibratory energy into a treatment region defined as that being between the balloons. The energy can thus act to mix or enhance penetration of a treatment held between the balloons in performing a vascular treatment procedure.—DRR

6,276,212

43.35.Zc ULTRASONIC TRANSDUCER

Stephen R. W. Cooper and Neil G. Murray, Jr., assignors to TRW, Incorporated
21 August 2001 (Class 73/632); filed 8 July 1999

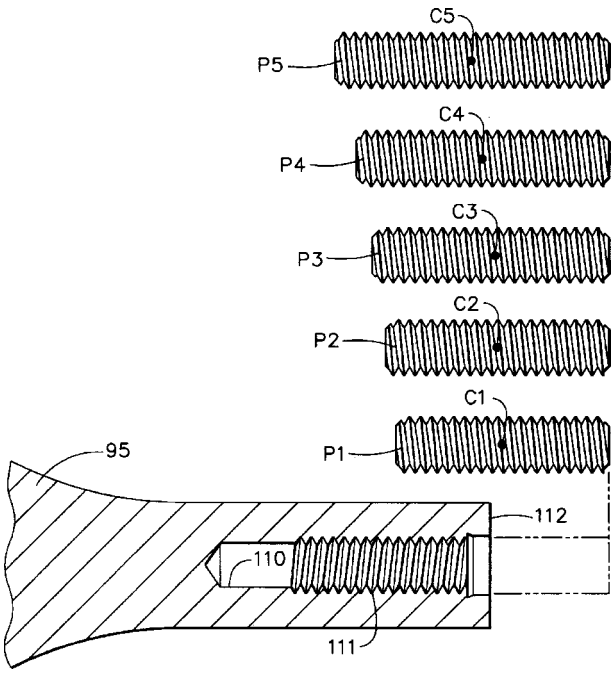
An ultrasonic vibrating mass and diaphragm are coupled to a relatively large mass, having sufficient inertia to reduce transmission of vibrations to an outer housing. A layer of insulating material separates the large mass from the housing.—IMH

6,278,218

43.35.Zc APPARATUS AND METHOD FOR TUNING ULTRASONIC TRANSDUCERS

Ashvani K. Madan *et al.*, assignors to Ethicon Endo-Surgery, Incorporated
21 August 2001 (Class 310/312); filed 15 April 1999

High-power ultrasound sources, especially sandwich-type transducers, and horns are tuned according to the invention by placing a threaded weight



into a threaded bore or opening of a resonator. Depending on the desired resonant frequency, an appropriate threaded element is selected and inserted into the threaded bore.—IMH

6,279,378

43.35.Zc ULTRASONIC GAS ANALYZER AND METHOD TO ANALYZE TRACE GASES

Shuh-Haw Sheen *et al.*, assignors to The University of Chicago
28 August 2001 (Class 73/24.01); filed 27 October 1999

A gas sample is drawn into a cavity accessible to a pair of ultrasonic transducers operating in transmit/receive mode. The gas is analyzed on the basis of time of flight and amplitude information received by the receiving transducer(s). A gas pump acts to draw a slow steady flow of the sample gas through the chamber. The system can determine the content of hydrogen or exhaust gases in air.—IMH

6,013,311

43.38.Fx USING MORPHOLOGICAL CHANGES TO MAKE PIEZOELECTRIC TRANSDUCERS

Dilip K. Chatterjee *et al.*, assignors to Eastman Kodak Company
11 January 2000 (Class 427/100); filed 8 June 1998

This patent discusses a method of forming a piezoelectric material in which the *d*-coefficient varies, say linearly, in one of the spatial directions. This would be useful, for example, in creating a bender transducer using a single piece of this material rather than using two pieces of conventional material having opposite polarity.—WT

6,275,448

43.38.Fx PRESSURE-COMPENSATED ACCELERATION-INSENSITIVE HYDROPHONE

Elmore Kittower and James W. Pell, Jr., assignors to L3 Communication
14 August 2001 (Class 367/155); filed 12 December 1977

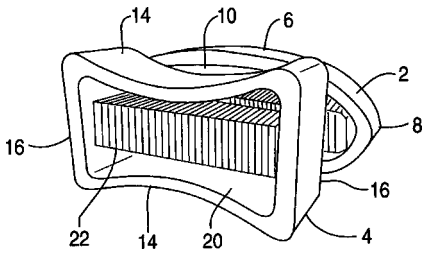
The patent covers the design of a mass-balanced, acceleration-canceling, dumbbell-type transducer with piezoceramic elements for use in towed line arrays.—WT

6,278,658

43.38.Fx SELF BIASED TRANSDUCER ASSEMBLY AND HIGH VOLTAGE DRIVE CIRCUIT

Colin W. Skinner and Qi-Chang Xu, assignors to L3 Communications Corporation
21 August 2001 (Class 367/161); filed 25 March 1999

A flextensional transducer features both a concave shell **14** and a convex shell **6**. Not shown are two end plates that close the front and back faces of the combination plus an internal bulkhead that closes the interior open faces of both shells. Both shells are set into vibration by means of two separate, parallel stacks of PMN piezoceramic such as **22**. The two driver stacks are operated 180 degrees out of phase from each other so that the resulting shell displacements are additive. The dc bias required by the PMN piezoceramic is configured to accomplish the 180-degree phase difference between the two stacks and also to eliminate the need for a large, heavy, expensive blocking capacitor to protect the power amplifier. The two shells have slightly different resonant frequencies resulting in increased bandwidth for the transmitting voltage response.—WT



6,279,397

43.38.Hz METHOD AND APPARATUS FOR FOCUSING PROPAGATING WAVE PATHS OF A PHASED ARRAY IN SPHERICALLY-BOUNDED MATERIALS

Roger Francis Dwyer, assignor to Westinghouse Electric Company LLC
28 August 2001 (Class 73/606); filed 17 November 1999

The disclosure describes an apparatus and method for focusing waves from phased array elements onto a point within a spherical lenslike object. The concept involves calculating the best phasing/timing for the array elements so that the resulting focus occurs within an area under the spherically bounded workpiece. The invention is directed to applications such as nondestructive flaw detection in nuclear reactor pressure vessels, for example.—IMH

6,223,853

43.38.Ja LOUDSPEAKER SYSTEM INCORPORATING ACOUSTIC WAVEGUIDE FILTERS AND METHOD OF CONSTRUCTION

Graeme John Huon *et al.*, Mt. Waverley, Victoria, Australia
1 May 2001 (Class 181/145); filed in Australia 23 December 1994

In designing vented or bandpass loudspeaker systems, lumped parameters are traditionally used. A vent is modeled as an acoustic mass. More recently, vent pipe resonances have been included (usually incorrectly) but these are assumed to be unwanted side-effects. The inventors argue that a sufficiently sophisticated computer program can include multiple pipe resonances in the predicted system response and use them as part of the design process. This U.S. patent was issued more than five years after the filing of its application. During that time, bandpass systems based on this design have been demonstrated.—GLA

6,237,715

43.38.Ja SUBWOOFER ASSEMBLY

Dennis A. Tracy, Culver City, California
29 May 2001 (Class 181/156); filed 1 December 1998

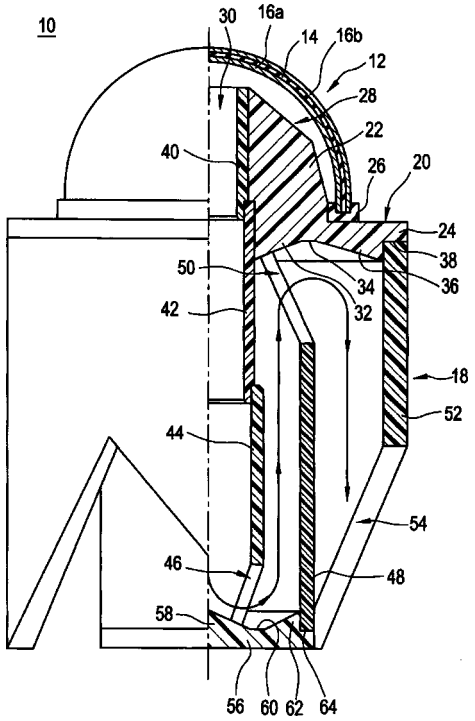
What the author has managed to patent is the concept of a bandpass subwoofer, ''...having a curved profile shaped to substantially conform to the wall of an aircraft fuselage.''—GLA

6,243,475

43.38.Ja SPEAKER

Takeshi Nakamura and Yoshiaki Heinouchi, assignors to Murata Manufacturing Company, Limited
5 June 2001 (Class 381/190); filed 28 May 1997

A hemispherical piezoelectric transducer **14** is loaded by a small coupling chamber **28** which drives horn throat **30**. Several triangular openings **54** form the mouth of the folded horn. This compact assembly is said to be ''...efficient in the low sound range and is unidirectional relative to the plane of installation or support of the speaker.'' If the illustration is drawn to scale,



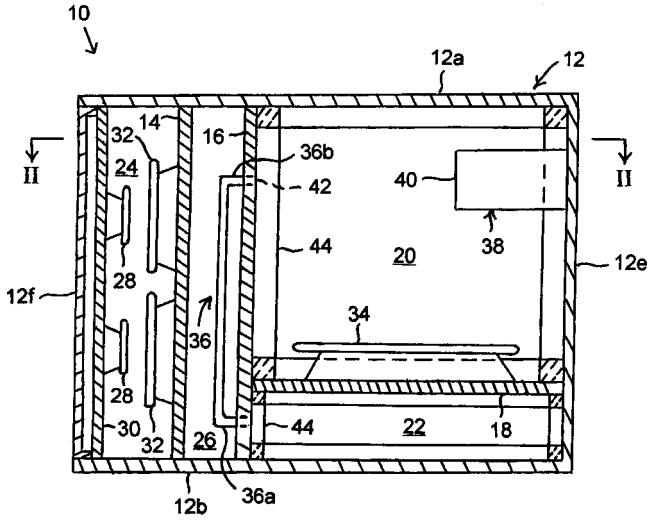
then the radiating area of the diaphragm is much larger than the horn throat and about twice as large as the horn mouth. No dimensions are given, but if we assume that a practical embodiment might be about 100 mm in diameter and 150 mm high, then its usable low-frequency limit is probably around 250 Hz (middle C). Above 300 Hz or so, response should exhibit the characteristic square-wave pattern of an undamped transmission line speaker.—GLA

6,243,477

43.38.Ja AUDIO SYSTEM WITH PARTITIONED INPUT AND OUTPUT COMPARTMENTS

Aldo M. Ruiz, Tucson, Arizona
5 June 2001 (Class 381/351); filed 26 May 1998

Woofer **34** is mounted in chamber **20**, which transmits sound energy to the outside world through vent **40**. The front of the woofer is loaded by compartment **22** which communicates with compartment **20** through multiple pipes **36**. Thus far, the geometry translates into a two-chamber bandpass system of known prior art. Tweeters **28** are conventionally mounted, but midrange speakers **32** face into compartment **26**. ''The sound emitted by the midrange speakers enters the atmosphere through the top wall **12a** and side walls **12c**, **12d** of the housing **10**.'' The system is said to provide ''enriched'' bass response and ''enhanced'' overall response.—GLA

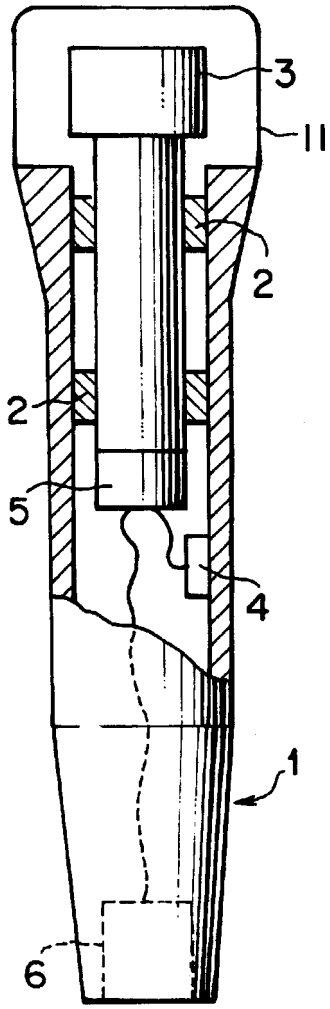


6,226,386

43.38.Kb MICROPHONE

Hiroshi Akino, assignor to Kabushiki Kaisha Audio-Technica
1 May 2001 (Class 381/173); filed in Japan 15 May 1998

Microphone element **3** is shock-mounted **2** within case **1**. Even so, mechanical shocks will produce noise in the microphone's output signal.



Piezoelectric sensor **4** picks up mechanical vibrations, which are amplified, processed, and mixed with the microphone output to cancel such unwanted noise.—GLA

6,243,138

43.38.Kb CAMERA INCLUDING MEANS FOR ACQUIRING BI-DIRECTIONAL SOUND

Richard Scott Keirsbilck, assignor to Eastman Kodak Company
5 June 2001 (Class 348/374); filed 4 August 1995

Miniature directional microphones can be fitted with extension tubes to improve their performance. Known prior art also includes embedding such a miniature microphone and extension tube within a larger object, such as a metal plate. Ah, but suppose the larger object is a hand-held camera with sound entry ports on front and rear surfaces? The geometry described constitutes the novel feature of this patent.—GLA

6,243,322

43.38.Si METHOD FOR ESTIMATING THE DISTANCE OF AN ACOUSTIC SIGNAL

Pierre Zakarauskas, assignor to WaveMakers Research, Incorporated
5 June 2001 (Class 367/127); filed 5 November 1999

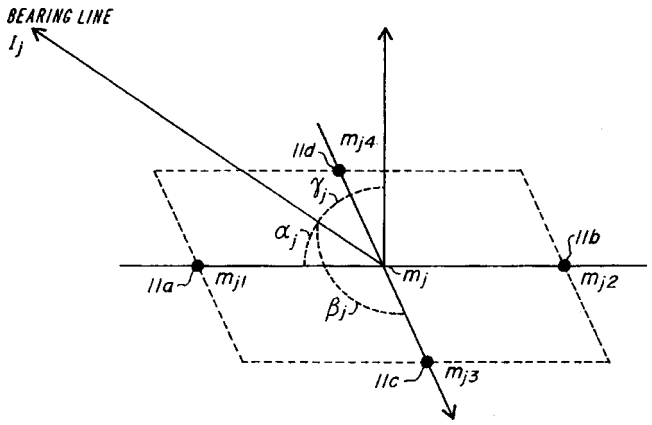
In boardroom sound reinforcement systems and teleconference systems it is desirable to energize only one or two microphones at any given time. In practice, multiple sound paths and reverberation confuse simple gating schemes. As part of a more sophisticated approach, the system should be able to discriminate between near and far sound sources at each microphone location, even in a reverberant room. The invention attempts this feat by electronically comparing signals from pairs of microphones and then calculating the ''...angular distribution of acoustic power.''—GLA

6,243,471

43.38.Si METHODS AND APPARATUS FOR SOURCE LOCATION ESTIMATION FROM MICROPHONE-ARRAY TIME-DELAY ESTIMATES

Michael S. Brandstein *et al.*, assignors to Brown University Research Foundation
5 June 2001 (Class 381/92); filed 27 March 1995

The invention is intended primarily as a video camera steering apparatus. Two pairs of microphones **11a-11b** and **11c-11d** might be located on the wall of a conference room. Sound from, say, a roving talker is picked up by all four microphones. Signals from each pair are processed to estimate



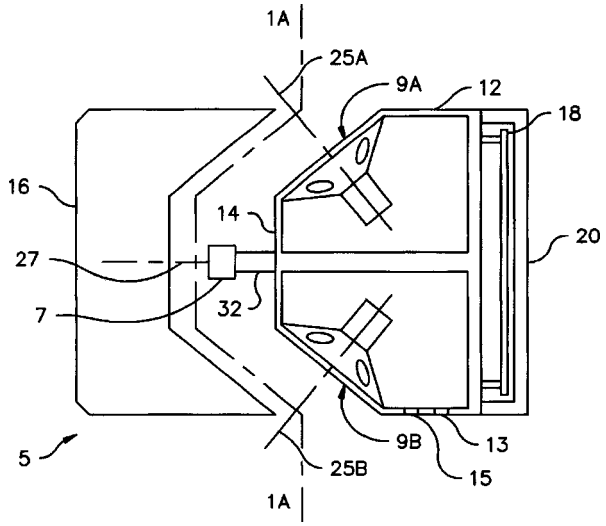
relative time delay and this information is used to compute the bearing line. Using two or more such arrays, the intersection of bearing lines locates the sound source in three dimensions. The patent also includes a simplified weighting function which improves target accuracy in the presence of noise.—GLA

6,275,580

43.38.Si TELECONFERENCING DEVICE HAVING ACOUSTIC TRANSDUCERS POSITIONED TO IMPROVE ACOUSTIC ECHO RETURN LOSS

Philip Faraci and Philip Lang, assignors to Tellabs Operations, Incorporated
14 August 2001 (Class 379/388); filed 7 July 1998

A device that minimizes acoustic coupling between the microphone and loudspeakers of a full-duplex teleconferencing system is described. A



directional microphone **7** is placed at the midpoint between two oppositely phased loudspeakers **9A** and **9B** so that the sound from the speakers destructively cancels at the microphone.—IMH

6,275,836

43.38.Si INTERPOLATION FILTER AND METHOD FOR SWITCHING BETWEEN INTEGER AND FRACTIONAL INTERPOLATION RATES

Jinghui Lu, assignor to Oak Technology, Incorporated
14 August 2001 (Class 708/313); filed 12 June 1998

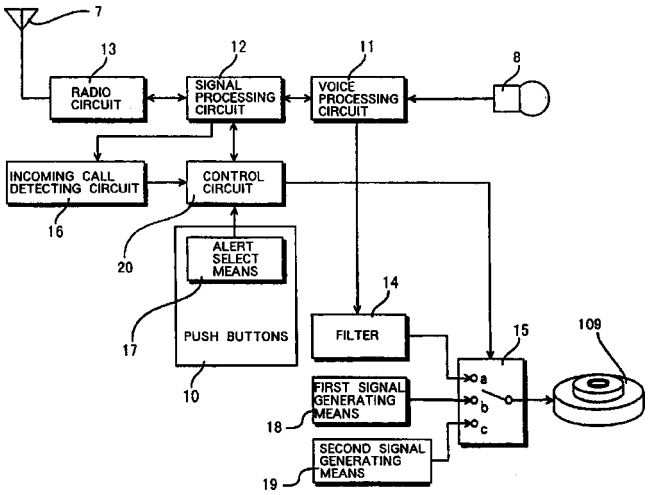
This telecommunications device consists of a variable sample rate interpolation filter coupled to a digital-to-analog converter. The filter accommodates differing input signals by varying the input sample rate as required or by switching between integer and fractional interpolation rates on-the-fly to produce a fixed oversampling output rate. The device is said to be able to accept a signal at virtually any digital signal frequency up to one-half of the analog conversion sample rate.—DRR

6,281,785

43.38.Si VIBRATION GENERATOR FOR NOTIFICATION AND PORTABLE COMMUNICATION DEVICE USING THE VIBRATION GENERATOR

Toshihide Hamaguchi, assignor to Sanyo Electric Company, Limited
28 August 2001 (Class 340/407.1); filed in Japan 21 March 1997

This device is a ''buzzer'' to be used in portable communication devices such as cell telephones and pagers. A vibration generator has a magnetic circuit composed of a permanent magnet, a coil, and a diaphragm



secured to fixed member **109** so as to mechanically resonate at a specific frequency. A filter which blocks that frequency may be switched in or out. When switched out, the buzzer operates to notify the user. When the filter is in the circuit, an incoming speech signal is applied to the voice coil.—DRR

6,229,899

43.38.Vk METHOD AND DEVICE FOR DEVELOPING A VIRTUAL SPEAKER DISTANT FROM THE SOUND SOURCE

Elwood G. Norris and James J. CroftIII, assignors to American Technology Corporation
8 May 2001 (Class 381/77); filed 17 July 1996

For more than 30 years inventors have experimented with the idea of generating a virtual sound source at the intersection of two ultrasonic beams, one of them modulated by an audio signal. More recently, parametric loudspeakers have been demonstrated. These use a single ultrasonic transducer whose beam is demodulated throughout a relatively large volume of air. This patent suggests that omnidirectional surround sound sources can be produced by bouncing parametric beams off reflective room surfaces. These are synchronized with sound from conventional, primary loudspeakers.—GLA

6,236,730

43.38.Vk FULL SOUND ENHANCEMENT USING MULTI-INPUT SOUND SIGNALS

Brian Cowieson *et al.*, assignors to QSound Labs, Incorporated
22 May 2001 (Class 381/18); filed 19 May 1997

Previously patented QSound circuitry can be used to create a virtual sound source at almost any location relative to two stereo loudspeakers. If one accepts this premise, then it follows that the multiple channels of a 5:1 surround sound system can be QSound-encoded into two channels and recreated through two stereo loudspeakers. The patent describes a number of ways in which this basic idea might be implemented.—GLA

6,239,348

43.38.Vk SOUND SYSTEM AND METHOD FOR CREATING A SOUND EVENT BASED ON A MODELED SOUND FIELD

Randall B. Metcalf, Cantonment, Florida
29 May 2001 (Class 84/723); filed 10 September 1999

During the past quarter-century a great amount of research has been devoted to the problem of effectively recording and reproducing a three-dimensional sound field. Numerous technical papers have been published and a number of patents have been issued. That body of work is largely ignored in this patent. It describes what seems to be a pure thought experiment, the novelty of which is difficult to see. Yes, a sound field can theoretically be ''modeled'' by sensing pressure and directional components at a great many points on an imaginary spherical surface and, yes, this process can be reversed to recreate the field.—GLA

6,243,476

43.38.Vk METHOD AND APPARATUS FOR PRODUCING BINAURAL AUDIO FOR A MOVING LISTENER

William G. Gardner, assignor to Massachusetts Institute of Technology
5 June 2001 (Class 381/303); filed 18 June 1997

The invention is one of many concerned with the problem of generating a realistic three-dimensional sound field from two loudspeakers. According to the patent. ''The present invention extends the concept of three-dimensional audio to a moving listener, allowing, in particular, for all types of head motions (including lateral and frontback motions and head rotations). This is accomplished by tracking head position and incorporating this parameter into an enhanced model of binaural synthesis.''—GLA

6,281,813

43.38.Wl CIRCUIT FOR DECODING AN ANALOG AUDIO SIGNAL

Matthias Vierthaler *et al.*, assignors to Micronas GmbH
28 August 2001 (Class 341/50); filed 9 July 1999

The patent describes a circuit for decoding a Sound Intercarrier Frequency (SIF) signal using the Broadcast Television System Committee (BTSC) standard that is prevalent in the United States for transmission of stereo sound in television programming. An integrated circuit includes a tuner that receives the analog audio signal and a decoder which provides the SIF signal. A digital demodulator digitizes the SIF signal and digitally demodulates it to yield a digitized multichannel television sound (MTS) demodulated signal. A BTSC-compatible decoder decodes the demodulated MTS signal and provides sum (L+R) and difference (L−R) audio output signals.—DRR

6,229,762

43.38.Zp ACOUSTIC SENSOR FOR A POINT IN SPACE

Stanley A. Fisher and Larry S. Chandler, assignors to the United States of America as represented by the Secretary of the Navy
8 May 2001 (Class 367/149); filed 26 August 1996

A fiber-optic sensing arrangement is mounted in an enclosure in the shape of a hemispherical shell or the like, which has an acoustic focus. The sensing arrangement consists of two optical cables, one of which is exposed to the acoustic field, and the other of which is shielded from that field so that it serves as a reference. Coherent light is injected into the two cables and the phase difference of the emerging light is used to detect acoustic signals at the enclosure's focus. The sensor is claimed to be useful for biological or medical applications.—EEU

6,234,021

43.40.Le ENHANCED DETECTION OF VIBRATION

Kenneth R. Piety *et al.*, assignors to CSI Technology, Incorporated
22 May 2001 (Class 73/592); filed 2 February 1999

The device described in this patent is intended to detect and monitor specific noise or vibration signals against a noisy background. Cited poten-

tial applications include the detection of leaks in pipes, of machinery defects, and of electrical arcing. The device in essence consists of a resonant sensor tuned to the desired frequency range, coupled to an electrical filter and automatic auto-ranging system. Potential sensors include microphones, ultrasonic sensors, and conventional accelerometers.—EEU

6,234,022

43.40.Le BEARING RIGIDITY EVALUATION APPARATUS

Hisakazu Tadokoro, assignor to NSK Limited
22 May 2001 (Class 73/593); filed in Japan 28 August 1998

The axial stiffness of a ball bearing is here determined from measurement of the transmissibility from one bearing race to the other. The bearing is mounted on a support by one of it races and that race is subjected axially to frequency-swept excitation via a shaker. The vibrations of both races are sensed and used to compute the corresponding transfer function. It is claimed that the bearing stiffness obtained by this means correlates well with data from static tests.—EEU

6,236,329

43.40.Le SLIDING VIBRATION WARNING SWITCH

Christos Kyrtsos, assignor to Meritor Heavy Vehicle Systems, LLC
22 May 2001 (Class 340/683); filed 23 March 1999

This patent pertains to a micro-machined sensor that provides an indication of excessive axial vibrations of a shaft. A small mass is held against a surface that is parallel to that of the shaft by centrifugal force and by a resilient element. Friction and an axially acting springlike element limit the mass's axial motion due to axial vibration. If the axial vibration exceeds a certain level, the axial excursion of the mass becomes large enough so that a protrusion on it enters an opening in the aforementioned surface and thus triggers a signal.—EEU

5,954,169

43.40.Tm ADAPTIVE TUNED VIBRATION ABSORBER, SYSTEM UTILIZING SAME AND METHOD OF CONTROLLING VIBRATION THEREWITH

William S. Jensen, assignor to Lord Corporation
21 September 1999 (Class 188/378); filed 24 October 1997

A dynamic vibration absorber is described consisting of a mass element suspended from, or otherwise supported by, one or more flexible plates. The resonant frequency of the absorber for either transverse or longitudinal motions of the mass element can be adjusted to match that of the vibrating system to which it is attached by either repositioning the mass to change the radius of gyration or statically stressing the flexible support plates to change the effective stiffness. The device incorporates suitable sensors, controllers, and actuators to monitor and effectuate these changes.—WT

6,223,672

43.40.Tm ULTRASHORT FAIRINGS FOR SUPPRESSING VORTEX-INDUCED-VIBRATION

Donald Wayne Allen and Dean Leroy Henning, assignors to Shell Oil Company
1 May 2001 (Class 114/243); filed 15 November 1996

The fairings described in this patent are intended for attachment to substantially cylindrical marine elements. A typical fairing here follows the element's circular profile for about 270 degrees or more, then departs from that profile and forms a trailing edge. The length of the fairing is between 1.1 and 1.2 times the diameter of the cylindrical element. Arrangements of such fairings with their trailing edges extending in various directions are claimed to be able to reduce substantially the vortex-induced vibrations resulting from flows impinging on a marine element from any direction.—EEU

6,223,601

43.40.Yq VIBRATION WAVE DETECTING METHOD AND VIBRATION WAVE DETECTOR

Muneo Harada and Naoki Ikeuchi, assignors to Sumitomo Metal Industries, Limited
1 May 2001 (Class 73/649); filed in Japan 22 May 1998

The vibration detector described in this patent is of the resonator array type. It consists of a number of cantilever beam elements, each having a different length and thus a different natural frequency, attached to a base element. A piezoresistor is installed on each beam element and these resistors are connected in parallel so as to provide an output corresponding to the sum of the vibrations of the beam elements. The resistors are placed at locations along the lengths of the beam elements so that the output signal from each element is set at a desired level.—EEU

6,276,209

43.40.Yq SYSTEM AND METHOD OF ASSESSING THE STRUCTURAL PROPERTIES OF WOODEN MEMBERS USING ULTRASOUND

Mark E. Schafer *et al.*, assignors to Perceptron, Incorporated; the United States of America as represented by the Secretary of Agriculture
21 August 2001 (Class 73/597); filed 30 September 1999

An ultrasonic transmitter is used to transmit longitudinal and/or shear waves in wood. A corresponding receiver analyzes the received waveforms and makes a determination of defects or grain characteristics in the wood. The transducers may be placed in rollers for stock processing capability. Various configurations, including through-transmission and multi-point sensing, are also presented.—IMH

6,224,020

43.50.Gf PAYLOAD FAIRING WITH IMPROVED ACOUSTIC SUPPRESSION

Bruce D. Hopkins *et al.*, assignors to Alliant Techsystems, Incorporated
1 May 2001 (Class 244/158 R); filed 24 December 1998

This patent pertains to fairings that enclose payloads, such as satellites, mounted atop a rocket launch vehicle, typically primarily to protect the payloads from aerodynamic forces. The fairings described in this patent are intended also to provide protection from noise during lift-off and flight. These fairings in essence consist of a honeycomb core, generally of aluminum, to which fiber-resin face sheets are bonded on both sides. The stiffness of this configuration is intended to provide noise reduction at low frequencies, and the damping of the face sheets is claimed to improve noise reduction at higher frequencies.—EEU

6,243,671

43.58.KR DEVICE AND METHOD FOR ANALYSIS AND FILTRATION OF SOUND

Thomas Lagö, 564 34 Bankeryd, and Sven Olsson, 226 51 Lund, both of Sweden
5 June 2001 (Class 704/200.1); filed in Sweden 3 July 1996

The invention might be used in speech recognition systems or acoustical analysis of mechanical systems. Two banks of parallel filters are used. One might be a conventional $\frac{1}{3}$-octave analyzer, while the other is made up of constant-bandwidth filters. The two sets of outputs are compared using weighting factors that may have been derived during a learning process. The system is said to provide more accurate analysis than prior art by emulating the human hearing mechanism.—GLA

6,278,366

43.60.Bf ADAPTIVE SOUND ACTUATED ILLUMINATION DEVICE FOR BATTERY OPERATION

Mark W. Fletcher *et al.*, Odessa, Florida
21 August 2001 (Class 340/542); filed 22 December 1999

This rather simple battery-powered device is intended to be mounted onto existing door lock hardware. It incorporates a sound detector that turns on a light and an optical sensor to detect any ambient light. An integrated circuit receives the inputs of the microphone and the optical sensor and turns on the light if sufficient ambient light is not detected. A timer automatically turns off the light after a predetermined period of time.—DRR

6,227,036

43.60.Gk MULTIPLE MICROPHONE PHOTOACOUSTIC LEAK DETECTION AND LOCALIZATION SYSTEM AND METHOD

Serdar H. Yonak and David R. Dowling, assignors to The Regents of the University of Michigan
8 May 2001 (Class 73/40.5A); filed 28 October 1998

Sulfur hexafloride gas is introduced into the article to be tested. A carbon dioxide laser directed at the article's exterior surface causes sound pulses to be generated where the laser interacts with leaked gas. These pulses are recorded across a broad bandwidth by multiple ultrasonic microphones and the signals are compared to the background noise. If leak-related pulses are detected, the recorded sound is processed using matched-field processing; the sounds recorded at the microphones are reversed in time by computer simulation to determine their origins.—EEU

6,240,051

43.60.Gk ACOUSTIC SURVEILLANCE APPARATUS AND METHOD

Carey D. Bunks, assignor to GTE Service Corporation
29 May 2001 (Class 367/127); filed 4 September 1998

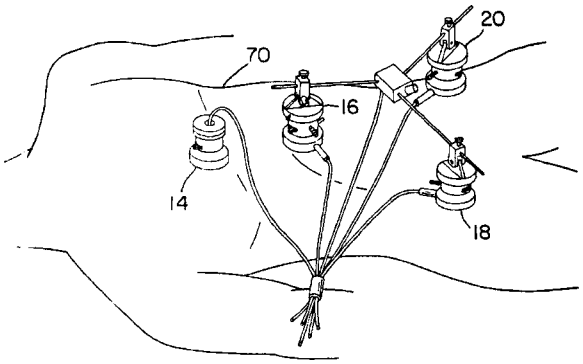
Suppose that a hostage is held in a large, open warehouse whose windows have been painted black. One of the bad guys fires a warning shot. Resulting vibrations in the glazing of several windows are detected by laser interferometers or other suitable sensors. Relative time delays between the individual signals are then used to calculate the location of the pistol at the time the shot was fired. For this particular scenario, it seems obvious that the method could be made to work. In real life, the acoustic excitation might be voices or footsteps in any one of several rooms behind a gypsum board partition. In this instance, flanking paths, reflections, structural vibrations, and bending waves in the partition all will contribute to the vibration of the gypsum board surface at any particular point. Geometrical reconstruction would seem to be much more difficult than the patent suggests.—GLA

6,287,266

43.60.Qv METHOD AND APPARATUS FOR CHARACTERIZING GASTROINTESTINAL SOUNDS

Richard H. Sandler and Hussein A. Mansy, assignors to Rush-Presbyterian-St. Lukes Medical Center
11 September 2001 (Class 600/593); filed 3 April 1996

A method and apparatus for characterizing gastrointestinal sounds includes a microphone array to be positioned on a body for picking up gastrointestinal sound signals. The signals are digitized, and a processor estab-



lishes their spectra and duration. A characterization of the state of the gastrointestinal tract is made on the basis of the spectra and duration of the sound or event.—DRR

6,259,951

43.66.Ts IMPLANTABLE COCHLEAR STIMULATOR SYSTEM INCORPORATING COMBINATION ELECTRODE/TRANSDUCER

Janusz A. Kuzma and William Vanbrooks Harrison, assignors to Advanced Bionics Corporation
10 July 2001 (Class 607/57); filed 14 May 1999

A hybrid cochlear implanted prothesis utilizes both the acoustic modulation of the fluid within the inner ear for stimulating low-to-mid frequencies as well as electrical stimulation of ganglion cells in the basal end of the cochlea for the high frequencies. An acoustic modulator is included in the electrode/transducer array. The system may also include an implantable microphone, speech processor, acoustic transducer, and coil. Electrical stimulation is accomplished by inserting a short electrode array into the basal region of the cochlea. In one embodiment, the acoustic modulator is used as a ''cochlear microphone'' sensor to pick up fluid modulation within the cochlea when the middle ear is functioning.—DAP

6,264,603

43.66.Ts MIDDLE EAR VIBRATION SENSOR USING MULTIPLE TRANSDUCERS

Joel A. Kennedy, assignor to St. Croix Medical, Incorporated
24 July 2001 (Class 600/25); filed 7 August 1997

A middle ear implant for hearing assistance is described that uses the vibrations of the malleus bone as an input signal sensor. Since the direction of motion of the malleus vibrations may vary between individuals, the system uses up to three independent transducers for sensing the vibration of the malleus in several directions. Output signals from the transducers are combined as a sum of squares of the transduced signals or a sum of filtered transduced signals.—DAP

6,227,054

43.72.Ar VIBRATION WAVE DETECTING METHOD AND VIBRATION WAVE DETECTOR

Shigeru Ando *et al.*, assignors to Sumitomo Metal Industries Limited
8 May 2001 (Class 73/651); filed in Japan 26 May 1997

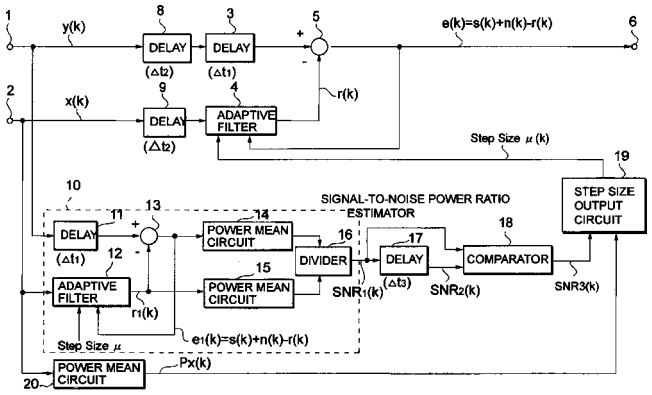
An array of cantilever resonators, each tuned to a particular predetermined frequency, is attached to a support rod. A detector associated with each resonator is used to determine the intensity of its response to sound. The system is designed to provide responses in the mel or Bark scales, to facilitate the efficient extraction of speech features.—EEU

6,266,422

43.72.Dv NOISE CANCELING METHOD AND APPARATUS FOR THE SAME

Shigeji Ikeda, assignor to NEC Corporation
24 July 2001 (Class 381/71.11); filed in Japan 29 January 1997

This patent describes a method of using an adaptive filter to cancel a background noise signal introduced into a speech signal via a microphone, a handset, or the like. The speech signal with background noise is applied to terminal **1**. A second signal from a microphone remote from the talker provides a reference signal **2**. The adaptive filter **4** filters the noise signal and



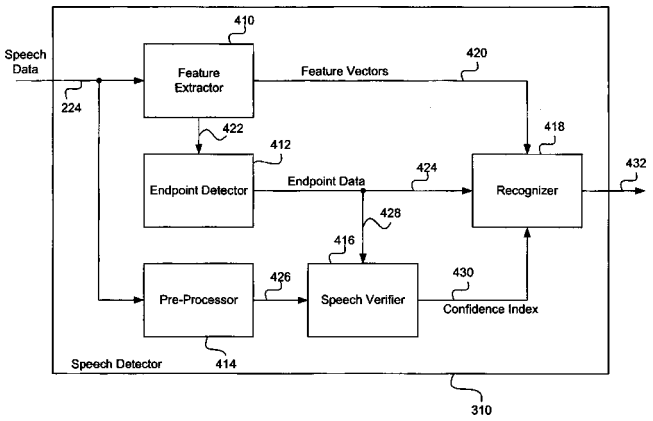
generates the pseudo noise signal, which is fed to the subtracter **5**. The coefficients of the filter are updated by use of the least mean squares algorithm. The filter **12** produces a second pseudo-noise signal and an error signal. A divider **16** calculates the signal-to-noise (SNR) power ratio. A step size circuit **19** corrects, based on the extended SNR power ratio **18** and from **20**, a step size **19** used to adaptively tune the coefficient of filter **4**.—HHN

6,272,460

43.72.Dv METHOD FOR IMPLEMENTING A SPEECH VERIFICATION SYSTEM FOR USE IN A NOISY ENVIRONMENT

Duanpei Wu *et al.*, assignors to Sony Corporation; Sony Electronics, Incorporated
7 August 2001 (Class 704/226); filed 10 September 1998

This patent describes a speech verification system for use in a noisy environment. The system includes a noise suppressor, a pitch detector, and a confidence determiner. The noise suppressor suppresses noise by summing a frequency spectrum for each frame of the speech signal. The pitch detector



calculates the correlation for each frame in the utterance. The confidence determiner evaluates the correlation values and uses the confidence measures to generate a confidence index indicating whether the utterance is or is not speech. This system also includes a speech detector unit as shown in the figure.—HHN

6,263,311

43.72.Fx METHOD AND SYSTEM FOR PROVIDING SECURITY USING VOICE RECOGNITION

Robert Dildy, assignor to Advanced Micro Devices, Incorporated
17 July 2001 (Class 704/273); filed 11 January 1999

This building security system listens for unauthorized voices. Presumably, you would have to say ''Good morning'' to the building when you arrive for work. Any unrecognized speaker produces one alarm level and the voices of particular unwanted people would be detected, producing a second alarm level. Certain keywords also allow authorized users to trigger other alarm levels or to dial 911.—DLR

6,260,017

43.72.Gy MULTIPULSE INTERPOLATIVE CODING OF TRANSITION SPEECH FRAMES

Amitava Das and Sharath Manjunath, assignors to Qualcomm, Incorporated
10 July 2001 (Class 704/265); filed 7 May 1999

One effective technique to encode speech efficiently at low bit rates is multimode coding. Conventional multimode coders apply different modes, or encoding–decoding algorithms, to different types of input speech frames such as voiced, unvoiced, and silence. This patent pertains to multipulse interpolative coding of transition speech frames between voiced and unvoiced segments. In various embodiments, the rate of data transmission is varied on a frame-to-frame basis from 13.2 kbs for voiced signals, 6.2 kbs for transition signals, 2.6 kbs for unvoiced signals, to 1 kbs for silence.—AAD

6,260,016

43.72.Ja SPEECH SYNTHESIS EMPLOYING PROSODY TEMPLATES

Frode Holm and Kazue Hata, assignors to Matsushita Electric Industrial Company, Limited
10 July 2001 (Class 704/260); filed 25 November 1998

In the quest for better text-to-speech quality, this synthesis system uses a word stress marking arrangement which governs the generation of prosodic structure. A phonetic spelling dictionary includes syllable boundaries and assigns a word stress level (none, primary, or secondary) to each syllable. A library of word prosody patterns is accessed by number of syllables

and the syllable stress values. The current version stores only a pitch pattern for the word, but other embodiments could include other information, such as amplitude and duration controls.—DLR

6,262,669

43.72.Ja VOICE DIAGNOSTIC APPARATUS FOR MAINBOARD

Sophie Leu, assignor to Hsin-Chan Lu
17 July 2001 (Class 340/635); filed 7 September 1999

You've heard of talking computers. This one gripes about its aches and pains. Sensors and other detectors on the motherboard detect failures, such as overheating or boot up failures. Built-in speech synthesis circuitry announces the problem using the computer's small loudspeaker.—DLR

6,266,637

43.72.Ja PHRASE SPLICING AND VARIABLE SUBSTITUTION USING A TRAINABLE SPEECH SYNTHESIZER

Robert E. Donovan *et al.*, assignors to International Business Machines Corporation
24 July 2001 (Class 704/258); filed 11 September 1998

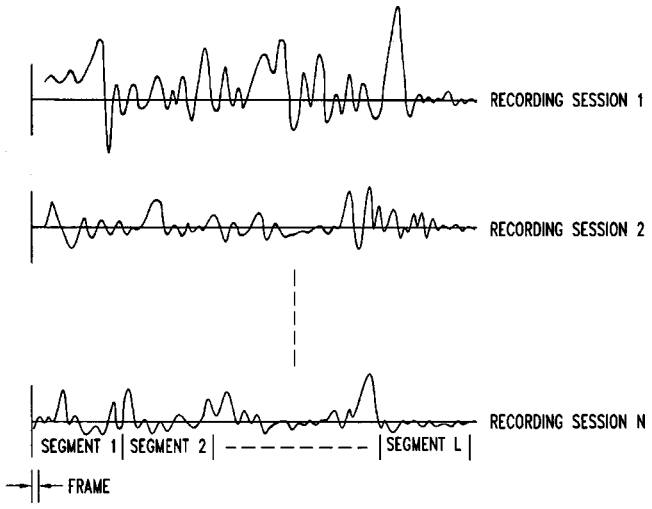
This patent describes another in a long line of techniques for combining prerecorded speech playback with true synthesis to get better quality speech output. Here, new vocabulary may be entered into the system as spoken words or phrases. The speech input is processed so as to recover prosody as well as phonetic structure. The new material is used to add pronunciation details to existing system dictionary entries or to create new entries, as needed. Synthesis may be guided by a system of hidden Markov models (although the Markov states are not really hidden, are they?) A system known as splice files provides additional information for better synthesis of particular word sequences.—DLR

6,266,638

43.72.Ja VOICE QUALITY COMPENSATION SYSTEM FOR SPEECH SYNTHESIS BASED ON UNIT-SELECTION SPEECH DATABASE

Ioannis G. Stylianou, assignor to AT&T Corporation
24 July 2001 (Class 704/266); filed 30 March 1999

In the design of this speech synthesis system, the decision was made to use as much prerecorded speech as possible in order to get the highest quality. But then the designers had to face the problem of nonuniformity in the pronunciations. Frames of the recorded speech are transformed into a typical form, such as cepstral coefficients. Gaussian mixture models are then used to choose one of multiple recordings of each word or phrase which is closest to having predetermined "preferred" voice characteristics. Synthesis segments are then prepared from the chosen recording.—DLR

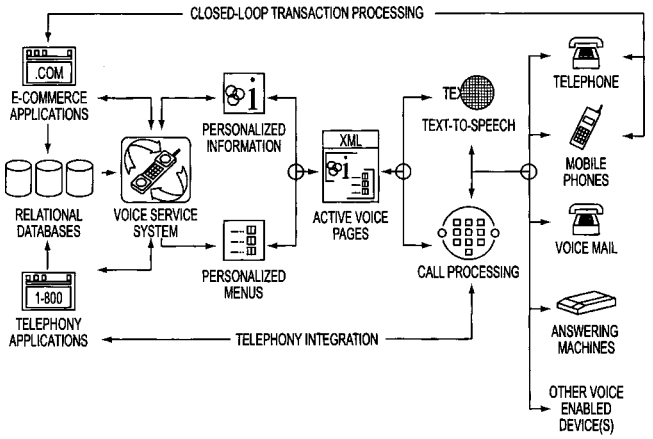


6,263,051

43.72.Kb SYSTEM AND METHOD FOR VOICE SERVICE BUREAU

Michael J. Saylor *et al.*, assignors to Microstrategy, Incorporated
17 July 2001 (Class 379/88.17); filed 13 September 1999

Primarily intended for small companies that cannot afford their own interactive telephone interface system, this voice interactive server provides speech services for either the telephone system or the Internet. Several speech dialog description languages are available, mostly based on the XML language, allowing a "voice webmaster" to define a user protocol. Menu and voice broadcast services are described.—DLR



6,266,617

43.72.Kb METHOD AND APPARATUS FOR AN AUTOMATIC VEHICLE LOCATION, COLLISION NOTIFICATION AND SYNTHETIC VOICE

Wayne W. Evans, Alpharetta, Georgia
24 July 2001 (Class 701/301); filed 10 June 1999

This vehicle voice module works together with a global positioning system (GPS) receiver and can notify a fixed station, such as a dispatcher, of the vehicle's movements. The system comes into its own, however, in the event of a collision. It then works like an on-line "black box," relaying vehicle location, condition, and precollision data, such as routes, speeds, etc., to the base station. The vehicle operator may or may not be notified of such communications.—DLR

6,266,642

43.72.Ne METHOD AND PORTABLE APPARATUS FOR PERFORMING SPOKEN LANGUAGE TRANSLATION

Alexander M. Franz and Keiko Horiguchi, assignors to Sony Corporation; Sony Electronics, Incorporated
24 July 2001 (Class 704/277); filed 29 January 1999

This spoken language translator includes a speech recognizer, a syntactic analyzer, a set of rules to transform the syntactic structure to the target language, a target lexicon with inflection structure, and an output synthe-

sizer. The recognition system uses syntactic and semantic consistency checks to select among multiple hypotheses and also requests a final confirmation from the user before starting the translation process.—DLR

6,269,334

43.72.Ne NONGAUSSIAN DENSITY ESTIMATION FOR THE CLASSIFICATION OF ACOUSTIC FEATURE VECTORS IN SPEECH RECOGNITION

Sankar Basu and Charles A. Micchelli, assignors to International Business Machines Corporation
31 July 2001 (Class 704/256); filed 25 June 1998

This patent introduces a new mixture of non-Gaussian statistical probability densities for modeling speech subword units, such as phonemes in a hidden Markov model. The mixture employs a parameter by which different types of mixture components are constructed. The functions are then used for maximum likelihood modeling of speech data. According to the patent, a mixture constructed by this method shows improved recognition accuracy.—AAD

6,277,084

43.80.Ev ULTRASONIC MEDICAL DEVICE

John E. Abele and Kevin R. Heath, assignors to Boston Scientific Corporation
21 August 2001 (Class 601/2); filed 31 March 1992

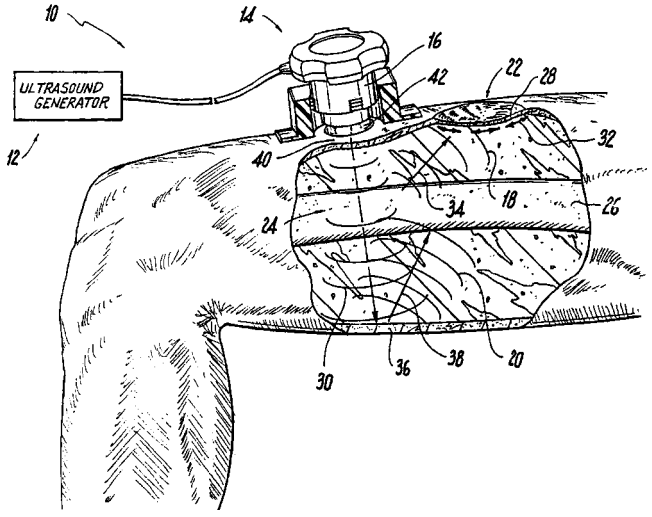
The patent emphasizes the use of special wires in a medical device for conducting ultrasound energy. These wires supposedly achieve optimum properties by means of a multiple coaxial construction. For example, in a particular embodiment it is desirable to have an elastic core of nitonol to conduct sonic or ultrasonic vibrations axially and a thin stiff cladding of stainless steel in order to minimize transverse vibrations that result in loss of energy.—DRR

6,273,864

43.80.Gx ULTRASONIC TREATMENT FOR WOUNDS

Luiz R. Duarte and Roger J. Talish, assignors to Exogen, Incorporated
14 August 2001 (Class 601/2); filed 14 February 1997

It is generally recognized that longitudinally propagating ultrasound provides effective healing of a wound. As an update to United States Patents 4,530,360 by Duarte and 5,003,965 and 5,211,150, both by Talish and Lifshey, this patent by two of the same authors concentrates on a portable device in which a transducer is deployed adjacent to the wound to promote healing. The apparatus deals with reflections of the ultrasound by bone tissue, skin layers, or by internally dispersed reflective media. A focusing element is used to direct the propagation of the sound at a predetermined angle toward the wound. The operating surface of the transducer may be annularly shaped and placed around the wound to better direct the energy. An optional housing and adjustable straps allow the transducer to be positioned near the wound, indenting the skin to form a cavity for the transducer, emitting the ultrasound toward an internal surface of the wound.—DRR

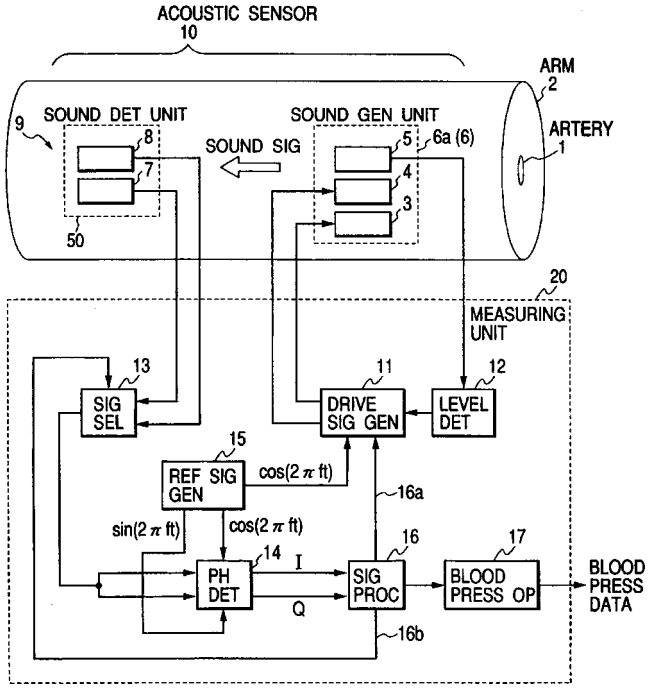


6,275,447

43.80.Gx SOUND GENERATING APPARATUS, A SOUND DETECTION APPARATUS, AN ACOUSTIC SENSOR, AND AN ACOUSTIC LIVING BODY MEASURING APPARATUS

Hiroshi Fukukita and Hisashi Hagiwara, assignors to Matsushita Electric Industrial Company, Limited
14 August 2001 (Class 367/137); filed in Japan 18 February 1998

The apparatus is a combined sound generating and sound detection device for parametric measurement of body tissues. The device is said to include ''a flexible substrate and a plurality of bimorphs arranged on a surface of the flexible substrate in a direction at a predetermined interval, each of the bimorphs having input/output terminals.'' One wonders exactly what materials the ''bimorph cells'' are made of. Each of the bimorphs is described as having ''a rectangular shape and its longitudinal direction is perpendicular to the direction,'' presumably the direction of sound propagation. The flexible substrate contains a slit between two adjacent bimorphs. Several alternate arrangements are described. An acoustic sensor includes the sound generating gear and the sound detection component connected to it by a flexible substrate, which may include a cable and a connector. In operation, a phase shift is detected between the sound signal induced by the bimorph and a reference drive signal. The amount of phase shift indicates the blood pressure. In summary, this is not a well-written patent.—DRR

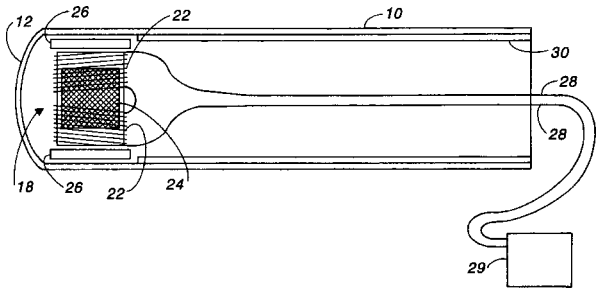


6,277,085

43.80.Gx SOLENOIDAL SKIN VIBRATOR ENERGIZED BY COMPLEX ELECTRIC WAVEFORMS

Michael P. Flynn, Ann Arbor, Michigan
21 August 2001 (Class 601/80); filed 9 June 1998

This device converts simple or complex electrical waveforms into corresponding simple or complex mechanical waveforms or vibrations that are applied to the skin and underlying tissues for therapy or pleasure. The complex waveforms may be completely random using a random electrical noise generator or nonrandom but complex using music or speech. The vibrator



consists of an oblong hollow body having a solenoid motor mounted inside. The solenoid is partially damped at both extremes of its stroke to mitigate the impact within the vibrator and thereby soften the vibration. Provisions are made to connect the vibrator to a signal generator, audio amplifier, or other source of electrical waveforms.—DRR

6,285,630

43.80.Nd AUTO-CONTROL BIRD-EXPELLING DEVICE

Te-Chin Jan, Taipei 110, Taiwan, Province of China
4 September 2001 (Class 367/139); filed in Taiwan, Province of China, 27 March 2000

The device covered by this patent is basically an electronic scarecrow that uses a PIR motion detector to detect specific types of birds. The detector senses heat to determine any intrusion within a specified region and uses a specialized signal generator to trigger a ''bird-expelling'' signal. The signal could be, for example, the sound of a bird of prey such as a hawk, an eagle, or a falcon. The sound of the bird of prey (i.e., a raptor) is stored in an audio-data memory and fed into an amplifier and speaker.—DRR

6,274,963

43.80.Sh METHODS AND DEVICES FOR CONTROLLING THE VIBRATION OF ULTRASONIC TRANSMISSION COMPONENTS

Brian Estabrook *et al.*, assignors to Ethicon Endo-Surgery, Incorporated
14 August 2001 (Class 310/316.02); filed 28 April 1997

A phase-locked loop and frequency control loop are used to control unwanted vibrations in an ultrasonic device, such as a handheld therapy wand applicator. Zero current crossings are detected and harmonics other than the desired fundamental are filtered out by applying the proper driving signals to sections of a multi-section transducer coupled to the wand.—IMH

Individual acoustic variation in Belding's ground squirrel alarm chirps in the High Sierra Nevada (L)

Brenda McCowan and Stacie L. Hooper
Behavioral Biology Laboratory, Veterinary Medicine Teaching and Research Center, School of Veterinary Medicine, University of California at Davis, 18830 Road 112, Tulare, California 93274

(Received 12 March 2001; revised 19 November 2001; accepted 27 November 2001)

The acoustic structure of calls within call types can vary as function of individual identity, sex, and social group membership and is important in kin and social group recognition. Belding's ground squirrels (*Spermophilus beldingi*) produce alarm chirps that function in predator avoidance but little is known about the acoustic variability of these alarm chirps. The purpose of this preliminary study was to analyze the acoustic structure of alarm chirps with respect to individual differences (e.g., signature information) from eight Belding's ground squirrels from four different lakes in the High Sierra Nevada. Results demonstrate that alarm chirps are individually distinctive, and that acoustic similarity among individuals may correspond to genetic similarity and thus dispersal patterns in this species. These data suggest, on a preliminary basis, that the acoustic structure of calls might be used as a bioacoustic tool for tracking individuals, dispersal, and other population dynamics in Belding's ground squirrels, and perhaps other vocal species. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1446048]

PACS numbers: 43.80.Ev [WA]

I. INTRODUCTION

Acoustic variation and distinctiveness is found in the vocal communication systems of all nonhuman animal species. The acoustic structure of calls within call types can vary as functions of individual identity, sex, and social group membership and is important in kin and social group recognition. Many species exhibit acoustic distinctiveness in various temporal and frequency parameters that likely provides the acoustic basis for individual or kin recognition (Chapman and Weary, 1990; Reby *et al.*, 1998a). In addition, dialects have been documented in many avian species (Adret, 1986; Sorjonen, 1987) and some cetacean species (pilot whales: Taruski, 1976, 1979; orcas: Ford and Fisher, 1983; Ford, 1991; bottlenose dolphins: McCowan *et al.*, 1998). Recent studies suggest that some primate species exhibit population and social group differences in call structure (chimpanzees: Mitani *et al.*, 1992; Mitani and Brandt, 1994; Marshall *et al.*, 1999; marmosets: Elowson and Snowdon, 1994) indicating a possible function for social recognition and/or cohesion. Several studies on avian species have reported dialectal differences in the song or calls of different populations, which are frequently consistent with differences in geographical location (Wright, 1996; Nelson, 1998).

Belding's ground squirrels (*Spermophilus beldingi*) are alpine-dwelling, social animals that live in colonies of related adult females and their dependent offspring (Armitage, 1981; Boellstorff and Owings, 1995). Sons appear to disperse from the natal burrow while daughters establish burrow systems adjacent to or overlapping that of their mother's. Adult males establish overlapping territories with those of the females before the breeding season and continue to defend these territories after the breeding season has ended.

Ground squirrels are vocal animals that respond to predators with alarm calls. These alarm calls have been variously named chatters, chats, whistles, squeals, chirps, and trills (Owings and Virginia, 1978; Owings and Leger, 1980; Leger *et al.*, 1984; Owings *et al.*, 1986). Whistles and chirps are harmonically structured narrow-band calls that have a relatively low fundamental frequency (given the body size of members of this species) between 2.8 and 5.1 kHz. Although the contexts and functions of ground squirrel alarm calls are relatively well known, the extent of acoustic variation in ground squirrel alarm vocalizations and whether such variation provides a basis for individual, kin, or group recognition remains mostly unexplored (Hare, 1999). The functions of chirps and whistles in predator and territory defense suggest that information on individual and social group identity might be present in these calls. Thus, the purpose of this study was to determine if the alarm chirps (whistles) of Belding's ground squirrels contained individual signature information that could provide the basis for individual and kin recognition. To address this goal, this preliminary study examined the alarm chirps of eight free-ranging adult female ground squirrels from four alpine lakes in the French Canyon of the High Central Sierra Nevada for individual differences and in relationship to the geographical locations of these ground squirrel populations.

II. METHODS

A. Subjects and study site

Subjects were eight free-ranging adult female ground squirrels. All subjects were identified as adults, determined by body size, and as females due to the proximity of dependent offspring. Subjects were recorded from four different alpine lakes in the French Canyon region of the High Sierra Nevada, approximately 12 miles west of Pine Creek Pack Station near Bishop, CA in the Owens valley. The lakes on which recordings of the eight different ground squirrels were conducted included Moon Lake ($n=3$), Elba Lake ($n=3$),

TABLE I. List of analyzed acoustic variables and their definitions.

Acoustic parameter	Description
Coefficient of frequency modulation McCowan and Reiss (1995) McCowan *et al.* (1998)	Calculated variable that represents the amount and magnitude of frequency modulation across a chirp, computed by summing the absolute values of the difference between sequential frequencies divided by 10 000.
Jitter factor Mitani and Brandt (1994)	Calculated variable that represents a weighted measure of the amount of frequency modulation, by calculating the sum of the absolute value of the difference between two sequential frequencies divided by the mean frequency. The sum result is then divided by the total number of points measured minus 1 and the final value is obtained by multiplying it by 100.
Frequency variability index Mitani and Brandt (1994)	Calculated variable that represents the magnitude of frequency modulation across a chirp, computed by dividing the variance in frequency by the square of the average frequency of a chirp and then multiplying the value by 10.
Minimum frequency	Lowest frequency attained by chirp, measured in Hz
Maximum frequency	Highest frequency attained by chirp, measured in Hz
Mean frequency	Calculated as a average frequency across chirp
Frequency range	Calculated as maximum frequency minus minimum frequency
Maximum frequency/Mean frequency	Calculated as maximum frequency divided by mean frequency
Mean frequency/Minimum frequency	Calculated as mean frequency divided by minimum frequency
Frequency at peak amplitude	Frequency at which the peak amplitude occurs in the chirp
Minimum frequency location	Location of minimum frequency in chirp, given as percentage of duration
Maximum frequency location	Location of maximum frequency in chirp, given as percentage of duration
Duration	Temporal distance of chirp, measured in ms
Location of peak amplitude	Location at which the peak amplitude occurs in the chirp, given as a % of duration
Start slope	Slope of the initial third of the chirp contour
Middle slope	Slope of the middle third of the chirp contour
Finish slope	Slope of the final third of the chirp contour

Alsace Lake ($n=1$), and L Lake ($n=1$). Each lake exhibited rocky montane terrain with intermittent evergreen trees, grass, and brush. Mountain passes, ranging from 0.7 to 1.8 km in distance and 152–305 m in altitude (which ranged from 3048 to 3505 m), separated each of the lakes.

B. Vocal recordings

One to three adult female ground squirrels from each of the four lakes were acoustically recorded on one of four days using an Audio-Technica AT4071 directional microphone (frequency response to 20 kHz) and a Sony D-10 Pro DAT Recorder (frequency response to 24 kHz). For this preliminary study, each lake was visited only once and recordings were conducted from 1 to 3 h at each site. To ensure that we recorded different individuals at each of the lakes, recordings were conducted near the burrow systems of individuals located at distinct locations around each lake. Each vocalization was individually identified by observing the animals vocalizing during call production (recordings were conducted within 6 m of each subject). The context of alarm calling for each individual was to presence of human observers as Belding's ground squirrels from this remote region are not habituated to humans. A total of 358 vocalizations were collected from the eight adult female ground squirrels at four alpine lakes during the study period (Moon: 93, 105, 14 vocalizations from three individuals, respectively, Elba: 67, 43, 5 vocalizations from three individuals, respectively, Alsace: 12 vocalizations from one individual, L Lake: 19 vocalizations from one individual). With the exception of one individual (Elba 3), multiple calling bouts ($n>2$), which were defined by an intersignal duration of at least 1 min, were recorded from each individual subject.

C. Acoustic analyses

All vocal recordings were digitized onto a Micron Pentium Computer using a Sound-Blaster soundcard (sampling rate up to 44.1 kHz) and Cool Edit Pro Signal Analysis software (sampling rate of 44.1 kHz and using 1024-point FFT with a Hamming filter). Acoustic files were filtered for background noise using standard parametric filtering in Cool Edit Pro on the Micron computer and cued for subsequent digital analysis.

The chirps recorded from each individual were measured using a modified version of the Contour Similarity Technique (for detailed descriptions of this technique, see McCowan, 1995; McCowan and Reiss, 2001). After call digitization and measurement were completed, several subsequent calculations were conducted. Several summary acoustic variables defining various call spectral, temporal, amplitude, and contour parameters (e.g., minimum frequency, maximum frequency, mean frequency, frequency range, duration, frequency, and location of the peak amplitude) were calculated from these measurements (see Table I for a list of analyzed parameters).

D. Statistics

The outcomes and covariates of the statistical tests were continuous in structure. Thus, discriminant function analysis and fixed effects linear regression were the statistical methods of choice (Pinheiro and Bates, 2000). Continuous variables were tested for normality. Because most variables required transformation, principal component analysis was conducted on the raw variables. All final statistical tests were conducted on the principal component values, which were

TABLE II. Percent correct classification from the discriminant analysis on the alarm chirps of individual Belding ground squirrels[a] from four high Sierra lakes.

		No. of cases classified into group								Comparison to random assignment (as "expected")		
Individual	% correct	Alsace1	Elba1	Elba2	Elba3	Moon1	Moon2	Moon3	L1	Total *N*	Fisher's exact	*p*
Alsace1	83	10	0	0	1	0	0	1	0	12	13.9	<0.0001
Elba1	54	0	25	7	1	12	1	1	0	47	33.8	<0.0001
Elba2	51	5	7	32	3	5	8	3	0	63	32.9	<0.0001
Elba3	80	0	0	1	4	0	0	0	0	5	19.2	<0.0001
Moon1	45	0	15	9	1	42	25	1	0	93	57.2	<0.0001
Moon2	64	0	12	18	0	8	67	0	0	105	90.8	<0.0001
Moon3	79	0	1	0	1	1	0	11	0	14	14.7	<0.0001
L1	100	0	0	0	0	0	0	0	19	19	25.9	<0.0001
										358		

[a]All factors in the order F1, F2, F3, F4 were entered into the discriminant function.

tested for and confirmed normality. Covariates included individual, calling bout, and "time of day" of the recording. "Time of day" and "calling bout" were insignificant for all analyses and thus removed from the models. All statistical tests were conducted using programmable S-Plus statistical software on a Pentium or Pentium II PC.

III. RESULTS

A. Principal components of ground squirrel chirp vocalizations

Principal component analysis on the chirp calls of the eight adult female Belding's ground squirrels generated 16 statistically independent components. Eigenvalues of the first four components met Kaiser's criterion of 1.00. These four components accounted for 82% of the variation in the original data set. Factor 1 represents measures of spectral shape (e.g., coefficient of frequency modulation, maximum frequency location, and minimum frequency location) and amplitudinal emphasis (e.g., the frequency and location of peak amplitude), as well as mean frequency and minimum frequency of chirps. Factor 2 represents another measure of spectral shape including the frequency range, jitter factor, maximum frequency/mean frequency, and the start slope of chirps. Factor 3 represents the duration and the frequency variability index. Factor 4 represents the middle and final slopes as well as the mean frequency/minimum frequency of chirps.

B. Individual differences in ground squirrel chirps

Belding's ground squirrels showed individual distinctiveness in the acoustic structure of their alarm chirps. Cross-validation discriminant analysis revealed that each individual could be reliably distinguished based upon the acoustic structure of their calls (see Table II, Fig. 1). The variables

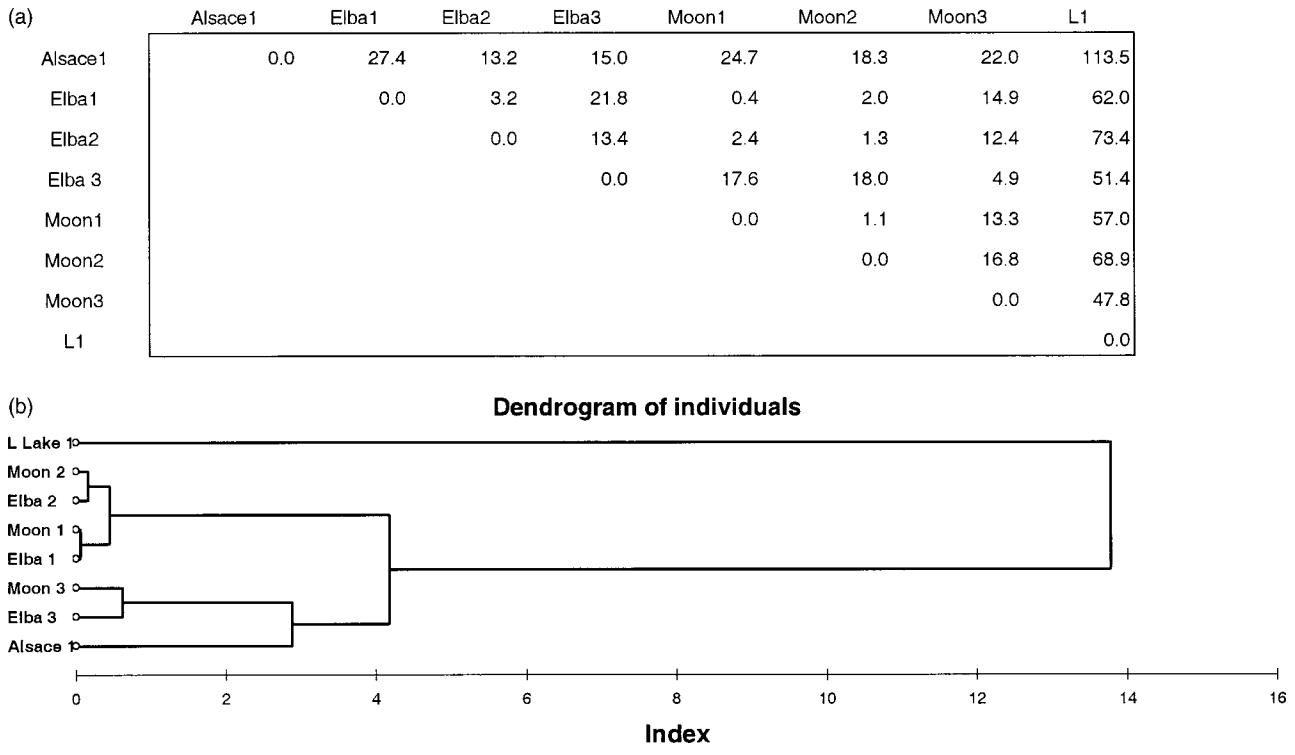


FIG. 1. (a) Mahalanobis distance generated from the discriminant function analysis between the alarm chirps of individual Belding's ground squirrels. (b) Dendrogram of Mahalanobis distance between the alarm chirps of individuals.

that contributed most to individual distinctiveness were factors 1 and 2, although all factors were entered into the discriminant function. Fixed effects linear regression confirmed the discriminant function analyses. Factor 2 significantly differed ($F_{7,350}=42.5$, $p<0.0001$) for each pairwise comparison at the 0.05 level using the Bonferroni method except Alsace1-Elba3, Alsace1-L1, Elba3-L1, Elba2-Moon2, Moon1-Moon2. Factor 1 significantly differed ($F_{7,350}=91.3$, $p<0.0001$) for three of the insignificant comparisons above: Alsace1-Elba3, Alsace1-L1, Elba3-L1 and factor 4 significantly differed ($F_{7,350}=30.2$, $p<0.0001$) for the remaining two insignificant comparisons: Elba2-Moon2, Moon1-Moon2.

In addition, a dendrogram generated from Mahalanobis distances from the discriminant analysis of individuals' alarm calls revealed that individuals from Elba Lake were more similar acoustically to individuals at Moon Lake than to each other, suggesting that acoustic similarity may represent a measure of genetic relatedness in this species (Fig. 1).

IV. DISCUSSION

Our preliminary results suggest that the alarm chirp vocalizations of adult female Belding's ground squirrels contain individual signature information which likely provides the basis for individual and kin recognition, as found in the vocalizations of several avian and mammalian species (Chapman and Weary, 1990; Reby *et al.*, 1998a). It is more likely that the acoustic differences found in this study are a result of genetic and not social influences, although the mechanisms underlying this acoustic variation will need to be evaluated in subsequent studies. Despite the mechanism(s), however, it might be possible to use this acoustic variation to track individuals and thus population dispersal patterns in this species using a noninvasive bioacoustic technique. Application of quantitative bioacoustic techniques might reveal subtle and important features of acoustic variation and patterns in free-ranging populations of many mammalian species (Reby *et al.*, 1998b). Therefore, we might effectively use bioacoustics as a tool for tracking population dispersal and dynamics, and possibly as a measure of genetic diversity, in other more threatened and endangered mammalian species, and thus in wildlife management and conservation.

ACKNOWLEDGMENTS

We would like to thank Rob Atwill, Kenneth Tate, Thomas Harter, and Bill Weitkamp for their help on the field portion of this project. This research was conducted under IACUC protocol No. 8607 approved by the IACUC of the University of California at Davis on 10 June 1999.

Armitage, K. B. (**1981**). ''Sociality as a life-history tactic of ground squirrels,'' Oecologia **48**, 36–49.

Adret-Hausberger, M. (**1986**). ''Temporal dynamics of dialects in the whistled songs of starlings,'' Ethology **71**, 140-152.

Boellstorff, D. E., and Owings, D. H. (**1995**). ''Home range, population structure, and spatial organization of California ground squirrels,'' J. Mammal. **76**, 551–561.

Chapman, C. A., and Weary, D. M. (**1990**). ''Variability in spider monkeys' vocalizations may provide basis for individual recognition,'' Am. J. Primatol **22**, 279–284.

Elowson, A. M., and Snowdon, C. T. (**1994**). ''Pgymy marmosets, *Cebulla pygmaea*, modify vocal structure in response to changed social environment,'' Animal Behaviour **47**, 1267–1277.

Ford, J. K. B. (**1991**). ''Vocal traditions among resident killer whales (*Orcinus orcinus*) in coastal waters of British Columbia,'' Ca. J. Zool. **69**, 1454–1483.

Ford, J. K. B., and Fisher, H. D. (**1983**). ''Group-specific dialects of killer whales (*Orcinus orca*) in British Columbia,'' in *Communication and Behavior of Whales*, edited by R. T. Payne (Westview Press, Colorado), pp. 129-161.

Hare, J. F. (**1999**). ''Juvenile Richardson's ground squirrels, *Spermophilus richardsonil*, discriminate among individual alarm callers,'' Anim. Behav. **55**, 451–460.

Leger, D. W., Berney-Key, S. D., and Sherman, P. W. (**1984**). ''Vocalizations of Belding's ground squirrels *Spermophilus beldingi*,'' Anim. Behav. **32**, 753–764.

Marshall, A. J., Wrangham, R. W., and Arcadi, A. C. (**1999**). ''Does llearning affect the structure of vocalizations in chimpanzees?'' Animal Behaviour **58**, 825–830.

McCowan, B. (**1995**). ''A new quantitative technique for categorizing whistles using simulated signals and whistles from captive bottlenose dolphins, *Delphindae Tursiops trucatus*,'' Ethology **100**, 177–193.

McCowan, B., and Reiss, D. (**2001**). ''The fallacy of 'signature whistles' in bottlenose dolphins: a comparative perspective of 'signature information' in animal vocalisations,'' Animal Behaviour **62**, 1151–1162.

McCowan, B., Reiss, D., and Gubbins, C. M. (**1998**). ''Social familiarity influences whistle acoustic structure in adult female bottlenose dolphins (*Tursiops truncatus*),'' Aquat. Mammals **24**, 21–40.

Mitani, J. C., and Brandt, K. (**1994**). ''Acoustic variability in the long distance calls of male chimpanzees,'' Ethology **96**, 233–252.

Mitani, J. C., Hasegawa, T., Gros-Louis, J., Marler, P., and Byrne, R. (**1992**). ''Dialects in chimpanzees?'' Am. J. Primatol. **27**, 233–243.

Nelson, D. A. (**1998**). ''Geographic variation in song of Gambel's white-crowned sparrow,'' Behaviour **135**, 321–342.

Owings, D. H., and Leger, D. W. (**1980**). ''Chatter vocalizations of California ground squirrels: predator and social-role specificity,'' Z. Tierpsychol. **54**, 163–184.

Owings, D. H., and Virginia, R. A. (**1978**). ''Alarm calls of California ground squirrels *Spermophilus beecheyi*,'' Z. Tierpsychol **46**, 58–70.

Owings, D. H., Hennessy, D. F., Leger. D. W., and Gladney, A. B. (**1986**). ''Different functions of ''alarm'' calling for different time scales: a preliminary report,'' Behaviour **99**, 101–116.

Pinheiro, J. C., and Bates, D. M. (**2000**). *Mixed-Effects Models in S and S-Plus* (Springer, New York).

Reby, D., Joachim, J., Lauga, J., Lek, S., and Aulagnier, S. (**1998a**). ''Individuality in the groans of fallow deer (*Dama dama) bucks*,'' J. Zool. **245**, 79–84.

Reby, D., Hewison, A. J. M., Cargnelutti, B., Angibault, J. M., and Vincent, J. P. (**1998b**). ''Use of vocalizations to estimate population size of roe deer,'' J. Wildlife Management **62**, 1341–1347.

Sorjonen, J., (**1987**). ''Temporal and spatial differences in the traditions and repertoires in the song of a thrush nightingale (Luscinia luscinia),'' Behaviour **102**, 196–212.

Tarurski, A. G., (**1976**). ''Whistle of the pilot whale Globicephala spp.: variations in whistling related to behavioral-environmental context, broadcast of underwater sounds, and geographic locatioin.'' Ph.D. dissertation, University of Rhode Island, Kingston.

Taruski, A. G. (**1979**). ''The whistle repertoire of the North Altantic pilot whale (Globicephala melaena) and its relationship to behavior and environment,'' in Behavior of Marine Animals, edited by H. E. Win and B. L. Olla, Vol. 3: Cetaceabs (Plenum Press, New York), pp. 345—368.

Wright, T. F. (**1996**). ''Regional dialects in the contact call of a parrot,'' Proc. R. Soc. London, Ser. B **263**, 867–872.

Optical and atomic force microscopic studies on sonoporation (L)

Joel P. Ross
Department of Chemistry and Physics, Saint Michael's College, Colchester, Vermont 05439

Xiane Cai
Department of Physics, University of Vermont, Burlington, Vermont 05405

Jen-Fu Chiu
Department of Biochemistry, University of Vermont, Burlington, Vermont 05405

J. Yang and Junru Wu
Department of Physics, University of Vermont, Burlington, Vermont 05405

(Received 19 June 2001; accepted for publication 30 November 2001)

Optical and atomic force microscopies were used for investigating morphological changes of Jurkat lymphocytes, which showed sonoporation after being exposed to a vibrating Mason horn tuned to 21.4 kHz for 2 min. The tip of the Mason horn had a diameter of 400 μm and its transverse vibration amplitude was 7.8 μm. It was found that immediately after sonification, some cells became irregular in shape but still more or less maintained their sizes, and other cells appeared to be much smaller than their normal sizes. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1448340]

PACS numbers: 43.80.Gx, 43.35.Ei [FD]

I. INTRODUCTION

Since an ultrasound field can be focused in space as desired, ultrasound-induced transfection, a relatively new technique, has the potential to become an effective site-specific *in vivo* tool for delivering drugs and foreign DNA strands in gene therapy trials. In many cases, gene therapy requires an introduction of foreign nucleic acids, most often DNA, containing corrective genes, into target cells where defective genes need to be corrected. However, the low permeability of the cell membrane results in an efficient barrier to entry of foreign DNA. Therefore, gene therapy requires techniques that promote the uptake and expression of foreign DNA by cells (transfection). Recent *in vitro* studies have shown[1–6] that the cell membranes can become temporarily ''open'' to small and large molecules and then ''resealed'' when relatively low-amplitude ultrasound is applied to a cell suspension that contains an ultrasound contrast agent. Consequently, foreign molecules are trapped in the cells. This phenomenon is called sonoporation.[1,4] The acoustic pressure amplitude threshold for generating the sonoporation was found to be quite modest in the presence of the ultrasound contrast agent; they were 0.12 MPa (Ref. 2) and 0.1 MPa (Ref. 1) for 1- and 2.25-MHz ultrasound, respectively. In contrast, the phenomenon in which a cell suffers lethal damage upon exposure to a sound field and its membrane, becomes permanently open is called cell lysis.

Sonoporation was also observed in Jurkat lymphocytes in a suspension exposed to a transversely vibrating Mason horn tuned to 21.4 kHz.[7] The tip of the Mason horn had a diameter of 400 μm and its transverse vibrating amplitude was 7.8 μm. It was found that the shear stress associated with microstreaming surrounding the Mason horn tip was the primary reason for the cell sonoporation. The threshold stress was determined to be 12 ± 4 Pa for up to 7 min exposure time.

Recently, Tachibana *et al.*[8] studied the cell-killing mechanism of HL-60 cells exposed to low-intensity ultrasound in the presence of photosensitive drugs. The morphology of HL-60 cells that had been exposed to continuous wave ultrasound in the presence of the photosensitive drug merocyanine 540 (frequency=255 KHz and acoustic pressure amplitude=0.11 MPa) for 30 s was investigated using a scanning electron microscope (SEM). Their SEM images showed that sonification in the presence of the photosensitive drug merocyanine 540 induced dimple-like craters and multiple surface pores of diameters up to 2–3 μm. The photosensitive drug merocyanine 540 or ultrasound alone could only cause minor surface disruption and flap-like wrinkles on the cell surface and the disappearance of the microvilli, originally being a part of the cell surface. They concluded that the rapid process of cell membrane porosity was the primary reason of the cell killing.

In this study, we have used both optical and atomic force microscopy to investigate morphological changes of Jurkat lymphocytes after being subjected to ultrasound radiation. It was found that immediately after sonification, some cells became irregular in shape but still more or less maintained their original sizes, and other cells appeared to be much smaller than their normal sizes. Our results suggest cells experiencing sonoporation after sonification may involve immediate morphological changes.

II. EXPERIMENTAL METHODS

A. Sample preparation and fluorescence microscopy

Jurkat lymphocytes were incubated in a 37 °C, 5% CO_2 atmosphere in 25-cm^2 growth area tissue culture flasks (Becton-Dickson, Oxnard, CA) in a solution of RPMI medium 1640 (GIBCOBRL, Grand Island, NY) supplemented with 10% fetal bovine serum (FBS). For exposures, the cells were diluted to a concentration of 4×10^5 cells/ml. For each trial, 0.5 ml of the cell suspension was placed in a plastic

culture tube of 12-mm diameter and 75-mm length (VWR Scientific, West Chester, PA) using aseptic technique.

B. Acoustic source

The vibrator used in this experiment was a Mason horn made in this laboratory. It consists of a stainless-steel stepped (exponentially tapered) bar glued to a cylindrical PZT tube transducer (3-mm thickness, 10.5-cm long, and 2.5-cm diameter). This vibrator was tuned to 21.4 kHz; at this frequency, a transverse vibration mode was excited. The tip of the bar had a diameter of 400 μm and the total length of the metallic bar was 17 cm. The Mason horn was driven by a Hewlett Packard 3314A function generator (Agilent Technologies, Palo Alto, CA) and an ENI 2100L RF power amplifier (ENI, Rochester, NY). The vibrating Mason horn was immersed in a test tube containing the suspension described earlier[7] and the sample was exposed to the sound field produced by the Mason horn for a specific time. The displacement amplitude of the Mason horn tip was measured using a calibrated microscope and found to be 7.8 μm. The shear stress produced near the tip of the Mason horn due to acoustic streaming was then calculated to be 12 Pa using the following equation:[9]

$$S = 2\pi^{3/2}\epsilon_0^2(\rho f^3 \eta)^{1/2}/r_0, \tag{1}$$

where η and ρ are, respectively, viscosity and density of the liquid, ϵ_0 is the displacement amplitude of the tip of a Mason horn of radius r_0, and f is frequency.

C. Optical and atomic force microscopy

Optical images were obtained using a Nikon Optiphot-2 microscope optically coupled with a CCD camera (GBC CCD-500E, CCTV Corp., South Hackensack, NJ). The calibration was achieved using a 10-μm standard (Digital Instruments, Inc., Santa Barbara, CA). Limited by spatial resolution of the optical images, damaged cells with diameters less than 5 μm were not measured.

Atomic force microscopic (AFM) images (512×512 pixels) were obtained using a NanoScope E AFM with a J head (170 μm) and oxide-sharpened Si_3N_4 tips (spring constant $k = 0.12$ N/m) (Digital Instruments, Inc., Santa Barbara, CA). The scanning speed was 0.5 line/s. In most cases the imaging required constant adjustment of imaging forces to compensate drifts due to softness of the cell; these drifts are much more severe and very irregular compared with in-solution imaging of supported membranes and mica-bound proteins. Thus, in order to image Jurkat lymphocytes, a piece of the 8×8-(mm) cover glass was silanized by coating γ-methacryl-oxypropyltrimethyl-silane (Sigma, St. Louis, MO) on its surfaces by first dipping the clean cover glass in a 10% silane solution in ethanol, followed by drying the cover glass in air for 15 min. The dried cover glass was then heated at 120 °C for 30 min. The resolution of the cell imaging was estimated to be about 100 nm. About 100 μl of prepared Jurkat lymphocyte suspension in phosphate-buffered saline (PBS) solution was applied to the cover glass and left there for 20 min. Afterwards, a fixation agent, glutaraldehyde at 2% by weight was introduced through solution exchange that ensured hydration of the sample in the entire process.

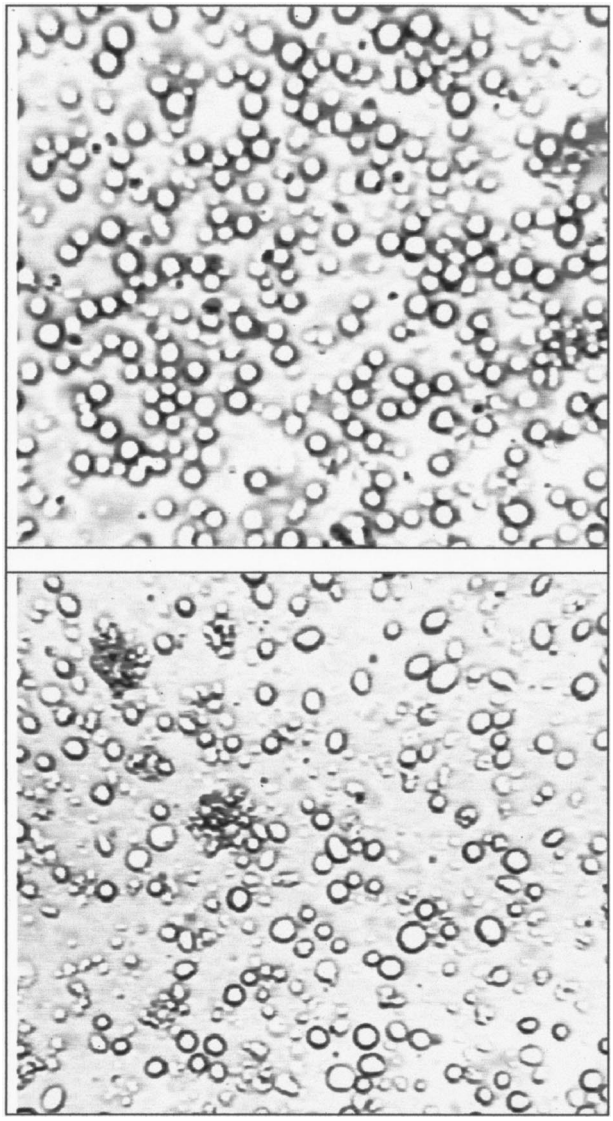


FIG. 1. The top panel is an image of unsonicated Jurkat lymphocytes and the bottom one is that after 2 min sonication. Their dimensions are: height ×width=353 μm×470 μm.

Glutaraldehyde also enabled a stronger binding of the fixed cells to the substrate. After the fixation procedure, the specimen was further washed with PBS and the adherent cells were imaged in PBS at room temperature.

III. EXPERIMENTAL RESULTS

Figure 1 contains two optical images: The top panel is an image of unsonicated Jurkat lymphocytes and the bottom one is that after 2 min sonification. Their dimensions are: height×width=353×470 μm. Obviously, the unsonicated cells are more uniform in size and shape, while the sonicated ones are somewhat irregular in shape and less uniform in size. Plots of the size distribution (number of cells versus cell diameter) for the cells shown in Fig. 1 are presented in Fig.

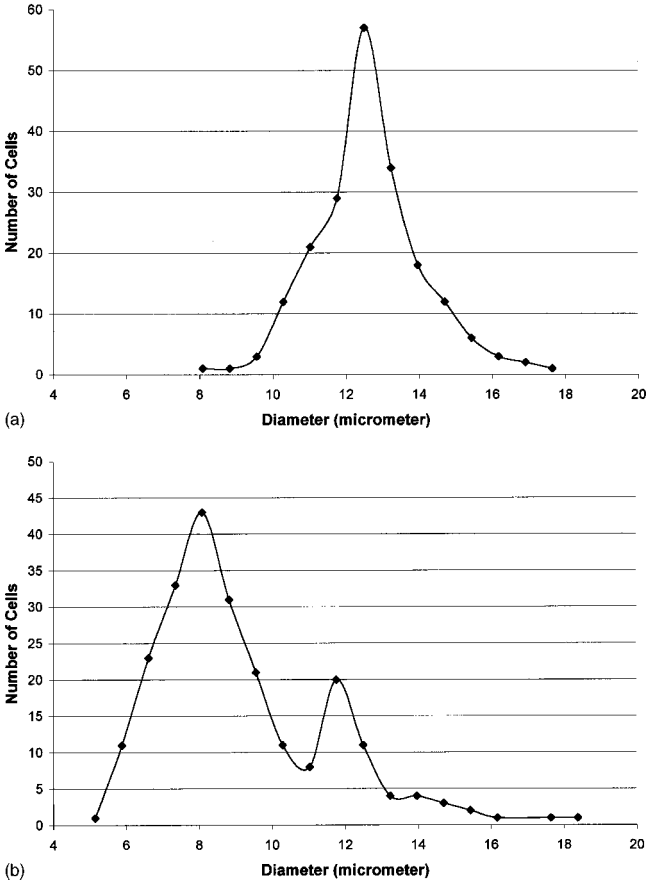


FIG. 2. Plots of the size distribution (number of cells versus cell diameter). (a) Unsonicated cells. (b) Sonicated cells. The diameter of the cell was estimated by the averaged value of the largest horizontal and vertical dimensions of each cell.

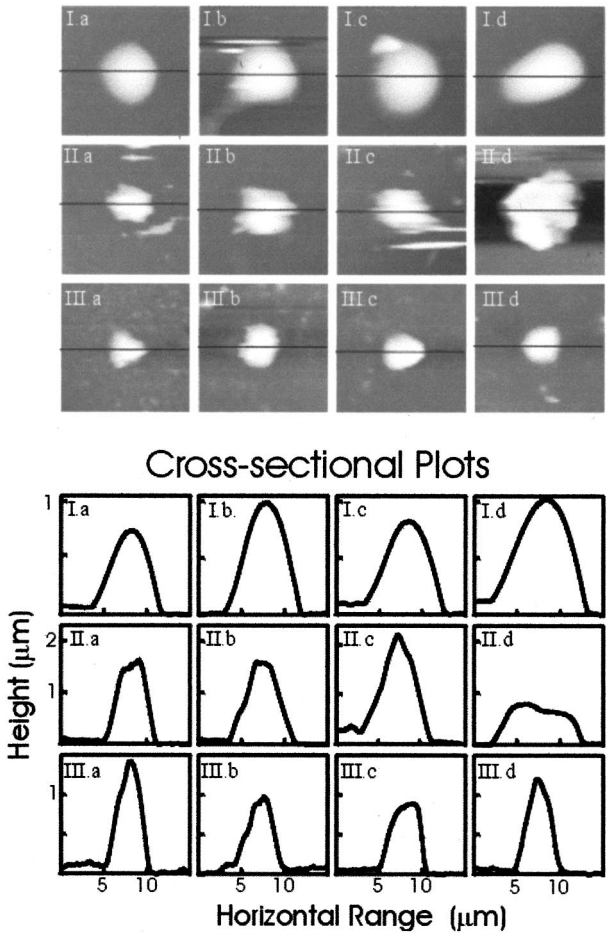


FIG. 3. Typical results of the atomic force microscope (AFM) study. The top 3×4 panels are AFM top view images of typical Jurkat lymphocytes and the bottom 3×4 panels are the cross-section images corresponding to the cells, respectively, denoted by the top 3×4 panels. These cross-section images taken along the solid lines marked in the top 3×4 panels describe the relation between the height and the linear location for the cells.

2; more than 200 cells were measured. The diameter of the cell was estimated by the averaged value of the largest horizontal and vertical dimensions of each cell.

It is quite evident that for the unsonicated cells [Fig. 2(a)], the distribution curve of cell size peaks at 12.5 mm and that for the sonicated cells [Fig. 2(b)] there are two peaks at 8.1 and 11.8 μm, respectively.

Figure 3 shows typical results of the atomic force microscopic (AFM) study. The top 3×4 panels are AFM top-view images of typical Jurkat lymphocytes and the bottom 3×4 panels are the cross-section images corresponding to the cells, respectively, denoted by the top 3×4 panels. These cross-section images taken along the solid lines marked in the top 3×4 panels describe the relation between the height and the linear location for the cells. The four images of the top row (row I) are for the unsonicated cells and those of rows II and III are for the sonicated cells. The unsonicated cells have smooth surfaces and the diameters of their cross sections are relatively large. The sonicated cells have irregular surfaces and smaller cross-section diameters. The sonicated cells of row II are relatively larger than those of row III. These findings were quite typical for the sonicated cells and consistent with what was found about the two-peak diameter distribution characteristics of the optical images shown in Fig. 2. In addition, within the resolution of AFM (100 nm), there was no indication that pores whose dimensions were greater than 100 nm existed on the membrane surfaces.

IV. DISCUSSION

The optical and AFM microscopic images have revealed that after sonification the overall morphology of Jurkat lymphocytes became irregular and some of them were severely distorted. The fact that no pore greater than 100 nm was observed cannot rule out the possibility that small pores might exist on the membrane surface. It is perceivable (without proof) that small pores might be the reason that sonoporation occurred. Even though most cells underwent morphological change shortly after the sonification, only a small percentage (about 10%) of them were notably sonoporated according to the fluorescence detection.[7] These results indicate that cellular morphological changes may or may not be accompanied by temporary poration that allows the incorporation of the fluorescent particles. The increase of the population of sonoporated cells in our previous experiments[7] as the sonication exposure time increased seems to suggest that sonoporation induced by cell morphological changes increased with time. A process called pinocytosis, which involves uptake of fluid-filled vesicles into cells, played a less

significant role in sonoporation, as it was demonstrated in previous experiments[1–4] that no FITC-dextran entered non-sonicated Jurkat lymphocytes.

[1] S. Bao, B. D. Thrall, and D. Miller, ''Transfection of a reporter plasmid into cultured cells by sonoporation *in vitro*,'' Ultrasound Med. Biol. **23**, 953–959 (1997).

[2] W. J. Greenleaf, M. E. Bolander, G. Sarkar, M. B. Goldring, and J. F. Greenleaf, ''Artificial cavitation nuclei significantly enhance acoustically induced cell transfections,'' Ultrasound Med. Biol. **24**, 587–595 (1998).

[3] M. Ward, J. Wu, and J. F. Chiu, ''Ultrasound-induced cell lysis and sonoporation enhanced by contrast agents,'' J. Acoust. Soc. Am. **105**, 2951–2957 (1999).

[4] M. Ward, J. Wu, and J. F. Chiu, ''Experimental study of the effects of Optison concentration on sonoporation *in vitro*,'' Ultrasound Med. Biol. **26**, 1169–1175 (2000).

[5] J. Wu, ''A possible physical mechanism of ultrasound-activated gene delivery: Shear stress generated by microstreaming,'' J. Acoust. Soc. Am. **108**, 2470(A) (2000).

[6] E. C. Unger, T. P. McCreery, and R. H. Switzer, ''Ultrasound enhanced gene expression of Liposomal transfection,'' Invest. Radiol. **32**, 723–727 (1997).

[7] J. Wu, J. P. Ross, and J. F. Chiu, ''Reparable sonoporation generated by microstreaming,'' J. Acoust. Soc. Am. (submitted).

[8] K. Tachibana, T. Uchida, K. Ogawa, N. Yamashita, and K. Tamura, ''Induction of cell-membrane porosity by ultrasound,'' Lancet **353**, 1409 (1999).

[9] J. A. Rooney, ''Shear as a mechanism for sonnically induced biological effects,'' J. Acoust. Soc. Am. **6**, 1718–1724 (1972).

Measurement of the scattering of a Lamb wave by a through hole in a plate

Paul Fromme[a] and Mahir B. Sayir
Institute of Mechanical Systems, ETH-Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland

(Received 17 May 2001; revised 15 November 2001; accepted 25 November 2001)

Flexural waves propagating in an aluminum plate containing a circular hole are studied. In the experiments the first antisymmetric Lamb wave mode A_0 is excited selectively by a piezoelectric transducer. The scattered field around a circular cavity is measured pointwise using a heterodyne laser interferometer. The measurements are compared with theoretical calculations. Different approximate analytical approaches, employing Kirchhoff and Mindlin types of plate theories to describe the scattered field, are used. Good agreement between the experimental data and the analytical solutions is found within the ranges of validity of the different models. Introduction of a small imperfection, like a notch, at the boundary of the cavity changes the measured scattered field significantly. The approach allows a fast measurement of large surfaces and might be useful for nondestructive testing purposes, e.g., the detection of cracks at fastener holes in airplane fuselage.
© *2002 Acoustical Society of America.* [DOI: 10.1121/1.1448338]

PACS numbers: 43.20.Gp, 43.20.Mv, 43.35.Cg [DEC]

I. INTRODUCTION

This paper deals with the scattering of flexural waves by an obstacle in isotropic, homogeneous plates. These waves can be described in three-dimensional theory as the first antisymmetric mode A_0 of Lamb[1] waves. The plate displacement associated with the wave is primarily of the bending type. For low frequencies, i.e., when the wavelength is large compared to the plate thickness, the propagation can be described approximately using classical plate theory (CPT), taking only bending stiffness and inertia into account. For higher frequencies, shear and rotatory inertia have to be considered according to the theory of Mindlin.[2] Alternatively, approximate theories can be derived from an asymptotic expansion of the full three-dimensional theory for different levels of accuracy, as done by Niordson.[3] A review of theories describing the motion of waves in plates is given for example by Graff.[4]

The scattering of flexural waves by an obstacle in a plate has been analyzed by Pao and Chao.[5] They solved the case of a circular cavity, i.e., a hole, using Mindlin's theory of plates, and derived three scattered waves to fulfill the boundary conditions at the hole. A similar analysis employing Kirchhoff-type boundary conditions was done by Staudenmann,[6] who also studied the scattering at different types of circular inclusions. Vemula and Norris[7,8] did a similar study for thin plates and Mindlin-type plates, using an optical theorem. Further analytical work on the scattering of the first symmetric Lamb mode S_0 by a circular hole was reported by Pao,[9] and McKeon and Hinders,[10] who also give a good review of other papers. Finite-element methods (FEM) combined with a wave function expansion were used by Paskaramoorthy, Shah, and Datta[11,12] to calculate the scattered field of a flexural wave by a circular cavity and a crack in a plate. Their method can be expanded to more complicated geometries, like irregularly shaped cavities or a crack at the hole boundary. Chang and Mal[13] investigated the difference in the scattered field of a longitudinal wave at a hole due to two cracks at opposite sides of the hole, also using a hybrid FEM modeling. A hybrid boundary element method was employed by Cho and Rose[14] to study the edge reflection and mode conversion of the lower Lamb modes above the lowest cutoff frequency. Wang and Ying[15] used a waveform expansion to calculate the reflection and transmission coefficients of the two fundamental Lamb wave modes S_0 and A_0, scattered by an elastic cylinder embedded in an isotropic plate. Recently, Valle *et al.*[16] used FEM modeling and digital signal processing to locate and size cracks in hollow isotropic cylinders by means of guided circumferential waves. For a similar geometry, nonaxisymmetric guided waves were studied by Li and Rose[17,18] experimentally and using normal-mode expansion.

However, few experimental papers on the scattering of Lamb waves by a hole in a plate can be found. Chang and Mal[13] compared their numerical data with experimental results, using a wedge transducer to excite the first symmetric mode S_0. Measurements were made at two points with contact-type transducers. They achieved good qualitative agreement between measured and calculated time series and power spectra, but did not get an exact match. This might be due to the large transducer size used, which has the same order of magnitude as the wavelength at the applied frequencies of about 0.5 MHz.

Malyarenko and Hinders[19] experimentally studied the scattering of the S_0 mode at a through hole in a plate in the frequency range between 1 and 2 MHz, below the cutoff frequencies of the higher modes, to reconstruct flaws using fan beam tomography. They used longitudinal contact transducers at various positions around the hole and analyzed the arrival times. Though both the S_0 and A_0 mode were excited, only the arrival times of the S_0 mode were used, as it propagates faster and is nondispersive.

Chan and Cawley[20] investigated the propagation of

[a]Electronic mail: fromme@imes.mavt.ethz.ch

higher Lamb modes in an attenuative plate. Selecting the angle of incidence of a water-coupled broadband transducer, desired higher Lamb modes in a polyethylene plate were excited and their group velocity and attenuation measured. Alleyne and Cawley[21] studied the reflection and transmission of the S_0 and A_1 mode at a notch in a steel plate experimentally, using similar transducers and comparing their results to numerical calculations using FEM.

The geometry of the setup used in the investigation described in this paper is not a conventional C-scan. The transducer exciting the wave is fixed in one position, while the scattered field is measured on a grid around the hole. Therefore, the incident wave for each measuring point is the same, as assumed in the analytical calculations. The scattered field can be compared with calculations using the different approximate theories mentioned above.

The first antisymmetric mode A_0 is excited selectively, working in a frequency range between 20 and 100 kHz, below the cutoff frequencies of the higher Lamb modes. The out-of-plane displacement is measured pointwise on a grid around the hole with a heterodyne laser interferometer. At each measuring point, a time signal of the velocity of the plate surface is stored. Using fast Fourier transform (FFT), amplitude and phase of the scattered wave are extracted and analyzed.

Small defects at the hole can be detected from a difference between experimental measurements and theoretical results for an undamaged structure. This is important for the nondestructive testing of airplanes, an important application of the measurement method. A common problem in aircraft maintenance is the emergence of fatigue and corrosion cracks at fasteners and rivet holes in the fuselage. One of the conventionally used methods for the detection of such defects is ultrasonic testing (UT). For the large body of an aircraft this is rather time-consuming, as the transducer has to be moved manually across the whole inspection area. An idea pursued by Bar-Cohen[22] is the use of robotic devices to automate this scanning procedure. In contrast to these ultrasonic techniques, in which the wavelengths are short compared to the plate thickness, guided waves have the advantage of propagating over large areas, thus omitting the scanning process. From the measurement of the guided wave at a few points of the surface of the structure, defects in a large area can be detected with a fast and cost-effective method. Rose and Soley[23] showed the practical applicability of guided waves for a variety of problems in aircraft components, e.g., crack detection in helicopter blades. However, to achieve a guided wave in a plate, the wavelength has to be significantly larger than in UT, and therefore the scattered field is less sensitive to small defects. To show the practical applicability of the method, the scattered field at a complicated geometry like a line of holes in a plate is measured and compared with a theoretical calculation. Good agreement is found, and a small notch introduced at one of the holes shows a significant effect on the scattered field and can thus be detected. Further measurements at a notched hole in a plate are described by Fromme and Sayir.[24]

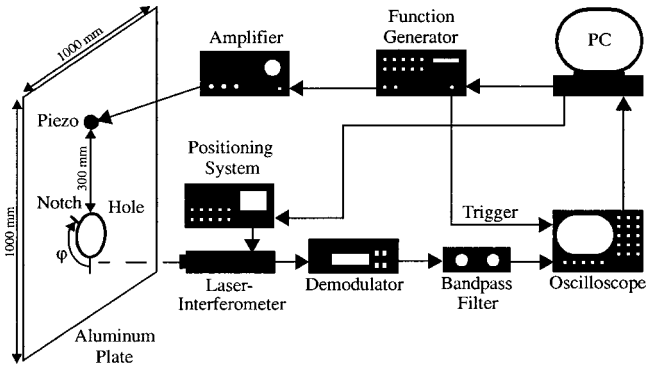


FIG. 1. Experimental setup; angle around the hole measured from vertical (direction of wave propagation).

II. EXPERIMENTAL SETUP

A. Specimen geometry

The specimen is a 1-mm-thick aluminum plate with a size of 1000 by 1000 mm. The large size permits a time separation between the wave scattered at the hole and the part scattered at the plate boundaries. The plate material is a commercially available aluminum alloy, having a Young's modulus E of 6.9×10^{10} N/m^2 and a density ρ of 2700 kg/m^3. Poisson's ratio ν is assumed as 0.31. One or three holes with a radius of 10 mm are drilled through the plate. The plate is suspended vertically to avoid static bending. A piezoelectric disk, polarized for thickness extension mode, of 10-mm diameter and 1-mm thickness, is glued to the plate 300 mm away from the hole using a two-component epoxy glue. To simulate a defect at the hole boundary, a fine saw blade is used to cut a through-thickness notch of about 0.2 mm width, 2 mm length, and a blunt tip into the plate at an angle φ_0 to the vertical (see Fig. 1).

B. Excitation signal

A sinusoid multiplied by a Hanning window is chosen as the excitation signal to achieve a short-time, narrow-band signal with the energy concentrated around the center frequency f_0. The signal has to be short in time to allow a time windowing that avoids the interference of incident wave and wave scattered at the hole and the notch from the reflection at the plate boundaries, arriving some time later at the measurement spot. The arrival times of the different pulses are calculated from the theoretical group velocities. However, the signal also has to have a narrow bandwidth to avoid extensive signal distortion due to the dispersive character of the A_0 mode. A measured time signal, having a center frequency of 20 kHz and 10 cycles, used in the experiments, can be seen in Fig. 2. The signal is generated in a programmable function generator and then amplified to 200 V peak to peak.

C. Transducer

When the voltage is applied to the piezoelectric transducer, the disk contracts and expands. This generates a vertical force to the plate surface and excites primarily the first antisymmetric mode A_0, as the resulting normal stress in the

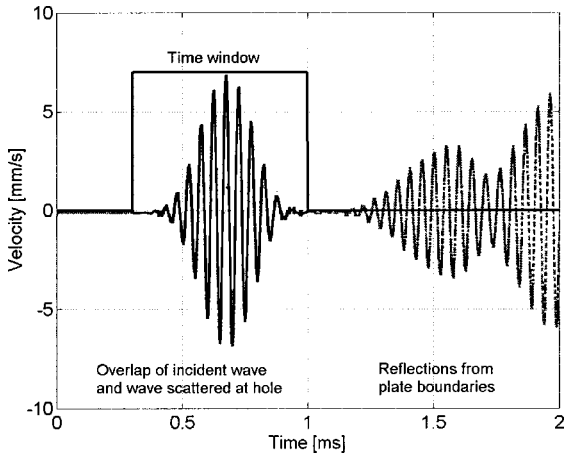


FIG. 2. Typical measured signal: sinusoid multiplied by a Hanning window, center frequency $f_0 = 20$ kHz, 10 cycles; dotted: measured time signal; solid: time window to cut off boundary reflections.

plate is also antisymmetric. For the frequencies used in this study, the energy transferred to the longitudinal mode S_0 and the shear-horizontal mode SH is negligible. As we work well below the cutoff frequencies for the higher wave modes, only the desired mode A_0 is excited. The approach shown here is feasible for frequencies below the cutoff frequencies, as no selection between different modes is necessary, avoiding the need for prescribing the wavelength at a given frequency by an angle of incidence as in classical UT. The excited frequencies of up to 100 kHz are well below the eigenfrequencies of the piezoelectric disk, so that a linear transfer curve is achieved. However, the amplitude of the excited wave is less than 1 μm.

D. Wave propagation

The piezoelectric disk acts as a point source for the resulting flexural wave, which propagates radially outwards. The distance between the transducer and hole of 300 mm is large compared to the hole radius of 10 mm. Therefore, the wavefront can be assumed to be planar when it reaches the hole, allowing a simpler theoretical simulation. The incident wave is scattered at the stress-free boundaries of the hole, and a scattered wave is generated. The scattered wave consists of two parts: a boundary layer close to the hole and a part propagating radially outwards from the hole. In the vicinity of the hole, incident and scattered wave overlap in time, so that only a single pulse is visible in Fig. 2. This is due to the length of the excitation signal and the low group velocity of the A_0 mode. The specimen size is selected large enough that the reflections from the plate boundaries reach the measurement area some time later. This way, as seen in Fig. 2, a time separation between the scattered field and the boundary reflections is achieved.

E. Measurement

The scattered field on a measurement grid around the hole is recorded using a commercially available heterodyne laser interferometer. The demodulator output is a voltage signal proportional to the velocity of the out-of-plane component of the displacement of the plate surface. The measurement spot is defined by the laser beam diameter, which is well below 0.1 mm. This allows a mostly pointwise measurement of variations in the scattered field, as no implicit average over a rather large surface of the measuring transducer is made. The laser interferometer is moved parallel to the plate on a positioning system, allowing a measurement in the vicinity of the defect without disturbance. A radial grid is used, moving the measurement spot on concentric circles around the hole and recording a time signal every 5 deg on radii 0.5 mm apart. For the experiment with three holes, a Cartesian grid with 2-mm step size in the horizontal and vertical direction is used. The measurements are very repeatable, with the largest cause of variation due to an inaccurate positioning of the laser beam relative to the hole center, which could only be achieved with an accuracy of about 0.1 mm.

The voltage signal is bandpass filtered around the center frequency f_0 and averaged over 25 measurements in a digital storage oscilloscope. The function generator triggers the oscilloscope, so that excitation and measurement start at the same time. The measured time series are then transferred to the computer for evaluation. At each point of the grid a time series with 10 000 values is stored. A time windowing is applied to cut off the reflections caused by the plate boundaries, which contain no information about the scattering at the hole. Fast Fourier transform (FFT) is applied and the amplitude and phase values at the center frequency f_0 are extracted. These values are the equivalent of the theoretical results, derived in Sec. III, where an infinite sinusoid is assumed for the incident wave. The amplitude of a typical scattered field around a hole is shown in Fig. 3(a) in Sec. IV.

III. THEORETICAL DESCRIPTION

The scattering of a flexural wave by a circular cavity with radius r_0 in an isotropic, homogeneous plate is studied. Flexural waves are the first antisymmetric mode A_0 of the Lamb waves in plates. Different approximative solutions can be used to achieve a faster computation; see, e.g., Graff.[4] The incident wave is taken as an infinite sinusoid, planar wave with unit amplitude, propagating in the y direction, and is given by

$$w_i = e^{i(\omega t - k_1 y)}, \tag{1}$$

with out-of-plane displacement w, wave number of a propagating flexural wave k_1, and angular frequency ω. The origin of the coordinate system is chosen at the center of the hole; cylindrical coordinates (r, φ, z) are introduced. To satisfy the stress-free boundary conditions at the hole

$$\sigma_{rr} = \sigma_{r\varphi} = \sigma_{rz} = 0, \quad r = r_0, \quad \varphi \in [-\pi, \pi], \tag{2}$$

a scattered wave must occur. This problem was studied by Norris and Vemula,[8] and Staudenmann[6] using CPT, which takes only inertia and bending stiffness into account. CPT is valid only at low frequencies, when the wavelength is large compared to the plate thickness (as shown by Niordson[3] using an asymptotic expansion). Following their approach, the scattered wave is assumed in the form of

$$w_s = \sum_{n=0}^{\infty} (a_{1n} H_n^{(2)}(k_1 r) + a_{2n} H_n^{(1)}(k_2 r)) \cos(n\varphi) e^{i\omega t}. \tag{3}$$

The first part describes a wave propagating outwards (Hankel function of the second kind), and the second part a boundary layer around the hole (Hankel function of the first kind). The angular dependence is given by the sum over the cosine functions. k_2 is the wave number of the nonpropagating flexural mode, and for CPT $k_2 = i \cdot k_1$.

With this approach, the boundary conditions can only be fulfilled in the Kirchhoff approximation

$$M_{rr} = 0, \quad Q_r + \frac{1}{r_0} M_{r\varphi,\varphi} = 0, \tag{4}$$

satisfying a combination of vertical force and the derivative of the twisting moment. The coefficients of the scattered wave are calculated from the substitution of the incident and scattered wave into the boundary conditions and the projection in tangential direction $[\cos(n\varphi)]$. For the numerical evaluation, the first 30 coefficients of the scattered wave are calculated. It is found numerically that the higher coefficients have a negligible influence on the scattered field, as they only describe very local oscillations and the coefficients converge quickly to zero. This can also be shown using the numerical check derived from an optical theorem by Norris and Vemula.[8]

For higher frequencies, corresponding to shorter wavelengths, shear and rotatory inertia have to be considered. Therefore, the theory of Mindlin[2] without normal pressure q on the plate faces is used. Following the work of Pao and Chao,[5] the boundary conditions can be fulfilled as an average over the plate thickness with

$$M_{rr} = 0, \quad M_{r\varphi} = 0, \quad Q_r = 0. \tag{5}$$

The scattered wave consists of the same flexural wave as in Eq. (3), and additionally a shear wave with the two components

$$\begin{aligned} \psi_r = & \frac{1}{r} \sum_{n=0}^{\infty} (a_{1n}(\sigma_1 - 1) H_n^{(2)}(k_1 r) + a_{2n}(\sigma_2 - 1) \\ & \times H_n^{(1)}(k_2 r) + a_{3n} H_n^{(1)}(k_3 r)) \cos(n\varphi) e^{i\omega t}, \\ \psi_\varphi = & \frac{1}{r} \sum_{n=0}^{\infty} (a_{1n}(1 - \sigma_1) H_n^{(2)}(k_1 r) + a_{2n}(1 - \sigma_2) \\ & \times H_n^{(1)}(k_2 r) - a_{3n} H_n^{(1)}(k_3 r)) \sin(n\varphi) e^{i\omega t}. \end{aligned} \tag{6}$$

Evaluating the boundary conditions in polar coordinates, the coefficients of the scattered waves are calculated. k_3 is the wave number of the shear wave; σ_1 and σ_2 give the relation between the wave numbers, as defined in Ref. 5.

IV. COMPARISON BETWEEN EXPERIMENTS AND THEORY

A typical measured scattered field around a hole in an aluminum plate for an excitation with a center frequency of 100 kHz can be seen in Fig. 3(a). The measurement is made on a circular grid around the hole, with a signal recorded every 5 deg on radii 0.5 mm apart. The incident wave with a nearly plane wavefront propagates along the direction of the positive y axis and is scattered at the hole boundary. Behind the hole, a dark area (low amplitude) can be seen, the so-called shadow area, where only little energy arrives. Directly at the hole a high amplitude (light) results from the scattering at the free surface. Further outwards a characteristic hill and valley pattern develops due to the constructive and destructive interference of incident and scattered waves. Rather strong amplitude variations over short distances are evident. In the backscattered field, every half-wavelength a high and low amplitude appears due to interference. This might pose a problem when measuring with contact-type transducers with a diameter about as large as the wavelength, as an implicit average over the contact surface is made.

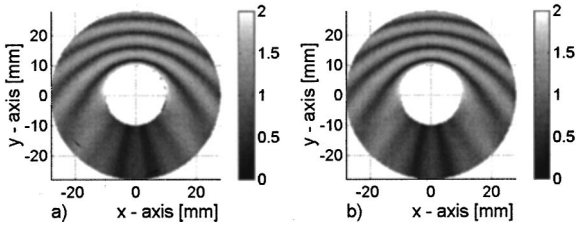


FIG. 3. Amplitude (normalized: $U_i = 1$) of the scattered field around the hole; plate thickness 1 mm, hole radius $r_0 = 10$ mm, frequency $f_0 = 100$ kHz, wavelength $\lambda = 10$ mm: (a) measured; (b) Mindlin's theory.

In Fig. 3(b) the scattered field, calculated using Mindlin's theory and following Pao and Chao,[5] is shown. Measurement and analytical solution show good qualitative agreement. To verify the validity of the various approximations, the scattered field around a hole is measured for different relations between wavelength, plate thickness, and hole radius. The measurements on a concentric circle around the hole are compared with analytical calculations using classical plate theory and Mindlin's theory.

Good quantitative agreement between both approximate theories and experiments is found, as shown in Fig. 4, for a center frequency f_0 of 20 kHz. This corresponds to a wavelength of 22 mm, 22 times the plate thickness of 1 mm and

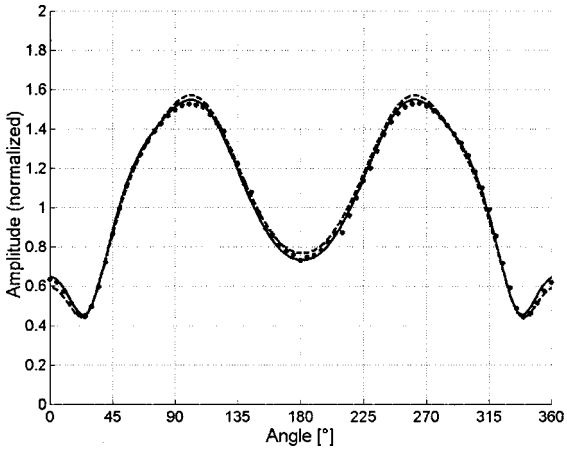


FIG. 4. Amplitude (normalized: $U_i = 1$) on a circle around the cavity; plate thickness 1 mm, hole radius $r_0 = 10$ mm, radius $r = 13$ mm, frequency $f_0 = 20$ kHz, wavelength $\lambda = 22$ mm: measured (dots), Mindlin's theory (solid), classical plate theory (dashed).

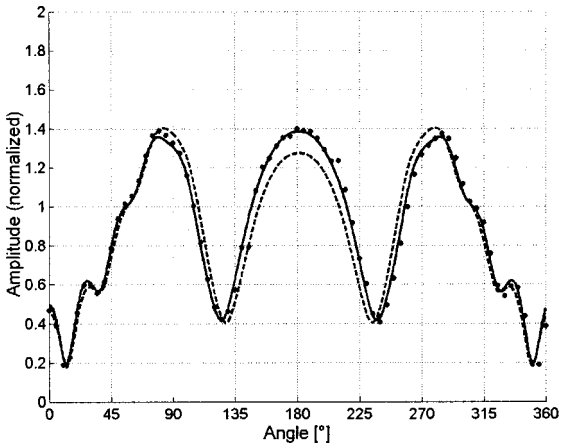


FIG. 5. Amplitude (normalized: $U_i=1$) on a circle around the cavity; plate thickness 1 mm, hole radius $r_0=10$ mm, radius $r=13$ mm, frequency $f_0=100$ kHz, wavelength $\lambda=10$ mm: measured (dots), Mindlin's theory (solid), classical plate theory (dashed).

about as large as the hole diameter of 20 mm. The experimental points, measured every 5 deg, and the curves for Mindlin (solid) and CPT (dashed) are normalized with the amplitude of the incident wave. For a higher frequency of 100 kHz with a wavelength of 10 mm, a deviation between CPT on one side and Mindlin and the experimental data on the other side, especially in the backscattered region (around 180°), is evident in Fig. 5. This is due to the effects of shear and rotatory inertia having a stronger influence at shorter wavelengths. Describing the plate according to Mindlin's theory leads to good agreement with the measured values. However, the difference of CPT is still rather small, so that it can be used as a first approximation for the case considered here, where wavelength and hole radius are large compared to the plate thickness.

A different case is shown in Fig. 6, with a hole radius of 5 mm and a plate thickness of 2 mm. Even for a low frequency of 20 kHz, a significant, systematic difference between measurement and Mindlin's theory on one side and CPT on the other side is evident. This results from an additional boundary layer close to the hole due to the effect of

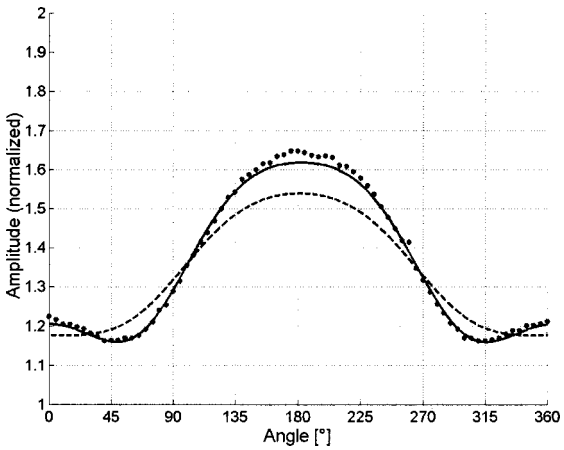


FIG. 6. Amplitude (normalized: $U_i=1$) on a circle around the cavity; plate thickness 2 mm, hole radius $r_0=5$ mm, radius $r=6$ mm, frequency $f_0=20$ kHz, wavelength $\lambda=31$ mm: measured (dots), Mindlin's theory (solid), classical plate theory (dashed).

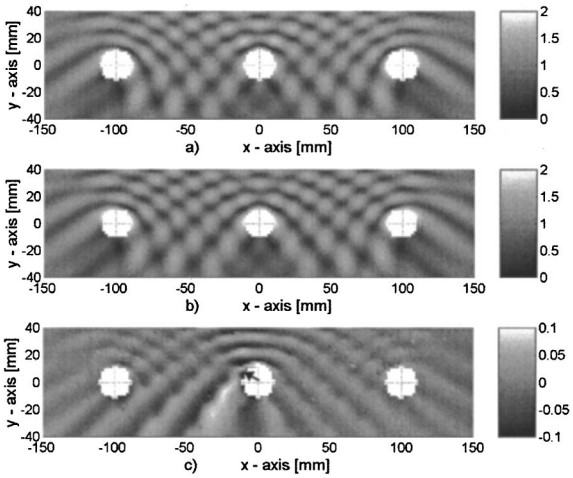


FIG. 7. Amplitude (normalized: $U_i=1$) of the scattered field around three holes; plate thickness 1 mm, hole radius $r_0=10$ mm, hole distance $d_0=100$ mm, resolution: $\Delta x=\Delta y=2$ mm, frequency $f_0=20$ kHz, $\lambda=22$ mm: (a) measured; (b) classical plate theory; (c) difference in measured amplitude due to a 2-mm through notch at the center hole ($\varphi_0=120°$).

shear and rotatory inertia, which are neglected in CPT, but are well described using Mindlin's theory. Here, the wavelength of 31 mm is large compared to both hole radius and plate thickness. The solution using CPT can be shown by an asymptotic expansion to be the first approximation, when both wavelength and hole radius are large compared to the plate thickness.

V. MEASUREMENTS AT MORE COMPLICATED GEOMETRIES

The case of a single hole in an infinite plate is the simplest case to test the accuracy and validity of the measurements and analytical descriptions. However, more complicated geometries are present in engineering applications, e.g., a line of rivet holes in an aircraft fuselage. In order to be able to detect defects in such samples, the combined scattered field has to be described and measured very accurately. A line of three holes is drilled into a plate and measurements on a Cartesian grid are performed before and after a notch is cut at the center hole. The measured combined scattered field can be seen in Fig. 7(a). At each hole, the incident wave has a different amplitude, phase, and angle of incidence, as it propagates radially outwards from the transducer, located 300 mm above the center hole. The chessboard-like pattern results from the interference of the different scattered waves. This case can be described well theoretically by a superposition of the three scattered fields at each hole according to CPT, without considering secondary scattering [Fig. 7(b)]. After the first measurement [Fig. 7(a)] a notch of 2-mm length is cut through the thickness of the plate at the center hole using a fine saw blade. The notch simulates a defect, like a fatigue or corrosion crack at a fastener in an airplane. The scattered field is measured again and the difference in amplitude due to the notch is shown in Fig. 7(c). At the free surfaces of the notch an additional scattered wave is generated, which changes the scattered field significantly up to about 10% in amplitude.

VI. CONCLUSIONS

Guided flexural waves in an aluminum plate are generated by a piezoelectric transducer. Driven below the cutoff frequencies of the higher wave modes, the first antisymmetric mode A_0 is excited selectively. The scattered field around a hole is measured using a heterodyne laser interferometer. Good spatial resolution of the measurement is achieved. The scattered field is described using approximate theories, namely classical plate theory and Mindlin's theory of plates. Good agreement between the measurements and analytical calculations is found. Care has to be taken to use approximations that are valid for the selected ratio of wavelength, hole radius, and plate thickness. The scattering at complicated geometries, like a line of holes, can be described accurately as well. The measured scattered field at the hole changes significantly with the presence of a notch, much smaller than the wavelength. This allows the use of the proposed method for the nondestructive testing of large structures. A wave mode traveling along the structure can be excited, and its scattering measured. From changes in the measured amplitudes, the development of defects, like fatigue or corrosion cracks in aircraft structures, can be detected. The expansion of the theory to describe the scattered field at a hole with a crack at the hole boundary would be an important tool to estimate the sensitivity of this approach.

ACKNOWLEDGMENTS

We would like to thank Georgios Kotsalis and Bernard Masserey for their work on the theoretical calculation of the scattering of a flexural wave by a hole.

[1] H. Lamb, ''On waves in an elastic plate,'' Proc. R. Soc. London, Ser. A **93**, 114–128 (1917).
[2] R. D. Mindlin, ''Influence of rotatory inertia and shear on flexural motions of isotropic, elastic plates,'' J. Appl. Mech. **18**, 31–38 (1951).
[3] F. I. Niordson, ''An asymptotic theory for vibrating plates,'' Int. J. Solids Struct. **15**(2), 167–181 (1979).
[4] K. F. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1991).
[5] Y.-H. Pao and C. C. Chao, ''Diffractions of flexural waves by a cavity in an elastic plate,'' AIAA J. **2**(11), 2004–2010 (1964).
[6] M. Staudenmann, ''Structural waves in nondestructive testing,'' Ph.D. thesis ETH No. 11394, Swiss Federal Institute of Technology, Library, CH-8092 Zürich, Switzerland (1995).
[7] C. Vemula and A. N. Norris, ''Flexural wave propagation and scattering on thin plates using Mindlin theory,'' Wave Motion **26**, 1–12 (1997).
[8] A. N. Norris and C. Vemula, ''Scattering of flexural waves on thin plates,'' J. Sound Vib. **181**(1), 115–125 (1995).
[9] Y.-H. Pao, ''Dynamical stress concentration in an elastic plate,'' J. Appl. Mech. **29**, 299–305 (1962).
[10] J. C. P. McKeon and M. K. Hinders, ''Lamb wave scattering from a through hole,'' J. Sound Vib. **224**(5), 843–862 (1999).
[11] R. Paskaramoorthy, A. H. Shah and S. K. Datta, ''Scattering of flexural waves by a crack in a plate,'' Eng. Fract. Mech. **33**(4), 589–598 (1989).
[12] R. Paskaramoorthy, A. H. Shah, and S. K. Datta, ''Scattering of flexural waves by cavities in a plate,'' Int. J. Solids Struct. **25**(10), 1177–1191 (1989).
[13] Z. Chang and A. Mal, ''Scattering of Lamb waves from a rivet hole with edge cracks,'' Mech. Mater. **31**, 197–204 (1999).
[14] Y. Cho and J. L. Rose, ''A boundary element solution for a mode conversion study on the edge reflection of Lamb waves,'' J. Acoust. Soc. Am. **99**, 2097–2109 (1996).
[15] X.-M. Wang and C. F. Ying, ''Scattering of Lamb waves by a circular cylinder,'' J. Acoust. Soc. Am. **110**, 1752–1763 (2001).
[16] C. Valle, M. Niethammer, J. Qu, and L. J. Jacobs, ''Crack characterization using guided circumferential waves,'' J. Acoust. Soc. Am. **110**, 1282–1290 (2001).
[17] J. Li and J. L. Rose, ''Excitation and propagation of nonaxisymmetric guided waves in a hollow cylinder,'' J. Acoust. Soc. Am. **109**, 457–464 (2001).
[18] J. Li and J. L. Rose, ''Implementing guided wave mode control by use of a phased transducer array,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**(3), 761–768 (2001).
[19] E. V. Malyarenko and M. K. Hinders, ''Fan beam and double crosshole Lamb wave tomography for mapping flaws in aging aircraft structures,'' J. Acoust. Soc. Am. **108**, 1631–1639 (2000).
[20] C. W. Chan and P. Cawley, ''Lamb waves in highly attenuative plates,'' J. Acoust. Soc. Am. **104**, 874–881 (1998).
[21] D. N. Alleyne and P. Cawley, ''The interaction of Lamb waves with defects,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **39**(3), 381–397 (1992).
[22] Y. Bar-Cohen, ''Emerging NDT technologies and challenges at the beginning of the third millenium. II,'' Mater. Eval. **58**(2), 141–150 (2000).
[23] J. L. Rose and L. E. Soley, ''Ultrasonic guided waves for anomaly detection in aircraft components,'' Mater. Eval. **58**(9), 1080–1086 (2000).
[24] P. Fromme and M. B. Sayir, ''Experimental detection of cracks at rivets using structural wave propagation,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, AIP Conference Proceedings 557, edited by D. O. Thompson and D. E. Chimenti (AIP, New York, 2001), Vol. 20B, pp. 1626–1633.

On the selection of loads in the multiload method for measuring the acoustic source parameters of duct systems

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

(Received 5 September 2001; revised 13 December 2001; accepted 17 December 2001)

The in-duct source can be characterized by two acoustical parameters such as the source strength and the source impedance, which permit the prediction of radiated sound pressure or insertion loss of the whole duct system. One-port acoustic characteristics of an in-duct source can be measured by the multiload method using an overdetermined set of open pipes or side-branch pipes with different lengths as applied loads. The input data, viz. load pressure and load impedance, are usually contaminated by measurement error in the actual measurements, which result in errors in the calculated source parameters. In this paper, the effects of the errors in the input data on the results have been studied numerically, varying the number of loads and their impedances in order to determine what combination of the loads will yield the best result. It is noted that, frequently, only a set of open pipes is used when applying the multiload method to the internal combustion engine sources. A set of pipe lengths, which cause the calculated results to be least sensitive to the input data error, can be found when using open pipe loads. The present work is intended to produce guidelines for preparing an appropriate load set in order to obtain accurate source properties of fluid machines. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1451064]

PACS numbers: 43.20.Mv, 43.20.Ye, 43.50.Gf [LLT]

I. INTRODUCTION

In order to predict the acoustic performance of a ducted silencing system, the information on the acoustic characteristics of the source as well as that of the silencer itself should be known. For a source located at one end of the duct, the acoustic one-port source model can be used by assuming plane wave propagation and a linear, time-invariant system. The in-duct source can be characterized by the source strength and the source impedance. In order to determine these quantities, various direct or indirect measurement techniques can be employed. For measuring the source properties on the exhaust side of internal combustion engines, only indirect measurements are possible due to the high gas temperature and the acidic, turbulent gas flow. Often the indirect method which is used is the multiload method with open pipes or side branches of different lengths as the loads.[1] A refined multiload method has been proposed to overcome the serious instability problem of the original four-load method.[2] The input data are acoustic impedance and radiated sound pressure for each load. In actual measurements, the input data are prone to add a certain amount of error.

In this paper, the effects of errors in the input data on the estimated source parameters have been studied numerically, varying the number of loads and their impedances in order to determine what combination of the loads will yield the best result. An error analysis is applied to each case of possible loads, which consist of open pipes. A guideline will be given for preparing a load set that would result in the measurement of accurate source properties of fluid machines.

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

II. THEORY

A. Multiload method

The electroacoustic analogous circuit for an acoustic source in a duct can be expressed as shown in Fig. 1, in which the linear and time-invariant characteristics are assumed. The frequency domain representation of the source–load interaction is given by

$$P = P_s Z_l / (Z_l + Z_s), \tag{1}$$

where P_s and Z_s are the source strength and the source impedance, respectively. P, V, and Z_l are the acoustic pressure, the acoustic volume velocity at the source–load junction, and the acoustic impedance of the load, respectively.

The refined multiload method[1] was proposed to overcome the instability problems of the original four-load method developed by Prasad[2] for the evaluation of the source properties. One can modify Eq. (1) as

$$|P_s|^2 = |P_n|^2 \frac{|Z_n + Z_s|^2}{|Z_n|^2} \quad (n = 1,2,3,...,N), \tag{2}$$

where P_n and Z_n denote the acoustic pressure and impedance of the nth load, respectively, and N is the number of applied loads. This equation deals with the energy quantity rather than the physical parameter itself, as employed in the conventional method. Because $|P_s|^2$ should be invariable no matter what load is connected, an error function $F(Z)$ can be derived from Eq. (2) as follows:

$$F(Z) = \sum_{m<n} \left\{ \frac{|P_m|^2}{|Z_m|^2} |Z_m + Z|^2 - \frac{|P_n|^2}{|Z_n|^2} |Z_n + Z|^2 \right\}^2 \quad (m,n = 1,2,3,...,N). \tag{3}$$

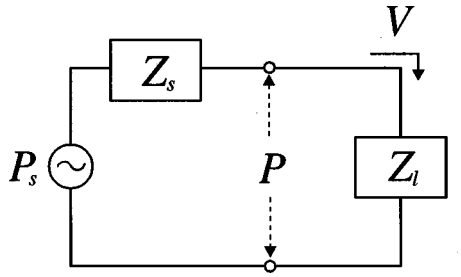


FIG. 1. Electroacoustic circuit for the source model.

In principle, the source impedance Z_s should be the value of Z which satisfies $F(Z)=0$. However, because the measured input data would always be deteriorated by small errors, in this case, the value of Z minimizing F is chosen approximately as Z_s. The impedance value Z can be separated into real and imaginary parts as $Z=x+iy$ and the necessary condition for the minimum F is given by

$$\frac{\partial F(x,y)}{\partial x}=\frac{\partial F(x,y)}{\partial y}=0. \tag{4}$$

The condition of Eq. (4) implies

$$G(Z)\equiv\sum_{m<n}\left\{\frac{|P_m|^4}{|Z_m|^4}|Z_m+Z|^2(Z_m+Z)-\frac{|P_m|^2|P_n|^2}{|Z_m|^2|Z_n|^2}\times|Z_n+Z|^2(Z_m+Z)+\frac{|P_n|^4}{|Z_n|^4}|Z_n+Z|^2(Z_n+Z)-\frac{|P_m|^2|P_n|^2}{|Z_m|^2|Z_n|^2}|Z_m+Z|^2(Z_n+Z)\right\}=0. \tag{5}$$

Equation (5) can be arranged into separate real and imaginary parts to give a pair of nonlinear cubic equations in the two variables of resistance and reactance. If the polar representation of the impedances is employed in order to get the suitable solution, one can obtain the two cubic polynomial equations after substituting $re^{i\theta}$ for Z, so that the angle θ and the corresponding magnitude r satisfying the equations can be found in an iterative manner. Because Eq. (5) is the necessary condition for F having a minimum value, the value of Z minimizing F among several solutions should be chosen as the source impedance Z_s. After obtaining the source impedance, the source strength can be easily obtained from Eq. (1) as

$$|P_s|=\frac{\sum_m |P_m|\left|\frac{Z_m}{Z_m+Z_s}\right|}{\sum_m\left|\frac{Z_m}{Z_m+Z_s}\right|^2}. \tag{6}$$

B. Acoustical load

An open pipe is often used as the acoustical load when applying the multiload method on the intake or exhaust sides of internal combustion engines because the load impedance of a uniform pipe can be accurately determined. When the load set consists of open pipes, the transfer matrix between inlet (1) and outlet (2) of the pipe can be written as

$$\begin{Bmatrix} p_1 \\ U_1 \end{Bmatrix}=e^{iMKl}\begin{bmatrix} \cos Kl & iZ_0\sin Kl \\ iZ_0^{-1}\sin Kl & \cos Kl \end{bmatrix}\begin{Bmatrix} p_2 \\ U_2 \end{Bmatrix}, \tag{7}$$

where U is the volume velocity, l is the length of the open pipe load, Z_0 denotes the characteristic impedance, and $K=(k-i\alpha)/(1-M^2)$ with free-space wave number k, attenuation constant α, and flow Mach number M.[3–5] If the aeroacoustic variables are used, the transfer matrix defined with respect to the convective pressure and velocity is the same as that defined in Eq. (7) for a pipe with nonviscous medium.[3] The load impedance Z_l is given by

$$Z_l=\frac{Z_r+iZ_0\tan Kl}{1+i(Z_r/Z_0)\tan Kl}, \tag{8}$$

where Z_r denote the radiation impedance. For negligibly small mean flow velocity, K becomes $(k-i\alpha)$ and, for low frequencies, α is negligible. Consequently, under these conditions, Eq. (8) can be rewritten by considering the radiation impedance of an unflanged open pipe as

$$Z_l\approx iZ_0\tan(kL_{\text{eff}}), \tag{9}$$

where $L_{\text{eff}}\cong l+0.6r$ and r is the radius of the pipe. Note that the approximation of $Z_r\approx iZ_0\tan(0.6kr)$ is employed in deriving Eq. (9) and $0.6r$ is the end correction. Because Eq. (9) is periodic, the following can be defined:

$$kL_{\text{eff}}=2\pi L_{\text{eff}}/\lambda=m\pi+\theta. \tag{10}$$

Here, m is 0 or a positive integer and $0\leqslant\theta<\pi$. Pipes with different lengths but similar θ values give almost the same load impedances. It is also noted that the variation of the real part of the load impedance between different loads is very small. This will make the problem ill-conditioned and may cause large errors in the estimated source parameters.[6] Therefore, the choice of a load set which is least sensitive to the input error is necessary for obtaining accurate results.

III. NUMERICAL SIMULATION AND DISCUSSION

A. Preliminary investigation on the choice of loads

During the source measurement by using the multiload method, the acousticians always want to know what combination of the load impedances will yield the best result. In order to answer this question, three types of source impedance are considered and different load sets are used for the simulation of resultant errors in the source impedance due to input errors.

For *a priori* known source parameters, the pressures at the source–load interface for different loads can be calculated and they are utilized as the input data to perform the subsequent simulations. The error simulation was performed in the following way. First, the radiated pressure spectrum $\hat{P}_l$ was calculated by using the given source impedance $\hat{Z}_s$ (hereafter, a variable with a caret denotes *a priori* given one), source strength $\hat{P}_s$, and load impedance $\hat{Z}_l$. Second, in order to generate a certain amount of relative input error, random noises with uniform distribution were added to the radiated sound spectrum ($\tilde{P}_l=\hat{P}_l+\epsilon_p$; hereafter, a variable with a

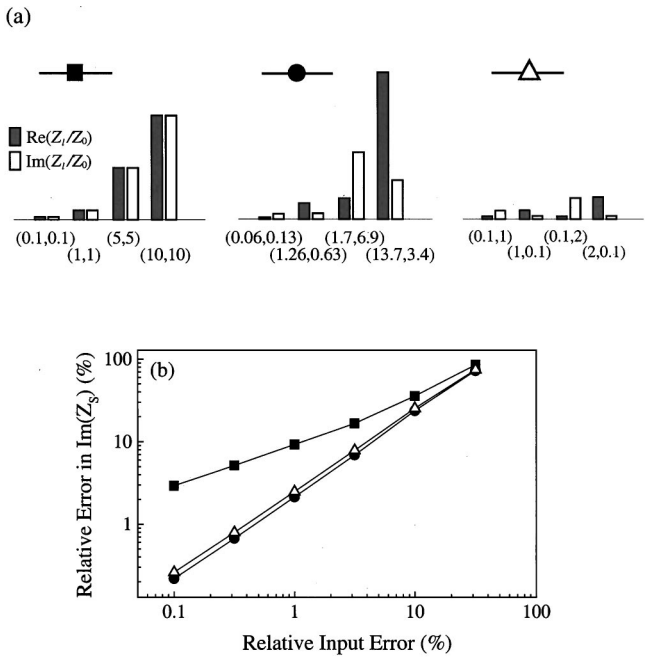


FIG. 2. Comparison of the error sensitivities of the resultant source impedance ($\hat{Z}_s/Z_0=1+1i$) to the relative input error of three types of loads. (a) Three different load sets (four loads for each set), (b) calculated error sensitivities.

tilde denotes that it is contaminated by the error) and to the load impedances ($\tilde{Z}_l=\hat{Z}_l+\epsilon_z$). Third, the load impedance data $\tilde{Z}_l$ and the corresponding radiated pressure spectra $\tilde{P}_l$, which are contaminated by input noise, were used in the multiload method and the source impedance $\tilde{Z}_s$ was obtained. By comparing with the *a priori* given source impedance $\hat{Z}_s$, the relative error $e_s(=\tilde{Z}_s-\hat{Z}_s)$ was calculated. The analyzed data from 100 random error samples were averaged in order to decrease the variation in the result caused by the stochastic variation and the average relative errors in the source impedance results were finally obtained. The same source strength $\hat{P}_s$ was used for all calculations and the load impedance data $\hat{Z}_l$ were normalized by the characteristic impedance.

To start with, the source impedance having the same magnitude of real and imaginary part is considered, in which it is set as $\hat{Z}_s/Z_0=1+1i$, for an example. The first set of loads in Fig. 2(a) is consisted of four loads that have the same magnitude of real and imaginary parts for each load, but the magnitude of impedance is different for four loads. The second set of loads in Fig. 2(a) is consisted of four loads that have quite different magnitudes of real and imaginary parts and the magnitude of impedance is different for four loads. The third set of loads in Fig. 2(a) again consists of four loads that have different magnitudes of real and imaginary parts. In this data set, two groups of load data exist: each group consists of two load impedances that the magnitudes of real and imaginary parts of one load are the same with the magnitudes of imaginary and real parts of another load, respectively. It is noted that the second set of loads has more pronounced difference in the magnitude of load impedances than the third set of loads. Also, it should be mentioned that the imaginary parts of load impedances are all positive and their real parts are of course positive, which is natural considering the characteristics of actual open pipe loads. Figure 2(b) shows the results of source reactance. Actually, there exists very small difference less than about 10% between the relative (or percentage) errors of source resistance and reactance. This small difference is owing to the fact that the real and imaginary parts of source impedance are exactly the same and each load set has similar trends in real and imaginary parts. In this regard, the result of source resistance is omitted here. One can find that the resultant error in source impedance is large when the first set of loads is employed, i.e., when the multiload method employs the loads having nearly the same magnitudes of real and imaginary parts, the error in the resulting source impedance will be large. It is also found that the resultant error is irrelevant to the types of loads when the input error is larger than about 10%.

In the foregoing case, the load impedances were in the same phase plane. The next simulation case is for the load sets having different signs in some involved load reactances, but the magnitudes of real and imaginary parts are set different, based on the suggestion from the foregoing simulation. Two types of source impedance are considered and two different load sets are used for the simulation of resultant errors in the source impedance due to input errors. The first set of loads in Fig. 3(a) consists of four loads that have the same absolute magnitude in imaginary part and the magnitude of real parts are far larger than their imaginary parts. The two loads out of four have the same positive magnitude in real part, but the sign of the imaginary part is different from each other. The second set of loads in Fig. 3(a) consists of four loads that have the same positive magnitude in the real part and the magnitudes of imaginary parts are far larger than their real parts. The magnitudes of imaginary parts are all different and their signs are either positive or negative. Figure 3(b) shows the result for very resistive source impedance such as $\hat{Z}_s/Z_0=10+0.1i$, for an example. Except for the case when the input error is small, which is less than about 0.3%, the second load set produces far accurate results in both real and imaginary parts of source impedance. This might be due to the resistive nature of the source impedance as well as the alternating phases in the load set. It is quite hard to say what is the actual cause for this consistently low error in the resultant source impedance. Figure 3(c) can answer this question. In this case, the source impedance with very reactive nature is employed, i.e., $\hat{Z}_s/Z_0=0.1+10i$. As can be seen in Fig. 3(c), the second load set yields smaller error in the source impedance again and one can think that the alternating phases in the load set are important for yielding the accurate source impedance by using the experimental multiload method. Another fact one can find in this figure is that the percentage error of the source resistance is smaller than that of source reactance when the source impedance is resistive, and vice versa.

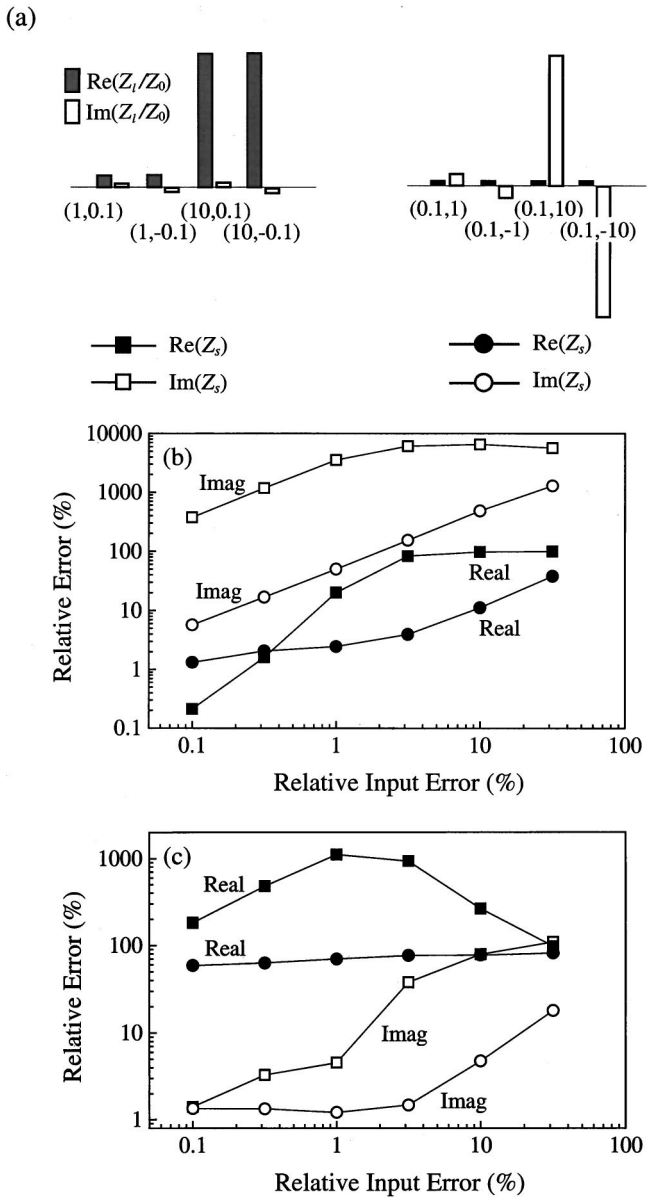


FIG. 3. Comparison of the error sensitivities of the resultant source impedance to the relative input error of two types of loads. (a) Two different load sets (four loads for each set), (b) calculated error sensitivity for $\hat{Z}_s/Z_0 = 10+0.1i$, (c) calculated error sensitivity for $\hat{Z}_s/Z_0 = 0.1+10i$.

B. Preferred load combination

In order to find a load set that will yield accurate source properties in the case of open pipe loads, the effects of input errors on the calculated results have been studied numerically by using the same method in the foregoing simulations. Here, the source impedance values that can be usually found in actual fluid machine sources were considered. Normalized, measured source resistance of practical fluid machines is found to be usually within the range of 0–2 and the normalized source reactance is in the ± 2 range.[1–3,6] It should be mentioned that, in applying the multiload method to the internal combustion engines or some of the compressors,[1,6] the negative source resistances have been measured, but the results are due to the violation of basic assumptions.[7,8] Ten evenly spaced values in each range were chosen and all possible combinations of the resistance and reactance were used

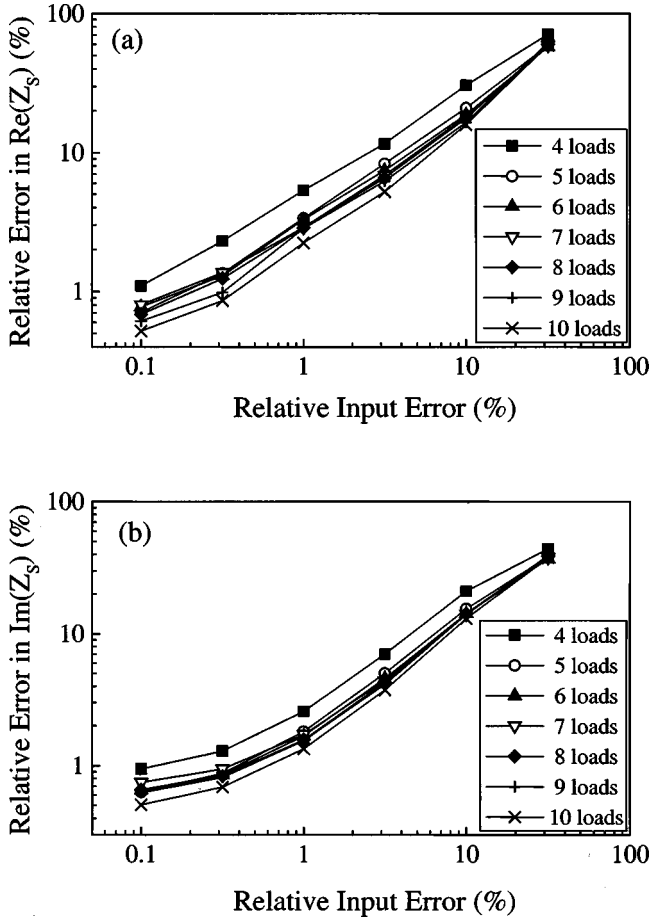
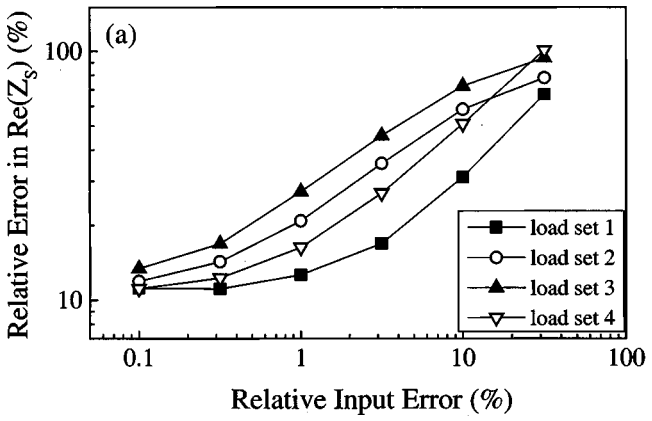


FIG. 4. Effect of the number of loads on the error sensitivity of the resultant source impedance to the relative input error. (a) Resistance, (b) reactance.

as was done for $\hat{Z}_s$ in the error analysis. In the following figures, therefore, each data point is the result obtained by using the 100 source impedance values and 100 random errors, i.e., averaging 10 000 samples.

In Fig. 4, a simulated result is shown for the effect of the number of loads on the error sensitivity of the source impedance to input error. Starting from the four loads that were selected arbitrarily from the usual range of tube impedance values, other loads are added one by one to the formerly chosen loads. The normalized load resistance was selected from the 0.1 to 0.2 range, which is a typical range of actual pipes in length range 2–4 m at frequencies 200–800 Hz, and the normalized load reactance from the ± 2 range. Although a definite improving trend in sensitivity can be observed with an increase of the number of loads, the amount of decrease in relative error is small for each increment in the number of loads. However, for a 10% input error and using five loads, the relative error in the source resistance is expected to be about 10% smaller than is the case using four loads. The reduction of error in reactance under the same condition is about 5%. It seems that the improvement in source impedance estimation is marginal when using more than five or six loads. The fluctuation of data with increase of the number of loads is within $\pm 0.5\%$ and it is thought that this is caused from the arbitrariness in choosing the loads and the statistical variation in the subsequent simulation.



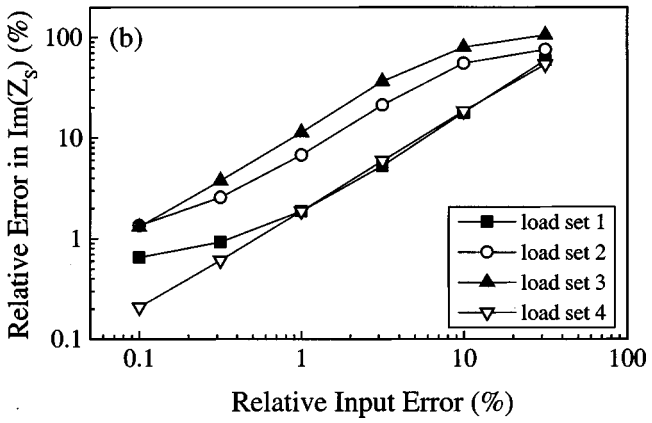


FIG. 5. Comparison of the error sensitivities of the resultant source impedance to the relative input error in applying different load sets. (a) Resistance, (b) reactance.

In Fig. 5, a comparison is given for error sensitivities of the resultant source impedance to the relative input error in applying different load sets. Each load set consists of three open pipes and the load impedances are determined by using Eqs. (9) and (10) varying the angle θ as in Table I. For the real part of the source impedance, the load set 1 yields a relative error, which is smaller than that for the other load sets. The resultant source impedance values from applying load sets 2 and 3 are more sensitive to the input error and each load set contains a pair of loads that have similar θ values. Load sets 1 and 4 yield far smaller error in the imaginary part of the source reactance than the other load sets. Actually the values of θ in load set 1 are distributed with a uniform interval, while the variance in the load reactance is largest in the case of load set 4.

Figure 6 shows the effect of the choice of loads on the error sensitivity of the resultant source impedance to the input error. Sets of load impedances are defined by n (n

TABLE I. Four sets of open pipe loads used for the numerical simulation.

Load set no.	1	2	3	4
$(\theta_1, \theta_2, \theta_3)$	$(0,\pi/3,2\pi/3)$	$(0,\pi/5,7\pi/9)$	$(0,\pi/9,\pi/3)$	$(0,5\pi/11,6\pi/11)$
$\tan\theta$	(0,1.73, −1.73)	(0,0.73, −0.84)	(0,0.36, 1.73)	(0,6.96, −6.96)

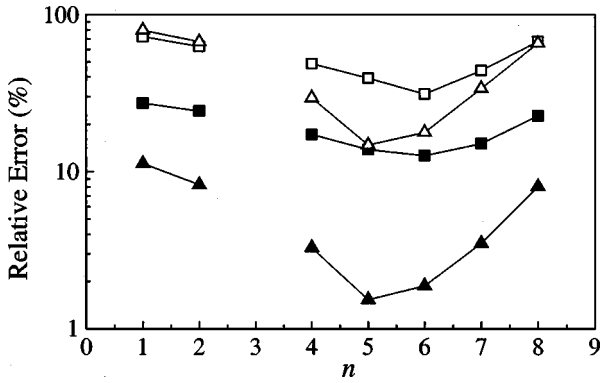


FIG. 6. A comparison of the error sensitivities of the resultant source impedance to the relative input error for varying the values of n: (■, □) the real part of the source impedance; (▲, △) the imaginary part of the source impedance: (—■—, —▲—), 1% relative input error; (—□—, —△—), 10% relative input error.

$=1,2,...,8$) and consist of the three loads $\theta=0,\pi/3, n\pi/9$. In this case, Eq. (10) can be modified as

$$2\pi L_{\text{eff}}/\lambda = m\pi + n\pi/9. \tag{11}$$

The results for $n=0$, 3, and 9 are omitted because these values give a pair of loads having identical impedance values. It is observed that the imaginary part of the source impedance is least sensitive to the input error for the $n=5$ case, where the variance of the load reactance is largest of all. Here, the ''variance'' means a squared sum of any possible difference of two impedance values, i.e., such a sum of three sets of samples. For example, a load combination of $\theta_2 = \pi/3$ and $\theta_3=5\pi/9$ has a large variance because $\tan\theta_3$ is a very large negative value ($\sim\infty$). Figure 7 illustrates the variation of $\tan\theta_3$ with the change of n. Compared to the reactance case, a somewhat different trend can be seen in the error sensitivity of the estimated source resistance, i.e., the relative errors are smallest for $n=6$ case, where the range of θ is distributed with a uniform interval. In order to obtain the real part of the source impedance accurately, the load set having uniformly distributed θ is preferable. It should be reminded that the parameter θ depends on the frequency and the optimal lengths of open pipe loads are given differently for each frequency. If the number of load sets is restricted as is usually practiced, the optimal load sets can be determined from the aforementioned guidelines at the major frequency components, viz., the fundamental firing frequency and its harmonics in the internal combustion engines.

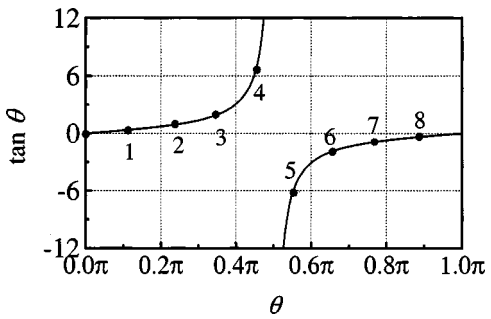


FIG. 7. Graphical representation of the load. The numbers denote n value in Eq. (11).

IV. CONCLUSIONS

The effects of input error on the estimated source parameters have been numerically studied for the case of the multiload method, varying the number of loads and their impedances in order to determine what combination of loads will yield the best result. It was shown that the improvement in estimating the source impedance is marginal when using more than six or seven loads, although a definite improving trend in error sensitivity can be observed with an increase of the number of loads. The effect of the choice of loads was studied with the parameter θ defined for describing the impedance of open pipe loads. It was found that a load set of uniformly distributed θ is preferable for accurate estimation of the source resistance. On the other hand, a load set in which the variance in the load reactance is large is preferable for obtaining the source reactance with precision. In addition, it is observed that the load combination having two different signs and nonuniform magnitude of load reactance values are preferable for the precision of the resultant source impedance. The results of the present study would be very useful in choosing the appropriate length of open pipe or side-branch loads in practical measurement situations.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Keith S. Peat at Loughborough University in the U.K. for helpful comments. This work has been partially supported by BK21 Project and NRL.

[1] S.-H. Jang and J.-G. Ih, ''Refined multiload method for measuring acoustical source characteristics of an intake or exhaust system,'' J. Acoust. Soc. Am. **107**, 3217–3225 (2000).

[2] M. G. Prasad, ''A four load method for evaluation of acoustical source impedance in a duct,'' J. Sound Vib. **114**, 347–356 (1987).

[3] M. L. Munjal, *Acoustics of Ducts and Mufflers* (Wiley Interscience, New York, 1987).

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

[5] R. M. Munt, ''Acoustic transmission properties of a jet pipe with subsonic jet flow I. The cold jet reflection coefficient,'' J. Sound Vib. **142**, 413–436 (1990).

[6] H. Bodén, ''On multi-load methods for determination of the source data of acoustic one-port sources,'' J. Sound Vib. **180**, 725–743 (1995).

[7] K. S. Peat and J.-G. Ih, ''An analytical investigation of the indirect measurement method of estimating the acoustic impedance of a time-varying source,'' J. Sound Vib. **244**, 821–835 (2001).

[8] J.-G. Ih and K. S. Peat, ''On the causes of negative impedance in the measurement of intake and exhaust noise sources,'' Appl. Acoust. **63**, 153–171 (2002).

An approximation to the far field and directivity of elastic wave transducers

Maarten C. M. Bakker[a)]
Section of Structural Mechanics, Faculty of Civil Engineering and Geosciences, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands

Martin D. Verweij[b)]
Laboratory of Electromagnetic Research, Faculty of Information Technology and Systems, Centre for Technical Geoscience, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands

(Received 30 April 2001; revised 16 October 2001; accepted 18 October 2001)

The far field of surface mounted elastic wave transducers, as used in Non-Destructive Testing (NDT), is analyzed in order to understand the mechanisms that determine the directivity of these devices. To model the physical configuration, a moving distribution of normal traction, acting at the free surface of an elastic half-space, is used. The exact field expression consists of a spatial superposition of the elementary waves emitted by the hypothetical point sources that make up the finite source. The current aim is to derive approximate far field expressions that are physically instructive and numerically fast. This is achieved through a paraxial approximation of the elementary waves and a Taylor approximation of their arrival times. The approximate expressions thus obtained, consist of a time convolution of a paraxial wave and a transfer function. The latter represents the influence of the source geometry. Besides elucidating the mechanisms of directivity, the analysis renders a rule of thumb for synthesizing a source with a desired directivity pattern. Numerical experiments show that the approximations reduce the computational effort by a factor of 1000, and that the accuracy of the approximate far field and directivity proves excellent at distances beyond a few times the characteristic source size. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1428542]

PACS numbers: 43.20.Px, 43.20.Rz, 43.40.Le [ANN]

I. INTRODUCTION

The wave field of a finite (not infinitely small) source of normal traction, acting on an elastic half-space, has been the subject of many investigations, often in the context of Non-Destructive Testing (NDT). Much research has been focused on deriving an approximation of the exact expression in order to reduce the computational effort.[1,2] To gain some physical insight into the problem, plane wave approximations[3] or a high frequency approximation[4,5] may be used. Exact analytical expressions have been derived for a number of nonmoving sources, for example, a point source,[6] a line segment source,[7,8] a circular source,[9–12] an annular source,[13] or a rectangular source.[7,8,14] Exact analytical expressions have also been derived for a source moving with a constant speed[6,8,15,16] or expanding at a constant rate.[17]

The exact expression for a finite source of normal traction is in general obtained as a spatial superposition of the elementary waves emitted by the point sources that may be thought to form the finite source. A direct numerical evaluation of the resulting double integral over the source surface has two major drawbacks. First, it offers no direct insight into the nature of the wave field, or how properties like the directivity depend on the source parameters or the material parameters. Second, the numerical evaluation may be extremely time consuming. This is especially the case when relatively high frequencies are involved, as the discretization of the spatial integrals is linked to the time sampling, which must be relatively fine in view of Shannon's sampling theorem.

In this paper we consider the space–time domain wave field due to a moving distribution of normal traction, acting on the otherwise traction free surface of an elastic half-space. This type of source models the effect of an angle beam transducer, as frequently applied in NDT.[18] For this configuration we aim to derive approximate far field expressions that are physically instructive and numerically fast. To achieve this, we subsequently subject the individual point source solutions to a paraxial approximation and an arrival time approximation. This enables us to replace the spatial superposition by a time convolution of (1) a paraxial wave and (2) a transfer function that represents the influence of the geometry of the source. If the traction is uniformly distributed over the source surface, the transfer function may often be obtained by analytical means. In case of a nonuniformly distributed traction, evaluation of the transfer function requires a low cost single spatial integration.

In our analysis we will prove that the directivity of the considered source type is largely dependent on the maximum arrival time difference. Since this parameter depends directly on the source geometry, this aspect of the analysis renders it valuable for synthesis purposes, i.e., for designing a source with a desired directivity pattern. On the other hand, our theory implies that a measured wave field from the finite source may be transformed into the wave field of a point

[a)]Electronic mail: m.c.m.bakker@ct.tudelft.nl
[b)]Electronic mail: m.d.verweij@its.tudelft.nl

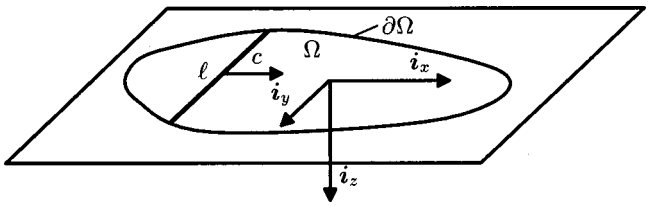


FIG. 1. An arbitrarily shaped, fixed source domain Ω with edge $\partial\Omega$ that determines the actual length of the moving line segment ℓ. The fixed Cartesian reference frame is used to indicate the position $\mathbf{x}=x\mathbf{i}_x+y\mathbf{i}_y+z\mathbf{i}_z$ of a point in the half-space.

source by deconvolution with the transfer function. This offers interesting applications in the field of signal processing in NDT practice. Our final expressions allow a very fast numerical evaluation of the complete space–time domain elastic wave field. We will show that the accuracy of the approximate far field expressions is excellent for distances that are much larger than the dimensions of the source, while the accuracy remains quite good at distances that are in the order of the source dimensions.

We start in Sec. II with the description of the configuration and the presentation of the exact space–time domain expression. Section III deals with the far field approximations and leads in a natural way to our definition of the exact transfer function. Section IV contains an approximation of the arrival time, which we apply in Sec. V to obtain a convenient approximation for the transfer function. In Sec. VI our analysis is applied to a line segment load moving within an elliptical source domain. Here, numerical results for the approximate wave field and the approximate directivity are compared with those for the exact case. Next, in Sec. VII the accuracy of the theory is related to the error in the maximum arrival time difference. We discuss the relation between this quantity and the directivity in Sec. VIII. Finally, in Sec. IX we draw some conclusions.

II. CONFIGURATION AND EXACT SPACE–TIME DOMAIN EXPRESSION

First we model the effect of an angle beam transducer with an aperture Ω that is coupled to the surface of an elastic half-space through a thin layer of lowly viscous fluid. For this purpose we consider a homogeneous, isotropic, elastic half-space $z\geqslant 0$ that is loaded with a line segment ℓ of normal traction, see Fig. 1.

The line load is directed parallel to the y axis and moves along the x axis with a constant speed c. At each time instant t the length of ℓ is such that it exactly fits inside a fixed source domain Ω with edge $\partial\Omega$. The line load becomes active at $t=0$, when it has entered Ω at $x=-a$. At that instant, it begins to move across Ω. The line becomes inactive again at $t=T^c=2a/c$, when it will leave Ω at $x=a$. The traction of this moving line segment load is given by the normal component

$$T_z(\mathbf{x},t)=[H(t)-H(t-T^c)]\,g^{\mathrm{amp}}(x,y)\,\delta[x-x'(t)]. \tag{1}$$

This models the normal component of the traction imposed on the half-space by the transducer. Since pressure is supported by the coupling fluid, the normal traction is continuous when going from the transducer to the half-space. Both transversal components of the traction imposed on the half-space are zero since shear is not supported by the lowly viscous coupling fluid.[18] In the x direction the position of ℓ is given by the source coordinate $x'(t)=ct-a$. To accommodate local variations of the imposed traction, we have introduced the traction amplitude $g^{\mathrm{amp}}(x,y)$, with $g^{\mathrm{amp}}(x,y)=0$ for (x,y) outside Ω. The source in Eq. (1) represents a transducer emitting an impulsive wave. To model a transducer that emits a nonimpulsive wave we change Eq. (1) into

$$T_z(\mathbf{x},t)=g(t)*_t\{[H(t)-H(t-T^c)]\,g^{\mathrm{amp}}(x,y)\,\delta[x-x'(t)]\}. \tag{2}$$

Here, we have introduced the source signature $g(t)$, with $g(t)=0$ for $t<0$. The symbol $*_t$ denotes a time convolution. In fact, the moving line segment of normal traction described by Eq. (1) has broadened to a moving distribution of normal traction in Eq. (2).

Inside the elastic half-space, the particle velocity $\mathbf{v}=v_x\mathbf{i}_x+v_y\mathbf{i}_y+v_z\mathbf{i}_z$ of the generated wave field satisfies the differential equation

$$(\lambda+\mu)\boldsymbol{\nabla}\boldsymbol{\nabla}\cdot\mathbf{v}+\mu\boldsymbol{\nabla}^2\mathbf{v}=\rho\partial_t^2\mathbf{v}, \tag{3}$$

where λ and μ are the Lamé coefficients of the half-space and ρ is its mass density. Assume that we know the space–time domain Green's function $G_k(\mathbf{x},t)$ that renders, for each amplitude J of a point source of normal traction $T_z=J\,\delta(t)\,\delta(x)\,\delta(y)$, the particle velocity as $v_k=J\,G_k$. Then the exact space–time domain expression for the particle velocity due to a transducer with a normal traction as in Eq. (2) is[8]

$$v_k(\mathbf{x},t)=g(t)*_t\int_0^{\min\{t,T^c\}}\int_{\ell(\tau')}g^{\mathrm{amp}}(x',y')\times G_k(\mathbf{x}-\mathbf{x}',t-\tau')\,dy'\,d\tau'. \tag{4}$$

Here, k denotes either x, y or z, and $\ell(\tau')$ is the line segment at instant τ'. Further, $\mathbf{x}'=(x',y',0)$ indicates the position of the points on ℓ, while dy' is equal to $d\ell$ in the chosen reference frame. From now on we replace the upper integration limit by T^c, which is usually allowed in the far field.

III. DECOUPLED WAVES AND PARAXIAL APPROXIMATION

A. Decoupled wave fields

First consider a point source located on the surface of the elastic half-space. Several wavelengths away from the surface, the wave field inside the half-space may locally be decomposed into independently propagating plane wave constituents of compression, shear and head wave type. In the surface layer these constituents are no longer independent due to wave coupling at the surface. However, several wavelengths away from the source, the wave field in this layer is dominated by the Rayleigh wave, which may be regarded as yet another type of wave. Consequently, it is safe to assume that the different wave constituents generated by the finite

source are also decoupled in the far field and may be considered separately. Due to the small wavelengths commonly employed in NDT, this already applies for relatively small distances.

This analysis is facilitated by the Cagniard–De Hoop method,[8,19,20] which allows a unique decomposition of the exact Green's function according to

$$G_k(\mathbf{x},t)=G_k^P(\mathbf{x},t)+G_k^S(\mathbf{x},t)+G_k^H(\mathbf{x},t). \tag{5}$$

The process to obtain the exact Green's functions G_k^P, G_k^S, and G_k^H is explained in the Appendix. In view of Eq. (5) the particle velocity may be written as

$$v_k(\mathbf{x},t)=v_k^P(\mathbf{x},t)+v_k^S(\mathbf{x},t)+v_k^H(\mathbf{x},t). \tag{6}$$

The superscripts *P*, *S*, and *H* refer to a compression wave, a shear wave, and a head wave, respectively. Although these waves cannot be generated separately, we will assume that for our purpose these may be considered independently. Then, G_k^P is representative only for the compression waves, G_k^S is representative only for the shear waves and G_k^H is representative only for the head waves. The results will show that in the far field these assumptions are accurate. The Rayleigh waves that occur are incorporated in G_k^P, G_k^S, and G_k^H. Therefore, the Rayleigh waves need not be addressed separately. Since the remaining treatment applies to either type of wave, from now on we will omit the superscripts *P*, *S*, and *H* and we will describe our theory for one general type of bulk wave.

B. Paraxial approximation

For the type of wave under consideration, the particle velocity component v_k due to a point source $T_z=J\,g(t)\,\delta(x')\,\delta(y')$ is

$$v_k(\mathbf{x},t)=J\,g(t)*_t G_k(\mathbf{x}-\mathbf{x}',t). \tag{7}$$

For each source point the length and the angle of the path to the point of observation is different. As a result, both the shape and the arrival time of the waves differ slightly for different source points. In the far field, the differences between the various waves diminish and we may replace all the waves by one paraxial wave W_k^{par} due to a point source at a particular position $\mathbf{x}^p=(x^p,y^p,0)$. For this position we take the source point of the wave with the minimum arrival time. For each type of wave the point $\mathbf{x}^p$ may be determined by numerical means with little effort. Replacing in Eq. (4) the Green's function $G_k(\mathbf{x}-\mathbf{x}',t)$ by the paraxial Green's function $G_k(\mathbf{x}-\mathbf{x}^p,t)$, we find the preliminary paraxial approximation

$$v_k(\mathbf{x},t)\approx g(t)*_t G_k(\mathbf{x}-\mathbf{x}^p,t)*_t\int_0^{T^c}\int_{\mathscr{C}(\tau')} g^{\mathrm{amp}}(x',y')\times\delta(t-\tau')\,dy'\,d\tau', \tag{8}$$

in which the first part is the paraxial wave

$$W_k^{\mathrm{par}}(\mathbf{x},t)=g(t)*_t G_k(\mathbf{x}-\mathbf{x}^p,t). \tag{9}$$

In this paper we will employ the exact paraxial Green's function. This means that our paraxial approximation *only* involves the replacement of $G_k(\mathbf{x}-\mathbf{x}',t)$ by $G_k(\mathbf{x}-\mathbf{x}^p,t)$ and does *not* imply a further approximation of $G_k(\mathbf{x}-\mathbf{x}^p,t)$. The paraxial approximation just introduced does no longer account for the correct arrival time of the wave from an individual point source. To improve this aspect, we introduce the additional time delay $T(\mathbf{x}',\mathbf{x}^p)$ that, for the given point of observation $\mathbf{x}$, exists between the wave from a source point $\mathbf{x}'$ and the paraxial point $\mathbf{x}^p$. Doing so, we obtain the paraxial approximation

$$v_k(\mathbf{x},t)\approx g(t)*_t G_k(\mathbf{x}-\mathbf{x}^p,t)*_t\int_0^{T^c}\int_{\mathscr{C}(\tau')} g^{\mathrm{amp}}(x',y')\times\delta[t-\tau'-T(\mathbf{x}',\mathbf{x}^p)]\,dy'\,d\tau'. \tag{10}$$

In this approximation, the aspects that are related to the propagation of the waves in the elastic half-space, such as the geometrical damping, are accounted for by the paraxial Green's function. The aspects that are related to the geometry of the source, in particular the difference in arrival time and traction amplitude between the source points, is accounted for by the double integral. This is exactly the role we have in mind for the transfer function $M(\mathbf{x},t)$, so we define

$$M(\mathbf{x},t)=\int_0^{T^c}\int_{\mathscr{C}(\tau')} g^{\mathrm{amp}}(x',y')\times\delta[t-\tau'-T(\mathbf{x}',\mathbf{x}^p)]\,dy'\,d\tau'. \tag{11}$$

Due to the choice of the paraxial source point we have obtained a causal transfer function, i.e., $M(\mathbf{x},t)=0$ for $t<0$. Once M is known for a given source, the paraxial approximation for the particle velocity in the far field is obtained as

$$v_k(\mathbf{x},t)\approx W_k^{\mathrm{par}}(\mathbf{x},t)*_t M(\mathbf{x},t). \tag{12}$$

Due to the occurrence of the double integral in Eq. (11), the evaluation of the exact function M is time consuming. In Sec. V we will show that this problem may be eliminated by an appropriate transformation of the integration variables.

IV. ARRIVAL TIME ANALYSIS

A. Exact arrival times

In a given point of observation $\mathbf{x}$ in the half-space, the absolute value of the difference between the arrival times of two waves emitted from the source points $\mathbf{x}^a=(x^a,y^a,0)$ and $\mathbf{x}^b=(x^b,y^b,0)$ is

$$T(\mathbf{x}^a,\mathbf{x}^b)=\frac{\|\mathbf{x}-\mathbf{x}^a\|-\|\mathbf{x}-\mathbf{x}^b\|}{C}+\frac{x^a-x^b}{c}, \tag{13}$$

where C represents the wave speed of the considered type of wave. The path traveled by the two waves is depicted in Fig. 2.

The second term in the right-hand side of Eq. (13) is due to the motion of the distributed load. Without any restriction for the analysis we may assume $x^a\geqslant x^b$, as depicted in Fig. 2. When $T=0$, the two waves have the same arrival time, which means they are *focused*. By setting T in Eq. (13) equal

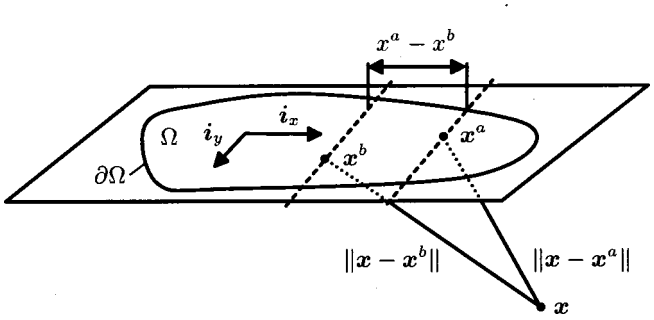


FIG. 2. Two point sources emitting waves that are received at **x**. The horizontal distance $x^a - x^b \geqslant 0$ determines the difference in delay caused by the motion of the distributed load.

to zero, we can find the locations of all source points yielding focused waves at a given arrival time and at a given point of observation. The collection of all these source points forms (one or more parts of) an elliptical curve within the source domain. The maximum value of T, denoted as ΔT, gives the time difference between the first and last arriving waves.

B. Approximate arrival times

The analysis is facilitated by introducing spherical coordinates (R,θ,ϕ) for the point of observation (x,y,z) according to

$$x = R\ \cos(\phi)\sin(\theta), \tag{14}$$

$$y = R\ \sin(\phi)\sin(\theta), \tag{15}$$

$$z = R\ \cos(\theta), \tag{16}$$

where $R=\sqrt{x^2+y^2+z^2}$ is the radius, $0 \leqslant \theta \leqslant \pi/2$ is the vertical angle of observation measured from the positive z axis and $-\pi < \phi \leqslant \pi$ is the horizontal angle of observation measured from the positive x axis. Now we are able to approximate the arrival time difference T in Eq. (13) for large R. By taking the first-order Taylor approximation around $\mathbf{x}^a = \mathbf{0}$ for the square root in $\|\mathbf{x}-\mathbf{x}^a\|$ and around $\mathbf{x}^b = \mathbf{0}$ for the square root in $\|\mathbf{x}-\mathbf{x}^b\|$, we obtain the far field approximation

$$T(\mathbf{x}^a,\mathbf{x}^b) \approx T^{\mathrm{ff}}(\mathbf{x}^a,\mathbf{x}^b) = \left\{ \Delta x \left[\frac{1}{c} - \frac{\sin(\theta)\cos(\phi)}{C} \right] - \Delta y\ \frac{\sin(\theta)\sin(\phi)}{C} \right\}, \tag{17}$$

where $\Delta x = x^a - x^b$ and $\Delta y = y^a - y^b$. The maximum value of T^{ff} is denoted as ΔT^{ff}. By setting T^{ff} in Eq. (17) equal to zero, we can find the approximate locations of all source points yielding focused waves. The elliptical curves from the preceding section now turn out to be approximated by straight lines that make an angle α with the x axis. This angle satisfies

$$\tan(\alpha) = \frac{\Delta y}{\Delta x} = \frac{C/c - \sin(\theta)\cos(\phi)}{\sin(\theta)\sin(\phi)}. \tag{18}$$

As a special case we consider observation points in the vertical half-plane $y=0$, $z \geqslant 0$, in which $\phi = 0$ or $\phi = \pi$, and $0 \leqslant \theta \leqslant \pi/2$. From Eq. (17) we obtain

$$T^{\mathrm{ff}}(\mathbf{x}^a,\mathbf{x}^b) = \begin{cases} \Delta x \left[\dfrac{1}{c} - \dfrac{\sin(\theta)}{C} \right], & \phi = 0, \\ \Delta x \left[\dfrac{1}{c} + \dfrac{\sin(\theta)}{C} \right], & \phi = \pi. \end{cases} \tag{19}$$

Setting $T^{\mathrm{ff}} = 0$, we find straight lines that satisfy $\Delta x = 0$ and $\alpha = \pm \pi/2$, i.e., these lines run parallel to the y axis. Further, we find that all waves emitted by the source are focused if $\phi = 0$ and $\sin(\theta) = C/c$. If such an angle θ exists, it is referred to as the *(vertical) main beam angle* θ^b. For the approximate maximum arrival time difference ΔT^{ff} we find

$$\Delta T^{\mathrm{ff}}(\mathbf{x}^a,\mathbf{x}^b) = \begin{cases} 2a \left| \dfrac{1}{c} - \dfrac{\sin(\theta)}{C} \right|, & \phi = 0, \\ 2a \left| \dfrac{1}{c} + \dfrac{\sin(\theta)}{C} \right|, & \phi = \pi, \end{cases} \tag{20}$$

where $2a$ is the maximum length of the source domain in the x direction.

V. TRANSFER FUNCTION FOR A FINITE SOURCE

A. Exact transfer function

Suppose that we are able to find the elliptical curve $\mathscr{C}^e$ of all source points $\mathbf{x}'$ that emit waves with the same time delay $\tau = T(\mathbf{x}',\mathbf{x}^p)$ with respect to the paraxial wave. Once we know $\mathscr{C}^e$ for each τ, it is easy to span the entire source domain by a continuum of curves $\mathscr{C}^e(\tau)$. This enables us to replace the double integration in Eq. (4), and consequently in Eq. (11), by an integration over the curves and an integration over the time delay.[2,21–23] In that case Eq. (11) becomes

$$M(\mathbf{x},t) = \int_0^{\Delta T} \delta(t-\tau) \times \int_{\mathscr{C}^e(\tau)} g^{\mathrm{amp}}(x',y')\, \frac{\partial(y',\tau')}{\partial(p,\tau)}\, dp\, d\tau. \tag{21}$$

In this equation, the variables x', y', and τ' depend on the new variables p and τ, where p is used to parametrize $\mathscr{C}^e$. Further, $\partial(y',\tau')/\partial(p,\tau)$ is the Jacobian of the transformation. The upper limit of integration ΔT is the time lapse between the first and last arriving waves. The outer integration can be evaluated analytically, so we obtain

$$M(\mathbf{x},t) = [H(t) - H(t-\Delta T)] \times \int_{\mathscr{C}^e(t)} g^{\mathrm{amp}}(x',y')\, \left. \frac{\partial(y',\tau')}{\partial(p,\tau)} \right|_{\tau=t} dp. \tag{22}$$

Although the number of integrations has now been reduced to one, the additional task of determining the elliptical curves in general has to be performed by numerical means. This considerably reduces the computational advantage of Eq. (22) over Eq. (11). Also, the approach reveals little about the driving mechanisms of directivity. This can be overcome by using the approximate curves from Sec. IV B.

B. Approximate transfer function

The approximate curves in Sec. IV B are straight lines ℓ^l that make an angle α with respect to the x axis. To facilitate the integration over the source domain, we first rotate the coordinates (x',y') over the angle α to obtain the new coordinates (p,q) as

$$p = x' \cos(\alpha) + y' \sin(\alpha), \tag{23}$$

$$q = -x' \sin(\alpha) + y' \cos(\alpha). \tag{24}$$

In this way, the position along ℓ^l is indicated by p, while the location of such a line in the source domain is described by q. Subsequently, we relate the location q of ℓ^l with the time delay τ, where $0 \leqslant \tau \leqslant \Delta T^{\mathrm{ff}}$. Taking into account that $x'(\tau') = c\tau' - a$, we find that the Jacobian of the transformation from (y',τ') to (p,τ) equals

$$\frac{\partial(y',\tau')}{\partial(p,\tau)} = \frac{\partial_\tau q}{c}. \tag{25}$$

Replacing in Eq. (21) the exact curves ℓ^e by the approximate curves ℓ^l, and employing the coordinate transformation just described, we obtain the far field approximation

$$M(\mathbf{x},t) \approx M^{\mathrm{ff}}(\mathbf{x},t) = \int_0^{\Delta T^{\mathrm{ff}}} \delta(t-\tau) \times \frac{\partial_\tau q}{c} \int_{\ell^l(\tau)} g^{\mathrm{amp}}(x',y')\, dp\, d\tau, \tag{26}$$

where it has been used that the Jacobian does not depend on p. Evaluating the outer integral analytically, we find the approximate far field transfer function of the finite source as

$$M^{\mathrm{ff}}(\mathbf{x},t) = [H(t) - H(t-\Delta T^{\mathrm{ff}})] \times \frac{\partial_t q}{c} \int_{\ell^l(t)} g^{\mathrm{amp}}(x',y')\, dp. \tag{27}$$

By combining Eqs. (12) and (27), our final approximation for the velocity in the far field is obtained as

$$v_k(\mathbf{x},t) \approx v_k^{\mathrm{ff}}(\mathbf{x},t) = W_k^{\mathrm{par}}(\mathbf{x},t) *_t M^{\mathrm{ff}}(\mathbf{x},t). \tag{28}$$

VI. EXAMPLE: LOAD MOVING WITHIN AN ELLIPTICAL SOURCE DOMAIN

As an example, we determine the approximate far field transfer function M^{ff} for a line segment ℓ moving at a constant speed c and having a length determined by the width of an elliptical source domain Ω, as shown in Fig. 3.

The shape of the source domain and the type of loading are representative for an angle beam transducer on a surface, as commonly encountered in NDT.[3,8,18] For convenience we take g^{amp} constant inside Ω, although in a practical situation it is likely that g^{amp} is nonuniform and falls off at the edges.[24] As a consequence, the normal traction of the source inside Ω is

$$T_z(\mathbf{x},t) = g(t) *_t \{[H(t) - H(t-T^c)]\, g^{\mathrm{amp}}\, \delta(x+a-ct)\}, \tag{29}$$

where a is half the length of the elliptical domain. The load appears at $x=-a$ and vanishes beyond $x=a$. The length of the line segment at $t=(x+a)/c$ follows from

$$\|\ell(x)\| = 2b \sqrt{1 - \frac{x^2}{a^2}}, \tag{30}$$

where b is half the width of the elliptical domain. The approximate far field transfer function for this source may be obtained analytically. With Eq. (27) we directly find

$$M^{\mathrm{ff}}(\mathbf{x},t) = [H(t) - H(t-\Delta T^{\mathrm{ff}})] \frac{\partial_t q}{c} \int_{\ell^l(t)} g^{\mathrm{amp}}\, dp = [H(t) - H(t-\Delta T^{\mathrm{ff}})] \frac{\partial_t q}{c} g^{\mathrm{amp}} \|\ell^l(t)\|. \tag{31}$$

Evaluation of this expression involves a straightforward geometrical procedure, which we demonstrate next.

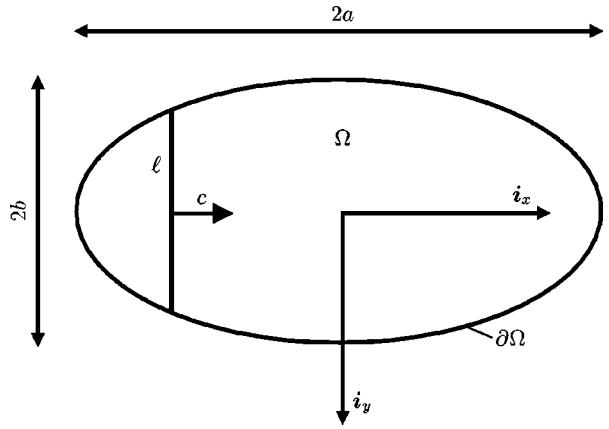


FIG. 3. A line segment load ℓ moving with a constant speed c within an elliptical source domain Ω.

A. Approximate far field transfer function for an elliptical source domain

In terms of the coordinates (x,y), a parametrization of the elliptical boundary $\partial\Omega$ of the source is

$$\left(\frac{x}{a}\right)^2 + \left(\frac{y}{b}\right)^2 = 1. \tag{32}$$

The two points of intersection (x_1,y_1) and (x_2,y_2) of a straight line ℓ^l with the ellipse, see Fig. 4(a), are

$$x_{1,2} = \frac{ab}{b^2 + a^2 \tan^2(\alpha)} \left[\frac{y_0 a \tan(\alpha)}{b} \pm \sqrt{b^2 + a^2 \tan^2(\alpha) - y_0^2} \right], \tag{33}$$

$$y_{1,2} = x_{1,2} \tan(\alpha) + y_0, \tag{34}$$

where y_0 is the offset of the line in the y direction. The length of a line of focal source points is therefore

$$\|\ell^l(y_0)\| = \sqrt{(x_1-x_2)^2 + (y_1-y_2)^2} = \frac{2ab}{Q} \sqrt{1 - \frac{y_0^2 \cos^2(\alpha)}{Q^2}}, \tag{35}$$

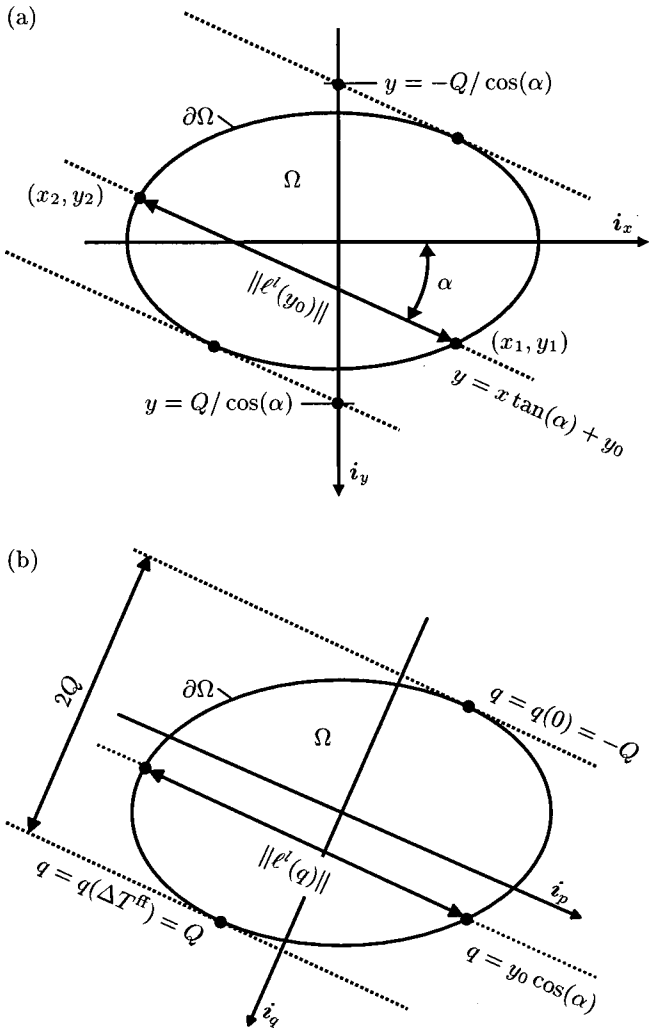


FIG. 4. (a) A line at an angle α and with a given offset y_0 intersects the ellipse $\partial\Omega$ in two points, except if it forms a tangent of the ellipse. (b) After rotating the reference frame through the angle α, the distance between the two points is a function of q and gives the length $\|\nearrow^l(q)\|$ of the interval of integration for p. When q is related with t, one tangent point on the ellipse is associated with $t=0$ and the other one with $t=\Delta T^{\rm ff}$.

where

$$Q=\cos(\alpha)\,\sqrt{b^2+a^2\tan^2(\alpha)}. \tag{36}$$

If $y_0=\pm Q/\cos(\alpha)$ the line is tangent to the ellipse, as depicted in Fig. 4(a). By changing to the coordinates (p,q) defined in Eqs. (23) and (24), we rotate the reference frame in Fig. 4(a) over an angle α, as shown in Fig. 4(b). For $q=y_0\cos(\alpha)$ we find that the length of the line of focal source points is

$$\|\nearrow^l(q)\|=\frac{2ab}{Q}\sqrt{1-\frac{q^2}{Q^2}},\quad -Q\leqslant q\leqslant Q. \tag{37}$$

Further, it is easily found that the relation between q and t is

$$q(t)=\left(\frac{2t}{\Delta T^{\rm ff}}-1\right)Q,\quad 0\leqslant t\leqslant\Delta T^{\rm ff}, \tag{38}$$

so an integration over q runs from $q(0)=-Q$ to $q(\Delta T^{\rm ff})=Q$, see Fig. 4(b). Combining Eqs. (31), (37), and (38), the approximate far field transfer function is found as

$$M^{\rm ff}(\mathbf{x},t)=[H(t)-H(t-\Delta T^{\rm ff})]\times\frac{4ab\,g^{\rm amp}}{c\Delta T^{\rm ff}}\sqrt{1-\left(\frac{2t}{\Delta T^{\rm ff}}-1\right)^2}. \tag{39}$$

B. Further approximations and limiting cases

Suppose that $\Delta T^{\rm ff}$ has been determined for an elliptical source domain. Then we may further approximate $M^{\rm ff}$ in Eq. (39) by the transfer function of a matching rectangular source domain with area πab. This gives

$$M^{\rm ff}(\mathbf{x},t)\approx\frac{\pi ab\,g^{\rm amp}}{c}\,\frac{H(t)-H(t-\Delta T^{\rm ff})}{\Delta T^{\rm ff}}. \tag{40}$$

The benefit is that the dependence on t is now much simpler and we obtain the approximate far field velocities from Eqs. (9) and (28) as

$$v_k^{\rm ff}(\mathbf{x},t)\approx\frac{\pi ab\,g^{\rm amp}}{c\Delta T^{\rm ff}}\,g(t)*_t\int_0^{\Delta T^{\rm ff}}G_k(\mathbf{x}-\mathbf{x}^p,t-\tau)\,d\tau. \tag{41}$$

This implies a time averaging of the constituting waves.

In case of a caustic ($\Delta T^{\rm ff}\to 0$), both the transfer functions in Eqs. (39) and (40) become $(\pi ab\,g^{\rm amp}/c)\,\delta(t)$, and from Eqs. (9) and (28) we find

$$v_k^{\rm ff}(\mathbf{x},t)=\frac{\pi ab\,g^{\rm amp}}{c}\,g(t)*_t G(\mathbf{x}-\mathbf{x}^p,t). \tag{42}$$

If $g(t)=H(t)$ and G_k may be approximated by $\delta'(t-R^p/C)\,\gamma_k$, where $R^p=\|\mathbf{x}-\mathbf{x}^p\|$ and γ_k is a constant vector, Eq. (41) becomes

$$v_k^{\rm ff}(\mathbf{x},t)\approx\frac{\pi ab\,g^{\rm amp}}{c}\,\frac{H(t-R^p/C)-H(t-R^p/C-\Delta T^{\rm ff})}{\Delta T^{\rm ff}}\,\gamma_k. \tag{43}$$

This approximation has first been proposed in a similar form by Chinn,[25] who has validated it with measurements.

C. Numerical results for the half-plane $y=0$, $z\geqslant 0$

To give an impression of the performance of the paraxial and arrival time approximations, for the current example we present some numerical results for a practical 54° angle beam shear wave transducer. The observation points are chosen in the vertical half-plane $y=0$, $z\geqslant 0$, where the origin of the reference frame is chosen in the center of the elliptical source domain that forms the transduction surface of the transducer. The dimensions of the elliptical domain are $a=22.5$ mm and $b=14$ mm, the speed of the line load moving inside the elliptical domain is $c=4000$ m/s, and the traction amplitude is $g^{\rm amp}=100$ N/m. The bulk wave speeds in the half-space are $C^S=3250$ m/s and $C^P=5900$ m/s, and the mass density is $\rho=7800$ kg/m^3. These values are representative for steel. The accompanying Rayleigh wave speed is $C^R=3004$ m/s. The distance R from the origin to the point of observation is expressed in terms of the characteristic source size $R^{\rm src}=\sqrt{(2a)^2+(2b)^2}=53$ mm. In most cases the resulting distance between the source and the points of observation

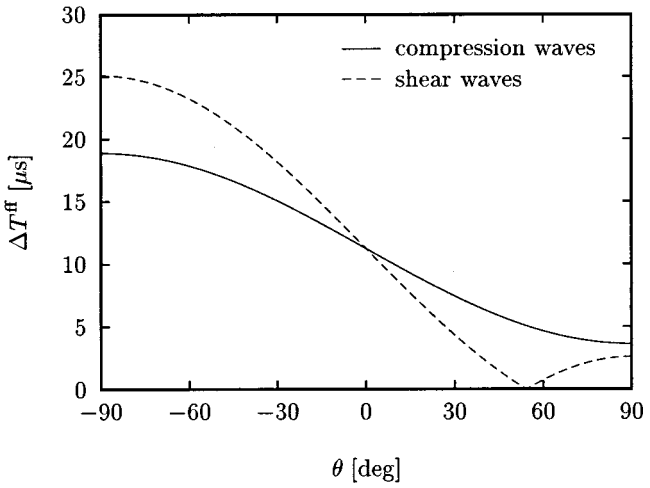


FIG. 5. The approximate maximum arrival time difference ΔT^{ff} for compression waves and shear waves, for points in the half-plane $y=0$, $z\geqslant 0$ at an observation distance $R=5R^{\mathrm{src}}$.

is given a small value in comparison to traditional far field calculations. By doing so, we wish to show the performance of our approximations under unfavorable conditions.

The approximate maximum arrival time difference ΔT^{ff} has been given in Eq. (20) and is depicted in Fig. 5 for both compression waves and shear waves. The figure applies to an observation distance $R=5R^{\mathrm{src}}$. For shear waves ΔT^{ff} vanishes at the main beam angle $\theta^b=54.3°$. When the approximate values are compared with the exact maximum arrival time difference ΔT, our approximate ΔT^{ff} has a relative accuracy of 0.2%. At an even shorter observation distance of $R=2R^{\mathrm{src}}$ the relative accuracy of ΔT^{ff} is still better than 0.9%.

For the next numerical results a specific source signature g is introduced, which is shown in Fig. 6(a). This is a typical pulse shape that has been measured for the practical 54° angle beam shear wave transducer. The center frequency of the pulse is 4.52 MHz and its bandwidth is 4.25 MHz (94% bandwidth).

The other panels in Fig. 6 show a graphical representation of the constituents of Eq. (28) in case of a shear wave. The vertical paraxial shear wave $W_z^{\mathrm{par}}=g*_t G_z^S(\mathbf{x}-\mathbf{x}^p)$, the approximate far field transfer function M^{ff} from Eq. (39), and the approximate vertical shear particle velocity v_z^{ff} from Eq. (28) are shown in Figs. 6(b), (c), and (d), respectively, for a vertical observation angle $\theta=53°$ and an observation distance $R=5R^{\mathrm{src}}$. Figures 6(e), (f), and (g) show the equivalent graphs for a vertical observation angle $\theta=21°$. The graphs in Figs. 6(b) and (e) have been determined using the exact paraxial Green's function G_z^S (see the Appendix). It turns out that the shape of W_z^{par} in Figs. 6(b) and (e) closely resembles the shape of $\partial_t g$. This indicates that, apart from a constant factor, G_z^S approaches a function $\delta'(t-R^p/C^S)$. Further, it is easy to explain the shape of v_z^{ff}. The value of ΔT^{ff} for $\theta=53°$, as plotted in Fig. 5, is quite small in comparison to the duration of the source signature. Therefore, apart from a constant factor, M^{ff} approaches $\delta(t)$ and, in turn, v_z^{ff} closely resembles W_z^{par}. On the other hand, the value of ΔT^{ff} for $\theta=21°$, as also plotted in Fig. 5, is quite large in comparison

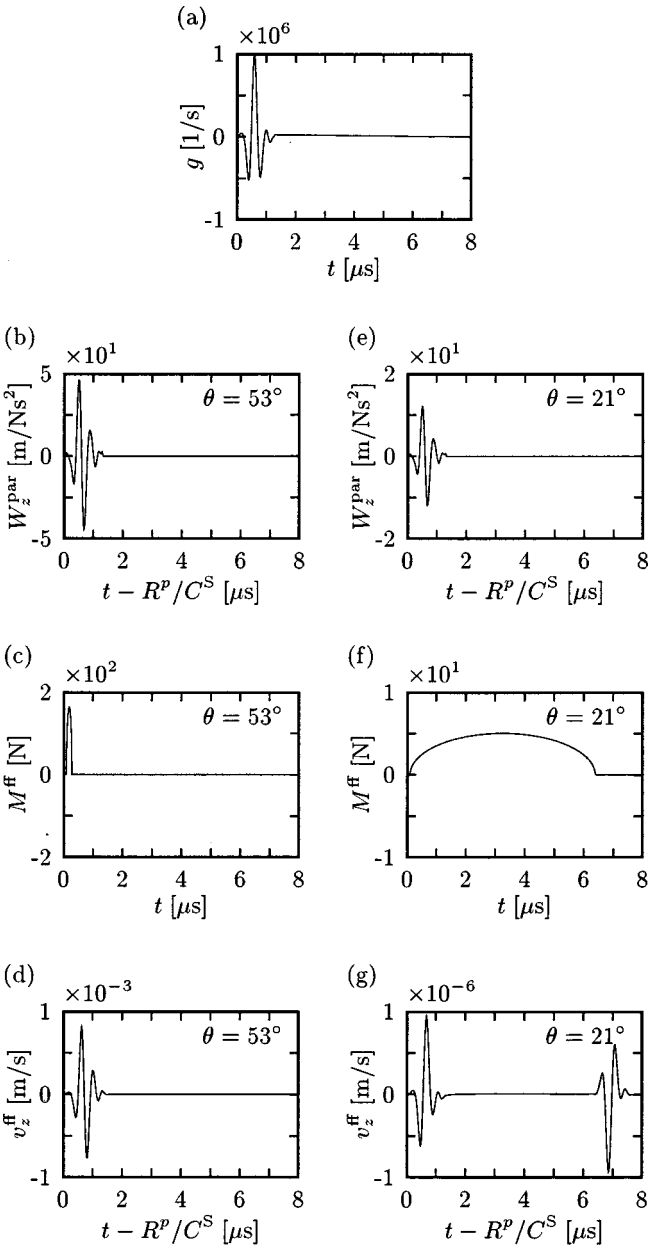


FIG. 6. Numerical results for the half-plane $y=0$, $z\geqslant 0$, an observation distance $R=5R^{\mathrm{src}}$, and a vertical observation angle $\theta=53°$: (a) the source signature g; (b) the vertical paraxial shear wave W_z^{par}; (c) the approximate far field transfer function M^{ff}; (d) the approximate vertical shear particle velocity v_z^{ff}. The plots in (e)–(g) correspond to the plots in (b)–(d), except that now $\theta=21°$.

to the duration of the source signature. As a consequence, M^{ff} is a wide function, causing v_z^{ff} to consist of two more or less separate events. These events are ΔT^{ff} apart in time, which agrees with the fact that they are produced at and near the points of the source domain associated with $t=0$ and $t=\Delta T^{\mathrm{ff}}$. Note that the maximum values of v_z^{ff} for the cases $\theta=21°$ and $\theta=53°$ differ three orders of magnitude.

The influence of the paraxial and arrival time approximations on the waveform is shown in Fig. 7. At the top of this figure the approximate vertical shear particle velocity v_z^{ff} and the exact vertical shear particle velocity v_z are given for a vertical observation angle $\theta=21°$ and an observation distance $R=5R^{\mathrm{src}}$. The middle and bottom parts of the figure

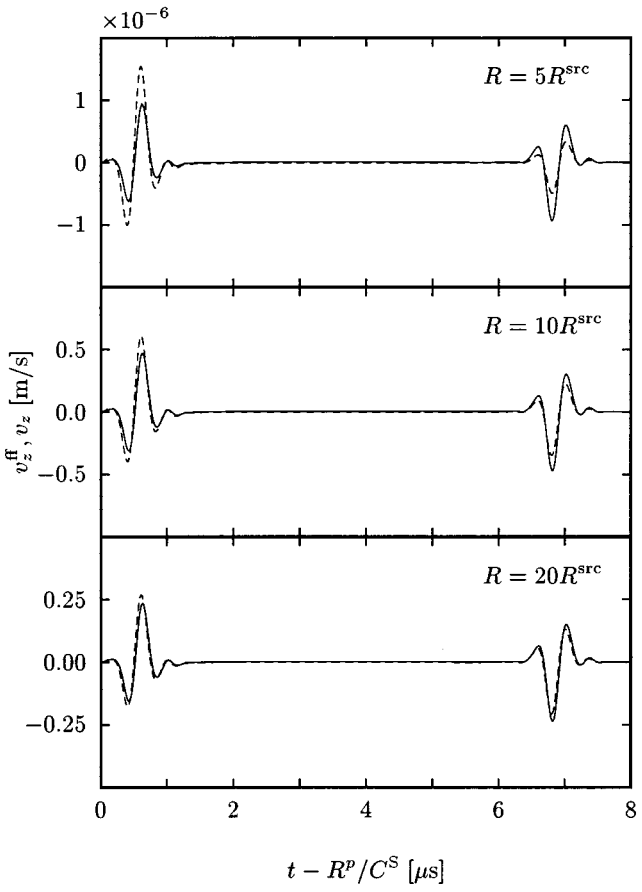


FIG. 7. Numerical results for the half-plane $y=0$, $z\geqslant 0$ and a vertical observation angle $\theta=21°$: the approximate vertical shear particle velocity v_z^{ff} (solid) and the exact vertical shear particle velocity v_z (dashed) for an observation distance of (top) $R=5R^{\mathrm{src}}$; (middle) $R=10R^{\mathrm{src}}$; (bottom) $R=20R^{\mathrm{src}}$.

show the equivalent graphs for $R=10R^{\mathrm{src}}$ and $R=20R^{\mathrm{src}}$, respectively. The differences between v_z^{ff} and v_z are due mainly to the arrival time approximation leading to M^{ff}, i.e., the approximation of the elliptical curves of focal points by straight lines. Going from the top to the bottom, the improvement of the accuracy of v_z^{ff} for increasing R is clearly demonstrated.

The top four panels of Fig. 8 show the influence of the paraxial and arrival time approximations on the arrival times and the spatial structure of the wave field. These panels display the relative norm of the approximate far field particle velocity $\mathbf{v}^{\mathrm{ff}}$ and the exact particle velocity $\mathbf{v}$ as a function of the observation angle θ and the time t. The relative norm of a particle velocity is its magnitude for a specific combination of θ and t, divided by its maximum magnitude over all values of θ and t. The source signature is g from Fig. 6(a). To this end we have determined the exact particle velocity using an implementation of the Cagniard–De Hoop technique.[8] All wave types, i.e., shear waves, compression waves, head waves and Rayleigh waves, have been incorporated. Figure 8(a) shows the approximate result from Eqs. (9), (28), and (39), and Fig. 8(b) shows the exact result from Eq. (4), both for $R=2R^{\mathrm{src}}$. Figures 8(d) and (e) show the corresponding results for $R=5R^{\mathrm{src}}$. Comparison of Fig. 8(a) with Fig. 8(b) and Fig. 8(d) with Fig. 8(e) demonstrates that as far as the arrival times are concerned, the approximations remain fairly good even at small observation distances.

Figures 8(a) and (d) have been determined using the exact paraxial Green's functions G_k^P, G_k^S, and G_k^H (see the Appendix). Using a PC with a Pentium 500 MHz processor, the calculation time required for each plot is in the order of a minute for the approximate result and a day for the exact result, both with a relative accuracy of 10^{-6}.

The irregularities in Fig. 8(a) for the head waves around $\theta=\pm 45°$ are due to the removal of a disturbing event in the paraxial head wave at the arrival time of the paraxial shear wave. Here our assumption that different types of waves may be considered separately is not valid, which is not surprising in view of the relatively short observation distance. From the above observations we may conclude that the far field assumption of decoupled waves in Eq. (5) is more restrictive for head waves in relation with shear waves than for the other combinations of wave types. From Fig. 8(d) we observe that these irregularities, which now occur around $\theta=\pm 40°$, diminish with increasing observation distance. The splitting of the Rayleigh wave events in Figs. 8(a) and (d) is due to the fact that major Rayleigh wave contributions come from both G_k^P and G_k^S, which are subject to different approximations.

We define the directivity of the source as

$$D(R,\theta,\phi)=\frac{\displaystyle\int_0^\infty \|\mathbf{v}(R,\theta,\phi,t)\|^2\,dt}{\max\left\{\displaystyle\int_0^\infty \|\mathbf{v}(R,\theta,\phi,t)\|^2\,dt\right\}}. \tag{44}$$

The directivity is, for a given value of R, a measure of the energy transmitted in a specific direction (θ,ϕ), divided by the maximum of the energy over all directions (θ,ϕ). In Figs. 8(c) and (f) we show polar plots of D using both the approximate far field particle velocity $\mathbf{v}^{\mathrm{ff}}$ and the exact particle velocity $\mathbf{v}$, for $R=2R^{\mathrm{src}}$ and $R=5R^{\mathrm{src}}$, respectively. The focusing of the shear waves around the main beam angle θ^b is clearly the dominant feature. It is further observed that our approximations yield better predictions of the directivity when the observation distance becomes larger.

VII. ACCURACY OF PARAXIAL AND ARRIVAL TIME APPROXIMATIONS

In general, the paraxial approximation in Eq. (10) and the arrival time approximation in Eq. (17) are accurate if the observation distance R complies with

$$R\gg R^{\mathrm{src}}=\sqrt{(2a)^2+(2b)^2}. \tag{45}$$

If this condition is not satisfied, the paraxial approximation may not be accurate enough. In that case we can improve the paraxial approximation by subdividing the source domain into N subdomains, treating each subdomain as a separate source, and adding the results.

Moreover, if Eq. (45) is not satisfied, the arrival time approximation may give rise to unsatisfactory results. To assess the effect of the error in the approximate far field arrival time T^{ff}, we have to compare this error with the duration

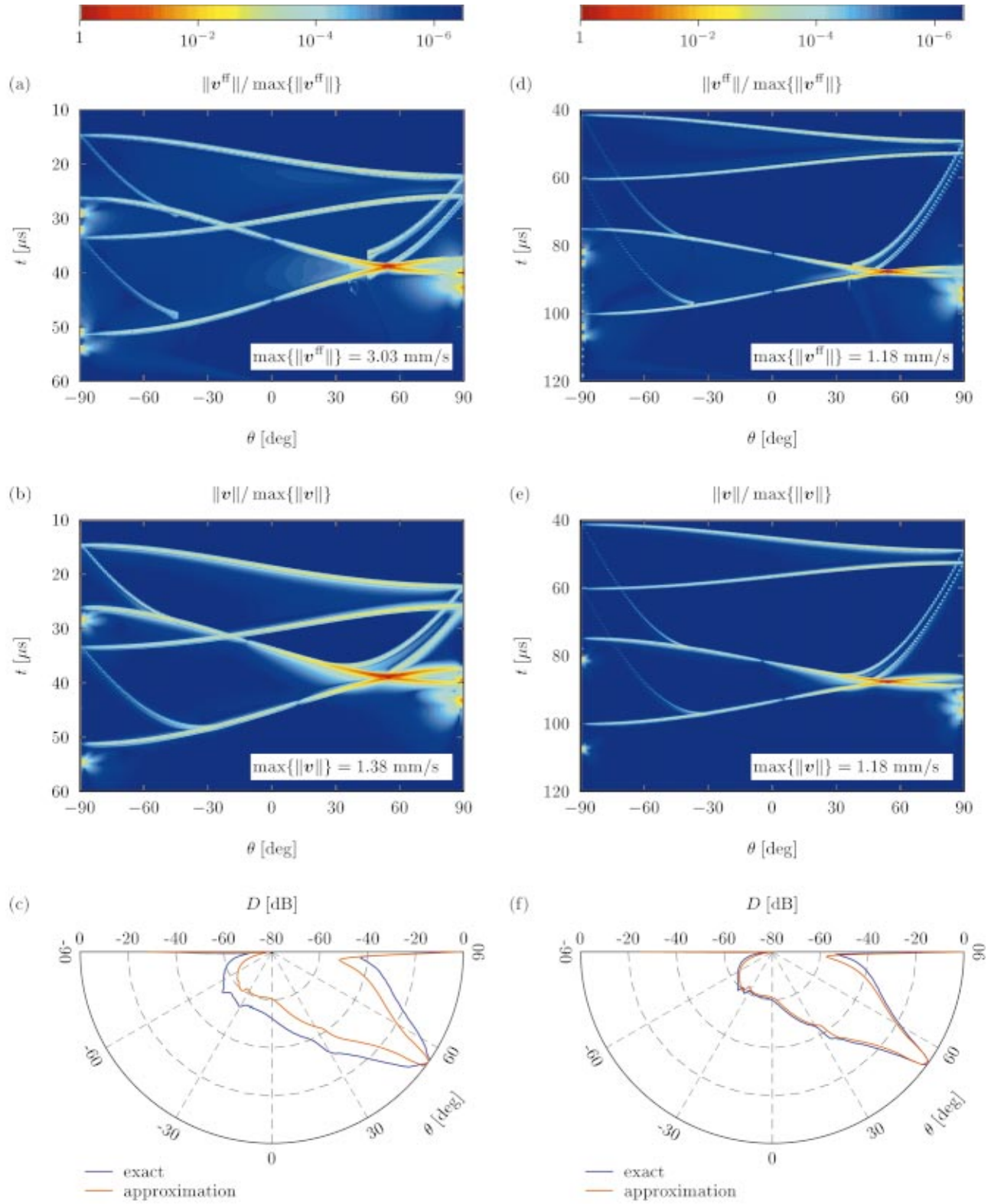


FIG. 8. Numerical results for the half-plane $y=0$, $z\geqslant 0$ and an observation distance $R=2R^{\mathrm{src}}$: (a) the relative norm $\|\mathbf{v}^{\mathrm{ff}}\|/\max\{\|\mathbf{v}^{\mathrm{ff}}\|\}$ of the approximate far field particle velocity versus the vertical observation angle θ and the time t; (b) the relative norm $\|\mathbf{v}\|/\|\max\{\|\mathbf{v}\|\}$ of the exact particle velocity vs θ and t; (c) the directivity D vs θ. The plots in (d)–(f) correspond to the plots in (a)–(c), except that now $R=5R^{\mathrm{src}}$.

ΔT^{src} of the main feature of the source signature g. To simplify the matter, we investigate the error in ΔT^{ff}, which represents the worst case.

In Fig. 9(a) we have depicted an example of a polar plot of ΔT^{ff} versus the horizontal observation angle ϕ, for constant R and θ.

We define the *(horizontal) main beam region* by the condition $\Delta T^{\mathrm{src}} > \Delta T^{\mathrm{ff}}$. An error in ΔT^{ff} turns out to have a significant influence on the shape of the total wave if $\Delta T^{\mathrm{ff}} \approx \Delta T^{\mathrm{src}}$. This occurs around the two solid lines that mark the main beam region. At a sufficient distance from the main beam region, the condition $\Delta T^{\mathrm{src}} \ll \Delta T^{\mathrm{ff}}$ may occur. In this

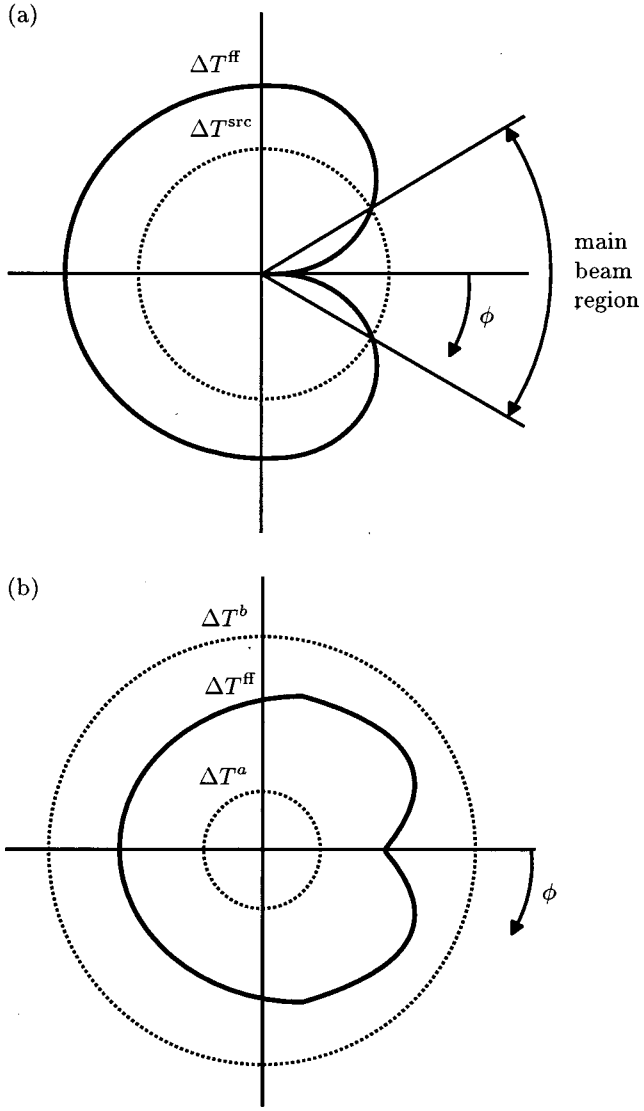


FIG. 9. The approximate maximum arrival time difference ΔT^{ff} as function of the horizontal observation angle ϕ, for constant R and θ: (a) the directivity changes considerably if ΔT^{ff} (solid curve) varies greatly around the duration ΔT^{src} (dotted circle) of the source signature; (b) no significant directivity changes are observed if ΔT^{ff} is much larger than ΔT^{src} ($\Delta T^{\mathrm{src}}=\Delta T^{a}$) or much smaller than ΔT^{src} ($\Delta T^{\mathrm{src}}=\Delta T^{b}$).

case, the elementary waves emitted by all source points average out, except those from the source areas near the points associated with $t=0$ and $t=\Delta T^{\mathrm{ff}}$, see Fig. 6(g). This may most clearly be seen from Eq. (41). Because in the far field approximation these points always lie on the boundary of the source domain, their combined contribution to the wave field is referred to as *edge diffraction* and the corresponding events are referred to as *edge waves*. The influence of the error in ΔT^{ff} on the total wave shape is negligible not only if $\Delta T^{\mathrm{src}} \gg \Delta T^{\mathrm{ff}}$, as in Fig. 6(d), but also if $\Delta T^{\mathrm{src}} \ll \Delta T^{\mathrm{ff}}$, as in Fig. 6(g), since in both cases it makes little difference how large ΔT^{ff} exactly is. The approximations T^{ff} and ΔT^{ff} thus remain sufficiently accurate if

$$\Delta T^{\mathrm{src}} \gg \Delta T^{\mathrm{ff}} \quad \text{or} \quad \Delta T^{\mathrm{src}} \ll \Delta T^{\mathrm{ff}}. \tag{46}$$

VIII. RULES OF THUMB FOR WAVE FIELD AND DIRECTIVITY

In this section we give a few rules of thumb that relate the size, shape and traction amplitude of a transducer to clearly observable trends in the spatial wave field and the directivity.

A. Influence of shape and size of source domain

Here we discuss the influence of the shape and size of a transducer while assuming a constant traction amplitude g^{amp}. Further, we assume that the transfer function M is approximated by a constant function as in Eq. (40). The shape and size of the source domain Ω directly determine the behavior of ΔT^{ff} as a function of the angles of observation θ and ϕ. Based on the reciprocal relation between v_k^{ff} and ΔT^{ff}, see for example Eq. (43), we already may infer the approximate directivity characteristics of the wave field from the behavior of ΔT^{ff}. In Fig. 9(a) the main beam region is determined by the horizontal angles ϕ where the dotted circle is further away from the origin than the solid curve. On the other hand, relatively little energy is expected for angles ϕ where the opposite occurs. It therefore seems that the task of synthesizing a source with a desired directivity may be replaced by the task of finding a source with a desired behavior of ΔT^{ff} as a function of the angles of observation. With this graphical interpretation of the directivity, we may immediately observe the most significant changes in the main beam region when the time ΔT^{src} is varied, i.e., when we vary the source signature g. If in Fig. 9(a) ΔT^{src} increases/decreases, the radius of the dotted circle will increase/decrease and the energy in the main beam region will be dispersed over a larger/smaller sector.

If we have $\Delta T^{\mathrm{src}}=\Delta T^{a} \ll \Delta T^{\mathrm{ff}}$ for all observation angles, as depicted in Fig. 9(b), the wave field in all directions will be dominated by edge diffraction. The energy in the diffracted wave field is usually small. If we have $\Delta T^{\mathrm{src}}=\Delta T^{b} \gg \Delta T^{\mathrm{ff}}$ for all angles of observation, as also depicted in Fig. 9(b), the finite size of the source will not impose any significant directivity to the wave field. In this case, the directivity of the finite source degenerates to the directivity of a single point source. This agrees with the fact that the characteristic wavelength $\lambda = C\Delta T^{\mathrm{src}}$ will be much larger than the dimensions of the finite source.

B. Influence of traction amplitude

Here we discuss the influence of the traction amplitude g^{amp} while assuming a fixed shape and size of the source domain Ω. Spatial variations in the traction amplitude cause additional observable features of the wave field. The amplitudes of these features tend to be larger when the variations are stronger. This is most obvious for the case of a uniform traction amplitude, in which case almost all significant waves seem to originate at the edges of the source domain, i.e., where the traction amplitude abruptly jumps to zero. Smoothing the traction amplitude near the edges will suppress these waves. An appropriately chosen, varying traction amplitude may improve the directivity characteristics of a transducer by suppressing the unwanted waves.[24] In prin-

ciple, the traction amplitude may be both positive (pressure) and negative (tension). A source which average traction amplitude equals zero (e.g., a dipole) has a different directivity and in most directions a smaller far field than a source of pure pressure or pure tension. This may be deduced from Eq. (27).

IX. CONCLUSIONS

A paraxial approximation and an arrival time approximation have successfully been applied to obtain approximate expressions for the far field of a finite source that consists of a moving distribution of normal traction. As a result, convenient far field approximations of the wave field have been obtained, in which the effect of the finite size of the source is represented by a transfer function. The determination of this transfer function involves only basic geometry. A simple approximation of the transfer function shows that the characteristic features of the directivity pattern depend on the maximum approximate arrival time difference ΔT^{ff}. This has led to a rule of thumb for synthesizing a source with a desired directivity pattern. The computational effort to evaluate the approximate wave field and directivity is 1000 times less than for the exact solution. The data scans produced for approximate wave fields prove to be quite accurate when these are compared with the exact wave fields. The approximations for the directivity curves are also quite accurate. The accuracy of the approximations improves when the observation distance becomes larger, and it is excellent if the observation distance is at least a few times the characteristic size of the source. Although the analysis is demonstrated for a normal load moving within an elliptical source domain, it may be easily adapted to other source domains and/or types of loading.

ACKNOWLEDGMENT

This research has been financially supported by the Dutch Technology Foundation (STW). This support is gratefully acknowledged.

APPENDIX: EXACT GREEN'S FUNCTIONS

A point source of normal traction acting on a homogeneous, isotropic elastic half-space excites in this half-space a compression wave and a vertical shear wave. In the temporal Laplace domain, the exact Green's functions for the velocity components of these waves are[8]

$$\hat{G}_k^P(\mathbf{x},s) = -\frac{s^2}{4\pi^2\mu}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\mathcal{N}_k^P}{\mathcal{R}} \times \exp[-s(\eta z + i\alpha x + i\beta y)]\, d\alpha\, d\beta, \tag{A1}$$

$$\hat{G}_k^{SV}(\mathbf{x},s) = -\frac{s^2}{4\pi^2\mu}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\mathcal{N}_k^{SV}}{\mathcal{R}} \times \exp[-s(\zeta z + i\alpha x + i\beta y)]\, d\alpha\, d\beta. \tag{A2}$$

The superscripts P and SV refer to the compression wave and the vertical shear wave, respectively. The symbol s denotes the parameter of the one-sided Laplace transformation, and μ is one of the Lamé coefficients of the elastic half-space. The components of $\mathcal{N}_k^P$ and $\mathcal{N}_k^{SV}$ are

$$\mathcal{N}_x^P = -[2\zeta^2 - (C^S)^{-2}]i\alpha, \tag{A3}$$

$$\mathcal{N}_y^P = -[2\zeta^2 - (C^S)^{-2}]i\beta, \tag{A4}$$

$$\mathcal{N}_z^P = -[2\zeta^2 - (C^S)^{-2}]\eta, \tag{A5}$$

and

$$\mathcal{N}_x^S = 2\eta\zeta i\alpha, \tag{A6}$$

$$\mathcal{N}_y^S = 2\eta\zeta i\beta, \tag{A7}$$

$$\mathcal{N}_z^S = 2\eta(\alpha^2+\beta^2), \tag{A8}$$

and $\mathcal{R}$ denotes the Rayleigh denominator

$$\mathcal{R} = [2\zeta^2 - (C^S)^{-2}]^2 - 4\eta\zeta(\alpha^2+\beta^2). \tag{A9}$$

In the equations above η and ζ are the vertical propagation coefficients of the compression wave and the shear wave, and are given by

$$\eta = \sqrt{\alpha^2+\beta^2+(C^P)^{-2}}, \tag{A10}$$

$$\zeta = \sqrt{\alpha^2+\beta^2+(C^S)^{-2}}. \tag{A11}$$

The symbols C^P and C^S denote the velocity of the compression wave and the shear wave.

The exact time domain counterparts of Eqs. (A1) and (A2) may be obtained with the aid of the Cagniard–De Hoop method.[8,19,20] This involves an integration over the so-called Cagniard contour in the complex slowness plane. For the compression wave, the Cagniard contour consists of a single branch of a hyperbola, and integration over this contour leads to the time domain result for G_k^P. For the shear wave, the situation may occur in which the Cagniard contour consists of two parts. One part is always present and consists of a single branch of a hyperbola, and the other part is conditionally present and consists of a detour around a branch cut. Integration over the hyperbolic part of the contour yields the time domain result for the body wave part of G_k^{SV}, which we denote by G_k^S, while integration over the detour part of the contour yields the time domain result for the head wave part of G_k^{SV}, which we denote by G_k^H.

[1] A. Lhémery, "A model for the transient ultrasonic field radiated by an arbitrary loading in a solid," J. Acoust. Soc. Am. **96**, 3776–3789 (1994).
[2] A. McNab, A. Cochran, and M. A. Campbell, "The calculation of acoustic fields in solids for transient normal surface force sources of arbitrary geometry and apodization," J. Acoust. Soc. Am. **87**, 1455–1465 (1990).
[3] H. Wustenberg, Ph.D. thesis, Techn. Univ. Berlin, 1972, pp. 38–48.
[4] L. W. Schmerr, Jr. and A. Sedov, "Elastodynamic model for compressional and shear wave transducers," J. Acoust. Soc. Am. **86**, 1988–1999 (1989).
[5] D. Gridin and L. J. Fradkin, "The high-frequency asymptotic description of pulses radiated by a circular normal transducer into an elastic half-space," J. Acoust. Soc. Am. **104**, 3190–3198 (1998).
[6] D. C. Gakenheimer and J. Miklowitz, "Transient excitation of an elastic half space by a point load traveling on the surface," J. Appl. Mech. **36**, 505–515 (1969).
[7] F. Guan and M. Novak, "Transient response of an elastic homogeneous half-space to suddenly applied rectangular loading," J. Appl. Mech. **61**, 256–263 (1994).
[8] M. C. M. Bakker, Ph.D. thesis, Delft Univ. of Techn., 2000, pp. 22–26, 47–125, 133–148.
[9] F. G. Laturelle, "The stresses produced in an elastic half-space by a nor-

mal step loading over a circular area, analytical and numerical results,'' Wave Motion **12**, 107–127 (1990).
[10] L. F. Bresse and D. A. Hutchins, ''Transient generation of elastic waves in solids by a disk-shaped normal force source,'' J. Acoust. Soc. Am. **86**, 810–817 (1989).
[11] H. Djelouah and J. C. Baboux, ''Transient ultrasonic field radiated by a circular transducer in a solid medium,'' J. Acoust. Soc. Am. **92**, 2932–2941 (1992).
[12] D. Gridin, ''On the radiation of ultrasound into an isotropic elastic half-space via wavefront expansions of the impulse response,'' J. Acoust. Soc. Am. **105**, 2565–2573 (1999).
[13] J. C. Baboux and R. Kažys, ''Analysis of the transient fields radiated in solids by circular and annular sources,'' J. Acoust. Soc. Am. **92**, 2942–2951 (1992).
[14] F. R. Norwood, ''Exact transient response of an elastic half space loaded over a rectangular region of its surface,'' J. Appl. Mech. **36**, 516–522 (1969).
[15] D. L. Lansing, ''The displacements in an elastic half-space due to a moving concentrated normal load,'' NASA technical report 238, 1966.
[16] M. C. M. Bakker, M. D. Verweij, B. J. Kooij, and H. A. Dieterman, ''The traveling point load revisited,'' Wave Motion **29**, 119–135 (1999).
[17] D. C. Gakenheimer, ''Response of an elastic half space to expanding surface loads,'' J. Appl. Mech. **38**, 99–110 (1971).
[18] T. P. Lerch, L. W. Schmerr, and A. Sedov, ''The wavefield of an ultrasonic angle beam shear wave transducer: an elastodynamic approach,'' in *Review of Progress in Quantitative Nondestructive Evaluation*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1997), Vol. 16, pp. 855–892.
[19] A. T. de Hoop, ''A modification of Cagniard's method for solving seismic pulse problems,'' Appl. Sci. Res., Sect. B **8**, 349–356 (1960).
[20] J. H. M. T. van der Hijden, *Propagation of Transient Elastic Waves in Stratified Anisotropic Media* (North-Holland, Amsterdam, The Netherlands, 1987), pp. 119–133.
[21] P. M. Morse, *Vibration and Sound*, 2nd ed. (McGraw-Hill, New York, 1948), pp. 84, 326–328, 344–346.
[22] P. R. Stepanishen, ''Transient radiation from pistons in an infinite planar baffle,'' J. Acoust. Soc. Am. **49**, 1629–1638 (1971).
[23] P. Kielczynski and W. Pajewski, ''Application of the transfer function method in calculations of the directivity pattern of ultrasonic transducer,'' J. Appl. Phys. **70**, 7257–7260 (1991).
[24] R. Stacey and J. P. Weight, ''Pulse-echo scattering in solids with nonuniform transducers,'' IEE Proc.: Sci., Meas. Technol. **141**, 363–368 (1994).
[25] D. J. Chinn, Ph.D. thesis, Delft Univ. of Techn., 1993, pp. 113–116.

Rotational aerophones

N. H. Fletcher[a)]
Research School of Physical Sciences and Engineering, Australian National University, Canberra 0200, Australia

A. Z. Tarnopolsky and J. C. S. Lai
School of Aerospace and Mechanical Engineering, University College, University of New South Wales, Australian Defence Force Academy, Canberra 2600, Australia

(Received 11 September 2001; accepted for publication 29 November 2001)

Free rotational aerophones such as the bullroarer, which consists of a wooden slat whirled around on the end of a string, and which emits a loud pulsating roar, have been used in many ancient and traditional societies for ceremonial purposes. This article presents an experimental and theoretical investigation of this instrument. The aerodynamics of rotational behavior is elucidated, and relates slat rotation frequency to slat width and velocity through the air. Analysis shows that sound production is due to generation of an oscillating-rotating dipole across the slat, the role of the vortices shed by the slat being relatively minor. Apparent discrepancies between the behavior of a bullroarer slat and a slat mounted on an axle in a wind tunnel are shown to be due to viscous friction in the bearings of the wind-tunnel experiment. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1446053]

PACS numbers: 43.28.Ra, 43.75.−z [MSH]

I. INTRODUCTION

The device generally called a ''bullroarer'' has an important place in many cultures,[1] ranging from predynastic Egypt of 3000 BC and the ancient tribal traditions of the Australian Aboriginal people, to the Inuit of Northern Canada. It consists simply of a thin slat of wood, typically 200–400 mm long, 30–60 mm wide and 5–10 mm thick, pierced with a small hole at one end through which is passed a thin cord about 1 m long, by means of which it is whirled in circles by the player. The slat may be simply rectangular and unadorned, or may have a more complex shape and be decorated with paintings of tribal gods or totems. An Australian example, the name of which is secret-sacred in Aboriginal languages, is shown in Fig. 1. The cross-section of this bullroarer is trapezoidal, as is common, rather than simply rectangular. During the whirling motion, the slat rotates about its long axis, producing a loud growling roar which pulsates at the rate of rotation, about 1 to 1.5 Hz, and also at about twice this rate. It is interesting that the frequency of the bullroarer drone is about 70 Hz, which is comparable with the frequency of the drone of the lip-blown didjeridu,[2] also used in ceremonies. One might feel with Marcuse[1] that bullroarers are not really musical instruments, but the bullroarer has in fact been used in at least one piece of classical music—the ballet Suite *Corroboree* by Australian composer John Antill[3]—to evoke the spirit of the Aboriginal people. We may term these devices ''free rotational aerophones'' and their investigation is the main subject of this article.

II. FREE AEROPHONE EXPERIMENTS

In an initial study, the bullroarer of Fig. 1 was whirled around in a vertical plane, as shown in Fig. 2, using a string of length about 70 cm, and an audio recording was made, along with a video record for later frame-by-frame analysis. At very low arm rotation speeds A or B, axial rotation C of the slat did not occur, but this began for arm rotation speeds of around 30 rpm (0.5 Hz) and was fully developed at the normal arm rotation speed of about 60–90 rpm (1–1.5 Hz). The pulsations synchronous with the hand motion are presumably associated with variations in the speed of rotation of the slat caused by uneven motion of the player's arm. The player's hand motion was generally elliptical, but the path of the slat was quite closely circular, with a radius of about 70 cm for slow arm motion and 80 cm for rapid rotation. The angular position of the slat in its orbit trailed that of the hand by an angle ϕ as shown in the second panel of Fig. 2. The variation of this trailing angle during one arm revolution is shown in Fig. 3. The angle ϕ increased with increased speed of arm rotation and oscillated quasi-periodically at twice the arm rotation speed, as might be expected from the relationship between circular slat motion and elliptical arm motion. The varying angle between the long axis of the slat and its direction of translational motion presumably contributes to sound pulsations.

In normal playing, the string linking the slat of the bullroarer to the player's hand traces out a conical surface with an apical semi-angle of about 80 degrees. In quasi-steady operation, the bullroarer rotates in one direction, C, about its axis for several seconds, then stops and reverses its motion, the orientation of the cone swept out by the string reversing simultaneously. This rotation reversal is clearly associated with progressive twisting of the string and depends upon its length and thickness, while the cone-axis reversal is apparently associated with a lift force, the direction of which depends upon the sense of the axial rotation, as for a rotating cylinder.[4]

In one set of experiments in an anechoic chamber, the

a)Electronic mail: neville.fletcher@anu.edu.au

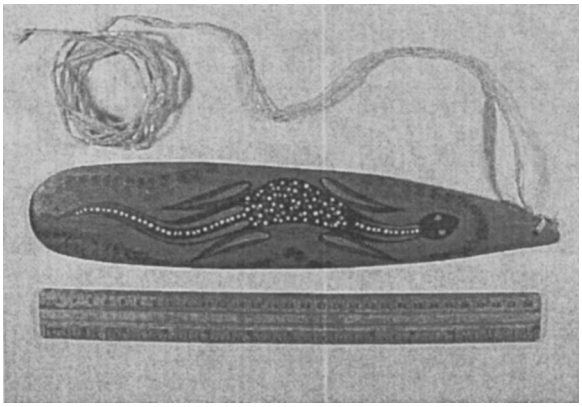


FIG. 1. A bullroarer, from Northern Australia. The ocher-painted design indicates the "dreaming" or totem of the maker.

microphone was positioned on the rotation axis and at a distance of about 1 m from the plane of rotation while the bullroarer was whirled at about 90 rpm (1.5 Hz). Figure 4(a) shows the signal recorded over a time of about 0.6 s. The generally sinusoidal waveform, with a frequency of about 70 Hz, is modulated at about 3 Hz in synchrony with the arm rotation. Because of the twofold rotation symmetry of the slat, the axial rotation speed is just half the acoustic frequency, and therefore about 35 Hz or 2100 rpm. The modulation depth and maximum sound output are both slightly different for the two senses of axial rotation of the slat, an effect presumably associated with its trapezoidal cross-section.

Figure 4(b) shows the spectrum of the radiated sound. It consists essentially of a single component at about 70 Hz, the width of the peak presumably being due to aerodynamic unsteadiness in frequency and to the amplitude variations shown in Fig. 4(a). The level of the second harmonic is about -30 dB relative to the fundamental, and the general noise level is low.

Because the genuine bullroarer slat is complicated geometrically, it was desirable to replace this by a simpler slat for more detailed analysis. Experiments similar to those carried out on the bullroarer were therefore performed using a simple rectangular slat of aluminum, 335 mm long, 40 mm wide and 1 mm thick, whirled by hand on a string of length 70 cm. The behavior was generally similar to that of the bullroarer except that the slat did not begin rotation spontaneously and had to be twisted initially. The instrument produced the same pulsating drone, and the sound spectrum and total radiated power were both also similar. The radiation pattern, measured in an open environment, was essentially isotropic to within about ± 3 dB, more accurate measurement being impossible because of variation in the sound signal.

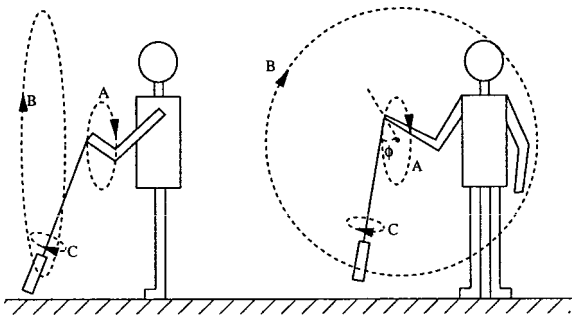


FIG. 2. Playing the bullroarer, with the rotational motions and the trailing angle ϕ defined. The string describes a conical path with apex directed successively towards and away from the player.

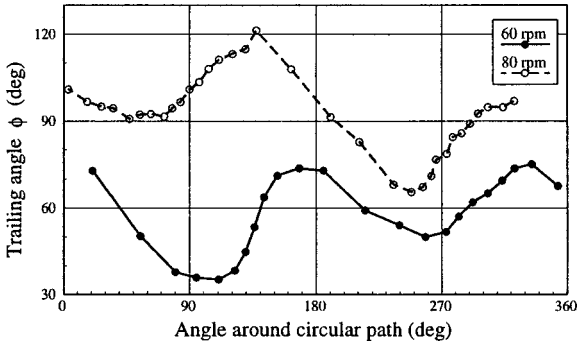


FIG. 3. Variation of the trailing angle ϕ between the projection of the player's arm on the rotation plane and the string direction, during one cycle of revolution. Measurements for two arm rotation rates are shown.

The measured dependence of the rotation rate f or angular rotation rate $\omega = 2\pi f$ of the slat (half the sound frequency, because of the twofold symmetry the slat) on arm rotation rate F or angular rate $\Omega = 2\pi F$, for slats of several widths, is shown in Fig. 5. As shown in the figure, the drone frequency is approximately proportional to the arm rotation rate and thus to the translational speed V of the centroid of the slat through the air, and approximately inversely proportional to the slat width W. The more precise result derived from the experimental regressions for a 70-cm string length is

$$\omega \approx 0.4\Omega^{0.9}W^{-1.3} \approx 0.5V^{0.9}W^{-1.3} \tag{1}$$

with an uncertainty of about ± 0.2 in the values of the exponents and of about $\pm 20\%$ in the value of the multiplying constant, which depends upon the length of the string. The equivalent range of airspeed V over which this relation was established was about $5-10\ \mathrm{m\,s^{-1}}$.

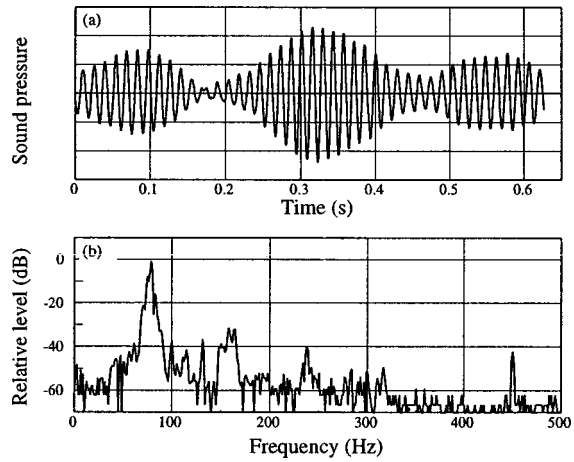


FIG. 4. (a) Pressure waveform of the sound produced by a bullroarer rotating at about 90 rpm. The sound frequency is about 70 Hz and the frequency of the pulsations about 3 Hz. The pulsations are more prominent for one direction of slat rotation than for the other. (b) Spectral analysis of the quasi-steady sound, which differs little for the two rotation directions.

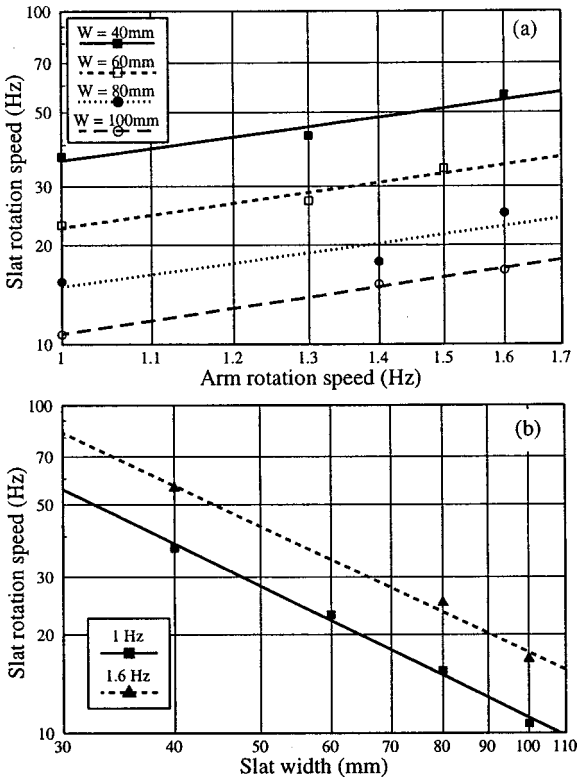


FIG. 5. Slat rotation speed f as a function of (a) arm rotation speed F, and (b) slat width W, for a hand-whirled slat on a string of length 70 cm.

An alternative analysis of the experimental data can be carried out by plotting f against F/W, as is suggested by later theory. The result, plotted in Fig. 6, gives a regression of the form

$$f \approx 1.6F/W - 5 \quad \text{or} \quad \omega \approx 1.6\Omega/W - 30 \tag{2}$$

for a string length of 70 cm. This expression can be rewritten more generally as

$$\omega \approx 1.1V/W - 30. \tag{3}$$

It is perhaps significant to note that the rotation speed is such that the edge of the slat is moving at about half the speed of the airstream.

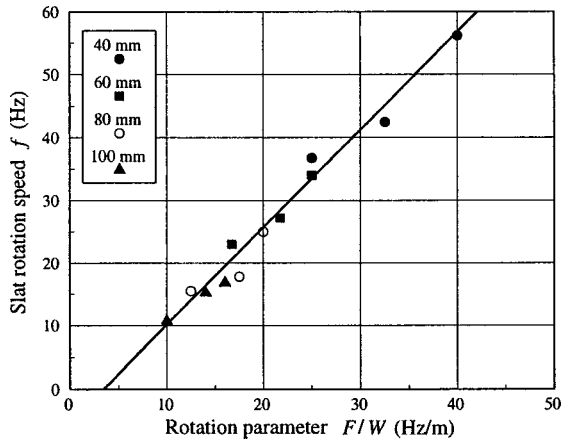


FIG. 6. Slat rotation speed f as a function of F/W for a slat of width W whirled at frequency F on a string of length 70 cm.

Because the available anechoic chamber was small and had a cut-off frequency above 100 Hz, which is well above the sound frequencies involved, it was necessary to perform acoustic measurements at an outdoor site, actually the middle of an isolated sports oval. Particular interest attaches to the radiated sound power and its dependence upon slat width and arm rotation speed. The measurements are complicated by the sound pulsations and by the slower rise and fall of sound level as the string is untwisted and then retwisted. There is, however, a plateau level in each long cycle during which fairly reliable measurement is possible.

Measurements were made using four rectangular aluminum slats 335 mm long and 1 mm thick, and ranging in width between 40 and 100 mm. The slats were whirled by hand using a string 70 cm in length. Sound pressure measurements were made on the rotation axis direction at a distance of 4 m from the player, this distance being dictated by the balance between sound output and ambient noise. Since the sound frequency in the experiments was in the range 40–160 Hz, this distance is less than the sound wavelength in some cases, and near-field effects must be considered. The pressure amplitude of a dipole, measured at distance r, exceeds the radiation contribution at this distance by a factor $[1+(jkr)^{-1}]$, where $k=2\pi/\lambda$ and λ is the sound wavelength.[5] Converting this to a sound pressure level, however, shows that the maximum correction is less than 1.3 dB, which is not significant in our measurements.

The ''linear'' setting (no frequency weighting) on a sound level meter with real-time $\frac{1}{3}$ octave frequency analysis was used, a spectral display being necessary in order to discriminate against wind noise and other environmental noise at these low frequencies. The results showed that the radiated sound level was completely independent of slat width over the width range tested and at arm rotation rates between 72 and 120 rpm ($7.5<\Omega<12.6\,\mathrm{s}^{-1}$ or $6<V<10\ \mathrm{m\,s}^{-1}$). The sound pressure level at the measuring position was 60 dB at 72 rpm and 72 dB at 120 rpm with an uncertainty of about ± 1 dB in each case.

A further measurement in which the player turned sequentially through 45-degree angles established that the radiation pattern was uniform to within an uncertainty of about ± 1 dB so that the SPL values can be converted to radiated acoustic power. Assuming good reflection at these frequencies from the grass field surface, this gives a radiated power of about 1 mW at an arm rotation speed of 120 rpm and about 60 μW at 72 rpm. The approximate behavior of radiated sound power was thus

$$P = \text{const} \times \Omega^{5.4}. \tag{4}$$

There is no dependence upon W and an uncertainty of perhaps ± 1 in the exponent. Since the sound source, being slaved to the rotation of the slat, is coherent along its length which is much smaller than the sound wavelength, we expect that the radiated power should be proportional to the square of slat length H, but this dependence was not examined experimentally.

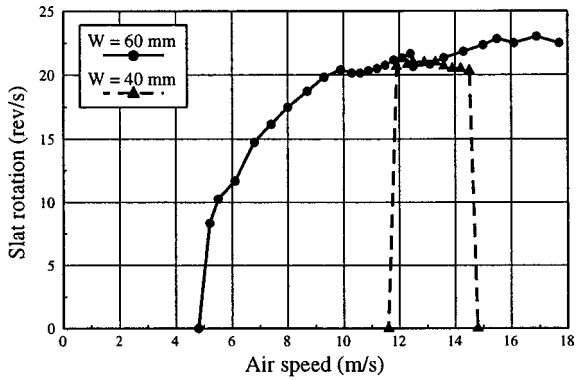


FIG. 7. Measured rotational frequency of the 60-mm slat as a function of airspeed in the wind tunnel. Also shown is the measured behavior of a 40-mm slat, which has a very limited range of airspeeds for rotation.

III. WIND-TUNNEL EXPERIMENTS

A fundamental question to emerge is that of the mechanism controlling the axial rotation of the slat. This was therefore investigated in a series of carefully controlled experiments in a wind tunnel. For this study, a simple rectangular steel slat 253 mm long, 1.5 mm thick and 20, 40 or 60 mm wide was mounted on Teflon bearings by means of an axle passing along its centerline. The bearings were lubricated with a light oil. The wind-tunnel cross-section was $255 \times 255\ \mathrm{mm}^2$ and the length of the slat extended almost completely across the width of the tunnel, so that there was a good approximation to two-dimensional unbounded flow. The air speed in the wind tunnel was measured using a pitot tube connected to a micro manometer, and the slat rotation speed was measured using an optical tachometer on the axle. The maximum attainable air speed was $18\ \mathrm{m\,s^{-1}}$.

The following conclusions were readily established experimentally:

(1) The slat will not generally begin rotation spontaneously under the influence of air flow, but requires to have an initial rotation imparted to it. Depending upon the airspeed, the slat rotation may then either settle down to a steady value or else stop.
(2) For any slat, there is a minimum air flow velocity below which rotation cannot be maintained. This critical velocity increases as the width of the slat is decreased. (In the experiment, it was impossible to maintain rotation of the 20-mm slat under any attainable air speed.)
(3) At least in the case of the 40-mm slat, there is also a maximum air flow velocity above which rotation cannot be maintained. No such upper limit was found for the wider slat up to the attainable wind tunnel air speed.
(4) The steady rotation frequency does not depend greatly upon the width of the slat in the two cases studied, and increases towards saturation with increasing air speed.

The experimental basis for these conclusions is demonstrated in Fig. 7. There are several features of the wind-tunnel experiments that are in conflict with observations on the bullroarer in normal operation. The first is the fact that the rotation speed of the 60-mm slat in the wind tunnel apparently reaches a plateau value once the airspeed exceeds about $10\ \mathrm{m\,s^{-1}}$, while the bullroarer rotation speed continues to increase about linearly with airspeed in the range 6 to $12\ \mathrm{m\,s^{-1}}$, as shown in Fig. 5 and Eq. (1). The second is that the 40-mm slat will rotate over only a limited airspeed range and ceases rotation if the airspeed is too high. Finally, the rotation speeds of the 40- and 60-mm slats are nearly the same in the airspeed range in which both rotate. The slats used for these determinations were very similar to those used in the free-aerophone experiments except for the mass, which should not be significant in wind-tunnel measurements. An explanation for these discrepancies must be sought, and is given in the next section.

IV. SIMPLE THEORY

This investigation scarcely warrants development of a detailed aerodynamic study, which would be a complex operation indeed. Fortunately it is straightforward to devise a simple theory that explains the experimental findings, at least in a semi-quantitative manner. A simple theory also exposes the physical principles involved in a more accessible way.

Consider a thin slat of width W, mounted as in the wind-tunnel experiment and rotating with angular velocity ω in an air flow of speed V. This assumption can later be specialized to apply to the cord-supported slat. The total torque acting on the slat can be divided conceptually into three parts, although it is not really possible to separate these rigorously because of the nonlinear nature of aerodynamic forces. First, there is the aerodynamic drag torque Γ_1 per unit length of slat arising from the rotation. On the assumption that this torque is proportional to the product of the square of the slat-edge velocity $W\omega/2$, the slat area per unit length W, and the slat radius of gyration $W/12^{1/2}$, this can be written to a first approximation as

$$\Gamma_1 \approx -\alpha \rho W^4 \omega^2, \tag{5}$$

where ρ is the density of air and α is a nondimensional constant.

Second, the frictional torque in the bearings must be taken into account. This torque depends upon the load borne by the bearings, which is largely the aerodynamic drag created by the interaction of the airstream with the rotating slat. While this varies periodically through the rotation as the angle of incidence changes, its average value is proportional to $\rho V^2 W$ per unit length. For generality the bearing torque should include a velocity-independent contribution from dry friction, a viscous friction contribution that is proportional to slat rotation speed, and a residual friction independent of the aerodynamic load. It is a reasonable approximation to assume that both the dry friction and the viscous friction are proportional to the total load on the bearings, and the resulting torque is also proportional to bearing radius r, so that the total bearing torque Γ_2 per unit length of slat can be written as

$$\Gamma_2 = -\Gamma_0 - \rho W r V^2 (\beta + \beta' \omega), \tag{6}$$

where Γ_0 is the frictional torque of the bearing under zero airflow conditions, β is a nondimensional constant, and β' is a constant with the dimensions of time.

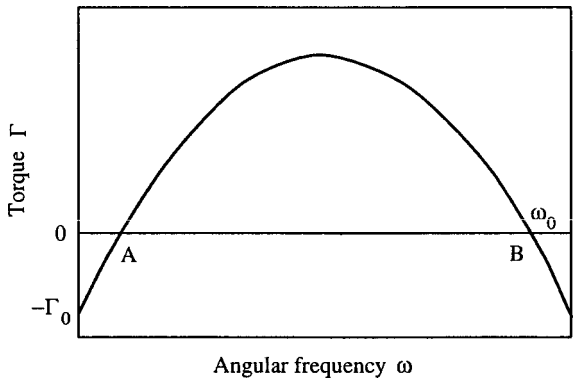


FIG. 8. Net accelerating torque Γ on the rotating slat as a function of rotation frequency ω, for a fixed airspeed, according to the simple theory. Γ_0 is the frictional torque. Note that the rotational speed must reach the point A if rotation is to be sustained; the rotation speed then increases to the steady value ω_0 at point B.

Finally, an expression must be derived for the torque Γ_3 per unit length of slat driving the rotation. Again, this torque clearly varies periodically during the rotation, and the concern here is simply with its average value. A qualitative argument suggests that the mass flow per unit length intercepted by the slat per unit time is proportional to ρWV, the imparted lateral speed due to slat rotation is proportional to $W\omega$, and the resultant moment is further proportional to W. This leads to the result, also supported by dimensional analysis, that

$$\Gamma_3 \approx \gamma \rho W^3 \omega V, \tag{7}$$

where γ is another nondimensional constant.

The complete first-order expression for the average torque per unit length acting on the slat is therefore

$$\Gamma = -\alpha\rho \ W^4\omega^2 - (\beta + \beta'\omega)\rho W r V^2 - \Gamma_0 + \gamma\rho W^3 \omega V. \tag{8}$$

If this expression for Γ is plotted as a function of ω for a fixed value of V, then it has the form shown in Fig. 8. Rotation can be maintained only over a limited range of angular velocity, explaining why the slat has to be set into rotation artificially and will not rotate spontaneously from rest. For angular velocities above the threshold A, the nett torque Γ is positive and the slat rotation speed increases until it comes to a steady state with rotation velocity ω_0 at the point B.

If the slat is to maintain its rotation, then it is clearly necessary that the rotation speed ω_0 for which $\Gamma = 0$ should be a real quantity—a statement equivalent to the requirement that the maximum value of Γ should be positive. Examination of the expression (8) shows that it is quadratic in ω and that the condition $\Gamma = 0$ is met when

$$\frac{W\omega}{V} = \left(\frac{\gamma}{2\alpha} - \frac{\beta' r V}{2\alpha W^2}\right) \pm \left[\left(\frac{\gamma}{2\alpha} - \frac{\beta' r V}{2\alpha W^2}\right)^2 - \frac{\beta r}{\alpha W} - \frac{\Gamma_0}{\alpha\rho W^2 V^2}\right]^{1/2}. \tag{9}$$

If this quantity is to be real, then the expression in square brackets must be positive, and this implies at least a minimum value for the airspeed V and at least a minimum value for the slat width W, both of these critical values depending upon the bearing coefficients Γ_0, β and β'. If these conditions are satisfied, then the steady rotational speed ω_0 of the slat will be given by (9) with the positive sign chosen.

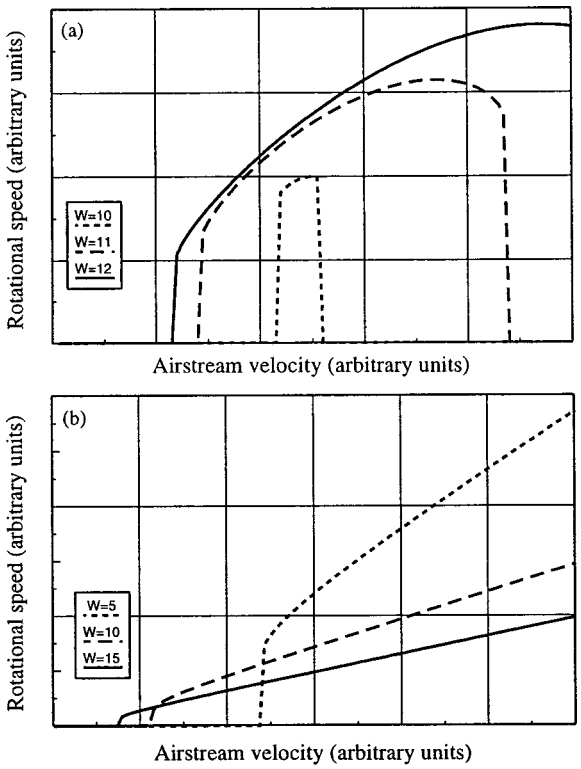


FIG. 9. (a) Predicted variation of steady rotational speed ω_0 with airspeed V, as given by Eq. (9), for a slat of width W with viscous torque in its bearings. (b) The same prediction for a slat with only static bearing torque Γ_0.

The predictions of this theory are most easily seen by considering a numerical example. For this purpose we choose $\alpha = \beta = \beta' = \gamma = 1$, $\Gamma_0 = 1$, $r = 1$, and $\rho = 1$ for simplicity. The results of the calculations are displayed in Fig. 9(a), which shows that, for a given slat width W, the rotation can be maintained only above a threshold airspeed determined by the frictional coefficients. There is also a maximum airspeed above which rotation ceases, and the airspeed region within which rotation can be maintained decreases in extent as the slat width W is reduced. Within this allowed region, the rotation speed tends to saturate and then decrease as the airspeed is increased, and the steady rotation speed for a given airspeed increases slowly with increasing slat width. All these predictions of the theory are in general qualitative agreement with the experimental wind-tunnel findings discussed in Sec. III, as can be seen by comparing Figs. 7 and 9(a).

In contrast, if the only bearing torque is a small constant value Γ_0, then once again this leads to threshold behavior, as shown in Fig. 9(b), but the rotational speed ω of the slat then increases quasi-linearly with airspeed, and no upper limit is predicted. The rotational speed is also predicted to decrease smoothly with increasing slat width, all in contrast with the predictions and experimental results for viscous bearing friction shown in Fig. 9(a).

When the theory is applied to the case of a free bullroarer slat whirled on the end of a string, the picture is quite

different because of the lack of bearing friction. Clearly $\beta = \beta' = 0$ and the static torque Γ_0 depends upon the twisting of the string and thus the integral $\int \omega dt$. Neglecting Γ_0 in (9), as is valid near the mid-period of a rotation, leads to the result

$$\omega_0 = \frac{\gamma V}{\alpha W} = \text{const} \times \frac{F}{W}. \tag{10}$$

This result is in qualitative agreement with the experimental results shown in Fig. 6 and Eq. (3). No upper or lower limit to operational airspeed is predicted, provided the supporting string introduces no axial torque.

As a final tentative deduction from this simple theory, we note that the average lift force normal to the airstream direction can be expected to behave very much like the torque Γ_3 except for a factor of order W^{-1}. This means that the sign of this force depends upon the sign of the rotation ω, which is in accord with the observed conical path swept out by the string of a bullroarer. The angle of the cone described by the string path depends upon the ratio between the mean lift force and the centrifugal force, the simple theory suggesting that this ratio, and thus the cone angle, should be nearly independent of arm rotation speed since both terms increase as V^2. Such a lift effect is, of course, well known in the case of a rotating cylinder, where a rigorous derivation is possible.[4] It is noted here just to show that the simple theory above is at least consistent with a more rigorous theory.

V. NUMERICAL ANALYSIS

Recognizing that the analysis above is at best semiquantitative, the aerodynamics underlying the behavior of rotational aerophones was further investigated by a numerical evaluation of the forces and torques acting on a rotating slat using the computational fluid dynamics software package Fluent 5.2.[6] This analysis was carried out only for the case of a freely rotating slat with no frictional forces.

Since the unsteady flow around a rotating plate can be turbulent, the Reynolds-averaged Navier–Stokes transport equations were used to represent the mean flow quantities, while the Reynolds stresses were modeled using the standard $k-\epsilon$ (turbulent kinetic energy–turbulent dissipation rate) turbulence model.[7] The computational domain was divided into two subdomains: a stationary wind-tunnel domain, and a rotating circular domain containing the slat. The two domains were created using an ''unstructured quadrilateral grid,'' as defined in the software, with a higher density of nodes in the mesh near the interface between the two domains. The computational domain comprised a total of 6568 nodes covering a physical domain of dimensions 200×500 mm, the longer dimension being in the flow direction. The slat, with a width W of only 40 mm, was located close to the input end of the domain. The rotational motion of the domain containing the slat was modeled using a sliding mesh technique.[6]

Calculations were carried out for a selected airspeed in the wind-tunnel and a given rotational speed of the slat, using a second-order upwind finite difference scheme in space and a second-order implicit finite difference scheme in time. Calculations were iterated until the flow around the slat became periodic in time, as confirmed by observing the history of the computed force on the slat. The convergence was then checked by repeating one cycle of the calculation with a halved time step, and no significant difference was found.

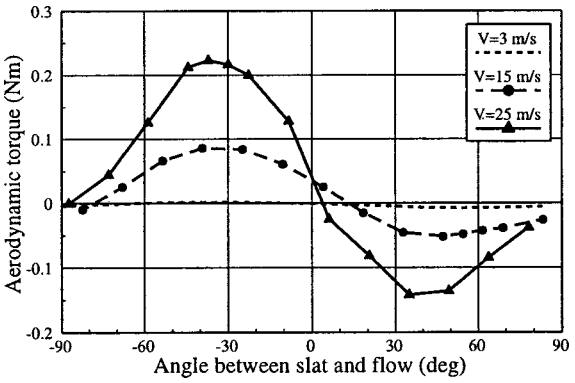


FIG. 10. Numerically calculated torque cycle on a slat of width 60 mm rotating at 50 revolutions per second in a stream with given speed. The second half of the cycle is identical. The mean accelerating torque Γ is the value of the signed area under the curve.

Figure 10 shows the calculated torque on the rotating slat as a function of angle of orientation for given air flow speeds and a constant slat rotation speed. It is clear that there is a net steady torque on the slat, evaluated by taking the signed area between the curve and the axis, and that this net torque increases with air speed. Figure 11 displays the torque as a function of slat orientation for a constant air speed and given slat rotation speeds. Here it is clear that the net torque is positive for small rotation speeds, but becomes negative for large rotation speeds.

The results of the calculation can be plotted in a generalized nondimensional fashion as shown in Fig. 12, the plotted quantity being the torque coefficient $C = 2\Gamma/\rho V^2$, where Γ is the torque and V is the air speed. This is a more refined version of the semi-quantitative curve depicted previously in Fig. 8. There is a threshold value of the parameter $\omega W/2V$, where ω is slat rotation speed, W is slat width, and V is air velocity, below which there is almost no net torque on the slat and so no induced rotation. With increasing values of the rotation speed ω, the torque rises to a maximum and then becomes negative for parameter values above about 0.6. This calculation, of course, takes no account of bearing friction Γ_0, and the calculated torque must exceed Γ_0 if rotation is to

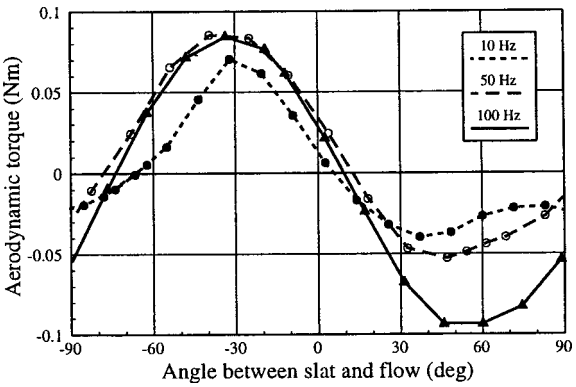


FIG. 11. Numerically calculated torque cycle on a slat of width 60 mm in a flow of 15 m s^{-1} for various rotation rates.

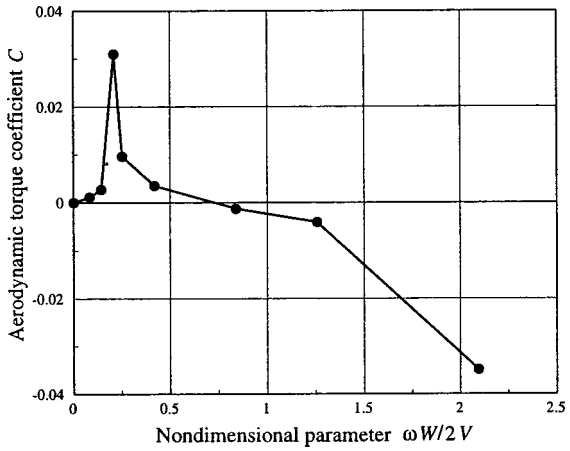


FIG. 12. Numerically calculated coefficient of aerodynamic torque as a function of the nondimensional parameter $\omega W/2V$, where ω is rotation frequency, W is slat width, and V is airstream velocity.

be sustained. This suggests a parameter value in the range 0.3–0.5 for zero net torque, depending on the magnitude of the bearing friction. The steady rotation speed ω_0 attained by the slat will be that for which the net torque is zero, and hence about

$$\omega_0 \approx (0.6 \text{ to } 1.0) V/W. \tag{11}$$

Converting this to apply to the whirled bullroarer strip with string length of about 70 cm and a width W of about 60 mm gives $\omega_0 \approx (0.4 \text{ to } 0.7)\Omega$, where Ω is the arm rotation rate. This conclusion is in good agreement with the measured result (2).

This numerical analysis, like the simple theory, does not reveal any upper limit to the air velocity V for which rotation will occur, and, as indicated in (11), suggests a slat rotation rate that increases linearly with airspeed. Both these findings are in agreement with measurements on the bullroarer (and incidentally with the simple theory of the previous section), confirming that the deviations from this behavior noted in wind-tunnel experiments can be attributed to bearing friction effects.

VI. SOUND PRODUCTION

As discussed in Sec. II and shown in Fig. 4, the radiated sound spectrum is almost a simple amplitude-modulated sine wave with little harmonic development. This suggests that the basic sound-production process is simple and regular. A simple mechanism suggests itself, as discussed below, but first a simple dimensional analysis establishes the expected behavior.

The sound power P conceivably depends upon most of the variables of the situation, so that we write

$$P = \text{const} \times \rho^m \mathrm{M}^n V^p W^q H^r, \tag{12}$$

where ρ is the air density, $\mathrm{M} = V/c$ is the Mach number, V is the translational velocity of the slat, W is its width, and H its length. Equating dimensions leads to the conclusion that $m = 1$, $p = 3$, and $q + r = 2$, but we expect from the fact that the source dimensions are small compared with the sound wavelength that $r = 2$, so that $q = 0$. The power n of the nondimensional Mach number M is undetermined, but we expect it to be non-negative. Indeed the dependence on M could well be more complicated than the simple power-law assumed. It can therefore be tentatively concluded that the radiated sound power is independent of the slat width W but varies as

$$P = \text{const} \times H^2 V^{3+n}, \tag{13}$$

where $n \geqslant 0$. This agrees with the experimental result (4) if $n \approx 2.4 \pm 1$.

In considering the sound-generation mechanism, it can be seen that the rotating slat moving through the still air clearly creates pressure and flow influences that have something of the character of a rotating and oscillating dipole source, with the positive pole of the dipole always directed upstream and having its maximum value when the slat is broadside on to its motion through the air. There are almost certainly quadrupole components as well, but these can be ignored to a first approximation because the power radiated by a macroscopic quadrupole is less than that from a dipole of the same monopole strength and source spacing W by a factor $(\omega W/c)^2$, which is of order 10^{-2} in the present case.

The magnitude of each monopole flow source in the dipole is approximately HWV or $HWL\Omega$, where H is the length of the slat and L is the length of the string. The separation between the two sources is approximately W, so the dipole magnitude is about $\mu \approx HW^2L\Omega$ and its acoustic frequency is $\omega_s = 2\omega \approx 3L\Omega/W$ by (2). The acoustic power radiated by such a dipole source is[5]

$$P = \frac{\omega_s^4 \rho \mu^2}{24\pi c^3}, \tag{14}$$

where c is the speed of sound. Inserting expressions for μ and ω_s gives

$$P \approx \frac{3\rho H^2 L^6 \Omega^6}{c^3} = \frac{3\rho H^2 V^6}{c^3}. \tag{15}$$

This result, which predicts that sound power is proportional to Ω^6, with no dependence on W, is in good agreement with the experimental result (4). Insertion of numerical values for the parameters of the slats used in the outdoor experiment gives a radiated power of order 1 mW, again in good agreement with measurements.

A related approach to the problem that has been well studied is the generation of sound by vortex shedding from a cylinder.[8–10] There are, of course, significant differences, in that vortex shedding in the cylinder case is correlated along its axis only by aerodynamic influences, giving a limited coherence length, while for a rotating slat coherence is enforced along its whole length by the slat rotation. Despite these differences, the cylinder case provides some guidance to the present problem. In particular, the frequency $f = \omega/2\pi$ of vortex shedding derived from equation (11.3.28) of Morse and Ingard or (3.1.11) of Howe is approximately $0.2V/d$, where d is the cylinder diameter, this result agreeing almost exactly with the experimental result (3) for the rotating slat.

If the axial correlation length for vortex shedding is Δ, then the radiated power per length Δ is given by equation (11.3.30) of Morse and Ingard which, in simplified form, is

$$P(\Delta) \approx 0.005 \rho V^3 \Delta^2 \mathrm{M}^3, \quad (16)$$

where M is the Mach number V/c. If Δ is much less than the cylinder length H, then there are H/Δ such sections along the cylinder and, since their phases are incoherent, their powers simply add, so that the total power radiated by the cylinder is given by (16) with Δ^2 replaced by $H\Delta$. This result is the same, except for a minor numerical multiplier, as that given by equation (3.1.13) of Howe,[10] which gives radiated intensity, including a directional factor $\cos^2\theta$.

For a rotating slat, however, the rotation ensures perfect correlation for aerodynamic behavior along the whole slat length H, so that Δ should be replaced by the slat length H, provided this is much less than the sound wavelength. Once again there is no dependence of radiated power on the slat width W, and the power is proportional to H^2V^6 as for the dipole model. There is, however, a great difference in the magnitude of the radiated power, that measured for the rotating slat being more than ten times the value calculated for an equivalent cylinder, even if coherence is assumed. A treatment related more directly to the slat geometry and including rotation could be expected to yield a higher power. The model underlying the aerodynamic generation of sound is, in fact, a refined version of the oscillating dipole model, so that the two are perhaps closely equivalent once the geometrical differences are taken into account.

The directionality of the sound radiation also requires some comment. In each of the explicit models discussed, the radiated sound is directed preferentially along the local direction of motion, which would be in the plane of rotation of the string if that motion were indeed planar. Because the dipole rotates as well as oscillating in sign, the radiation pattern is, however, much broader than for a simple dipole, and the conical form of the string path further broadens this distribution. There may also be some monopole component to the source, which would give a uniform angular distribution of the radiation. Taking these complications into account, it is not surprising that the radiation pattern is approximately uniform to within the modest accuracy of the experimental measurements.

VII. DISCUSSION

The aim of this analysis has been to illuminate the behavior of rotational aerophones of the bullroarer family. This has been achieved in part. The analysis and simplified experiment shows that rotation can occur only above a certain airspeed that is determined by the frictional torque of the bearings. In the case of a bullroarer there are no bearings but rather a string that is progressively twisted by the rotation of the slat, the resisting torque thus increasing steadily with time until it is adequate to stop the rotational motion. The torque then acts to accelerate rotation in the opposite sense, and the motion repeats itself. The period of this reversal, as judged by the sense of the cone described by the string or by the long-term behavior of sound level, is several seconds, though this is determined by the length and thickness of the string. The sound also pulsates at the arm rotation frequency and at twice that frequency, and the origin of these pulsations can be identified as the uneven rotational speed of the slat caused by elliptical arm motion.

Sound production is most easily visualized as the creation of an oscillating dipole across the slat as it rotates. Although such a dipole perturbs the downstream flow and generates a wake of vortices, the radiation from these is of relatively minor importance. A detailed treatment should, of course, include consideration of the aerodynamics of the wake, but largely in order to quantify more exactly the behavior of the attached dipole.

The behavior of rotational aerophones is much more complex to understand than is that of most other aerophones, because this behavior depends on aerodynamics and not on simple quasi-static pressure forces. The experiments and analysis detailed here do not by any means exhaust the possible study, but do provide an outline of the major phenomena and their likely explanation.

As a byproduct of the study, certain experimental problems associated with the apparently much simpler and better controlled environments provided by wind-tunnels and by anechoic rooms were identified and explained.

ACKNOWLEDGMENTS

This study was supported by a grant from the Australian Research Council. The authors are grateful to the Associate Editor (Michael Howe) for pointing out an error in the original draft of the article.

[1] S. Marcuse, *A Survey of Musical Instruments* (Harper & Row, New York, 1975), pp. 549–550.
[2] N. H. Fletcher, "The didjeridu (didgeridoo)," Acoustics Australia **24**, 11–15 (1996).
[3] J. Antill, Ballet Suite "Corroboree," Australian Broadcasting Commission (now Corporation), first British performance 1946.
[4] B. R. Munson, D. F. Young, and T. H. Okiishi, *Fundamentals of Fluid Mechanics* (Wiley, New York, 1998).
[5] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer-Verlag, New York, 1998), pp. 172–174.
[6] *FLUENT 5 User's Guide* (Fluent Inc., Lebanon, NH, 1998).
[7] B. E. Launder and D. B. Spalding, *Lectures on Mathematical Models of Turbulence* (Academic, London, 1972).
[8] R. D. Blevins, "Review of sound induced by vortex shedding from cylinders," J. Sound Vib. **92**, 455–470 (1984).
[9] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw–Hill, New York, 1968), pp. 717–777.
[10] M. S. Howe, *Acoustics of Fluid–Structure Interactions* (Cambridge U.P., Cambridge, 1998), Sec. 3.1.

Three-dimensional modeling of acoustic backscattering from fluid-like zooplankton

Andone C. Lavery, Timothy K. Stanton, Duncan E. McGehee,[a] and Dezhang Chu
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

(Received 23 March 2001; revised 19 October 2001; accepted 13 November 2001)

Scattering models that correctly incorporate organism size and shape are a critical component for the remote detection and classification of many marine organisms. In this work, an acoustic scattering model has been developed for fluid-like zooplankton that is based on the distorted wave Born approximation (DWBA) and that makes use of high-resolution three-dimensional measurements of the animal's outer boundary shape. High-resolution computerized tomography (CT) was used to determine the three-dimensional digitizations of animal shape. This study focuses on developing the methodology for incorporating high-resolution CT scans into a scattering model that is generally valid for any body with fluid-like material properties. The model predictions are compared to controlled laboratory measurements of the acoustic backscattering from live individual decapod shrimp. The frequency range used was 50 kHz to 1 MHz and the angular characteristics of the backscattering were investigated with up to a 1° angular resolution. The practical conditions under which it is necessary to make use of high-resolution digitizations of shape are assessed. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1433813]

PACS numbers: 43.30.Ft, 43.30.Sf, 43.20.Fn [DLB]

I. INTRODUCTION

Acoustic surveys are commonly used as a tool for assessing the distribution of zooplankton and other marine organisms that inhabit the water column. The primary advantage of this technique over other more conventional methods, such as net tows, is the possibility of performing synoptic surveys while also obtaining high-resolution data over relevant temporal and spatial scales (Medwin and Clay, 1998; Foote and Stanton, 2000). To accurately measure zooplankton biomass using this technique, it is necessary to first understand the process by which zooplankton scatter sound. Understanding the scattering of sound from zooplankton is a challenge for a number of reasons. First, many different species of zooplankton are typically present in the water column at any given time and location. Second, the scattering properties of these different species can vary enormously with organism size, shape, orientation, material properties, and acoustic frequency. Under many circumstances, successful interpretation of acoustic field data is limited in part by the availability of acoustic scattering models that take into account these diverse scattering properties. Though it is not feasible to develop a scattering model for each individual animal, it is possible to establish a categorization scheme based on general scattering properties and boundary conditions. One such scheme was developed by Stanton *et al.* (1994, 1998a, 1998b) in which three categories were identified; zooplankton with fluid-like acoustic characteristics (e.g., euphausiids, copepods), zooplankton with gas inclusions (e.g., siphonophores), and zooplankton with elastic shells (e.g., pteropods).

The investigation focuses on organisms with fluid-like material properties, meaning that shear waves are not included in the formalism. It is also assumed that there are no gas inclusions. Although the shape of the animals investigated in this paper is elongated, the model is not restricted to elongated scatterers, but is generally applicable to any fluid-like body, of any complex shape and size. Elongated fluid-like zooplankton are a possible application of this general model and are of particular interest due to the ecological significance and naturally high abundances found in many regions (Greene *et al.*, 1991; Hewitt and Demer, 1991). Specifically, in this study we will present data and modeling results obtained for common shore shrimp, *Palaeomonetes vulgaris*. This crustacean has fluid-like material properties and a similar shape and size to many elongated fluid-like zooplankton, and has the added advantage of being conveniently obtainable from local waters. In addition, previous models based on this local species have been adapted with some success to oceanic species. There have been a number of laboratory and field studies aimed at understanding acoustic scattering from fluid-like scatterers, concentrating predominantly on decapod shrimp and euphausiids (Greenlaw, 1977; Everson, 1982; Everson *et al.*, 1990; Foote *et al.*, 1990; Chu *et al.*, 1992, 1993; Martin Traykovski *et al.*, 1998; McGehee *et al.*, 1998; Stanton *et al.*, 1993a, 1993b, 1996, 1998b, 1998c, 2000).

In a recent paper, Stanton and Chu (2000) presented a comprehensive review of the scattering models currently available for predicting the scattering of sound from elongated fluid-like zooplankton. Briefly synthesizing this work, the Anderson (1950) homogeneous fluid-sphere model is the simplest possible representation for animal shape used to describe scattering from this class of zooplankton (Greenlaw, 1977; Holliday *et al.*, 1989; Holliday and Pieper, 1995). In a series of studies, cited in Stanton and Chu (2000), Stanton and others have developed a number of models that make use of finite-length deformed cylinders to describe animal shape, resulting in a more accurate description of the ob-

[a]Currently at BAE Systems, Analysis & Applied Research Division, San Diego, CA 92123.

served angular scattering characteristics. To date, the most sophisticated model available for this class of zooplankton, with the largest range of applicability, is a deformed-cylinder model based on the distorted wave Born approximation (DWBA) (Chu *et al.*, 1993; Stanton *et al.*, 1993b, 1998a, 1998b; McGehee *et al.*, 1998; Chu and Ye, 1999; Stanton and Chu, 2000). With this model it is possible to incorporate a relatively realistic animal shape, in addition to variations in the material properties, though a fundamental restriction of this model is the assumption of cylindrical symmetry. A result of this limitation is that the animal cross section is circular at every point along the lengthwise axis of the animal, and the material properties can only vary along the lengthwise axis.

In this study, we develop a DWBA-based scattering model that incorporates high-resolution, fully three-dimensional (3D) digitizations of animal shape obtained through the use of computerized tomography (CT). This model by-passes the limitations imposed by the assumption of cylindrical symmetry made in the finite-length deformed-cylinder model. We also compare the model predictions to acoustic backscattering laboratory measurements of both individual and aggregations of live decapod shrimp. The data presented in this paper were obtained during two separate experiments. The most recent data set, described in detail and for the first time in this paper, was obtained for live individuals under well-characterized and controlled conditions, using both single-frequency and broadband signals, with the animal orientation carefully controlled and with an angular resolution of 1° in two scattering planes. The second data set used in this paper to compare with the model predictions was previously published by Stanton *et al.* (1993b) and involves aggregations of live decapod shrimp. Though the acoustics and CT data were obtained on animals of the same species and of generally the same shape and size, for practical reasons the acoustics and CT data involved different specimens.

This paper is organized as follows. In Sec. II, two DWBA-based scattering models are described. One of these models makes use of 3D measurements of the animal shape and results in a volume-integral solution; the other model, the deformed-cylinder model, uses 2D measurements of animal shape and results in a line-integral solution. In Sec. III, the laboratory setup, methods used for data collection and analysis, and the methods used to obtain 2D and 3D measurements of animal shape are outlined. Intermodel comparisons are performed in Sec. IV. Scattering predictions are compared to the data in Sec. V, and the physics of the scattering process is discussed in Sec. VI. Finally, in Sec. VII, the results are summarized, and recommendations are made regarding conditions under which the various approaches should be used.

II. THEORY

A. DWBA-based scattering model

For a single insonification of a target of any complex shape and size, the incident and scattered sound wave are related by the scattering amplitude. The scattering amplitude, which has units of length, is a measure of the efficiency with which a target scatters sound, and is a function of the target's shape, size, orientation, material properties (mass density, ρ, and sound speed c) and the acoustic wave number, k, of the incident wave ($k=2\pi/\lambda$, where λ is the acoustic wavelength). The scattering amplitude for sound scattering from any weak scatterer with a finite body of volume V (such as the animals under investigation here, which have material properties that are similar to those of the surrounding medium) can be modeled in the far field using the Born approximation (Morse and Ingard, 1968). In fact, all formulations and measurements presented in this paper are for backscattering, in which case the backscattering amplitude, f_{bs}, is given by

$$f_{bs}=\frac{k_1^2}{4\pi}\int\int_V(\gamma_\kappa-\gamma_\rho)e^{2i\mathbf{k}_{2i}\cdot\mathbf{r}_v}dV, \tag{1}$$

where k_1 is the wave number of the incident sound ($k_1=|\mathbf{k}_{1i}|$), $\mathbf{k}_{2i}$ is the wave vector of the incident sound evaluated in the interior of the volume, $\mathbf{r}_v$ is the position vector of any volume element, $\gamma_\kappa=(\kappa_2-\kappa_1)/\kappa_1=(1-gh^2)/gh^2$, $\gamma_\rho=(\rho_2-\rho_1)/\rho_2=(1-g)/g$, and $g(=\rho_2/\rho_1)$ and $h(=c_2/c_1)$ are the density and sound-speed contrasts, respectively. $\kappa_1(=(\rho_1c_1^2)^{-1})$ and $\kappa_2(=(\rho_2c_2^2)^{-1})$ are the compressibilities in the surrounding medium and body interior. It should be noted that the wave number in the exponent has been evaluated within the body interior ($\mathbf{k}_{2i}$ instead of $\mathbf{k}_{1i}$ in the exponent). This modification to the Born approximation has been referred to as the distorted wave Born approximation (DWBA). Throughout this work, the subscript "1" indicates that the quantity referred to is evaluated in the surrounding medium, while the subscript "2" indicates that the quantity referred to is evaluated in the body interior. If the material properties are constant throughout the volume ($\gamma_\kappa-\gamma_\rho$) can be factored out of the integral, and, at very low frequencies, it is straightforward to show that f_{bs} is proportional to V ($f_{bs}=(k_1^2/4\pi)(\gamma_\kappa-\gamma_\rho)V$).

For an elongated body with a circular cross section at every point along its lengthwise axis, Eq. (1) can be further simplified to a one-dimensional line integral (Stanton *et al.*, 1998b), given by

$$f_{bs}=\frac{k_1}{4}\int_{r_p}(\gamma_\kappa-\gamma_\rho)a\frac{J_1(2k_2a\cos\beta_{\text{tilt}})}{\cos\beta_{\text{tilt}}}e^{2i\mathbf{k}_{2i}\cdot\mathbf{r}_p}|\mathbf{dr}_p|, \tag{2}$$

where $\mathbf{r}_p$ is the position vector of the body centerline, a is the radius, and β_{tilt} the local tilt angle of the body cross section relative to $\mathbf{k}_{1i}$ at each point on the axis. In this case, ($\gamma_k-\gamma_\rho$), a, and β_{tilt} can vary along the length of the body, but are restricted to a constant value in any given cross section.

For a small number of cases it is possible, by making very restrictive, but simplifying, assumptions about the shape and material properties (e.g., fluid-filled sphere or ellipsoid), to obtain a closed-form mathematical expression for the backscattering amplitude. In general, however, particularly when the conditions are as complex as those of the decapod shrimp we are attempting to model here, it is necessary to solve for the backscattering amplitude numerically.

The far-field backscattered energy is often expressed in terms of the target strength (TS) with units of decibel (dB) relative to 1 m^2 (Urick, 1983), and is given by

$$\mathrm{TS} = 10 \log \sigma_{bs} = 10 \log |f_{bs}|^2, \tag{3}$$

where $\sigma_{bs} = |f_{bs}|^2$ is the differential backscattering cross section, and differs from the often-used backscattering cross section σ by a factor of 4π ($\sigma = 4\pi\sigma_{bs}$). In order to compare scattering from objects of different sizes but similar proportions, TS is often normalized according to the square of some typical dimension, giving rise to the reduced target strength (RTS). For elongated zooplankton, of length L, the RTS is given by

$$\mathrm{RTS} = 10 \log \frac{\sigma_{bs}}{L^2} = 10 \log |f_{bs}|^2 - 10 \log L^2. \tag{4}$$

B. Modeling considerations

The most rigorous method of modeling the acoustic scattering from a complex body is to digitize the body shape and material properties within the body in increments significantly smaller than the wavelength. These high-resolution digitizations can then be used as input to a general 3D scattering model, such as that given by Eq. (1). Under most circumstances, measurements at a resolution of approximately $\lambda/10$ to $\lambda/20$ are required to accurately represent the animal's body, although the exact value may vary with the modeling approach (Stanton and Chu, 2000). This digitization criterion for the outer boundary shape is met for the range of frequencies used in these experiments when the animal shape is obtained from CT scans. However, if high-resolution CT scans are not available, it is possible to perform lower resolution 2D measurements of outer boundary shape, and use these digitizations as input to a simplified scattering model, such as that given by Eq. (2). As will be seen later, there are circumstances under which caution must be used when applying the line-integral model, although there are conditions when the results of the line-integral and volume-integral models are very similar. Comparison of the results obtained using these two approaches is one of the goals of this work.

High resolution digitizations of the material properties are also necessary to accurately evaluate f_{bs}. However, there are only a few studies in which the material properties of fluid-like zooplankton have been investigated (Greenlaw, 1977; Foote, 1990; Chu *et al.*, 2000) and in these cases, only the average quantities were studied. There is little information available as to the spatial variability of these properties throughout the body interior (Yayanos *et al.*, 1978; Foote, 1998). For most of the work presented here, g and h are held constant throughout the body interior, with values of $g = 1.0357$ and $h = 1.0279$ taken from Foote (1990), as measured for *Euphausia superba*.

III. EXPERIMENTAL METHODS

A. Acoustic data acquisition

The acoustic backscattering measurements involving individual decapod shrimp were made on ten specimens, but only data collected from animals that remained alive and in good condition throughout the data acquisition process are considered here, namely the four animals numbered 6, 8, 9, and 10. The experimental setup was very similar to that described in Stanton *et al.* (2000), and only a brief overview of the relevant equipment and methods will be described here together with a description of the difference between the two setups. The data presented in this study were collected in 1996 in the acoustics laboratory at the Woods Hole Oceanographic Institution.

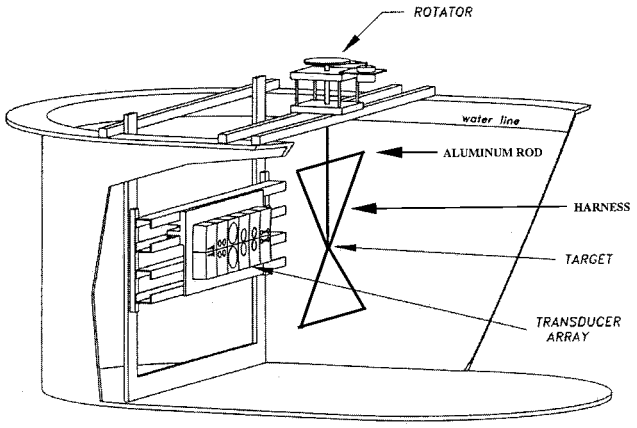


FIG. 1. Schematic of the experimental system used for acquiring acoustic backscattering data as a function of angle of orientation and frequency: tank, transducer array, stepper motor, and (acoustically transparent) harness system with horizontal aluminum rods used to tether the decapod shrimp.

The experiments were performed in a 1.5-m-deep by 3.7-m-long and 2.4-m-wide tank filled with filtered seawater (Fig. 1). An array of transducer pairs was mounted in the tank facing horizontally. A combination of single-frequency and broadband transducers was employed to collect the acoustic backscattering data. The transducers within each pair were placed in a bistatic configuration, as close as possible to each other so as to approximate a monostatic configuration. Six pairs of powerful narrow-band (NB) transducers, emitting 200-μs-long gated sine waves, were employed at the following frequencies: 50, 75, 120, 165, 200, and 258 kHz. In addition, two octave-bandwidth broadband (BB) transducers were also employed, emitting 200-μs-long "chirp" (linear frequency modulated) signals with center frequencies at 250 and 500 kHz. The transmitted voltage time series, $v_{bs}^T(t)$, and the received voltage time series, or equivalently, the backscattered return echo from the animal, $v_{bs}^R(t)$, were stored on a personal computer for later analysis. One of the main advantages of using broadband transducers is that large amounts of information can be gathered over a wide frequency range, allowing both temporal and spectral data analysis techniques to be exploited, maximizing the information that can be extracted about the sound scattering process. However, a considerable drawback of employing broadband signals is that less energy can be applied per frequency bin.

The animals were tethered in a two-point acoustically transparent harness made of 59-μm-diameter monofilament line (Fig. 1). This involved a loop around the thorax, and another line running through the midpoint of the abdomen [Fig. 2(a)]. The loop was loose enough to permit normal respiration and blood flow, but tight enough to prevent sub-

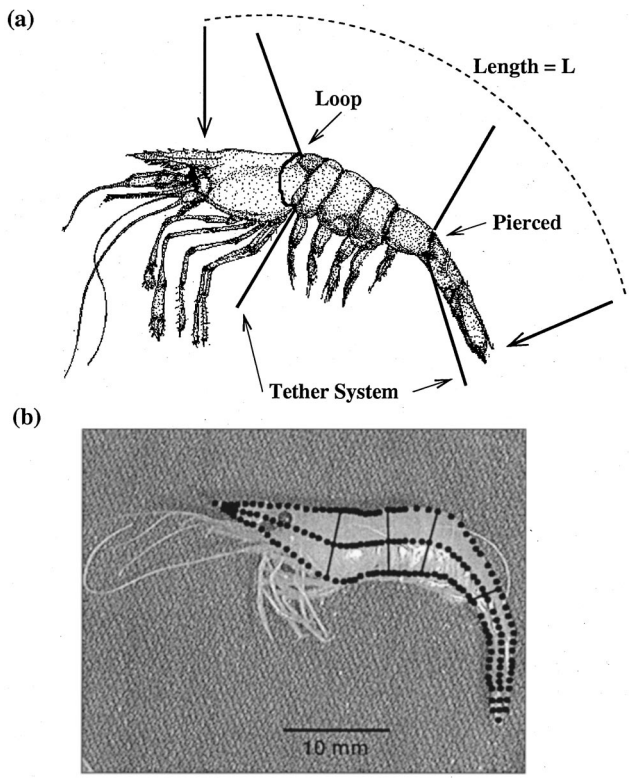


FIG. 2. (a) Schematic diagram of a decapod shrimp, showing the two-point tether. The animals were tethered with a loop about the thorax and were also pierced through the abdomen. The animal lengths (L) were measured from the photographic images. (b) The animal shape was digitized in 2D by obtaining measurements in the dorsal–ventral plane. The measurements are shown as black points along the outside contour of the animal. The calculated centerline is also shown (from which the length, L, was calculated), together with a selection of radii. For the scattering predictions made using the line-integral model, based on the 2D digitizations of animal shape, the animals were digitally straightened so that the shape of the animal more accurately reproduced the observed shape during the acoustic backscattering measurements. However, it was not necessary to digitally straighten the CT data since the animals were naturally straighter during the CT scanning process.

stantial movement. For the data reported here, the animals remained alive and in good condition throughout the data acquisition process. Tension was maintained in the harness by two horizontal aluminum rods. This kept the position and orientation of each animal stable. The animals were centered in the acoustic beam in the far field of the transducers at a range of 0.51 m. A computer-controlled stepper motor rotated the animals in 1° increments, in most cases through two full rotations (720°). The animals were insonofied with a single ping at each angle of orientation. Table I summarizes the frequencies and angles of orientations for which backscattering measurements were performed. Measurements of the animal lengths, radii, aspect ratios (L/a), and wet weights are summarized in Table II.

Animals 6 and 8 were tethered such that the dorsal–ventral aspect was approximately in the vertical plane, with an out-of-plane tilt of 25°–45°. In contrast, animals 9 and 10 were tethered such that the dorsal–ventral aspect of the animals was approximately in the horizontal plane, with an out-of-plane tilt of 25°–45°. There is some uncertainty in the tilt angle of the animals since there was no way to accurately quantify it and it also changed slightly as the animals were rotated.

TABLE I. Range of orientations and frequencies used to obtain acoustic backscattering measurements.

Frequency (kHz)	Animal no.	Range of orientations in 1° increments
50	6,8,10	0–720°
75	6,8,10	0–360°
120	6,8,9,10	0–720°
160	6,8,9,10	0–720°
200	6,8,9,10	0–720°
258	6,8,9,10	0–720°
250 (BB)	6,8,9,10	0–720°
500 (BB)	6,8,9,10	0–720°

The transducers were calibrated prior to the backscattering measurements following the procedure outlined in Stanton *et al.* (1998a). During the calibration, the receiver and transmitter were aimed at each other at a range $r_{\mathrm{cal}}=0.69$ m. For each transducer pair, ten 200-μs-long chirp signals were collected and averaged. The time series of the transmitted, $v^T_{\mathrm{cal}}(t)$, and received, $v^R_{\mathrm{cal}}(t)$, calibration voltages were stored for the purpose of postprocessing calibration of the scattering data. The background reverberation of the tank, without the animals present, was also measured for each transducer pair in the bistatic scattering configuration. This measurement involved averaging the background reverberation signals, for each set of transducers, over a large number of echoes. The average reverberation echo was stored on the data acquisition oscilloscope and subtracted in real time from the echoes obtained once the animals were placed in the tank. This procedure was repeated for each animal. Great care was taken throughout the data acquisition process to minimize the presence of bubbles on the tethers, since the acoustic signals from bubbles can greatly contaminate the acoustic returns of interest. For each acoustic return from each transducer pair

$$f_{bs}=\frac{V^R_{bs}}{V^R_{\mathrm{cal}}}\frac{V^T_{\mathrm{cal}}}{V^T_{bs}}\frac{r^2_{bs}}{r_{\mathrm{cal}}}, \tag{5}$$

where V^R_{bs}, V^T_{bs}, V^R_{cal}, and V^T_{cal} are the absolute values of the Fourier transforms of $v^R_{bs}(t)$, $v^T_{bs}(t)$, $v^R_{\mathrm{cal}}(t)$, and $v^T_{\mathrm{cal}}(t)$, respectively. Prior to evaluating the Fourier transforms of the voltage time series, a bandpass filter was applied to the data, to reduce out-of-band noise.

B. 2D measurements of animal shape

For application of the line-integral model it was necessary to estimate the cylindrical radius of each animal at a number of points along the lengthwise axis. These measurements were performed by first taking 2D photographic images of the preserved animals in the dorsal–ventral plane. As mentioned earlier, one of the underlying assumptions of this model is that the animal cross section is circular. Consequently, it is sufficient to obtain 2D measurements of the animal radius in one plane at a number of points along the centerline axis. The animal shapes were digitized by collecting the coordinates of a number of points along their outer contours. Animal centerlines and radii were then calculated from these measurements. To obtain the coordinates and radii

TABLE II. Animal lengths (defined in Fig. 2), average torso radii, wet weights, aspect ratio, number of data points along the lengthwise axis (including end points) taken from the 2D photographic images, and availability of acoustic and CT data. The average torso radii were calculated from the 2D photographic images by averaging over all the radii, excluding the ten points at each end of each animal.

Animal no.	Length (L) (mm)	Radius (a) (mm)	Wet weight (g)	Aspect ratio (L/a)	N	Acoustic: CT
6	30	1.5	0.17	20.5	49	Yes:No
8	25	1.35	0.10	17.7	46	Yes:No
9	25	1.6	0.10	15.6	45	Yes:No
10	26	1.4	0.13	18.1	47	Yes:No
11	39	2.2	0.48	17.9	57	No:Yes
13	35	2.3	0.45	16.9	43	No:Yes
14	30	1.65	0.19	18.2	41	No:Yes

of N points along the centerline axis, including two end points, it is necessary to collect $(2N\text{-}2)$ points along the outer contour of the animal (Table II). The result of this 2D digitization process is shown in Fig. 2(b) for animal 11. Since these measurements were performed in the dorsal–ventral plane of the animals, the results more closely represent a side view of the actual animal than a top view. Finally, since the digitization of animal shape was performed after the animals had been preserved, it was necessary to digitally straighten the animals so that they more closely represented the actual live animal shape.

C. 3D measurements of animal shape: CT scans

High-resolution computerized tomography (CT) scans were performed for three decapod shrimp (animals 11, 13, and 14) at the Massachusetts Eye and Ear Hospital using a spiral CT scanner. Scans of animals 13 and 14 were performed with the animals live (but not in water), while animal 11 was scanned after recent preservation. The CT scans were used to evaluate the 3D volume integral with the integration performed within MATLAB (version 5.3). Details of this procedure are described below.

The CT scans were performed along the lengthwise axis of the animals, with a full set of slices obtained in a 1–2 s time interval. The separation, dz, between the slices was 100 μm for animals 13 and 14, corresponding to the resolution limit of the CT scanner, while for animal 11 the separation between the slices was 200 μm. This resulted in N_z slices per animal, which varied according to the animal length (Table III). A 3D image of animal 11 is shown in Fig. 3(a), together with a number of representative CT slices. The resolution of the CT data along the lengthwise axis of the animal, determined by dz, and the resolution within any given slice, determined by the focus and zoom of the scanner, differ. However, for the CT data acquired during this study, both measures of resolution adequately meet the resolution criterion stated earlier, over the full range of acoustic wavelengths used for the data acquisition. For instance, at 500 kHz, 200 μm corresponds to approximately $\lambda/15$.

For each animal, the original CT slices contained 512 $\times$512 pixels. For ease of manipulation and numerical efficiency, it was possible to crop each CT slice (referenced to a common pixel to maintain overall alignment) to contain $N_p \times N_p$ pixels. Each CT slice was then converted to a binary

TABLE III. Animal number, number of CT slices (N_z), number of pixels per CT slice ($N_p \times N_p$), separation between CT slices (dz), and pixel area within each slice. Note that the length of animal 13 as measured from the photographic images (Table II) is smaller than that of animal 11. However, there are more CT slices of animal 13 since more of the antennae were captured during this set of scans. These animal parts are not expected to contribute significantly to the scattering; the lengths used to calculate the RTS values are based on Table II.

Animal no.	N_z	$N_p \times N_p$	dz (μm)	Pixel area (μm)^2
11	156	128×128	200	100×100
13	399	100×100	100	104×104
14	301	100×100	100	104×104

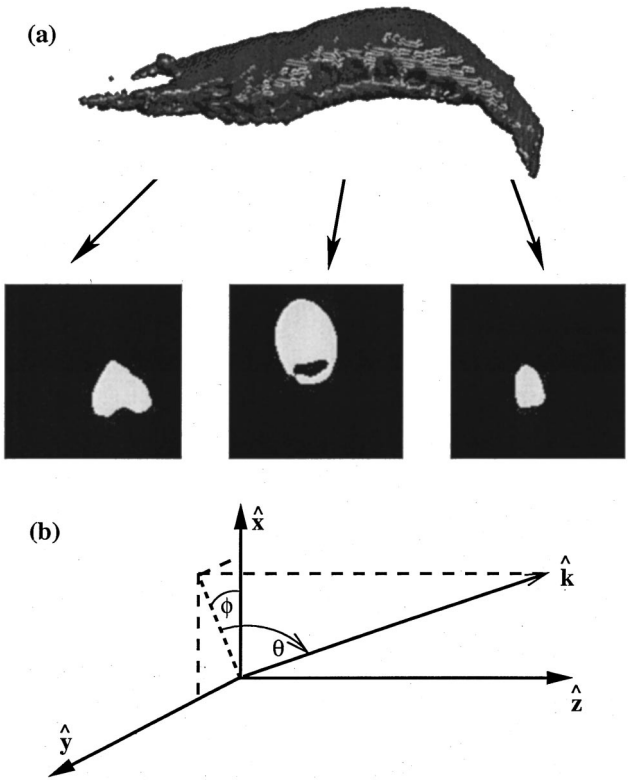


FIG. 3. (a) 3D image of animal 11, together with three representative CT slices. (b) Scattering geometry used in the scattering models. The $\hat{z}$ axis is aligned along the lengthwise axis of the animal. Broadside incidence corresponds to, approximately, θ=0° and θ=180°. Head-on incidence corresponds to $\theta=-90°$, or equivalently, θ=270°. End-on incidence corresponds to, approximately, θ=90°. The ϕ=0° and 180° planes correspond to the dorsal and ventral sides of the animals, while the $\phi=\pm 90°$ planes corresponds to the left and right sides of the animal.

matrix, $\mathbf{M}^l$, of size $N_p \times N_p$, where l runs from 1 to N_z. Matrix elements corresponding to pixels with an intensity above a certain threshold value contain ones and all others elements are zero. The matrices $\mathbf{M}^l$ were relatively insensitive to the value chosen for the threshold once it was set high enough to eliminate background diffraction effects. Thus, for each incident wave vector, the backscattering amplitude, f_{bs}, given as a volume-integral by Eq. (1), becomes

$$f_{bs} = \frac{k_1^2}{4\pi}(\gamma_\kappa - \gamma_\rho) \sum_{l=1}^{N_z} \sum_{j=1}^{N_p} \sum_{i=1}^{N_p} M_{ij}^l \times e^{2ik_2(\cos\theta\cos\phi X_{ij}^l + \cos\theta\sin\phi Y_{ij}^l + \sin\theta Z_{ij}^l)} dV_{ij}^l. \quad (6)$$

X_{ij}^l, Y_{ij}^1, and Z_{ij}^1 are the position vectors of each pixel, dV_{ij}^l corresponds to the volume associated to each pixel, and the material properties have been assumed constant throughout the body interior. The direction of the incident (and backscattered) wave vector is $\hat{k}_{li} = (\cos\theta\cos\phi\hat{x} + \cos\theta\sin\phi\hat{y} + \sin\theta\hat{z})$, where the coordinate system is aligned with the CT scanner coordinate system [Fig. 3(b)]. Thus, the $\hat{z}$ direction is, approximately, along the lengthwise axis of the animals. Broadside incidence corresponds to $\theta = 0°$ or $\theta = 180°$, in some ϕ plane, while $\theta = \pm 90°$ corresponds to end-on and head-on incidence, in any ϕ plane.

Simple variations in the material properties along the lengthwise axis of the animals can be incorporated into Eq. (6) by allowing the quantity $(\gamma_\kappa - \gamma_\rho)$ to vary between CT slices. Thus, Eq. (6) becomes

$$f_{bs} = \frac{k_1^2}{4\pi} \sum_{l=1}^{N_z} \sum_{j=1}^{N_p} \sum_{i=1}^{N_p} (M')_{ij}^l \times e^{2ik_2(\cos\theta\cos\phi X_{ij}^l + \cos\theta\sin\phi Y_{ij}^l + \sin\theta Z_{ij}^l)} dV_{ij}^l, \quad (7)$$

where $(M')_{ij}^l = (\gamma_\kappa - \gamma_\rho)^l M_{ij}^l$, and $(\gamma_k - \gamma_\rho)^l$ has a single value within any given CT slice. This is only an approximate method of incorporating lengthwise variations in the material properties of the animals since it does not take into account the degree of animal bend.

Full 3D variation in the material properties can also be easily incorporated in a similar manner. In this case, $(M')_{ij}^l = (\gamma_\kappa - \gamma_\rho)_{ij}^l M_{ij}^l$, where $(\gamma_\kappa - \gamma_\rho)_{ij}^l$ has a distinct value for each pixel. Finally, it is worth noting that it may be possible to extract density contrast values from the CT data, although it would be necessary first to calibrate the CT data.

D. Averages

In general, the scattering process is highly complex and can depend on many parameters simultaneously. Given the uncertainty in each parameter, predictions of a single scattering realization are often difficult to make. As a result, comparisons between predictions and data for single realizations are generally qualitative. Quantitative comparisons can be made for both the statistics of scattering and for averaged levels, where the averages can be over some distribution of parameters, such as a distribution of sizes or angles of orientation. There is also a more direct relationship between this approach and volume scattering data, such as that collected from aggregations involving a varied distribution of animal sizes and angles of orientation. It is possible to predict average scattering levels more accurately since the complex and stochastic structure of the scattering is smoothed during the averaging process. In this paper, the focus is on averages over orientation since the number of animals for which there are scattering data, and their size range, is relatively small.

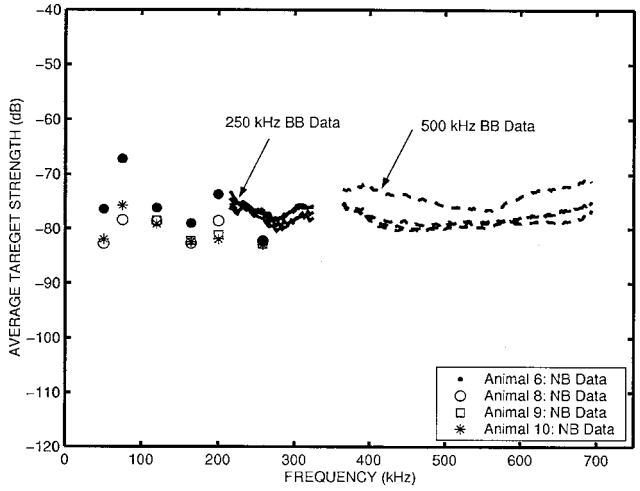


FIG. 4. TS averaged over orientation vs frequency for animals 6, 8, 9, and 10. The four solid lines represent data obtained using the 250-kHz broadband (BB) transducer. The four dashed lines represent data obtained using the 500-kHz broadband transducer. The symbols represent the average TS values obtained from the single-frequency, or narrow-band (NB), data. The averages are taken for all angles uniformly in one plane.

At sufficiently high frequencies, acoustic returns from aggregations of zooplankton tend to add incoherently. Thus, average TS ($= 10\log\overline{\sigma_{bs}}$) values were obtained by averaging the differential backscattering cross section, σ_{bs}, over all angles of orientation. Since there is little information available as to the natural orientation distributions of decapod shrimp, or any other elongated fluid-like animals, average TS values were calculated by assuming a uniform distribution of orientations from 0–360°.

The average TS values as a function of frequency were calculated for all four animals, for both single-frequency and broadband data, and these results are shown in Fig. 4. Once averages are taken, much of the fine structure observed in the data on a ping-by-ping basis is smoothed out. The TS values for animal 6 obtained from the 500-kHz broadband transducer were approximately 6 dB higher than those obtained for the other animals. Although no errors were noted during the data acquisition process for this animal, it is hypothesized that this anomaly could be due to an error in an oscilloscope setting.

IV. MODEL PREDICTIONS

A. Comparison of TS predictions versus angle

Though the TS values predicted by both models spanned a similar range for the frequencies and angles investigated, it was found that the differences between the predicted scattering levels were generally larger at angles close to end-on incidence than at angles close to broadside (Fig. 5), with the line-integral model predicting considerably higher values at angles close to end-on incidence. These differences were

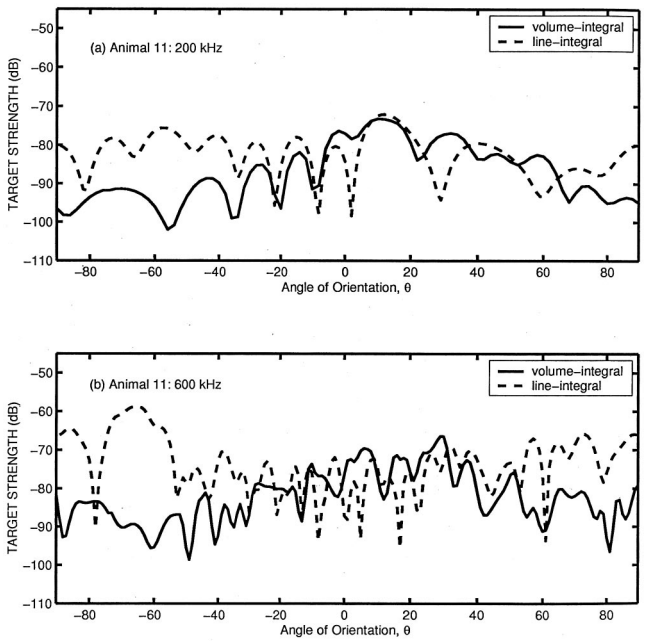


FIG. 5. Comparison of TS predictions for animal 11, in the $\phi=0°$ plane (dorsal–ventral), based on the line-integrated (dashed line) and volume-integral (solid lines) models. (a) 200 kHz and (b) 600 kHz. The largest differences observed between the two models occur at angles close to head-on and end-on incidence ($\theta=\pm 90°$) where the line-integral predicts higher backscattering levels, particularly at higher frequencies.

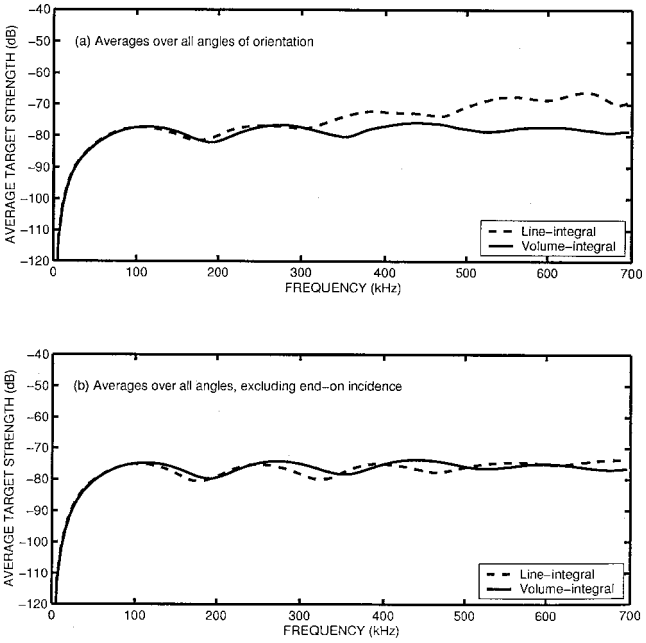


FIG. 6. Comparison of TS averaged over orientation vs frequency predicted by the line-integral (dashed lines) and the volume-integral (solid lines) models for animal 11 in the $\phi=0°$ plane (dorsal–ventral). (a) Averages performed over all angles of orientation uniformly. (b) Averages performed uniformly over all angles of orientation excluding the range of angles $\pm 25°$ centered on end-on incidence.

accentuated at the higher frequencies [Fig. 5(b)]. For ease of intermodel comparison, most scattering predictions were performed in the $\phi=0°$ plane. However, predictions were also made in the $\phi=45°$ and $\phi=90°$ planes, with trends in the results similar to those found for the $\phi=0°$ plane.

B. Comparison of average TS predictions

The average TS was initially calculated for averages taken over all angles of orientation uniformly distributed in the $\phi=0°$ plane [Fig. 6(a)]. The average TS values predicted by both models agree closely at the lower frequencies, but the agreement deteriorates with increasing frequency. At 700 kHz, the line-integral prediction lies almost 10 dB higher than the volume-integral prediction. At high frequencies, the volume-integral model predicts that the average TS attains a relatively constant value, whereas the line-integral model predicts that the average target strength slowly increases with increasing frequency, over the frequency range investigated. Average TS values were then calculated with the $\pm 25°$ centered about end-on incidence excluded [Fig. 6(b)]. In this case, both models predict very similar average TS levels. It can be seen that the differences between the model predictions are generally within several dB at most frequencies investigated. It is apparent from Figs. 5 and 6 that at higher frequencies the line-integral model is predicting significantly higher average TS levels when the averages include angles close to end-on incidence. This trend was observed for all ϕ planes investigated.

C. Comparison of average RTS predictions for different animals

The differences in the average RTS values obtained using the volume-integral model, for the three animals for which CT scans were available, are illustrated in Fig. 7. The average RTS values are plotted as a function of the dimensionless product *ka* so as to account for the different radii of the animals. (The values of *a* used to evaluate *ka* are those presented in Table II.) It can be seen that the structure of the

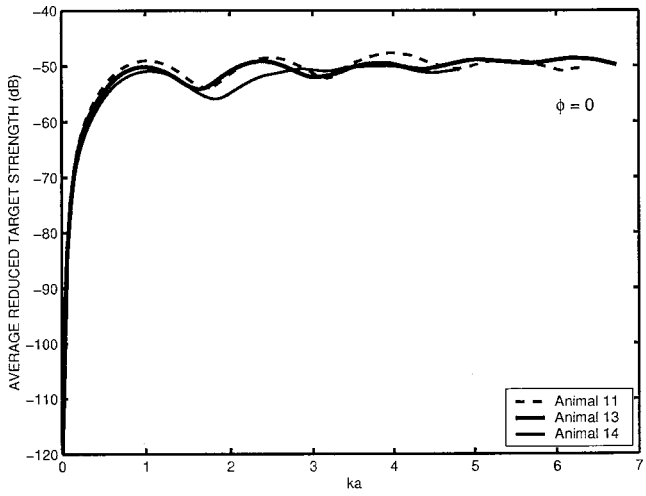


FIG. 7. Comparison of RTS averaged over orientation vs *ka* predicted by the volume-integral model for animals 11 (dashed line), 13 (thick solid line), and 14 (thin solid line), in the $\phi=0°$ plane (dorsal–ventral). The average torso radii, calculated from the 2D animal measurements (Table II), were used to evaluate *ka*.

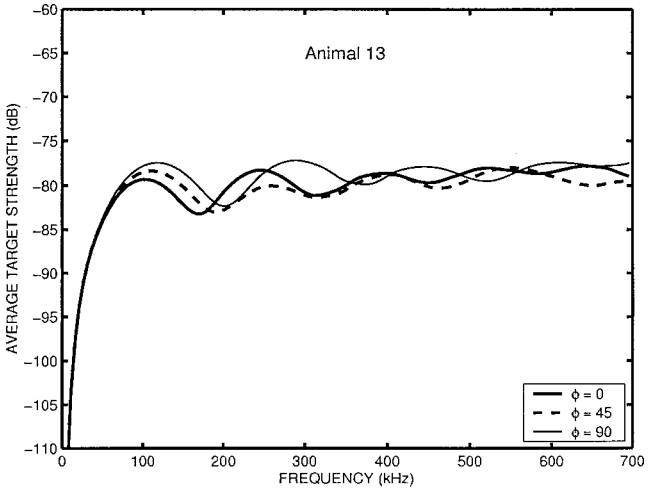


FIG. 8. Comparison of TS averaged over orientation vs frequency predicted by the volume-integral model for animal 13 in three different ϕ planes. The $\phi=0°$ plane corresponds to the dorsal side of the animal, while $\phi=90°$ corresponds to the left side of the animal.

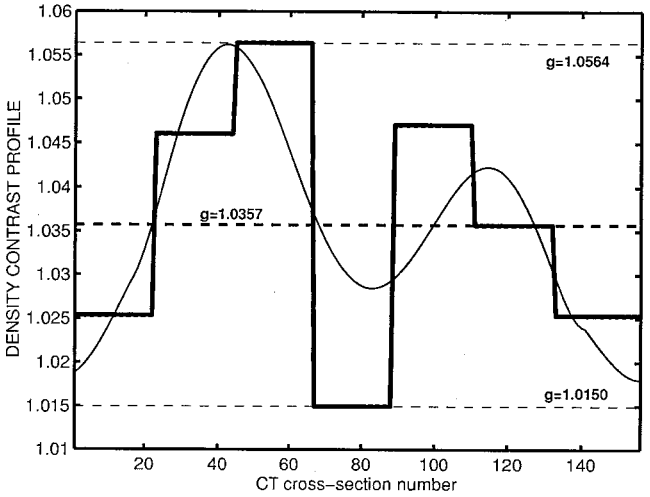


FIG. 9. Material property profiles. Three different uniform values of g were investigated (dashed lines): $g=1.0564$, $g=1.0357$, $g=1.015$. The thick solid line represents the profile with seven segments with a mean value of $g=1.0357$. The thin solid line represents the smoothly varying profile, with the mean value also kept at $g=1.0357$. The material profiles only varied along the lengthwise axis of the animal.

average RTS is very similar for all three animals, and the transition from Rayleigh to geometric scattering occurs at similar *ka* values.

D. Comparison of average TS predictions for different ϕ planes

To illustrate that the conclusions drawn above are not dependent on the ϕ plane chosen, average TS values for animal 13, calculated using the volume-integral model with $\phi=0°$, $\phi=45°$, and $\phi=90°$, are shown in Fig. 8. It can be seen that the average TS values in the different planes differ by no more than a few dB. There is a shift in the location of the first peak and null which is probably related to the fact that the average width of the animal is different in different ϕ planes.

E. Variations in material properties

The effect on backscattering due to changes in material properties was also briefly investigated. Two approaches were taken. The first approach involved studying the effects of changes in the values of the material properties, assuming homogeneous values throughout the body interior. The change in TS due to changes in the overall average levels of g and h is given by

$$\Delta \mathrm{TS} = \mathrm{TS}^{(2)} - \mathrm{TS}^{(1)} = 20 \log \frac{(\gamma_\kappa - \gamma_\rho)^{(2)}}{(\gamma_\kappa - \gamma_\rho)^{(1)}}, \tag{8}$$

where $\mathrm{TS}^{(1)}$ and $\mathrm{TS}^{(2)}$ are the target strength values obtained for (otherwise identical) animals with average material properties given by $(\gamma_\kappa - \gamma_\rho)^{(1)}$ and $(\gamma_\kappa - \gamma_\rho)^{(2)}$, respectively.

In the second approach, following the work by Stanton and Chu (2000), two different material property profiles were investigated, chosen to vary along the lengthwise axis of the animal only. Although it is possible that the actual material properties will vary in all directions, not just along the lengthwise axis, we limit this investigation to this particular scenario due to the large degree of speculation involved. In addition, since we are only illustrating the point that small changes in the material properties can cause large changes in the scattering, this study is limited to variations in the density contrast, g, keeping h constant at all times ($h=1.0279$). When dealing with inhomogeneous material properties, Eq. (8) is no longer applicable and the full solution for f_{bs} must be calculated before averages can be taken. Certainly, due to the lack of information available, this work can only be viewed as exploratory in nature.

The two material property profiles investigated involved

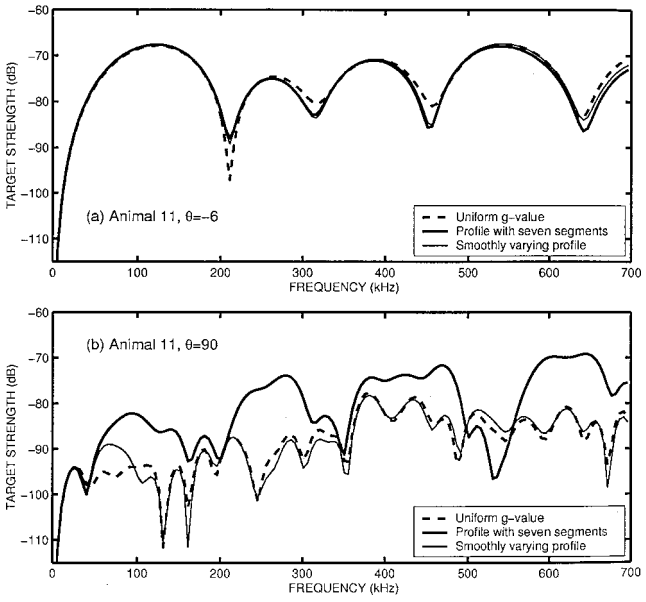


FIG. 10. Effects on TS due to changes in material property profiles. The predictions shown here are for a single realization based on the volume-integral model for animal 11. (a) TS vs frequency for $\theta=-6°$ and $\phi=0°$ ($\sim$broadside). (b) TS vs frequency for $\theta=90°$ and $\phi=0°$ ($\sim$end-on). The thick dashed lines correspond to a uniform value of $g=1.0357$, the thick solid lines correspond to the segmented profile, and the thin solid line correspond to the smoothly varying profile [almost indistinguishable from the thick solid line in panel (a)]. The sound-speed contrast was held constant with a value of $h=1.0279$ for all curves.

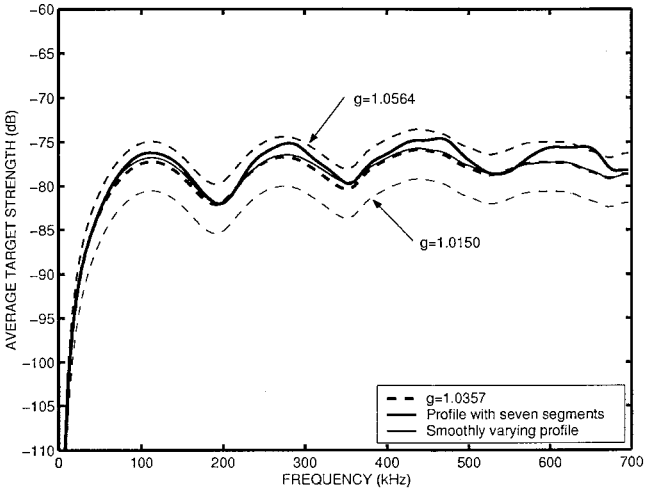


FIG. 11. Effects on TS, averaged over orientation due to changes in the material property profiles. The predictions shown here are based on the volume-integratal model for animal 11. The thin dashed lines correspond to uniform values of $g=1.0564$ (highest TS values) and $g=1.015$ (lowest TS values), respectively. The thick dashed line corresponds to a uniform value of $g=1.0357$, the thick solid line corresponds to the segmented profile, and the thin solid line corresponds to the smoothly varying profile. The sound-speed contrast was held constant with a value of $h=1.0279$ for all curves.

(1) a profile with seven segments and (2) a smoothly varying profile. These profiles are illustrated in Fig. 9. Each of these profiles contained N_z points, equal to the number of CT slices. This approach was chosen since incorporation of these profiles into the DWBA-based volume-integral model was then straightforward, using Eq. (7), as outlined in Sec. III C. The segmented profile was chosen to have seven equal length segments, each with a constant value of g. The value assigned to g in each segment was chosen randomly within the limits of $1.015 \leqslant g \leqslant 1.0564$, subject to the constraint that the average value of g remained at 1.0357. The smoothly varying profile was chosen such that it was varied, approximately, between the same constraints, with the same average value for g.

The effect on TS due to the different material property profiles was calculated for animal 11 at two specific angles of orientations [Figs. 10(a) and (b)]. The two scattering angles investigated were (a) $\theta=-6°$ (close to broadside), and (b) $\theta=90°$ (close to end-on incidence). $\theta=-6°$ was chosen instead of 0° since the peaks and nulls were more pronounced at this angle. Average TS, where the average is over all values of θ uniformly from 0 to 360° in the $\phi=0°$ plane, for the two material property profiles is shown in Fig. 11. The effect on backscattering when using the profile with seven segments is highly dependent on angle, and can be very large at angles close to end-on incidence. This is a good

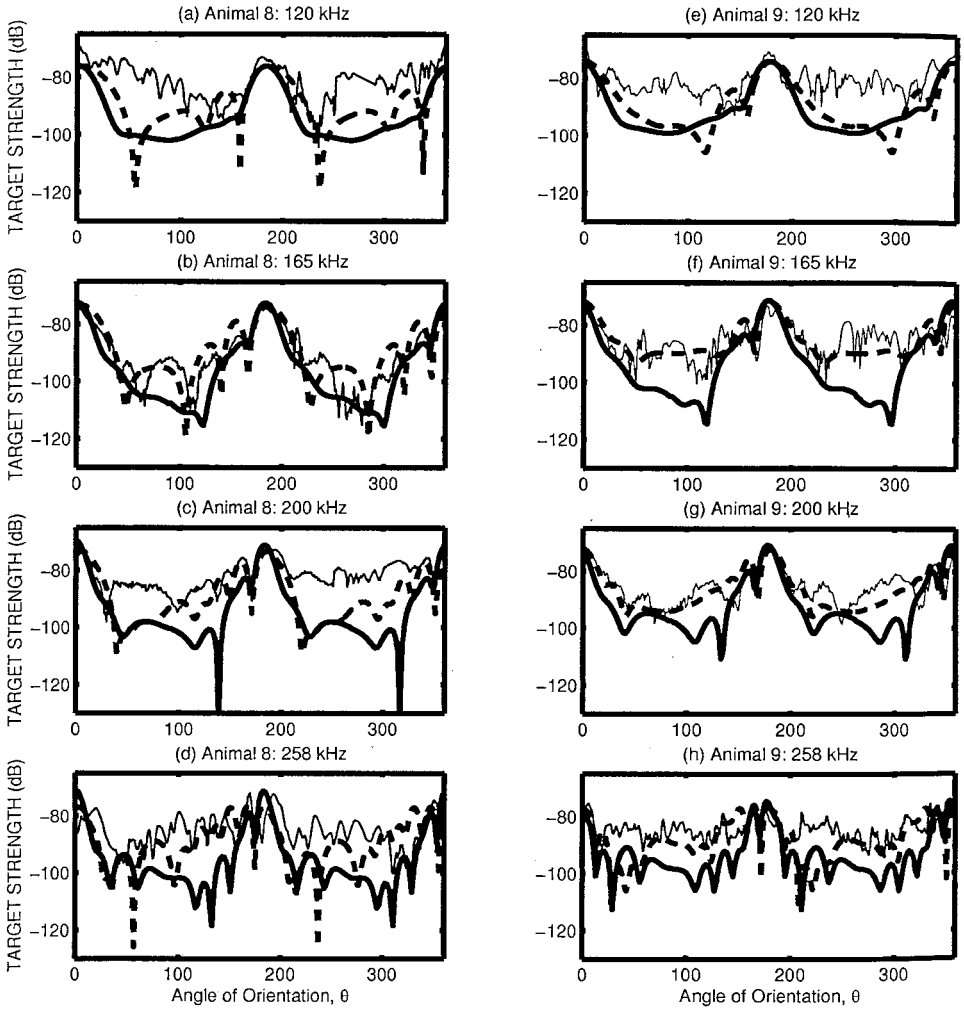


FIG. 12. TS vs θ for animals 8 (a)–(d) and 9 (e)–(h) at 120 kHz (a), (e), 165 kHz (b), (f), 200 kHz (c), (g), and 258 kHz (d), (h). The dashed lines correspond to line-integral predictions, the thick solid lines correspond to volume-integral predictions, and the thin lines correspond to data. Broadside incidence corresponds to, approximately, $\theta=0°$ and 180°. End-on scattering corresponds to $\theta=90°$, and head-on scattering corresponds to $\theta=270°$. The volume-integral predictions are based on the CT scans for animal 14, scaled to the size of animals 8 and 9, respectively. To better match the observed animal orientation during the data acquisition, the line-integral predictions with the 2D measurements were performed with a 30° out-of-plane tilt. However, the volume-integral predictions shown here, based on CT scans for animal 14, do not include any out-of-plane tilt. It is possible that the volume-integral predictions would better match the data if the dependence on tilt was further investigated.

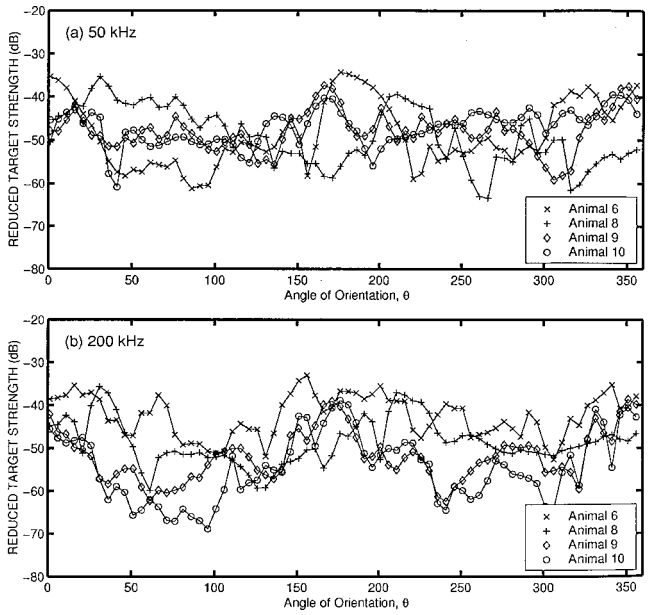


FIG. 13. Scatter plots of RTS vs θ for animals 6, 8, 9, and 10 at (a) 50 kHz and (b) 200 kHz. For all animals, broadside incidence corresponds to, approximately, $\theta=0°$ and 180°. End-on and head-on incidence correspond to $\theta=90°$ and $\theta=270°$, respectively. For animals 6 and 8, all data were collected with the dorsal–ventral aspects vertical (i.e., scattering in the side-aspect plane), while for animals 9 and 10 the dorsal–ventral aspect was horizontal. There was some degree of out-of-plane tilt for all animals, but typically $\leqslant 45°$. This figure is included to illustrate the variability in the data.

illustration of the heightened sensitivity of the scattering process to small inhomogeneities at angles close to end-on incidence. This dependence, however, almost completely disappears when averages over all angles of orientation are taken. Also included in Fig. 11 is the TS for different constant values of g: $g=1.015$, $g=1.0357$, and $g=1.0564$.

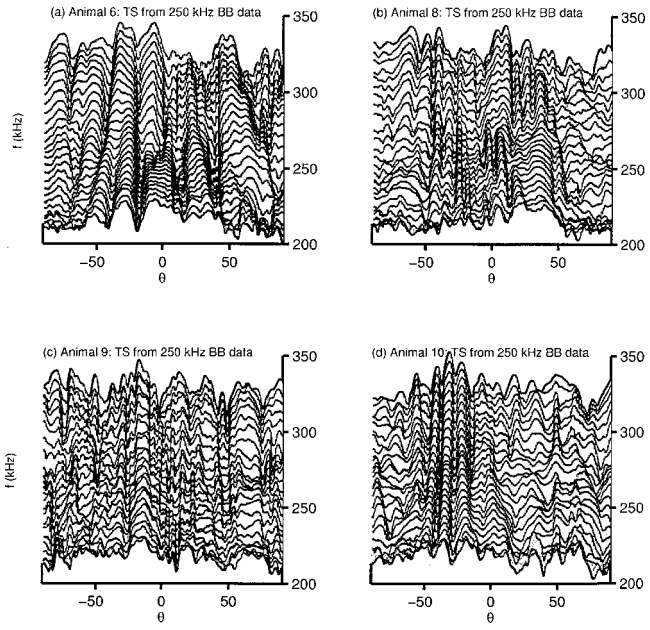


FIG. 14. TS vs frequency and θ obtained from the 250-kHz broadband transducer for (a) animal 6; (b) animal 8; (c) animal 9; and (d) animal 10. For clarity, only the first 180° of the data are shown here. The data have been averaged over a small frequency band to reduce the high levels of variability.

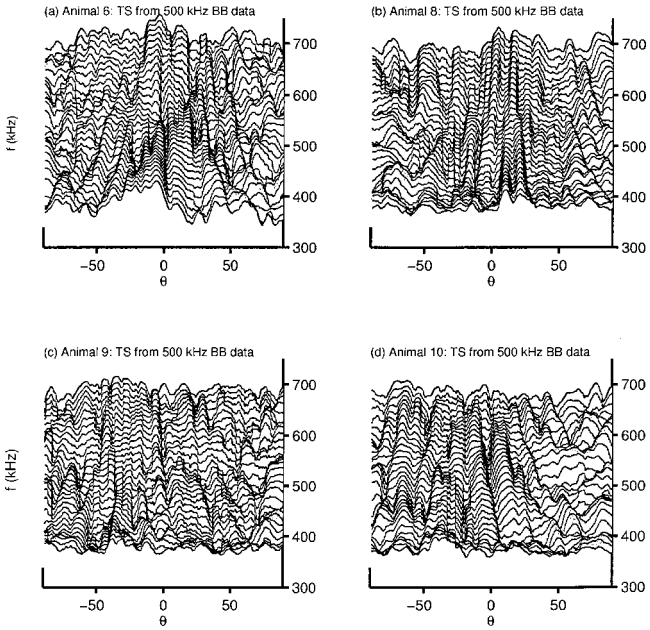


FIG. 15. TS vs frequency and θ obtained from the 500-kHz broadband transducer for (a) animal 6; (b) animal 8; (c) animal 9; and (d) animal 10. For clarity, only the first 180° of the data are shown here. The data have been averaged over a small frequency band to reduce the high levels of variability.

It is well known that scattering is highly sensitive to small changes in the average values of the material properties, in agreement with the modeling predictions presented here. Furthermore, it appears that the effects on backscattering (before average TS values are calculated) due to changes in the material property profiles have a relatively strong dependence on orientation, with the largest effects occurring at angles of orientation close to end-on incidence. Once average TS values are calculated, however, the differences in the scattering between the different material property profiles were almost completely eliminated. This is further evidence that reproducing scattering on a ping-by-ping basis is more difficult than predicting average quantities.

V. MODEL AND DATA COMPARISONS

A. Ping-by-ping

For all animals, it was found that the acoustic backscattering data were strongly dependent on both frequency and angle of orientation. To illustrate this, TS as a function of angle of orientation for animal 8 and 9 are shown in Fig. 12 at a selection of single frequencies. For comparison, predictions made with the line- and volume-integral models are also included. Since CT scans were not available for animals 8 and 9 as input into the volume-integral model, the CT data for animal 14 were scaled to the same aspect ratio as that of animals 8 and 9. It can be seen that the orientation dependence of the backscattering increases with increasing frequency, and the agreement between both the line- and volume-integral predictions and the data is relatively good at angles close to broadside. However, the agreement is not as good at other angles, with differences of up to 20 dB observed at 120 kHz for animal 9 between both models and the data. The general agreement between the line- and volume-

integral predictions is also better close to broadside than at other angles. These observations indicate that the level of sophistication of the model is less critical at angles close to broadside. As will be discussed later, at these lower frequencies the line-integral model is expected to reproduce the data to a level of accuracy comparable to the volume-integral model. It was also found that, for all animals, the data were generally reproducible over the two full rotations, though the variability in the backscattering between the different animals at any particular frequency was significant (Fig. 13).

The broadband data were generally highly variable on a frequency-by-frequency and angle-by-angle basis, exhibiting much fine structure, particularly at the higher frequencies. To illustrate this variability, TS values over a wide range of frequencies and angles of orientation are shown in Figs. 14 and 15, corresponding to data acquired with the 250- and 500-kHz broadband transducers, respectively. To eliminate some of this variability and allow the prominent features in the data to be more easily observed, the data were averaged over a narrow range of frequencies. However, it can be seen that there is still a strong dependence on the frequency, the animal number, and the angle of orientation.

B. Averages

Given the challenges associated with comparisons of data and model predictions on a ping-by-ping basis, we also assess the model performance based on comparisons of averages taken over a uniform distribution of orientations. The predicted average RTS values for animal 11 are compared to the acoustic data obtained for animals 8 and 10, which had similar aspect ratios, in Fig. 16. The predictions based on both the line- and volume-integral models are included. Also included are data taken from Fig. 7 from Stanton *et al.* (1993b). These data were obtained for backscattering from aggregations of live tethered decapod shrimp (the same species as that used in the present study) over a wide range of frequencies from 50 kHz to 1.2 MHz. Stanton *et al.* (1993b) also compared these data to a ray-based bent cylinder model in which averages over length and a uniform distribution of orientations were taken.

The line- and volume-integral models were used to make predictions for two different sets of g and h values [Figs. 16(a) and (b)]. Predictions made using $g=1.0357$ and $h=1.0279$ are presented in Fig. 16(a). These values of g and h were determined by Foote (1990) for *Euphausia superba*, and have been used in the literature for common shore shrimp. These values have been used for most of the predictions presented in this paper. However, Stanton *et al.* (1993b) found better agreement with the data when $g=h=1.06$ were used to make predictions with the ray-based model. The line- and volume-integral predictions using $g=h=1.06$ are shown in Fig. 16(b). Recent measurements of the material properties of decapod shrimp have been performed by Chu *et al.* (2000), resulting in values for g ($=1.043$) and h ($=1.0649$ to 1.0736) that are close to the range of values investigated in this study and by Stanton *et al.* (1993b).

It can be seen that, for a particular choice of g and h both the line- and volume-integral models predict similar average RTS values for low ka values. However, at higher ka values the line-integral model predicts significantly higher values for the average RTS than the volume-integral model, with difference between the two models of almost 10 dB at $ka=7$. As was observed by Stanton *et al.* (1993b), the agreement with the data is better when $g=h=1.06$ [Fig. 16(b)]. With these values of g and h, the predicted average RTS values based on the volume-integral model for animal 11 generally agree well with the data over the range of ka values investigated here.

It should be noted that the positions of the first two dips and nulls predicted by the models is highly sensitive to the value used for the mean radius. The mean radii shown in Table II were calculated from the 2D photographic images using all the measurements of radii along the lengthwise axis

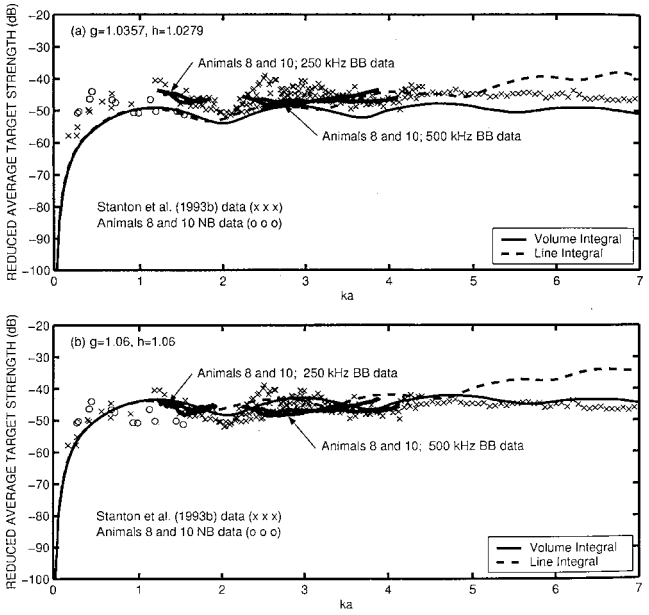


FIG. 16. Comparison of predicted and measured average RTS vs ka. In both panels, the thick solid lines correspond to the 250-kHz and 500-kHz broadband (BB) data obtained for animals 8 and 10. The open circles correspond to the data obtained for animal 8 and 10 from the single-frequency, or narrow-band (NB), transducers. Data taken over a broad frequency range (50 kHz–1.2 MHz) for live tethered decapod shrimp aggregations [Fig. 7 from Stanton *et al.* (1993b)] are also included (crosses). This is the same species of animal as used throughout the current study. The dashed line corresponds to predictions for animal 11 based on the line-integral model, while the solid line corresponds to predictions based on the volume-integral scattering model. Animals 8, 10, and 11 had similar aspect ratios. Two different sets of values for g and h were investigated. (a) $g=1.0357$ and $h=1.0279$. These values are taken from Foote (1990) and are commonly used in the literature. (b) $g=h=1.06$. These values were used in Fig. 7 of Stanton *et al.* (1993b) to compare the predictions of an averaged ray-based model to the data included here. The agreement between the data and the models is better when $g=h=1.06$, particularly at low frequencies and close to the first null, which occurs at approximately $ka=2$. In addition, the volume-integral model also reproduces the location of a second null apparent in the data at approximately $ka=3.8$. The exact position of the nulls is very sensitive to the value used for the mean radius. The values for the mean radii for each animal shown in Table II were calculated from the 2D photographic images using all the radii except the ten points close to each end of the animal. However, the difference between the mean radius calculated this way and the maximum radius can be as large as 30%. In fact, it was found that the best agreement between the model predictions (based on animal 11) and the Stanton *et al.* (1993b) data, in terms of the location of the first two nulls, occurred when the calculation of the mean radii included only the 10 points closest to the center of the animal, which is the dominant source of scattering.

of the animals, excluding the ten points at each end of each animal (since the animals were strongly tapered there). However, these mean radii differed by as much as 30% from the maximum radius measured for each animal. It was found here that the positions of the first two nulls observed in Stanton *et al.* (1993b) data could be best reproduced by using the volume-integral model with a mean radius calculated from the 10 points centered around the location of the maximum radius (Chu *et al.*, 1992). That is, the mean radius was calculated using the main part of the body only (i.e., the thorax section), which is similar to the mean radius used by Stanton *et al.* (1993b). For animal 11 this resulted in a mean thorax radius of 2.5 mm instead of 2.2 mm (an increase of approximately 14%). With this value of mean radius, it can be seen that the position of the first two nulls in the Stanton *et al.* (1993b) data, at $ka=2$ and $ka=3.8$, are reproduced very well by the volume-integral model for animal 11.

Just as we have observed here for predictions based on volume-integral model, Stanton *et al.* (1993b) also found good agreement between their ray-based predictions and the data, in terms of the locations of the first two nulls at $ka=2$ and $ka=3.8$. However, the ray-based model predicted a deeper null at $ka=2$ than was observed in the data, with a difference between the averaged RTS values at the locations of the first peak and first null of approximately 12 dB. The difference between the averaged RTS values at the location of the first peak and first null is approximately 5 dB for the volume-integral model which is in better agreement with the data. In fact, it can be seen by comparing the predictions made by Stanton *et al.* (1993b) for the ray-based model to the volume-integral model predictions shown here, that the oscillations in the average RTS versus *ka* are more accentuated for the ray-based model over the entire *ka* range investigated.

VI. DISCUSSION

One of the main findings of this modeling study is that scattering from elongated fluid-like zooplankton at angles close to end-on incidence is significantly more sensitive to small changes in material properties and roughness, than scattering at angles close to broadside. Furthermore, this sensitivity is accentuated with increasing frequency. As a result, it is correspondingly harder to accurately predict scattering at angles close to end-on incidence.

At angles close to broadside, much of the structure observed in the target strength as a function of frequency, e.g., position and number of the peaks and nulls in the frequency spectra, can be relatively well accounted for by considering a simple two-ray model (Stanton *et al.*, 1993a, 1993b; Stanton and Chu, 2000). For weak scatterers a large fraction of the incident energy is transmitted through the front interface. Consequently, constructive and destructive interference between the two primary rays that scatter from the front and back interfaces is principally responsible for the observed peaks and nulls. The phase difference between these two primary rays is determined by parameters such as the frequency and animal radius. Effects due to scattering from volume inhomogeneities and surface roughness are apparently small relative to the large primary returns from the front and back interfaces, explaining the relative lack of sensitivity of scattering at broadside incidence, as well as explaining the agreement observed between the different scattering models.

In contrast, at angles close to end-on incidence, the scattering is not dominated by two large returns from the front and back interfaces, and small-scale variability such as surface roughness and volume inhomogeneities play a more significant role. Thus, to accurately model scattering at angles close to end-on incidence it is necessary to have a more detailed knowledge of the small-scale variability. It is now possible to understand why predictions based on the line-integral model do not agree with predictions based on the volume-integral scattering model at angles close to end-on incidence. There are a number of assumptions underlying the line-integral model. A key aspect of this formulation is that the animals are assumed to have a circular cross section at every point along their lengthwise axis. This assumption of cylindrical symmetry artificially introduces elevated roughness levels, through the creation of *facets*, which will be most highly accentuated at angles close to end-on incidence. This problem has been identified by Stanton and Chu (2000), and results in increased scattering levels due to scattering from these facets. Clearly, this problem is magnified with increasing frequency as it becomes increasingly harder to satisfy the $\lambda/10-\lambda/20$ resolution criterion mentioned earlier. Although it may be possible to reduce the effect of facets by increasing the resolution of the 2D measurements, it is an intrinsic problem associated with this model and cannot be eliminated entirely. Stanton and Chu (2000) also found that the problem could be reduced by smoothing the surface roughness profiles. However, the procedure was necessarily subjective. The more rigorous method to eliminate the problem of facets is to obtain 3D digitizations of animal shape and roughness, as was done in this study using high-resolution CT scans. Finally, it should be noted that although predictions based on the volume-integral model are computationally more intensive than those based on the line-integral model, the difficulty involved in obtaining 3D measurements of animal shape, and incorporating these measurements into the model, is not significantly greater than for the 2D measurements.

Various conclusions can be drawn from this study regarding the practical conditions under which it is necessary to use the DWBA-based volume-integral scattering model with high-resolution 3-D digitizations of animal shape. Before discussing the range of practical conditions in further detail it should be noted that many of these conclusions are based on predictions of average quantities. Once averages are taken, either over a distribution of angles, lengths, shapes, or sizes, much of the structure seen in the individual pings will tend to be washed out. Consequently, the model predictions tend to agree better with each other.

As was mentioned earlier, little is known about the natural orientation distribution of elongated fluid-like zooplankton in their natural environment. Kils (1982), Sameoto (1980), and Endo (1993) have described tank measurements of the body orientations of Antarctic krill (*Euphausia superba*), and there have been similar studies (Miyashita *et al.*,

1996) performed on *Euphausia pacifica*. In a more recent study, Benfield *et al.* (2000) have found that the natural body orientation of *Calanus finmarchicus*, another member of the elongated fluid-like class of zooplankton, tends to be naturally oriented with the animal body axis vertically up, with a standard deviation of about 30°. Thus, based on the performance of the volume-integral model as compared to the line-integral model, it can be concluded that if the natural distribution of orientations includes angles of incidence close to end-on, more accurate scattering predictions are possible with the volume-integral model. Likewise, there is little information currently available as to the variations in the material properties throughout the body volume (Foote, 1998; Yayanos *et al.*, 1978). However, it is possible that the material properties will vary both along the lengthwise axis of the animals, as well as radially. Once quantified, it would be straightforward to include these volume inhomogeneities in the volume-integral model.

It is apparent from this investigation that the DWBA-based volume-integral scattering model, using high-resolution 3D measurements of shape, reproduces the laboratory data for common shore shrimp under certain conditions. However, there are models that make use of simpler representations of animal shape (Stanton and Chu, 2000) that are also accurate over a narrow range of conditions. Whether or not it is necessary to make use of a sophisticated, but more complex, scattering model that incorporates high-resolution measurements of animal shape may depend on the particular application. In fact, based on the results of the volume-integral model presented earlier, and depending on the specific applications, the largest sources of error in predicting scattering in realistic field situations may be dominated by the uncertainties in the material property and orientation distributions. Finally, if the animal proportions (i.e., the general shape and aspect ratio) scale with the animal size, it may be possible to create a database of scattering predictions from 3D measurements of a relatively small number of individuals.

VII. SUMMARY AND CONCLUSIONS

In summary, an acoustic scattering model for fluid-like scatterers with complex shapes, which is based on the distorted wave Born approximation and incorporates high-resolution three-dimensional digitizations of shape, has been developed. This model has a wide range of applicability, and has been applied in this study to the specific case of decapod shrimp (*Palaeomonetes vulgaris*), an elongated fluid-like scatterer, which closely resembles many members of the elongated fluid-like class of zooplankton. Detailed acoustic scattering measurements of live individual animals have been compared to the model predictions. These data, as well as the modeling results and predictions, have shown that the scattering process is highly complex and sensitive to the shape and size of the organisms, in addition to the angle of orientation, material properties, and acoustic frequency. This high level of variability indicates that sophisticated models are necessary to reproduce all aspects of the scattering process on a ping-by-ping basis.

High resolution CT scans have been used to accurately incorporate the full three-dimensional animal shape into the DWBA-based volume-integral scattering model. Comparisons have been made between the volume-integrated model and a DWBA-based line-integral scattering model that uses a 2D representation of animal shape. The model predictions agree reasonably well with each other, and with the ping-by-ping low-frequency narrow-band laboratory data, at angles close to broadside (Fig. 12). However, at angles close to end-on incidence, neither model successfully predicted TS value close to those observed in the ping-by-ping data, even at the lower frequencies (Fig. 12). It should be noted that CT scans were not available for the same animals from which the acoustic scattering data were collected; consequently, the volume-integral predictions were based on measurements of a scaled version of a different animal (scaled by aspect ratio). It has also been shown here that at higher frequencies the line-integral model predicts higher average scattering levels than those predicted by the volume-integral model or seen in the averaged data (Fig. 16), when averages are taken over all angles of orientation (0–360°) in one plane. The higher average scattering levels can be understood in terms of the creation of facets during the 2D digitization process. This was not the case with the volume-integral approach in which 3D CT scans were used to obtain the digitizations of animal shape. We conclude that care must be used with the line-integral model, particularly at higher frequencies. In contrast, the DWBA-based volume-integral scattering model, which makes use of the high-resolution 3D measurements of animal shape, compared better with the averaged data (Fig. 16) over the full frequency range investigated.

One of the many aims of studies such as this one is to improve estimates of the numerical abundances of zooplankton as extracted from acoustic data obtained during field surveys, as well as to perform classification of zooplankton aggregations. However, accurately extracting this information from the acoustic field data remains a challenging problem. Investigations such as the one presented here, which combine a general scattering formulation with high-resolution morphological information and high-quality laboratory data, are key to the quantitative use of acoustics in studying zooplankton in the ocean.

ACKNOWLEDGMENTS

The authors would like to thank Darlene Ketten and Scott Cramer for performing the CT scans of the decapod shrimp, as well as to Hanu Singh and Ryan Eustice for help with the image analysis of CT images. This work was supported in part by the Woods Hole Oceanographic Institution Education Office. This is Woods Hole Oceanographic Institution Contribution Number 10425.

Anderson, V. C. (**1950**). ''Sound scattering from a fluid-sphere,'' J. Acoust. Soc. Am. **22**, 426–431.

Benfield, M. C., Davis, C. S., and Gallager, S. M. (**2000**). ''Estimating the *in situ* orientation of *Calanus finmarchicus* on Georges Bank using the Video Plankton Recorder,'' Plankton Biol. Ecol. **47**, 69–72.

Chu, D., Stanton, T. K., and Wiebe, P. H. (**1992**). ''Frequency dependence of sound backscattering from live individual zooplankton,'' ICES J. Mar. Sci. **49**, 97–106.

Chu, D., Foote, K. G., and Stanton, T. K. (**1993**). ''Further analysis of target strength measurements of Antarctic krill at 38 kHz and 120 kHz: Comparison with deformed-cylinder model and inference of orientation distribution,'' J. Acoust. Soc. Am. **93**, 2985–2988.

Chu, D., and Ye, Z. (**1999**). ''A phase-conjugated distorted wave Born approximation representation of the bistatic scattering by weakly scattering objects: Application to zooplankton,'' J. Acoust. Soc. Am. **106**, 1732–1743.

Chu, D., Wiebe, P. H., and Copley, N. (**2000**). ''Inference of material properties of zooplankton from acoustic and resistivity measurements,'' ICES J. Mar. Sci. **57**, 1128–1142.

Endo, Y. (**1993**). ''Orientation of Antarctic krill in an aquarium,'' Nippon Suisan Gakkaishi **59**, 465–468.

Everson, I. (**1982**). ''Diurnal variations in mean volume backscattering strengths of an Antarctic krill (*Euphausia superba*) patch,'' J. Plankton Res. **4**, 155–162.

Everson, I., Watkins, J. L., Bone, D. G., and Foote, K. G. (**1990**). ''Implications of new acoustic target strengths for abundance estimates of Antarctic krill,'' Nature (London) **345**, 338–340.

Foote, K. G., Everson, I., Watkins, J. L., and Bone, D. G. (**1990**). ''Target strengths of Antarctic krill (*Euphausia superba*) at 38 kHz and 120 kHz,'' J. Acoust. Soc. Am. **87**, 16–24.

Foote, K. G., and Stanton, T. K. (**2000**). ''Acoustical Methods,'' in *ICES Zooplankton Methodology Manual*, edited by R. Harris, P. Wiebe, J. Lenz, H. Skjoldal, and M. Huntley (Academic, London), Chap. 6, pp. 223–258.

Foote, K. G. (**1990**). ''Speed of sound in *Euphausia superba*,'' J. Acoust. Soc. Am. **87**, 1405–1408.

Foote, K. G. (**1998**). ''Measurement of morphology and physical properties of zooplankton,'' in *Digest of Colloquium on 'Recent advances in sonar applied to biological oceanography'* (IEEE, Savoy Place, UK).

Greene, C. H., Stanton, T. K., Wiebe, P. H., and McClatchie, S. (**1991**). ''Acoustic estimates of Antarctic krill,'' Nature (London) **349**, 110.

Greenlaw, C. F. (**1977**). ''Backscattering spectra of preserved zooplankton,'' J. Acoust. Soc. Am. **62**, 44–52.

Hewitt, R. P. and Demer, D. A. (**1991**). ''Krill abundance,'' Nature (London) **353**, 310.

Holliday, D. V., Pieper, R. E., and Kleppel, G. S. (**1989**). ''Determination of zooplankton size and distribution with multi-frequency acoustic technology,'' J. Cons., Cons. Int. Explor. Mer **46**, 52–61.

Holliday, D. V., and Pieper, R. E. (**1995**). ''Bioacoustical oceanography at high frequencies,'' ICES J. Mar. Sci. **52**, 279–296.

Kils, U. (**1981**) ''The swimming behavior, swimming performance and energy balance of Antarctic krill, *Euphausia superba*,'' BIOMASS Scientific Series No. 3, 122.

Martin Traykovski, L. V., O'Driscoll, R. L., and McGehee, D. M. (**1998**). ''Effect of orientation on broadband acoustic scattering of Antarctic krill *Euphausia superba*: Implications for inverting zooplankton spectral acoustic signatures for angle of orientation,'' J. Acoust. Soc. Am. **104**, 2121–2135.

McGehee, D. M., O'Driscoll, R. L., and Martin Traykovski, L. V. (**1998**). ''Effect of orientation on acoustic scattering from Antarctic krill at 120 kHz,'' Deep-Sea Res., Part II **45**, 1273–1294.

Medwin H., and Clay, S. C. (**1998**). *Fundamentals of Acoustical Oceanography* (Academic, Boston).

Miyashita, K., Aoki, I., and Inagaki, T. (**1996**). ''Swimming behavior and target strength of isada krill (*Euphausia pacifica*),'' ICES J. Mar. Sci. **53**, 303–308.

Morse, P. M., and Ingard, K. U. (**1968**). *Theoretical Acoustics* (Princeton University, Princeton).

Sameoto, D. D. (**1980**). ''Quantitative measurement of euphausiids using 120-kHz sounder and their *in situ* orientation,'' Can. J. Fish. Aquat. Sci. **37**, 693–702.

Stanton, T. K., Clay, C. S., and Chu, D. (**1993a**). ''Ray representation of sound scattering by weakly scattering deformed fluid cylinders: Simple physics and application to zooplankton,'' J. Acoust. Soc. Am. **94**, 3454–3462.

Stanton, T. K., Chu, D., Wiebe, P. H., and Clay, C. S. (**1993b**). ''Average echoes from randomly oriented random-length finite cylinders: Zooplankton models,'' J. Acoust. Soc. Am. **94**, 3463–3472.

Stanton, T. K., Wiebe, P. H., Chu, D., Benfield, M. C., Scalon, L., Martin, L. V., and Eastwood, R. L. (**1994**). ''On acoustic estimates of zooplankton biomass,'' ICES J. Mar. Sci. **51**, 505–512.

Stanton, T. K., Chu, D., and Wiebe, P H. (**1996**). ''Sound scattering characteristics of several zooplankton groups,'' ICES J. Mar. Sci. **53**, 289–295.

Stanton, T. K., Chu, D., Wiebe, P. H., Martin, L. V., and Eastwood, R. L. (**1998a**). ''Sound scattering by several zooplankton groups. I. Experimental determination of dominant scattering mechanisms,'' J. Acoust. Soc. Am. **103**, 225–235.

Stanton, T. K., Chu, D., and Wiebe, P. H. (**1998b**). ''Sound scattering by several zooplankton groups. II. Scattering models,'' J. Acoust. Soc. Am. **103**, 236–253.

Stanton, T. K., Wiebe, P. H., and Chu, D., (**1998c**). ''Differences between sound scattering by weakly scattering spheres and finite-length cylinders with applications to sound scattering by zooplankton,'' J. Acoust. Soc. Am. **103**, 254–264.

Stanton, T. K., and Chu, D. (**2000**). ''Review and recommendations for modeling of acoustic scattering by fluid-like elongated zooplankton: Euphausiids and copepods,'' ICES J. Mar. Sci. **57**, 793–807.

Stanton, T. K., Chu, D., Wiebe, P. H., Eastwood, R. L., and Warren, J. D. (**2000**). ''Acoustic scattering by benthic and planktonic shelled animals,'' J. Acoust. Soc. Am. **108**, 535–550.

Urick, R. J. (**1983**). *Principles of Underwater Sound* (McGraw-Hill, New York).

Yayanos, A. A., Benson, A. A., and Nevenzal, J. C. (**1978**). ''The pressure–volume–temperature (PVT) properties of a lipid mixture from a marine copepod, *Calanus plumchrus*; Implications for buoyancy and sound scattering,'' Deep-Sea Res. **25**, 257–268.

An ensemble source spectra model for merchant ship-radiated noise

Stephen C. Wales and Richard M. Heitmeyer
Code 7120, Acoustics Division, Naval Research Laboratory, Washington, DC 20375-5350

(Received 14 May 1999; revised 8 June 2001; accepted 17 September 2001)

This paper presents an evaluation of the classical model for determining an ensemble of the broadband source spectra of the sound generated by individual ships and proposes an alternate model to overcome the deficiencies in the classical model. The classical model, proposed by Ross [*Mechanics of Underwater Noise* (Pergamon, New York, 1976)] postulates that the source spectrum for an individual ship is proportional to a baseline spectrum with the constant of proportionality determined by a power-law relationship on the ship speed and length. The model evaluation, conducted on an ensemble of 54 source spectra over a 30–1200-Hz to 1200-Hz frequency band, shows that this assumption yields large rms errors in the broadband source level for the individual ships and significantly overestimates the variability in the source level across the ensemble of source spectra. These deficiencies are a consequence of the negligible correlation between the source level and the ship speed and the source level and the ship length. The alternate model proposed here represents the individual ship spectra by a modified rational spectrum where the poles and zeros are restricted to the real axis and the exponents of the terms are not restricted to integer values. An evaluation of this model on the source spectra ensemble indicates that the rms errors are significantly less than those obtained with any model where the frequency dependence is represented by a single baseline spectrum. Furthermore, at high frequencies (400 to 1200 Hz), a single-term rational spectrum model is sufficient to describe the frequency dependence and, at the low frequencies (30 to 400 Hz), there is only a modest reduction in the rms error for a higher order model. Finally, a joint probability density on the two parameters of the single term model based on the measured histograms of these parameters is proposed. This probability density provides a mechanism for generating an ensemble of ship spectra. [DOI: 10.1121/1.1427355]

PACS numbers: 43.30.Nb, 43.30.Zk [SAC-B]

I. INTRODUCTION

Ambient noise constitutes the primary limitation on the performance of passive surveillance systems. Among the different sources of ambient noise, merchant shipping is the principal contributor over the frequency band of 5 Hz to as high as 1 kHz.[1] Within this band, the shipping noise level can be as much as 40 dB higher than the local wind-noise levels. Accordingly, considerable effort has been directed towards developing numerical models that describe properties of ship-induced noise observed by the system of interest. These noise models consist of a source model which describes both the sound generated by the individual ships and the locations or tracks of those ships, and a propagation model which describes the propagation of the sound generated by the individual ships to the region occupied by the system. In most of these models, the sound radiated by a ship is represented by associating a source spectrum with a monopole source located at some nominal depth; the propagation model then computes the Green's function from the source location to the region occupied by the system. Clearly, the integrity of these noise models is critically dependent on the correctness of the ship source level model.

Qualitatively, the spectrum of the sound generated by a ship consists of a broadband component, generated by propeller cavitation, and a number of narrow-band lines generated by both propeller cavitation (the blade lines) and internal machinery (engine, generators, pumps, etc.). The broadband sound generated by propeller cavitation results from the collapse of cavitation voids generated on the front (lifting) side of each propeller blade as that blade passes through the low-flow velocity region near the top of its cycle of rotation. The blade lines in the propeller cavitation spectrum are produced by the changing volume of the region occupied by the cavitation voids. This volume varies in a cyclic fashion, reaching a peak as each propeller blade nears the top of its rotation.[2,3] Much of this knowledge is summarized in the texts by Ross,[4] Blake,[5] and Breslin and Anderson.[6]

Models for the hydrodynamic processes related to the generation of the cavitation volumes and the resulting pressure fields have been developed. In particular, there is a significant body of work directed towards estimating the size of the cavitation volume on the blade in the presence of unsteady flow.[7,8] These models do not describe the distribution of the individual voids within the cavitation volume. Moreover, these models, which are largely numerical, require such a detailed knowledge of the physical parameters of the ship and the inflow to the propeller disk as to be currently impractical for predicting the cavitation volume across a large ensemble of diverse ships. Much of this work has been motivated by an effort to determine the pressures produced on the hull by the intermittent cavitation of the propeller.

To the authors' knowledge only one model for propeller-generated noise has been proposed that directly incorporates

a description of the hydrodynamic cavitation processes.[9] In this model the far-field acoustic pressure is viewed as the superposition of the contributions of the collapse of the individual voids that constitute the cavitation volume. To the authors' knowledge this model has not been tested experimentally. Furthermore, this model does not include the near-field acoustic propagation effects and requires a knowledge of quantities such as the flow inhomogeneity and the distribution of initial cavitation bubble sizes.

Structural models related to the coupling of machinery vibrations through the fluid-loaded hull have been developed. Predictions can be made for the noise levels radiated from the hull from a known internal source, if a detailed description of the ship's structure is available.

The relation of the pressure in the source generation region to the far-field acoustic pressure is governed by the Green's function. A complete calculation of the Green's function in the near-field environment of a ship would include: the interaction of the sound with the hull and the propeller, reradiation by the propeller wake, a time-varying surface overhead, propagation of the generated sound through the hull wake, and reflection off of a rough sea surface which includes the ship's bow wake. This is a complex four-dimensional problem in space and time. To the authors' knowledge there has been no work done on this problem starting from the source generation region on the front surface of the rotating propeller.

In the absence of a quantitative physical theory, the ambient noise modeling community has adopted an empirical model developed by Ross in his seminal work on ship-radiated noise[4] conducted up to the early 1960's. This model postulates that the source spectrum for an individual ship is proportional to a baseline spectrum with the constant of proportionality determined by a power-law assumption on specific parameters of the individual ships. Among the different formulations of the power-law model that have been proposed, the one based on the ship speed and length is the most common, since these parameters can be derived from ship's registry data, which are readily available. The ensemble of source spectra required in noise model applications can then obtained from a joint probability density on the ship speeds and lengths which, itself, can be determined from ship's registry data. Alternatively, the parameters can be determined from independent measurements of specific ship speeds and lengths in a region of interest.

In this paper, we present an analysis conducted on an ensemble of measured ship source spectra which shows that these spectra are not well approximated by a power-law model based on the ship speed and length. In particular, it is shown that large errors in the spectra of individual ships result from a power-law model of the form proposed by Ross. Furthermore, the power-law model overestimates the variability in an ensemble of source spectra. These deficiencies are largely due to the fact that the source level of individual ships is not well correlated with its speed and length. In conducting this analysis, the temporal properties of the spectrum are ignored and will be treated in a further paper.

In the light of these deficiencies, we propose an alternate model that approximates the individual ship spectra by a modified rational spectrum. A probability law on the parameters of the rational spectrum model, derived from the data, provides the mechanism for generating an ensemble of ship spectra for use in noise models. While this model does not relate the individual spectra to ship registry parameters, as does the Ross model, it does provide a more accurate approximation to both the individual spectra and the variability in the ensemble of ship spectra. Neither the proposed model nor Ross' power-law model seeks to describe the line structure of the source spectrum.

The paper is organized as follows. In the next section, we present an overview of the methodology used to acquire the ship-radiated noise data and to derive the estimates of the source spectra from that data. In Sec. III, we describe the variability in the source spectra ensemble and identify frequency bands where the nature of that variability is different. In Sec. IV, we present an evaluation of the historical Ross model on the current source spectra ensemble and discuss the deficiencies in the power-law assumption by comparing results obtained with other models of the same general form. Section V presents an analysis of the rational spectrum model proposed in this paper and an approximation to the joint probability density on the parameters of this model. Finally, a summary and a discussion of the results are presented in Sec. VI.

II. MEASUREMENT AND PROCESSING METHODOLOGY

The results presented in this paper are based on ship-radiated noise (SRN) measurements conducted over a 7-year period, 1986–1992, on 272 ships in the Mediterranean Sea and the Eastern Atlantic Ocean. The measurements for each year were obtained over a several day period in the same geographical area. All of these areas were located in deep water over an abyssal plain except one, which extended over a continental shelf. For each ship, the acoustic data was acquired on two sonobuoys deployed on either side of the ship at a depth of 300 m and at a closest-point-of-arrival (CPA) range that varied from about 360 m to about 1800 m, depending on the ship. About 10 minutes of data were acquired on each of the two sonobuoys. In addition to the acoustic data, both environmental and ship parameter data were obtained for each ship. The environmental data included AXBT measurements of the water temperature, wind speed and direction data, and estimates of the number of other ships in the vicinity of the measured ship. The ship parameter data included the ship name, type, and speed. Additional ship parameters were obtained from Lloyd's Registry of Shipping[10] for 200 of the 272 ships. This database was provided to the authors under an international data exchange agreement. To our knowledge, it is the largest collection of deep-water merchant ship-radiated noise measurements that is currently available.

The source spectra were obtained from the time series of the ship-radiated-noise spectra after dividing by the range–frequency propagation function. The complex SRN spectrograms were computed as 2-s Fourier transforms using Hann shading to provide a 0.75-Hz analysis bandwidth and 50-percent overlap to provide a 1-s sampling interval. The SRN

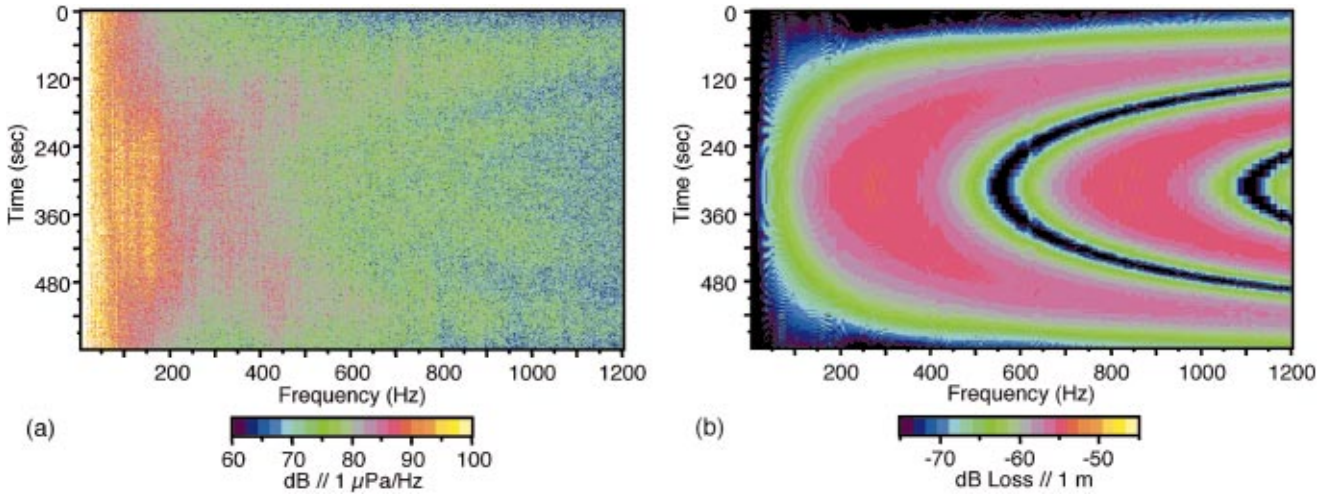


FIG. 1. (a) Measured received level spectrogram for the SHINOBU. (b) Time-frequency dependence of the point source transmission loss for (a) using Gray and Greeley's source depth.

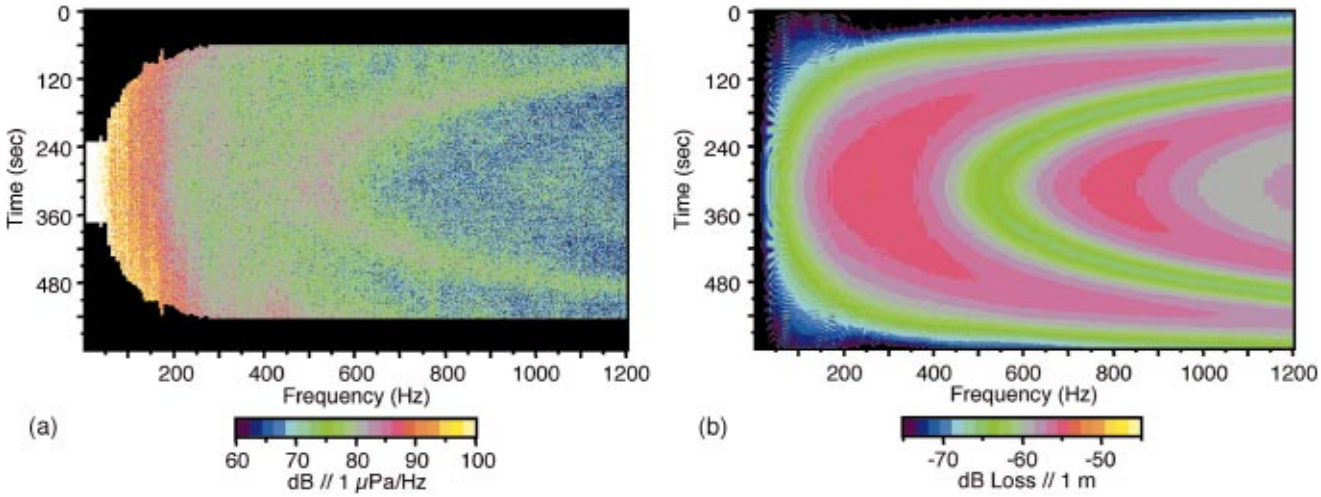


FIG. 2. (a) Source level spectrogram for the SHINOBU calculated from Fig. 1(a) using an upper-quadrant representation for the source distribution. (b) Time-frequency dependence of the upper quadrant source distribution transmission loss used to calculate (a).

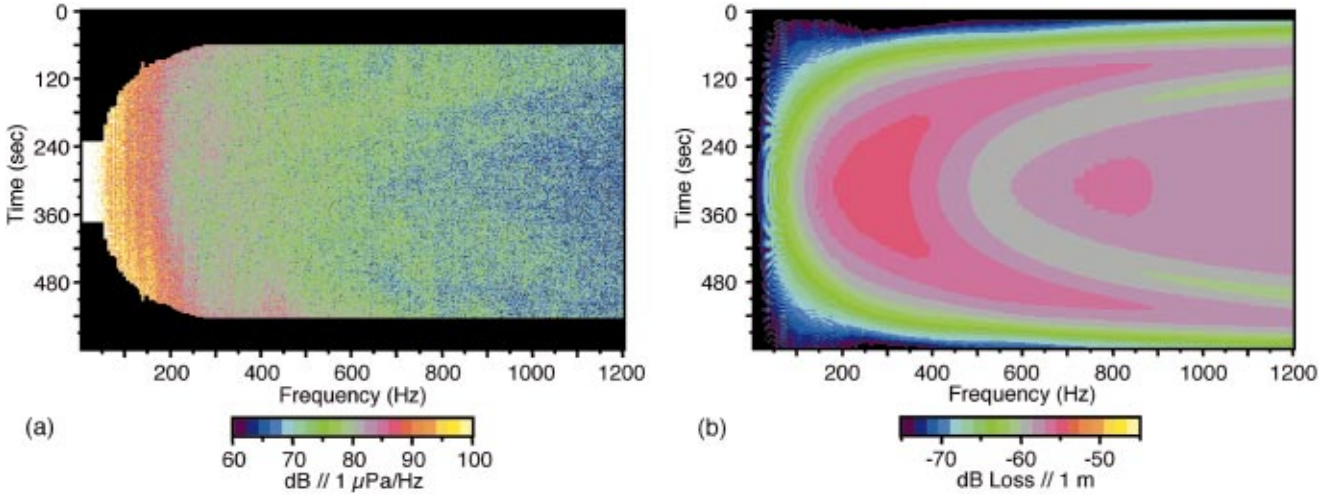


FIG. 3. (a) Source level spectrogram for the SHINOBU calculated from Fig. 1(a) using a Gaussian source distribution, standard deviation equal to one-half of the propeller radius. (b) Time-frequency dependence of the Gaussian source distribution transmission loss used to calculate (a).

spectrograms were computed over a 30–1200-Hz frequency band with the upper frequency determined by sampling rate limitations and the lower frequency determined by uncertainties in the hydrophone calibration at low frequencies. The propagation function for each ship was computed by representing the source region as a vertical line distribution of incoherent point sources. A wave number integration model[11,12] was used with a representative acoustic environment for each of the measurement areas to compute the Green's function from each of the point sources. The source spectrogram obtained from the SRN spectrogram and the propagation function was corrected for the time-dependent Doppler associated with the ship's passage and expressed in amplitude units of dB *re*: 1 μPa/Hz at 1 m. Finally, the source spectrum itself was computed from the source spectrogram as a time average of the amplitude squared over a 2-min interval centered about the CPA time.

The specific source distribution used to compute the propagation function was selected to both account for the finite extent of the cavitation volume for each ship and to provide a rough approximation for the propagation effects associated with that cavitation volume. The considerations leading to the choice of the source distribution are illustrated in Figs. 1, 2, and 3. Figure 1 shows a spectrogram of the ship-radiated noise for a specific ship along with the point-source transmission loss computed along the source–receiver track for that ship. The specific example shown is a large tanker (the SHINOBU) with a 14.6-m draft that passed by the sonobuoy with a CPA of 1090 m. The source depth for the calculation, 7 m, was based on a formula provided by Gray and Greeley,[13,14] which estimates the center of the cavitation volume to be below the top of the propeller blade arc by an amount equal to 15 % of the propeller diameter. There are two dominant features in the transmission loss plot of Fig. 1(b) the high loss region (black) located at the time–frequency boundary of the plot and the ''horseshoe''-shaped pattern located in the center of the plot. By construction, both of these features are symmetric with respect to the CPA time. The high loss region is due to the propagation induced shadow zone that occurs near the extremes of the time interval and due to mode cutoff at low frequencies. The horseshoe-shaped pattern is due to the interaction of the direct and the surface reflected propagation paths as the point source moves along the source–receiver track, i.e., the Lloyd's mirror effect.

The measured SRN spectrogram [Fig. 1(a)] shows vestiges of the horseshoe pattern seen in the point source propagation function of Fig. 1(b). Moreover, the locations of the low-level contours in the SRN spectrogram correspond closely to the high loss contours in the propagation spectrogram. However, the extremely low levels in the SRN spectrogram that would result from the high loss contours in the propagation spectrogram are not evident, even after accounting for the increase in the levels due to the ambient noise background. It is also noted that, in addition to the horseshoe structure, the measured SRN spectrogram shows asymmetries with respect to the CPA time. These asymmetries are likely due to asymmetries in the source directionality that result from the differences in the factors that affect the local propagation of the cavitation-induced energy in the forward and aft directions, e.g., hull interaction in the forward directions, propeller and hull wake interaction in the aft directions.[15] It is, of course, not possible to unambiguously distinguish the source directionality features in the SRN spectrogram from the propagation-induced features. Nevertheless, the correspondence between Figs. 1(a) and (b) strongly suggests that at least part of the symmetrical horseshoe-shaped feature in the SRN spectrogram can be interpreted as the result of propagation from a localized source located below the ocean surface. However, the disparity in the extent of the decrease in level in the SRN spectrogram and the extent of the increase of the loss in the propagation function suggest that this source cannot be approximated as a point source.

To obtain further insight into the nature of the source distribution we have computed the source spectrogram corresponding to the SRN spectrogram of Fig. 1 for two assumed point-source distributions. The results, along with the corresponding propagation spectrograms, are shown in Figs. 2 and 3. Figure 2 show the propagation spectrogram for a distribution that is intended to describe only the extent of the cavitation volume and the source spectrogram that results from that propagation spectrogram. The source distribution for the propagation functions has a vertical extent that spans the upper half of the propeller blade arc as estimated from the Gray and Greeley formula. The source-weighting function for this source distribution was derived under the assumption that the cavitation volume is uniformly distributed across the upper quadrant of the region swept out by the propeller. This is a reasonable depth-weighting approximation to the observed cavitation volumes reported in the literature.[7,13,16] A comparison of the ''upper quadrant propagation'' function [Fig. 2(b)] with the point-source propagation spectrogram [Fig. 1(b)] indicates that the effect of the distributed source is to reduce the levels in the high loss regions of the horseshoe pattern in the point-source spectrogram. However, the amount of this reduction is not sufficient to remove the apparent propagation-induced artifact that appears as the horseshoe pattern in the source spectrogram of Fig 2(a). This is not surprising since, even if the upper quadrant representation provided an acceptable approximation to the cavitation volume, the resulting propagation spectrogram assumes that the only important effect of the propagation of the upward component of that energy is an interaction with a flat, purely reflecting ocean surface in a bubble-free water column. Within the constraints of the point-source distributions proposed here (a vertical line of weighted point sources), this artifact can only be reduced by increasing the vertical extent of the distribution and/or altering the weighting function.

Figure 3(b) shows the propagation spectrogram for one such source distribution assumed and the source spectrogram that results from that propagation spectrogram. The point-source distribution for the propagation spectrogram is a Gaussian weighting function with a standard deviation equal to one-quarter of the propeller diameter.[17] As seen in the comparison of the propagation spectrogram plots [Figs. 3(b) and 2(b)], the increase in the vertical extent of the source

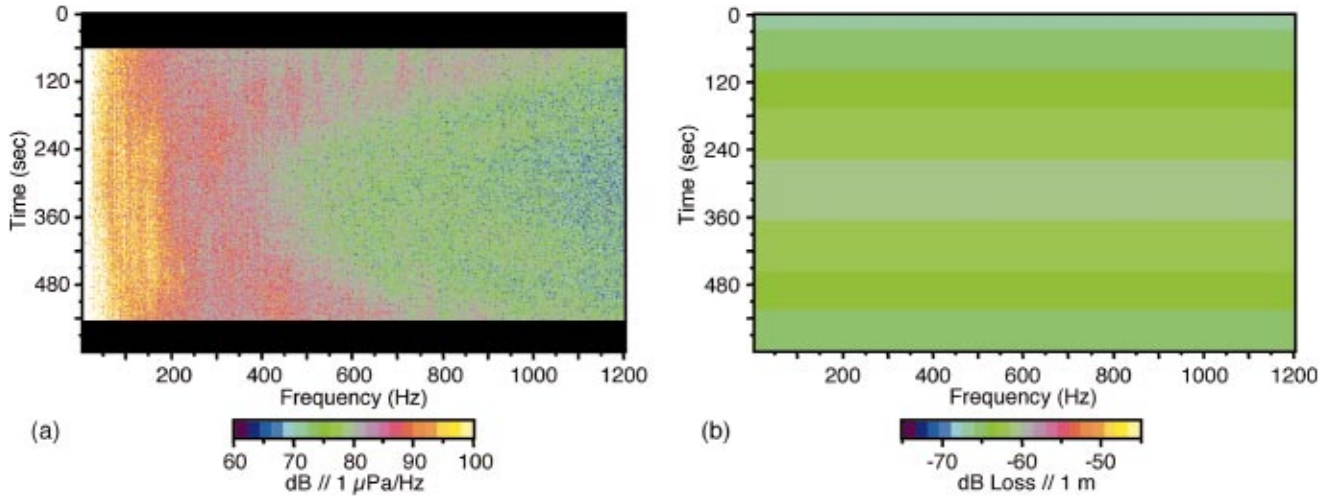


FIG. 4. (a) Source level spectrogram for the SHINOBU calculated from Fig. 1(a) using a surface monopole representation. (b) Time-frequency dependence of the surface monopole transmission loss used to calculate (a).

distribution has further reduced the loss in the horseshoe pattern and the increase in the depth of the centroid of the weighting function has slightly changed the location and shape of the pattern. An inspection of the source spectrogram for the ''Gaussian'' source distribution [Fig. 3(a)] indicates that these changes have essentially eliminated the propagation artifact in the source spectrogram.

Based on the considerations illustrated in the example of Figs. 1, 2, and 3, we have used the Gaussian source distribution to compute the source spectra discussed in this paper. We do not assert, however, that this source distribution properly represents the finite extent of the cavitation volume nor the physics of the local propagation of the energy radiated from that region. Indeed, the size of the cavitation volume can vary from a small region near the upper tip of the propeller blade arc for some ships to the full arc for other ships. Furthermore, a rigorous analysis of the local propagation effects would require a detailed knowledge of the shape of the hull and the propeller blade, the location and shape of the hull and the propeller wakes, the bubble radii distributions both within and outside of these wakes, and the effect of these wakes on the time-varying shape of the ocean surface. It is likely that a detailed analysis of both the size and the location of the cavitation volume and the local propagation effects would lead to a considerably more complex model for the local acoustic environment and the source distribution. Such a model would be required to estimate the full elevation and aspect angle dependence of the source directionality from SRN measurements.[18] However, for the near-beam aspect angles and the limited elevation angles associated with the source spectra of interest here,[19] the Gaussian source distribution is believed to be adequate.

We conclude this section with some observations on the effect of the source distribution assumption that will be useful in the discussion of the historical Ross model in Section IV. To this end, we note that in the early work by Ross and others, the source spectra were obtained by assuming a point source located at the ocean surface and a propagation loss determined by spherical spreading at a range given by the distance to the receiver, the slant range. As such, the resulting source spectra correspond to a monopole source located at the ocean surface in a homogeneous space. The implications of this assumption on the propagation and the corresponding source spectrograms for the SHINOBU are shown in Fig. 4. By assumption, the propagation spectrogram [Fig. 4(b)] is independent of frequency and hence does not show either the horseshoe pattern associated with the Lloyd's mirror effect nor the low-frequency roll-off associated with

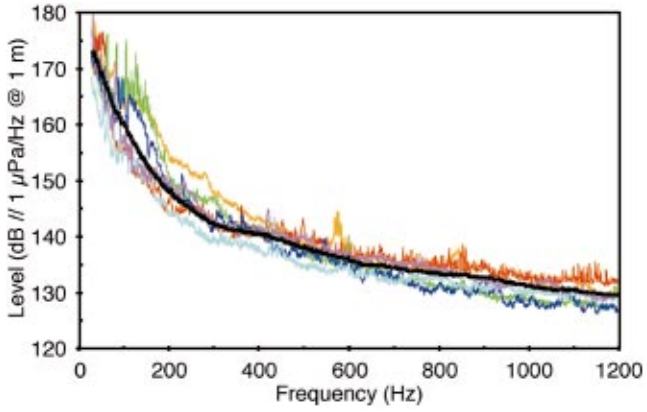


FIG. 5. Selected source spectra (colored curves) and the ensemble average spectrum (black curve).

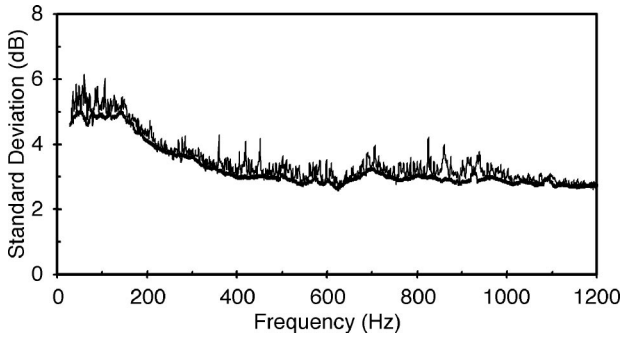


FIG. 6. Frequency-dependent ensemble standard deviation curves for the total spectrum level (thin curve) and the broadband curve (thick curve).

mode cutoff that appear in the propagation spectrograms of Figs. 1(b), 2(b), and 3(b). Accordingly, vestiges of the horseshoe pattern remain in the corresponding source spectrogram and the low-frequency levels, <80 Hz, are lower than those in the source spectrograms of Figs. 2 and 3.

The difference in the low-frequency levels across the ensemble of source spectra is of particular interest in the analysis of Sec. IV. To provide a basis for the discussions of Section IV, consider the ''finite dipole'' model of Fig. 1(b), where a single point source is located at a depth, d, below a flat ocean surface. Neglecting bottom interaction and sound-speed refraction, the acoustic intensity resulting from this source distribution is approximately given by $(d/\lambda)^2$ for $d/\lambda \ll (2 \sin \theta)^{-1}$, where θ is the depression angle to the receiver position. Accordingly, for a given ship (fixed d), a constant amplitude SRN spectrum would yield a source spectrum with an f^{-2} dependence at the low frequencies where the approximation is valid. Moreover, for an ensemble of ships, a constant amplitude SRN spectrum would introduce a d^{-2} dependence in the ensemble of source spectra. Now, if the source depth is proportional to the propeller depth and if the latter is proportional to the ship length, L, then the d^{-2} dependence is equivalent to L^{-2} dependence. The monopole source model has neither a frequency dependence nor a ship length dependence. Thus, for low frequencies, not only the form of an individual source spectra but possibly the ship length dependence across an ensemble of spectra are fundamentally different for the monopole and the finite dipole source representations. These remarks extend to the distributed source model assumed in the results presented here.

III. MEASURED SOURCE SPECTRA

In this section, we describe the variability in the source spectra and identify two low-frequency bands and a high-frequency band where the nature of that variability is different.

We begin with a qualitative description of the source spectra ensemble. Figure 5 shows a representative sample of the individual source spectra from the ensemble, together with the mean spectrum computed as a decibel average over all spectra in the ensemble.[20] It is seen in the figure that at the lower frequencies, less than 400 Hz, the individual spectra exhibit a complicated frequency structure with considerable variation in that structure from one spectra to another. This is due, in part, to the presence of the blade and machinery lines (cyan, magenta, and green spectra) and to the variability in the broadband components of the individual spectra (blue, red, and green spectra). Furthermore, many of the spectra in the ensemble exhibit one or more broad peaks (orange, green, blue, and red spectra), with some of those local peaks extending as high as 400 Hz (orange and green spectra). On the other hand, at frequencies above 400 Hz there are considerably fewer lines and the broadband components of the spectra have a much simpler structure. In fact, excluding the spectral lines, most of the individual spectra are well approximated by a f^{-p} dependence for values of p near 2. A small fraction of the ships (5%–10%) have a complex line structure that extends into the high-frequency band similar to the spectrum shown in red.

Finally, it is seen that the mean spectrum is a comparatively smooth curve, as might be expected, since the number of spectra in the ensemble is sufficient to average out the line contributions from the individual spectra. Using the rational spectrum approximation described in Section VII, it can be shown that the mean spectrum can be approximated by

$$\overline{S}(f) = 230.0 - 10\log(f^{3.594}) + 10\log\left(\left(1 + \left(\frac{f}{340}\right)^2\right)^{0.917}\right),$$

to 0.6 dB rms over the 30–1200-Hz frequency band. An inspection of this equation indicates that at both low and high frequencies, the spectrum can be approximated by a simple power law. The low-frequency exponent of this power law is -3.6; the high-frequency exponent is -1.76. The breakpoint frequency between the high-frequency and the low-frequency asymptotes is 340 Hz.

The preceding discussion asserts that the 30–1200-Hz frequency band can be partitioned into two frequency bands: a low-frequency band (30–400 Hz) and a high-frequency band (400–1200 Hz). In the high-frequency band, the individual spectra exhibit a comparatively simple frequency dependence and there is only a modest variability in the spectra across the ensemble. In the low-frequency band, the frequency dependence of the individual spectra is considerably more complex and the variability across the ensemble of spectra is significantly greater. Furthermore, the complexity in the frequency dependence of the individual spectra in the low-frequency band is due in part to the large number of strong spectral lines present in this band.

To provide quantitative measures of the variability, we have decomposed each spectrum into a line component and a background component. This was done using a line detection algorithm that computes the background component using the median filter algorithm of Wolcin.[21,22] The detection threshold for this algorithm was set to detect all lines that exceeded the background component by 50%. This threshold corresponds to a minimum line-to-background ratio of -3.0 dB. The background component was computed at a more stringent 42.5%, rejecting possible lines of -3.7 dB.

We consider first the variability in the spectra across the ensemble. To this end, we have computed the ensemble standard deviation of both the total spectrum level and the background level as a function of frequency. The results are plotted in Fig. 6. As seen in the figure, the standard deviation of the total spectra varies about a nominal value of about 5.3 dB for frequencies below about 150 Hz and then decreases to a nominal value of about 3.1 dB for frequencies greater than 400 Hz. This indicates that the complexity of the source spectra in the low-frequency band due the variability across the ensemble is limited to frequencies below 400 Hz. Moreover, the comparatively small values of the total standard deviation above 400 Hz are consistent with the observation that there is only a modest ensemble variability in the high-frequency band. The standard deviation of the background component exhibits the same frequency dependence as the standard deviation for the total spectrum. Furthermore, it differs from the nominal value of the total standard deviation

only for frequencies below about 100–150 Hz and this difference is less than about 0.5 dB. Taken together, these results indicate that the ensemble variability in the total spectrum is not due to the presence of spectral lines except for frequencies below about 150 Hz, and for these frequencies the effect of the spectral lines is minimal.

To obtain a higher-order description of the ensemble variability, it is useful to restrict the analysis to frequency bands where the nominal standard deviation is approximately constant, permitting an examination of the distributions. This requires a second division of the total frequency band into a low-frequency and a high-frequency band in addition to the partition based on the form of the individual spectra. The discussion of Fig. 6 suggests a 30–150-Hz low-frequency band where the standard deviation is approximately 5.3 dB and a 400–1200-Hz high-frequency band where the standard deviation is about 3.1 dB. To simplify the nomenclature in the remainder of the paper, we restrict the high-frequency band based on the nominal standard deviation to 400–1200 Hz so that it coincides with the high-frequency band based on the form of the spectra. However, for reasons that will be discussed later, we will preserve the distinction between the 30–150-Hz band based on the nominal standard deviation and the 30–400-Hz band based on the form of the individual spectra. We will distinguish between the two low-frequency bands by referring to the 30–150-Hz low-frequency band as the constant-sigma low-frequency band and to the entire 30–400-Hz low-frequency band as simply the low-frequency band. Comparisons with data distributions, either as histograms or CDFs, will only be carried out in the regions of constant sigma, 30–150 Hz and 400–1200 Hz.

With this nomenclature at hand, we describe the ensemble variability in terms of histograms of the relative source level. These histograms, which are shown in Fig. 7, are computed over the constant variability low-frequency and high-frequency bands identified above. Each of these histograms was obtained by subtracting the mean spectrum from the total spectrum at each frequency in the band and then pooling the relative levels over that frequency band. This is equivalent to computing the relative level histogram at each frequency in the band and then averaging the results. Furthermore, since the nominal standard deviation is constant over the band, each histogram provides an estimate of the probability density for any frequency in that band, provided that that underlying density depends only on its moments up to order two. Also shown in each plot is a Gaussian probability density determined by the measured standard deviation for each band. As seen in the figure, the relative source level histograms are reasonably well approximated by a Gaussian probability density with a frequency-dependent standard deviation.

Next, we consider the effect of the lines on the complexity of the individual spectra. To provide a measure of this effect, we have computed both the average number of lines and the average of the ratio of the line energy to the total energy of those lines as a function of frequency. The line number averages are shown in Fig. 8(a) in contiguous 10-Hz frequency intervals and normalized to unit Hz. As such, an average line number value at a given frequency can be interpreted as the probability of observing a line in a 1-Hz frequency interval on a ship picked at random from the ensemble, provided that the line occurrences are statistically independent and identically distributed across the ensemble of ships. With this interpretation, the plot of Fig. 8(a) indicates that the line occurrence probability drops rapidly from a nominal value of about 0.15 at 30 Hz to a nominal value of about 0.04 at 150 Hz. For frequencies greater than about 150 Hz, the line occurrence probability decreases over the frequency band with the exception of a slight peak in the vicinity of 400 Hz. As noted earlier, the large line occurrence probabilities at the low frequencies are primarily due to the blade lines which are evident in most, but not all, of the spectra in the ensemble.

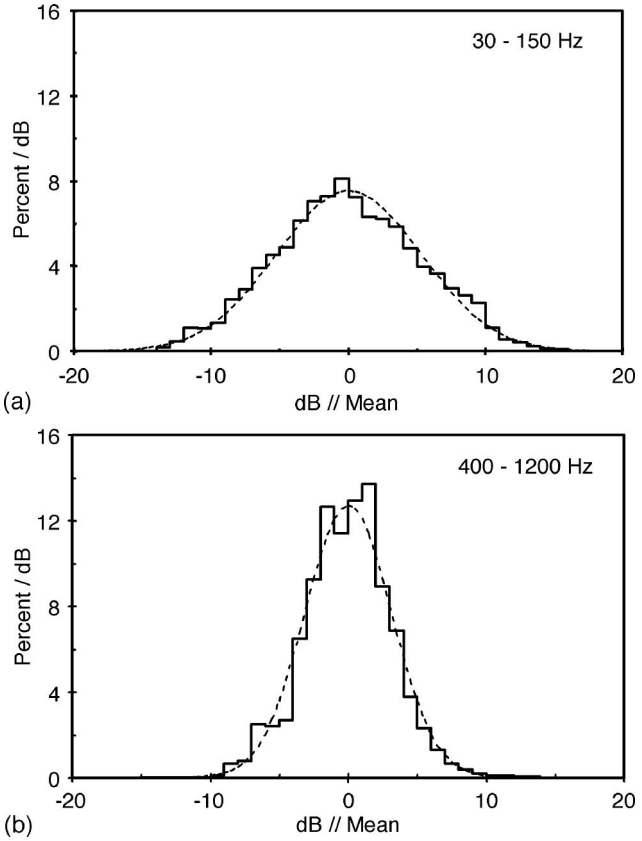


FIG. 7. The relative source level histograms and normal probability densities: (a) 30–150-Hz band and (b) 400–1200-Hz band.

The line energy ratios, Fig. 8(b), were also computed as averages from the lines detected in contiguous 10-Hz frequency bands. A comparison of this plot with the average line number plot of Fig. 8(a) indicates that not only are there a larger number of lines in the 30–100-Hz frequency band, but the lines in that band contain a significantly higher proportion of the total energy than at the higher frequencies. Furthermore, the larger line occurrence probabilities in the vicinity of 400 Hz are seen to be coupled with slightly larger line energy ratios in this band. An analysis of the line structure of the individual spectra indicates that most of the comparatively large lines in the 30–100-Hz band are blade lines, whereas, many of the lines in the vicinity of 400 Hz are engine-related lines.

The results presented in Fig. 8 suggest that the effect of

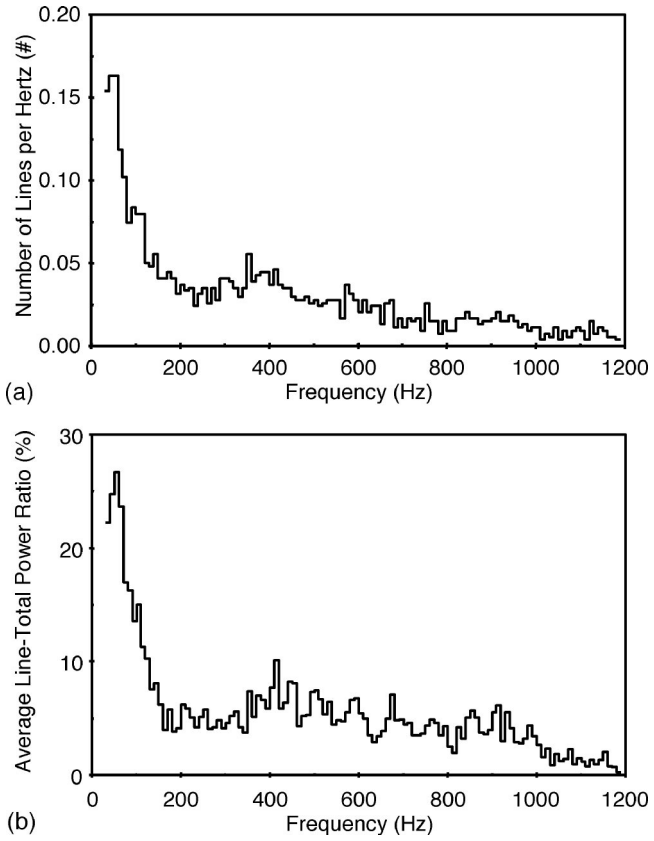


FIG. 8. Spectral line properties: (a) Average number of lines per hertz as a function of frequency. (b) Average line-to-total energy ratio as a function of frequency.

the lines on the complexity of the individual spectra is also largely limited to the 30–150-Hz frequency band. Beyond these frequencies, both the line occurrence probabilities and the line energy ratios are small. However, we note in passing that there is a large variability in the number of lines detected on the individual ships. A detailed count of the number of lines detected in the 30–1200-Hz band on each ship indicated that 43 percent of the ships have 20 lines or less, and 9 percent have over 160 lines in the band. The maximum number of lines detected on a single ship was 187; the average number of lines on a ship was 38. Clearly, these numbers depend on the specific detection threshold used in the line detection algorithm.

We conclude this section with a brief comparison of the source spectra reported by Scrimger and Heitmeyer[23] with those presented here. The Scrimger and Heitmeyer spectra were obtained for 50 ships over a 70–700-Hz frequency band from comparatively long-range measurements (2.5–35 km) using a towed array to resolve the individual ship contributions. The source spectra were derived from a transmission loss computed using a PE propagation model (PAREQ) for a point source at a common depth of 6 m for all ships. The resulting source spectra are similar in form to those of Fig. 5, but the levels are significantly larger, by 7–10 dB. Furthermore, at the lower frequencies the variability in their source spectra across the ensemble is comparable to that reported here. However, at the high frequencies there is considerably more variability in their spectra as evidenced by their larger standard deviation (σ=6.8 dB, 400–700 Hz) and the increased spread in their relative level histograms. It is possible that the increased variability at the high frequencies is due to range-frequency variability introduced in the long-range transmission loss calculations. The causes of the discrepancy in the mean spectra are not known.

IV. THE POWER-LAW ASSUMPTION

In this section, we present an evaluation of both the historical Ross model and an optimum model of the same form as the Ross model using the source spectra data set described in the preceding section. To provide a basis for the analysis presented in this and the following section, we first introduce some formal definitions.

A. Background definitions

A ship spectra model is a function of the form $\hat{S}(f,\boldsymbol{\alpha})$ that determines a source spectrum, $\hat{S}_i(f)=\hat{S}(f,\boldsymbol{\alpha}_i)$, in terms of a P-dimensional parameter vector, $\boldsymbol{\alpha}_i=(\alpha_{i,1},\ldots,\alpha_{i,P})$. The model can be either a physics-based or empirical model. In the physics-based model, the source spectrum function, $\hat{S}(f,\boldsymbol{\alpha})$, is determined from the physical laws governing the sound generation processes and the parameter vector, $\boldsymbol{\alpha}$, describes the parameters of those processes. In the empirical model, the form of the source spectrum function is postulated based on physical or functional approximation arguments. In this case, the source spectrum function also depends on a vector of model constants, **c**, which must be determined from an ensemble of measured source spectra, $\{S_i(f), i=1,\ldots,N_S\}$. In either case, the source spectra model can be used to determine an ensemble of source spectra for the ships in a region if the parameter vectors for those ships are known. For certain parametrizations, the parameter vectors can be determined through direct measurements on the individual ships that are actually present in a region during a time interval of interest. In these cases, an ensemble of source spectra for those specific ships can be generated from the source spectrum function and the measured parameter vectors. Alternatively, the parameter vectors can be described by a probability density, $p_{\boldsymbol{\alpha}}(\alpha_1,\ldots,\alpha_P)$, which can be determined from archival measurements or from ship-routing models. In these cases, it is only possible to associate the statistics of the ensemble of source spectra with those of a ''representative'' ensemble of ships.

The correctness of the model is determined by the accuracy to which it represents the individual source spectra in an ensemble of spectra. None of the models considered in this paper seeks to provide a totally accurate description of the spectrum, since none of them attempts to describe the spectral line structure. Given this limitation, a reasonable measure of the accuracy of the source spectra models is provided by the ensemble average of mean-square error

$$\varepsilon^2=\langle\varepsilon_i^2\rangle_I, \tag{1a}$$

where

$$\varepsilon_i^2=\langle(S_i(f_j)-\hat{S}_i(f_j,\boldsymbol{\alpha}_i))^2\rangle_F \tag{1b}$$

is the mean-square error for the ith ship, where $\langle\cdot\rangle_I$ denotes an average over the ensemble of ships and $\langle\cdot\rangle_F$ denotes a frequency average over the band of interest. Less-stringent accuracy measures arise in applications where there is no regional dependence on the ship parameters, or alternatively, where the regional dependence is not known. In these cases, a source spectra model that generates certain statistics that agree with the ensemble statistics may suffice. For ambient noise models that are intended to compute only the mean or the cumulative distribution function of the noise in these region-independent environments, it suffices to accurately predict only the mean or the cumulative distribution function of the source spectra ensemble. Thus, we will also consider these statistics as measures of the source spectra model accuracy.

The historical Ross model represents a source spectrum as a ''baseline'' spectrum together with a ship-dependent displacement to that spectrum that is determined as a linear combination of the logarithms of Ships Registry parameters. Thus, the Ross model can be considered as a special case of a model of the form

$$\hat{S}(f,\delta)=S_o(f)+\delta, \tag{2}$$

where $S_o(f)$ is the baseline spectrum and the displacement parameter, δ is given by

$$\delta=\sum_{p=1}^{P} c_p\alpha_p, \tag{3a}$$

where

$$\alpha_p=10\log(\theta_p/\overline{\theta}_p), \tag{3b}$$

and $\overline{\theta}_p$ is the geometric mean of the Ship Registry parameter, θ_p. Note that these equations state that, on a linear power scale, the source spectrum is the baseline spectrum scaled by a ship-dependent constant, where that constant has a power-law dependence on the ship parameters, i.e.,

$$10^{\delta/10}=\prod_{p=1}^{P}(\theta_p/\overline{\theta}_p)^{c_p}. \tag{4}$$

Given an ensemble of source spectra measurements and its associated Ships Registry parameters, the model constants, c_p, and the baseline spectrum, $S_o(f)$, can be determined to minimize the mean-square error for that ensemble. In the following, we refer to such a model as the optimum power-law model.

To provide a basis for interpreting the integrity of the power-law assumption, we introduce two additional models that also represent a source spectrum in terms of a baseline spectrum. In the first of these, the ship-dependent displacements, δ_i, in Eq. (1), as well as the baseline spectrum, are determined by minimizing the mean-square error over the ensemble of ship spectra. We refer to this model as the optimum displacement model. The second model is a degenerate model, which represents each spectrum by the baseline spectrum itself, i.e., $\delta_i=0$ in Eq. (1). It is easily shown that, given a source spectrum ensemble, the baseline spectrum that minimizes the minimum mean-square error is simply the mean spectrum of that ensemble. Thus, we refer to this model simply as the mean spectrum model.

The three ''baseline spectrum'' models are linear in the unknown quantities. Consequently, explicit expressions for the minimum mean-square error and for the quantities that achieve that error can be obtained using well-known linear least-squares arguments. In the Appendix, we outline the approach to obtaining the solutions for these three models. The solutions themselves are presented in Sec. IV C.

The utility of the optimum displacement and the mean spectrum models arises through the comparison of the mean-square errors for all three models when these errors are computed on the same ensemble of source spectra as that used to determine the model parameters. Under these conditions, it can be shown that $\varepsilon^2_{\mathrm{OD}}\leqslant\varepsilon^2_{\mathrm{OPL}}$, where $\varepsilon^2_{\mathrm{OD}}$ is the mean-square error for the optimum displacement model and $\varepsilon^2_{\mathrm{OPL}}$ is the mean-square error for the optimum power-law model. Furthermore, equality is obtained if and only if the optimum displacements, $\{\delta_i\}$, and the ship registry parameter vectors, $\{\boldsymbol{\alpha}_i\}$ satisfy the linear model of Eq. (3a) with zero error. Consequently, the difference between the minimum mean-square error for the two models provides a quantitative measure of the correctness of the power-law assumption on the ship registry parameters. Second, the mean spectrum model will have a larger minimum mean-square error than the optimum power-law model, $\varepsilon^2_{\mathrm{MS}}>\varepsilon^2_{\mathrm{OPL}}$, if there is any correlation between the ship registry parameters and the displacement to the baseline spectrum. As such, it provides a second measure of the integrity of the power-law assumption.

For later reference, we note that for any of these models the histogram of the predicted relative source level is identical to the histogram of the displacement parameter. As such, these histograms are independent of frequency for all frequencies in the band over which the displacement parameter has been determined. These properties, which follow immediately from Eq. (2), also apply to the cumulative distribution function of the relative source level.

Finally, we note that for all of these models, the model parameters obtained through mean-square error minimization and the resulting minimum mean-square errors are strongly dependent on the frequency band over which the spectrum is to be approximated. In this paper, we develop separate models for the low-frequency (30–400 Hz) and the high-frequency (400–1200 Hz) bands identified as the frequency-dependent partition in Sec. III. In addition, we develop a third model by conducting the minimization over the full 30–1200-Hz band. This full-band model is useful in its own right, since the subband models yield spectra that are discontinuous at the 400-Hz boundary of the subband models. Furthermore, the mean-square error for the full-band model provides a useful measure of the improvement obtained by the frequency partitioning. In particular, it is shown in the Appendix that the full-band mean-square error provides an upper bound to the band-weighted average of the mean-square errors for the subbands with equality if and only if the there is no improvement in the mean-square error obtained through subband partitioning.

B. The historical Ross model

The classical model proposed by Ross represents the spectrum of an individual ship by a function of the form of Eqs. (2) and (3).[24,25] As such it is a power-law model. A number of different ship registry parameters have been assumed in the different formulations. These include the ship speed, V, the length of the ship, L, its horsepower, and its displacement tonnage. Among these different formulations, the model that depends on the ship speed and the ship length has become the most widely used model due to the comparative ease with which these parameters can be obtained through direct measurements. For this choice, the model spectrum equation is customarily written in the form

$$S(f,V,L)=S_o(f)+c_V\,10\log(V/V_o)+c_L\,10\log(L/L_o), \tag{5}$$

where V_o and L_o are the reference speed and the reference length. The constants, c_V and c_L, were given as 6 and 2, respectively.[26] However, in other work[24] a range of values between 5 and 6 is given for c_V and the length term is replaced by a displacement tonnage term. The coefficients for the tonnage term vary between 0.9 and 1.5. These coefficients were determined over a number of years from 1/3-octave measurements made during and after World War II. Most of these measurements were made in shallow-water sheltered environments. As noted in Sec. II, the transmission loss was determined from the slant range assuming spherical spreading.

Since these models were originally proposed, they have undergone substantial ''revisions,'' primarily to the baseline spectra. In particular, it was recognized that the use of the surface monopole representation was not convenient for use with transmission loss models. Additionally, it was recognized that the design of ships was changing and that the spectra determined from the World War II data might not be representative of modern ships. Rather than undertake a major measurement program, efforts were made to ''calibrate'' the baseline spectrum by comparing predicted and measured ship-radiated noise and adjusting the baseline spectra to obtain a best fit. In the computations presented in this section, we take the baseline spectrum as the spectrum developed for the RANDI ambient noise model over a period of some years.[27] This spectrum includes an additional length-dependent term, $g(f,L)$, as a correction to the Ross formula.[28] This term is nonzero only below 192 Hz, where it increases with increasing ship length and decreasing frequency; at 30 Hz, $g(f,L)$ is just over 2 dB for the average ship and 6.4 dB for the largest ship. The speed and length coefficients were taken as 6 and 2, respectively.

Figure 9 shows the measured mean spectrum together with the mean spectra computed from the Ross model using the velocity and length parameters for each ship. Except at the very lowest frequencies, below 50 Hz, the Ross mean lies 2–4 dB below the measured mean at frequencies up to 350 Hz, and then drops to about 3–6 dB below the measured mean for frequencies above 400 Hz. For low frequencies this can be attributed to source depth differences, as a computation based on the center of the propeller yielded similar answers in the low-frequency range. The differences in source depth do not affect the high-frequency results. Note that the lack of agreement at high frequencies occurs despite the fact that the high-frequency asymptotes of the Ross mean and the measured mean are both power laws with virtually identical exponents (Ross, 1.8; measured, 1.76).[29]

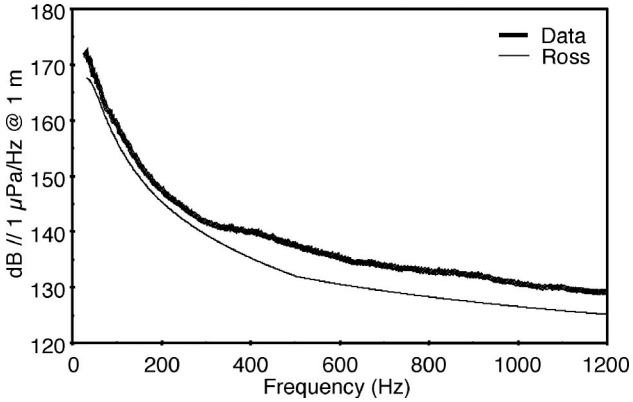


FIG. 9. The measured average source level spectra (thick curve) and predicted mean source spectrum for the historical Ross model (thin curve).

Figure 10 shows cumulative distribution functions of the relative source level computed from both the measured spectra and from the Ross model for the same low-frequency and high-frequency bands of Fig. 7 in the preceding section. As for Fig. 7, each cumulative distribution function (CDF) was obtained by subtracting the mean spectrum from the total spectrum at each frequency in the band and then pooling the

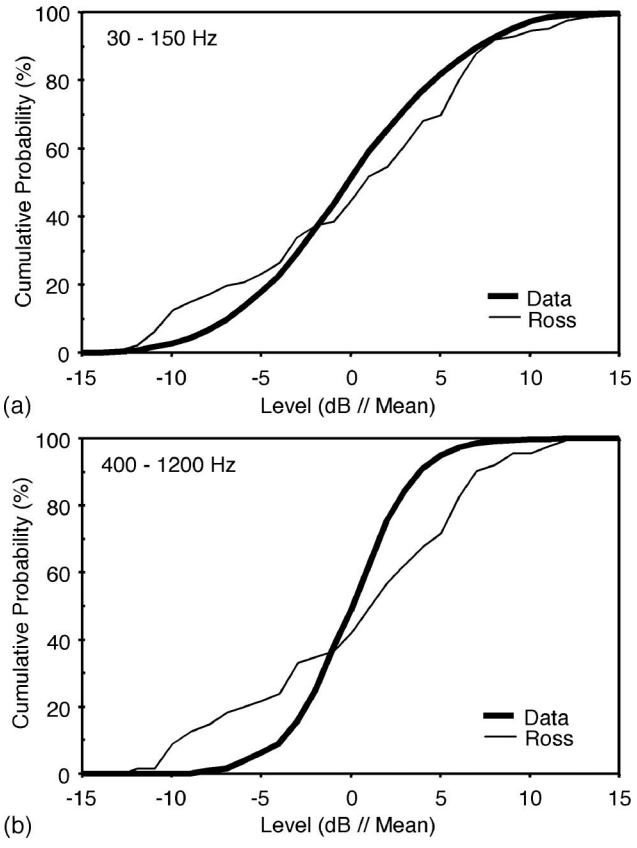


FIG. 10. Measured cumulative distribution functions of the relative source level (thick curve) and computed cumulative distribution functions for the historical Ross model (thin solid): (a) 30–150 Hz-band and (b) 400–1200-Hz band.

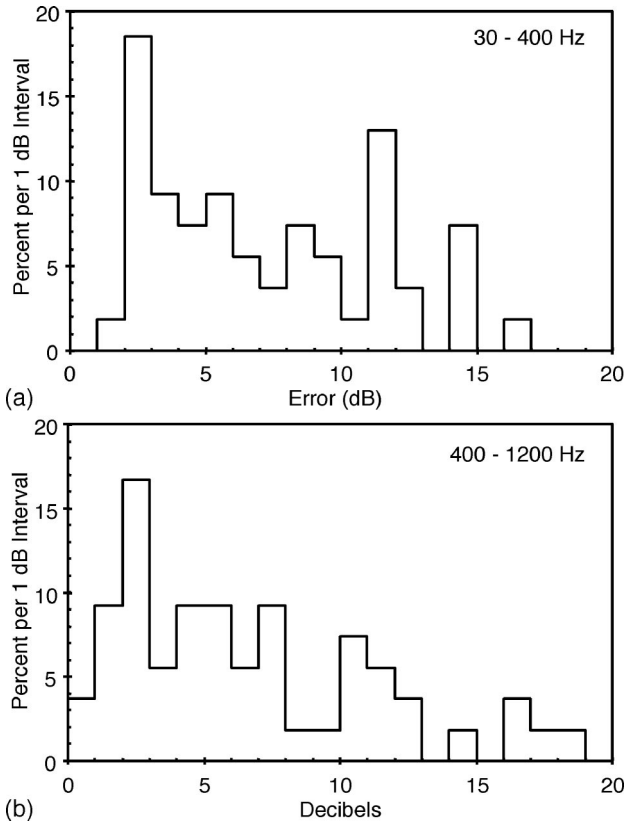


FIG. 11. Error histograms for the Ross model: (a) 30–400-Hz band and (b) 400–1200-Hz band.

relative levels over that frequency band. As seen in the figure, the Ross model CDF is only roughly comparable to the measured CDF in the low-frequency band and it significantly overestimates the spread of the histogram in the high-frequency band. The fact that the Ross model cannot provide a reasonable approximation in both bands is a consequence of the fact that the Ross model is a baseline spectrum model; hence, its relative level CDF is frequency independent (see Sec. IV A). Because of this the Ross curve is relative jagged as it is effectively determined by only 54 independent samples in each band, whereas the data has much more variability between each of the independent spectra and the mean and is hence smoother. The slight differences between the two Ross curves in the different bands are due to the length-dependent term mentioned above.

Figure 11 presents histograms of the rms spectra errors, ε_i, obtained from the Ross model for both the low- and high-frequency bands. Contrary to the other models presented in this paper, the parameters of the Ross model are the same for both frequency bands. The low-frequency plot, Fig. 11(a), indicates that the rms error for about 31 percent of the ships exceeds 10 dB and that there is only one ship with less than a 2-dB error. The ensemble rms error, ε, for the low-frequency band is 9.38 dB. The high-frequency histogram, Fig. 11(b), shows only a slight reduction in the rms errors. The percentage of the ships with rms errors greater than 10 dB has decreased to 28 percent and the percentage of the ships that have rms errors less than 2 dB has increased to 13 percent. Furthermore, the ensemble rms error for the high-frequency band has decreased to 8.54 dB. Nevertheless, the histograms of Fig. 11 indicate that the historical Ross model does not provide reasonable approximations to the observed spectra of individual ships in either frequency band.

C. The optimum power-law model

The preceding results indicate that for this source spectra ensemble, the Ross model does not provide an accurate description of either the relative source level CDF or acceptably small rms spectra errors. One possible cause of this discrepancy is that the coefficients for the Ross model were determined from ship-radiated noise data obtained during and shortly after World War II. Furthermore, the source spectra used to determine the Ross model coefficients were obtained from the monopole source assumption. A second possible cause is that the power-law assumption itself is not satisfied by the current data set. To test this second hypothesis, we have derived the coefficients for the optimum power-law model from the current source spectra ensemble. If the power-law assumption is satisfied by the current database, the errors obtained from the optimum power-law model should be comparable to those obtained from an optimum displacement model and considerably smaller than those obtained with the mean spectrum model.

To provide a background for the discussion, we present a summary of the solutions to the linear least-square problem for the three baseline spectrum models. For each of these models, the optimum baseline spectrum is the mean spectrum of the ensemble, $S_o(f)=\bar{S}(f)=\langle S_i(f)\rangle_I$. Furthermore, for each model the minimum mean-square error can be expressed in the form

$$\varepsilon_o^2=\sigma_S^2(1-\rho_S^2), \tag{6}$$

where ρ_S is the model-dependent multivariate correlation coefficient and σ_S^2 is the frequency average of the frequency-dependent ensemble variances, i.e., $\sigma_S^2=\langle\sigma_S^2(f)\rangle_F$, where $\sigma_S^2(f)=\langle(S_i(f)-\bar{S}(f))^2\rangle_I$. Note that for a given frequency band, σ_S^2 is the "nominal" standard deviation referred to in Sec. III.

The specific expressions for the displacements and the correlation coefficients for the three baseline spectrum models are summarized in Table I. As seen in the table, the degenerate mean spectrum model has a zero correlation coefficient so that the mean-square error is simply the ensemble variance. The optimum displacement model has displacements given by the difference between the average spectrum level for the ship, $S_i^b=\langle S_i(f_j)\rangle_F$, and the ensemble average of the average spectrum level, $\bar{S}^b=\langle S_i^b\rangle_I$. The square of the correlation coefficient for this model is the ensemble average of the displacements squared, $\sigma_\delta^2=\langle(S_i^b-\bar{S}^b)^2\rangle_I$, divided by the ensemble variance. For the optimum power-law model, the quantities, σ_V and σ_L are the standard deviations of $10\log(V/\bar{V})$ and $10\log(L/\bar{L})$, respectively. Similarly, $\rho_{V,S}$ and $\rho_{L,S}$ are the cross-correlation coefficients between S^b and $10\log(V/\bar{V})$ and between S^b and $10\log(L/\bar{L})$, respectively, and $\rho_{V,L}$ is the cross-correlation between $10\log(V/\bar{V})$ and $10\log(L/\bar{L})$.

TABLE I. Baseline spectrum model solutions.

Model	Displacement coefficients	ρ_s
Mean spectrum	$\delta_i = 0$	0
Optimum displacement	$\delta_i = S_i^b - \overline{S^b}$	σ_δ/σ_s
Optimum power law	$c_V = \frac{\sigma_S}{\sigma_V}\frac{(\rho_{V,S}-\rho_{V,L}\rho_{L,S})}{(1-\rho_{V,L}^2)}$	$\frac{\sigma_\delta}{\sigma_s}\mu$,
	$c_L = \frac{\sigma_S}{\sigma_L}\frac{(\rho_{L,S}-\rho_{V,L}\rho_{V,S})}{(1-\rho_{V,L}^2)}$	$\mu = \sqrt{\frac{\rho_{V,S}^2 - 2\rho_{V,L}\rho_{L,S}\rho_{V,S} + \rho_{L,S}^2}{(1-\rho_{V,L}^2)}}$

The ensemble rms errors for the three baseline spectrum models are presented in Table II. For reference purposes, we have also included the rms errors for the historical Ross model.

As seen in Table II, all three baseline spectrum models yield substantially smaller errors than the historical Ross model for all three frequency bands. Furthermore, the largest errors occur in the low-frequency band and the smallest errors occur in the high-frequency band, as expected from the discussion in Sec. III. In the full frequency band, the Ross model error of 8.8 dB reflects the inability of this model to accurately represent the individual source spectrum in the current database, as was seen in the error histograms of Fig. 11. Among the baseline spectrum models in this band, the mean spectrum model has the largest error, 3.6 dB, and the optimum power-law model has an error that is less than 0.1 dB smaller than the mean spectrum model. The optimum displacement model is significantly better than either of the other baseline spectrum models with a rms error of only 2.6 dB. The fact that the mean spectrum model is the worst of the three baseline spectrum models is not surprising since there are no displacements to the baseline spectrum to reduce the error. However, it is surprising that the optimum power-law model is essentially the same as the mean spectrum model and almost 1.0 dB worse than the optimum displacement model, since the optimum power-law model determines both the power-law exponents and the baseline spectrum to minimize the mean-square error for the source spectra ensemble. Thus, the results of Table II indicate that the assumption of a power-law dependence on the ship speed and length is not satisfied for this data set. This result is consistent with the comparatively large errors obtained for the Ross model.

Table III provides additional insight into the deficiencies of the power-law assumption. Here, we have listed the values of the power-law coefficients, the constituent correlation coefficients, and the multivariate correlation coefficients for each of the three frequency bands. As seen in the table, the small values of $\rho_{V,L}$ indicate that the ship speeds and lengths are essentially uncorrelated. Thus, the correlations between the average spectrum level and the decibel ship speeds and the lengths, $\rho_{V,S}$ and $\rho_{L,S}$, each act independently to reduce the mean-square error (see Table I). The values of these correlation coefficients, however, are so small that for μ, and hence ρ_s, there is very little reduction in the mean-square error. The only non-negligible reduction occurs in the low-frequency band, where $\rho_{L,S} = -0.427$ leads to a value $\rho_s = 0.427$, which corresponds to a modest 10% reduction in the rms error for the optimum power-law model over the mean spectrum model. Interestingly, the negative value of $\rho_{L,S}$ indicates that long ships do not necessarily generate more sound than short ships. The small values of $\rho_{V,S}$ and $\rho_{L,S}$ also account for the small values of the power-law coefficients, c_V and c_L, which provides an alternate statement of the deficiencies in the power-law assumption.

The lack of a functional dependence on the ship speed or length is illustrated in the scatter plots of Figs. 12 and 13. In each of these plots, the average source level, S_i^b, is plotted against one of the ship parameters for the subset of ships where the other ship parameter is restricted to a limited range.

Figure 12 shows scatter diagrams of the average source level plotted against ship speed for ships with lengths between 140–200 m. For this length range, the 60-m variation in length would result in, at most, a ± 1.55-dB variation in the broadband levels if the Ross assumption of a square-law dependency on the ship length were correct. Thus, within these margins of error the scatter diagrams of Figs. 12(a) and (b) illustrate the length dependence of the broadband level for a constant ship length. Clearly, no adjustments to the broadband levels of these magnitudes would result in a strongly correlated speed dependence of any simple functional form for either frequency band. Certainly, the 24-dB variation determined by a $60\log(V)$ dependence over the span of the speed axis is not apparent in the plots.

The lack of a speed dependence in Fig. 12 is not inconsistent with the well-known fact that a given ship will radiate

TABLE II. The ensemble rms errors for the baseline spectrum models and the historical Ross model.

Models	30–400 Hz	400–1200 Hz	30–1200 Hz
Historical Ross model	9.38	8.54	8.82
Baseline spectrum models			
Mean spectrum	4.41	3.14	3.59
Optimum power law	3.99	3.13	3.51
Optimum displacement	2.44	1.63	2.55

TABLE III. Coefficients of the optimum power-law model.

Coefficients	30–400 Hz	400–1200 Hz	30–1200 Hz
C_V	0.03	−0.07	−0.05
C_L	−0.83	−0.09	−0.33
ρ_{V,S^b}	−0.012	−0.016	−0.014
ρ_{L,S^b}	−0.427	−0.067	−0.206
$\rho_{V,L}$	0.019	0.019	0.019
ρ_S	0.427	0.068	0.206

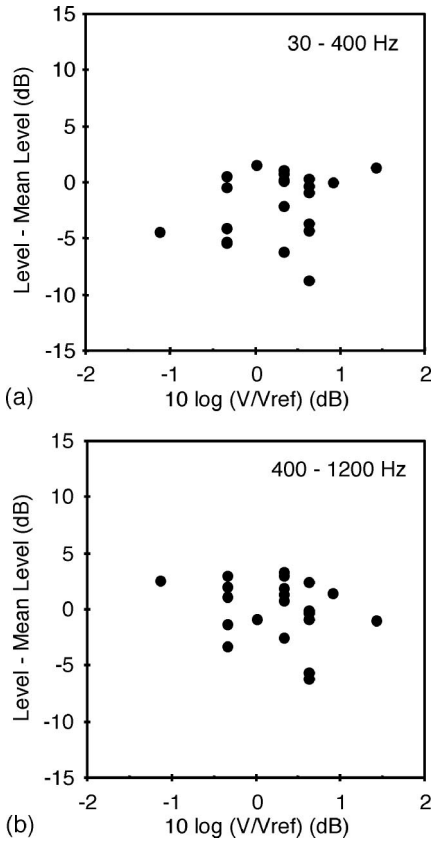


FIG. 12. Scatter plot of averaged source levels versus ship speed for ship lengths between 140–200 m. (a) 30–400 Hz; (b) 400–1200 Hz.

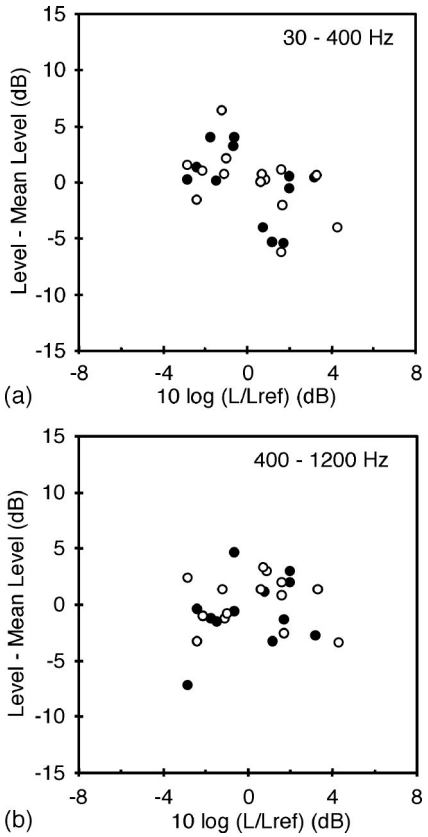


FIG. 13. Scatter plot of averaged source levels versus ship length for ship speeds at 12 knots (solid) and 14 knots (open): (a) 30–400 Hz; (b) 400–1200 Hz.

more sound at higher speeds than at lower speeds. In fact, the power-law relationship proposed by Ross may be a good approximation for a specific ship. However, it does not follow that a power-law dependence for a given ship extends to a power-law dependence for an ensemble of ships, since each ship in that ensemble is traveling at its own cruising speed. It is not clear from the available literature why the Ross data support a power-law dependence and the current data do not. One possible reason is that in the results reported by Ross[24] there are data points corresponding to the same ship traveling at different speeds as well as data points corresponding to individual ships. Including these multiple-speed data points may have given the appearance of a power-law dependence across the ensemble.

Figure 13 shows scatter diagrams of the average spectrum level plotted against ship length for ship speeds of 12 knots (open dots) and 14 knots (closed dots). For the 12-knot nominal ship speed, an uncertainty in the speed of ± 1 knot would result in at most a ± 2.17-dB variation in the broadband level if, in fact, the strong sixth-power dependence assumed by Ross governed the speed dependence. For the 14-knot nominal speed, the broadband level variation due to ship speed variation would be ± 1.86 dB. Thus, within these margins of error, the scatter diagrams of Fig. 13 illustrate the length dependence of the broadband level for a constant speed. Clearly, no adjustments to the broadband levels of these magnitudes would result in a strongly correlated length dependence of any simple form for either frequency band.

As with the ship speed dependence, it is also not clear why there is a disparity in the length dependence between the current results and the result reported by Ross. One possible explanation lies in the difference in the source distribution used to derive the source spectra in the two analyses. As noted in Sec. III, a point source located below a flat surface will introduce an L^{-2} dependence into the source spectra at low frequencies, provided that the length of the ship is proportional to the depth of the source. This remains true even for the distributed source model for sufficiently low frequencies. In the source spectra derived here under the distributed source assumption, there is at most a weak length dependence. Consequently, the L^2 dependence observed by Ross and the lack of a length dependence observed in the current results could be due simply to the difference in the source distribution assumed in the two analyses.

To test the source distribution hypothesis, we have recomputed the parameters of the optimum power-law model using source spectra derived under the monopole assumption. The results are summarized in Table IV. As seen in the table, there no significant correlations between the average spectrum level and the speed and length parameters, $\rho_{V,S}$ and $\rho_{L,S}$, and no significant increase in ρ_s. Thus, the deficiency in the power-law assumption cannot be attributed to the assumption used to calculate the source spectra from the measured ship-radiated noise. We note that for both the low-frequency band and the total band, the values of $\rho_{L,S}$ obtained under the monopole assumption are slightly positive, indicating a slight tendency for longer ships to generate more sound. The value of $\rho_{L,S}$ in Table III, which was ob-

TABLE IV. Coefficients of the optimum power-law model derived from the monopole source spectra.

Coefficients	30–400 Hz	400–1200 Hz	30–1200 Hz
C_V	0.23	0.11	0.15
C_L	0.23	0.20	0.21
ρ_{V,S^b}	0.049	0.027	0.035
ρ_{L,S^b}	0.157	0.148	0.151
$\rho_{V,L}$	0.019	0.019	0.019
ρ_s	0.164	0.150	0.154

tained under the distributed source region assumption, is slightly negative.

A qualitative comparison with the results of Wright and Cybulski[30] supports the observed lack of a speed and length dependence. The Wright and Cybulski source spectra were obtained for 14 ships over a 2–128-Hz frequency band using sonobuoy measurements of the ship-radiated noise. The transmission loss was computed using an FFP propagation model with a point source located below the surface at a ship-dependent depth halfway between the propeller shaft and the top of the propeller. The data were restricted to frequencies and ranges below the first Lloyd's mirror null. The resulting spectra are similar in level as well as in form to those of Fig. 5 over the common 30–128-Hz frequency band. This agreement is somewhat surprising since the average ship length for the Wright and Cybulski measurements is more than twice that for the current ensemble. Furthermore, seven of the Wright and Cybulski ships were supertankers, whereas only one of the ships in the current database is a supertanker. Given the relative size and speed differences, the Wright and Cybulski spectra should exceed the present mean spectra by approximately 8 dB, were Eq. (5) to hold. In fact, their results are distributed both above and below the mean spectrum and appear to be virtually identical to the current results.

Scrimger and Heitmeyer did not investigate the ship speed or ship length dependence. They did, however, compare the ensemble mean spectra for three ship classes, ferries/passenger ships, cargo ships, and oil/chemical tankers. Their results indicated no statistically significant differences between the mean spectra for the ship classes, with the mean spectra for each ship class lying ''essentially within one-half of a standard deviation of the full ship sample.''[23]

In light of the preceding results, it is certainly possible that the failure of the power-law assumption on the current data set is due, at least in part, to fundamental changes in the ships that have occurred since World War II. One possible difference lies in the advances made in the design of the ship propellers. Most World War II designs resulted in propellers that were fully cavitating. In the intervening years, considerable effort has been directed to designing propellers that could deliver the thrust required to meet the ever-increasing speed and length requirements and yet operate in less than a fully cavitating mode. This effort has not been motivated by acoustic considerations, but by the desire to reduce the hull vibrations induced by cavitation and to increase the propeller efficiency. The result has been a significant increase in the fraction of the world merchant fleet that does not operate in a fully cavitating mode. Since the sound radiated by a ship is an increasing function of the size of the cavitation region, this reduction might contribute to the smaller values of $\rho_{V,S}$ and $\rho_{L,S}$ observed in the current data set. Another notable difference change in the world merchant fleet has been in the shift to diesel engines. In the database used by Ross to derive the source spectra, all of the ships were steam driven. In the current database, only two ships are steam driven. The remaining are diesel driven. Other possible causes of the power-law assumption discrepancy might be found in differences in the measurement scenario (shallow water versus deep water) and in the processing procedure (1/3-octave filters versus FFT processing).

V. THE RATIONAL SPECTRUM MODEL

In the preceding section, it was seen that the basic assumption underlying the historical Ross model was not satisfied by the current data set. This was demonstrated, in part, by the large reduction in the rms error obtained for the optimum displacement model over that obtained for the optimum power-law model. The latter determines the displacements to a baseline spectrum that minimize the mean-square error subject to a linear constraint on the speed and length of the ship; the former determines the displacements without any constraint. In order to obtain a model that accurately determines the spectrum of an individual ship, it is necessary to have a deeper understanding of the physical processes that determine the noise radiated by a ship than is currently available. Moreover, if this ''ship-specific'' source model is to be applicable to specific ensembles of ships, it must be possible to determine the physical parameters of those processes, e.g., propeller speed and shape, hull shape, flow rates, etc., for a significant sample of the world shipping fleet.

In the absence of a credible ship-specific model, it is reasonable to consider a second type of model where the parameter vector, $\boldsymbol{\alpha}$, is not related to the physical parameters of individual ships. Rather, these vectors serve only to shape the individual source spectrum. Clearly such a model, which we refer to as a generic ship source model, cannot generate a spectrum that can be associated with a particular ship, and hence, it cannot generate an ensemble of source spectra that corresponds to the specific ships present in a region. Such an ensemble can only be determined by a source-specific model as noted in the preceding section. However, in the cases where the parameter vectors for the specific ships present in the region of interest cannot be determined, a ship-specific model offers no advantage over a generic source model.

The specification of a generic source model consists of both the source spectrum function, $\hat{S}(f,\boldsymbol{\alpha})$, and the parameter vector probability density $p_{\boldsymbol{\alpha}}(\alpha_1,\ldots,\alpha_P)$. The optimum displacement model considered in the preceding section is one example of such a model. By construction, the optimum displacement model achieves the smallest mean-square error among all models that represent the frequency dependence of the individual source spectra by a single baseline spectrum. However, the variability in the frequency dependence of the individual spectra seen in Fig. 5 suggests that these errors can be further reduced by a model where the form of the

individual spectrum varies from one ship to another. Such a model can be developed by replacing the baseline spectrum with a ship-dependent spectrum that consists of a linear combination of approximating functions. Accordingly, unlike the baseline models the form of the spectrum is different for each ship. In this section, we define such a model, compare the errors obtained with this model with those obtained from the optimum displacement model, and propose probability densities on the model parameter for both models.

We refer to the model proposed here as the rational spectrum model. This model has a spectrum function of the form

$$\hat{S}(f,\delta,\boldsymbol{\nu},\boldsymbol{\eta})=\delta+\sum_{p=1}^{P-1}\nu_p\left[\beta_p\left(\frac{f}{\eta_p}\right)-\left\langle\beta_p\left(\frac{f}{\eta_p}\right)\right\rangle_F\right], \quad (7)$$

where the parameter vector $\boldsymbol{\nu}=(\nu_1,\ldots,\nu_P)$ specifies the weights in the linear combination of approximating functions, $\beta_p(f)$, and the vector $\boldsymbol{\eta}=(\eta_1,\ldots,\eta_P)$ specifies location frequencies for those approximating functions. Such a model is a generic model with $\boldsymbol{\nu}$, $\boldsymbol{\eta}$, and δ determined for each source spectrum. In this paper we take the approximation functions to be

$$\beta_p(x)=\begin{cases}10\log(x^2), & p=1\\ 10\log(1+x^2), & p>1.\end{cases} \quad (8)$$

For this choice, the baseline spectrum, expressed on a linear power scale, has the form

$$10^{\hat{S}(f,\delta,\boldsymbol{\nu},\boldsymbol{\eta})/10}=10^{\delta'/10}\frac{\Pi_{n=1}^{M_N}(1+(f/\eta_n)^2)^{\nu_n}}{f^{2|\nu_1|}\Pi_{m=1}^{M_D}(1+(f/\eta_m)^2)^{|\nu_m|}}, \quad (9)$$

where M_N is the number of terms with positive coefficients, M_D is the number of terms with negative coefficients, and $M_N+M_D=P-1$. In Eq. (9), we have restricted the coefficient, ν_1, to be less than or equal to zero based on the form of the individual source spectra discussed in Sec. III. Further, we have absorbed the location frequency η_1 and the frequency averages of the approximating functions into the displacement parameter, δ'. With these adjustments, the right-hand side of Eq. (9) is recognized as the power spectrum associated with a system described by ordinary linear differential equations with constant coefficients. As such, the underlying complex spectrum has the form of the rational spectrum that forms the basis of the commonly used ARMA model. The spectrum of Eq. (9) differs from the rational spectrum in two ways. First, we have limited the poles and zeros, $2\pi\eta_m$ and $2\pi\eta_n$, to be real rather than allowing them to occur in complex pairs. The complex poles were eliminated to preclude approximating the narrow-band lines in the source spectrum, which are described by complex poles close to the imaginary axis. The complex zeros representing narrow minima in the spectra were also eliminated. Accordingly, the spectra can take on only positive values and will represent the broadband features and not attempt to replicate any narrow-band features. The second difference is that we allow the powers of the factors in Eq. (9) to be real rather than restricting them to integers. This extension was motivated by the noninteger powers observed in the low-frequency and the high-frequency asymptotes for many of the source spectra.

The rational spectrum model is not a linear model since the spectrum is not a linear function of the location frequencies. Consequently, explicit expressions for the minimum ensemble error, ε_o^2, and the parameter values that achieve that minimum are not available. However, for each source spectrum, $\hat{S}_i(f)$, the model is linear in the spectrum displacement, δ_i, and the exponent coefficient vector, $\boldsymbol{\nu}_i$, for a fixed frequency parameter vector, $\boldsymbol{\eta}_i$. Thus, for a fixed frequency parameter vector, $\boldsymbol{\eta}_i'$, explicit expressions can be obtained for the minimum mean-square error, $\varepsilon_{o,i}^2(\boldsymbol{\eta}_i')$ and for the linear parameters that achieve that error, $\boldsymbol{\nu}_i(\boldsymbol{\eta}_i')$ and $\delta_i(\boldsymbol{\eta}_i')$. The expression for the $\varepsilon_{o,i}^2(\boldsymbol{\eta}_i')$ can then be minimized numerically to provide the absolute minimum mean-square error for that spectrum, $\varepsilon_{o,i}^2$ and the value of the location frequency that achieves that minimum, $\tilde{\boldsymbol{\eta}}_i$. The linear parameters corresponding to the absolute minimum can then be determined as $\tilde{\boldsymbol{\nu}}=\boldsymbol{\nu}_i(\tilde{\boldsymbol{\eta}}_i)$ and $\tilde{\delta}_i=\delta_i(\boldsymbol{\eta}_i')$. Thus, the numerical minimization must be carried out only over the $P-1$ dimensional space of unknown location frequencies, rather than the full set of $2P$ unknown parameters. The minimum ensemble error, ε_o^2, must be determined numerically as the ensemble average of the $\varepsilon_{o,i}^2$. Expressions for the minimum mean-square error and the linear coefficients are presented in the Appendix.

The results presented here are obtained for three cases, a one-term, a two-term, and a three-term approximation. The one-term approximation represents the frequency dependence of the source spectrum as a simple power law. As such, it cannot approximate local minima or maxima in the spectrum. The three-term approximation can represent a function with a single local minimum and a single local maximum as well as a monotonic function and a function with a single local minimum. Thus, one would expect the three-term model to achieve substantial improvements for spectra with strong local minima or maxima. For brevity we hereafter refer to the one-term, the two-term, and the three-term models as the RS-1, RS-2, and RS-3 models, respectively.

The error histograms for the RS-1 and the RS-3 models are illustrated in Fig. 14, and the ensemble errors for the RS-1, RS-2, and the RS-3 models are presented in Table V. For comparison purposes we have also included the errors for the optimum displacement model in both the figure and the table.

A comparison of the errors between the optimum displacement model and the RS-1 model provides a measure of the importance of the variability in the form of the individual spectra across the ensemble. As seen in the table, the RS-1 model achieves a lower ensemble error than the optimum displacement model in both frequency bands—0.20 dB smaller in the low-frequency band and 0.31 dB smaller in the high-frequency band, or about 8% and 19%, respectively. While these are apparently small decreases, the error histograms of Fig. 14 indicate that these smaller ensemble errors appear as a significant increase in the percentage of ships with small rms errors. For example, in the low-frequency band, the percentage of ships with rms errors less than 2 dB

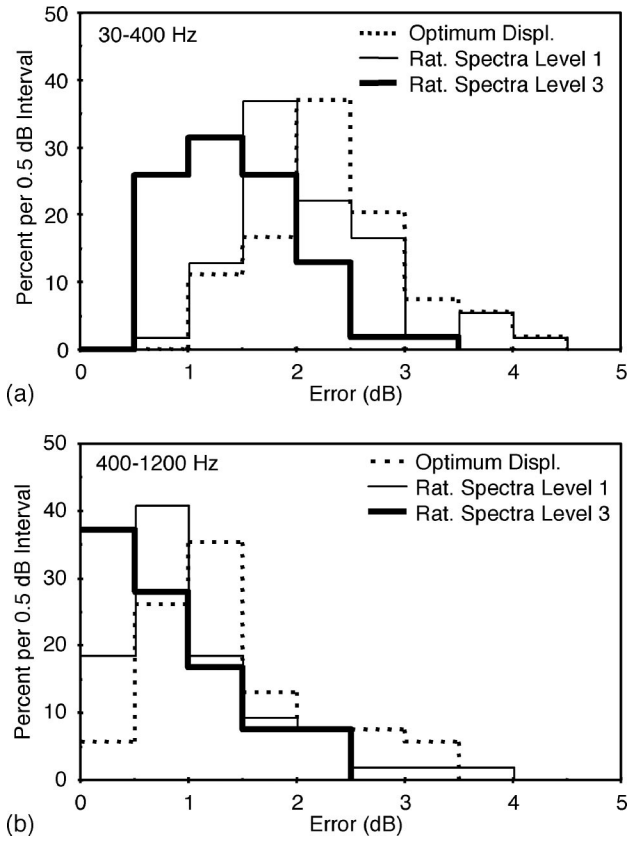


FIG. 14. Error histograms for the optimum displacement (thick dotted red lines) model, the level 1 rational spectrum model (thin solid blue lines), and the level 3 rational spectrum model (thick solid green lines): (a) 30–400-Hz band; (b) 400–1200-Hz band.

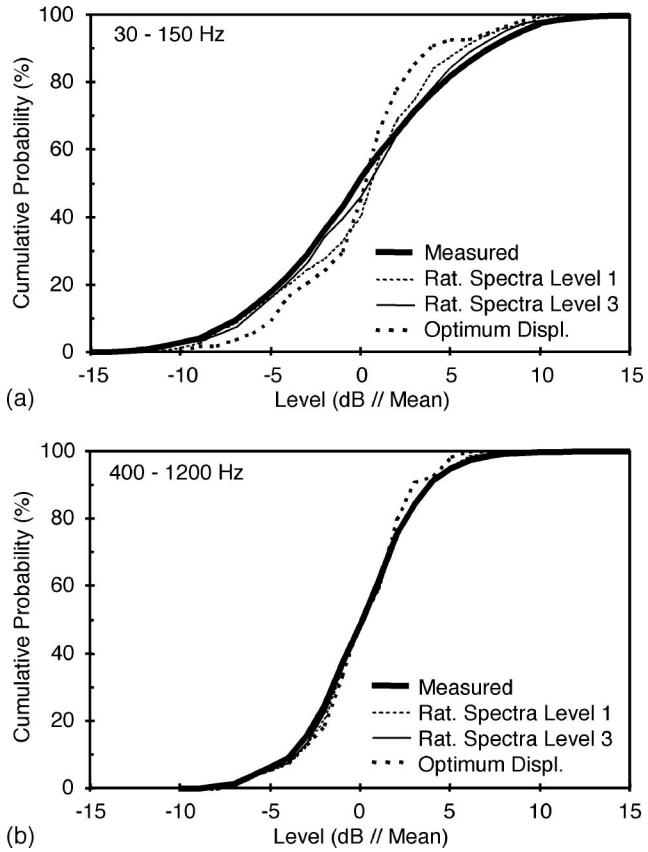


FIG. 15. Measured cumulative distribution functions of the relative source level (thick solid curve) and computed cumulative distribution functions for the optimum displacement model (thick dashed curve), the one-term rational spectrum model (thin dotted curve), and the three-term rational spectrum model (thin solid curve): (a) 30–150-Hz band; (b) 400–1200-Hz band.

increases from 27.8% for the optimum displacement model to 51.9% for the RS-1 model. Similarly, in the high-frequency band, the percentage of ships with rms errors less than 1 dB increases from 31.5% for the optimum displacement model to 59.3% for the RS-1 model. These results indicate that allowing the individual source spectra to have a ship-dependent frequency dependence results in significantly better estimates of the individual spectra.

A comparison of the errors between the RS-1 and the RS-3 models provides a measure of the importance of the form of the frequency dependence for the individual source spectra. In the low-frequency band, it is seen in the table that the ensemble error for the RS-3 model is about one-third less, 0.7 dB, than that for the RS-1 model. The error histograms of Fig. 14 indicate that this decrease in the ensemble error is accompanied by a significant increase in the percentage of ships with very small rms errors. For example, the percentage of ships with rms errors less than 1.5 dB increases from 14.8% for the RS-1 model to 57.4% for the RS-3 model. These results indicate that the higher-order rational spectrum model can significantly reduce the error for the ships with a complex frequency dependence such as exists in the low-frequency band. Finally, in the high-frequency band, both the ensemble error and the error histogram for the RS-3 are essentially the same as those for the RS-1 model, with only a modest 0.14-dB decrease. This indicates that the three-term rational spectrum approximation offers essentially no improvement over the power-law frequency dependence of the one-term rational spectrum approximation. These results are consistent with the qualitative description of the high-frequency structure of the source spectra presented in Sec. III.

TABLE V. Root-mean-square errors for the optimum displacement model and the rational spectrum models.

Models	30–400 Hz	400–1200 Hz	30–1200 Hz
Optimum displacement	2.44	1.63	2.55
Rational spectra			
One term	2.24	1.32	2.20
Two term	1.89	1.23	1.70
Three term	1.54	1.18	1.50

Figure 15 shows the cumulative distribution functions of the relative source level determined by the three generic models. Also shown are the measured CDFs from Fig. 10. As is the case for the histograms of Fig. 7 and the CDFs of Fig. 10, the upper frequency of the low-frequency band is limited to 150 Hz in recognition of the frequency dependence of the nominal standard deviation (see Fig. 6). As seen in the plots of Fig. 15, the CDFs for each of the models provide a much better approximation to the measured CDF than the historical Ross model (see Fig. 10) in both the low-frequency and the high-frequency bands. This is not surprising in light of the deficiencies in the power-law assumption and the possible

impact on the radiated noise due to changes in ship and propeller design since World War II. In the high-frequency band, the rational spectrum CDFs are almost indistinguishable from the measured CDF, and the optimum displacement CDF shows only a modest departure from the measured CDF. At least part of the discrepancy between the optimum displacement and the measured CDFs is due to sample-size effects since; as noted earlier, the relative source level for any baseline model is frequency independent, and hence, its CDF is effectively determined by only one sample per source spectrum. In the constant-sigma low-frequency band [Fig. 15(a)], there are significant discrepancies between the modeled CDFs and the measured CDF. These discrepancies are largely due to the fact that the CDFs in Fig. 15, both measured and model, are computed only over the constant-sigma portion of the low-frequency band, 30–150 Hz, whereas the model parameters are determined to describe the spectra over the full low-frequency band, 30–400 Hz. As was seen in Fig. 6, the nominal standard deviation drops from about 5.3 dB below 150 Hz to about 3.1 dB at about 400 Hz. Thus, the model parameters have been selected to describe an ensemble of spectra where the variability across the ensemble is a strong function of frequency. It is not surprising that the optimum displacement model, for which the relative spectrum level is frequency independent, is the least successful.

The CDF was computed using the model parameters derived from the data set. A rigorous comparison would need to be conducted on a separate data set using the parameter probability densities derived from this data set. A partial test can be obtained by using the current data and the parameter probability densities derived from the data set. Calculations, not shown, using the parameter probability densities derived immediately below indicated only minor differences for the optimal displacement and the RS-1 models and matched the data CDF equally as well.

We turn next to the probability densities on the parameters of the generic models. Because of the comparatively high dimensionality of the parameter space for the RS-3 model (six parameters), we limit the discussion to the optimum displacement and the RS-1 models.

The optimum displacement model represents each spectrum by the baseline spectrum shifted by a displacement, $\delta_i = S_i^b - \overline{S^b}$. Since the baseline spectrum is the average spectrum computed over the ensemble, it suffices to specify the probability density on the displacements, $p_\delta(\delta)$. Note that by construction, the ensemble average of the displacements is zero so that the probability density on the displacements have a zero mean. Further, since the optimum displacement model is a baseline spectrum model, the displacement for the *i*th spectrum is identical to the relative source level predicted by that model at all frequencies in the band. Thus, the specification of the probability density, $p_\delta(\delta)$, is equivalent to the specification of the CDF on the relative source level.

Figure 16 shows plots of the histograms of the displacements for the optimum displacement model along with normal probability densities determined by the standard deviation of the displacements. The displacement histogram for the high-frequency band [Fig. 16(b)] is similar to the relative source level histogram for the high-frequency band [Fig. 7(b)]. This is to be expected from the observation in the preceding paragraph and the good agreement between the optimum displacement model CDF and the measured CDF in the high-frequency band seen in Fig. 15. By the same token, it is not surprising that the spread of the displacement histogram for the 30–400-Hz, low-frequency band [Fig. 16(a)] is noticeably smaller than that of the relative source level histogram for the 30–150-Hz band [Fig. 7(a)]. The standard deviations of the displacement histograms in the low-frequency and the high-frequency bands are 3.71 and 2.71 dB, respectively.

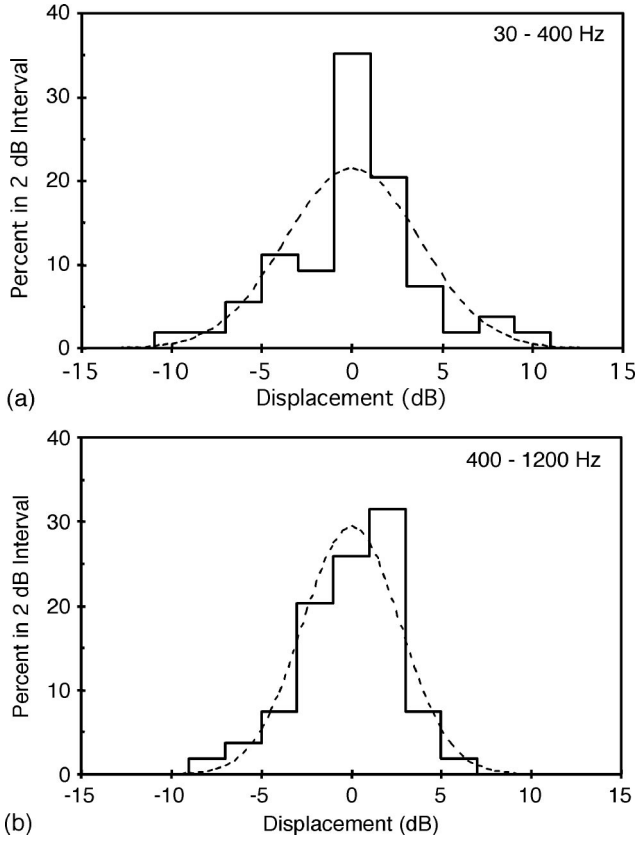


FIG. 16. Histograms of the displacement parameter for the optimum displacement model (solid lines) and an approximating normal distribution (dashed curves): (a) 30–400-Hz band; (b) 400–1200 Hz.

The RS-1 model is specified by the joint probability density on the displacement parameter, $\delta_i = S_i^b$, and the exponent coefficient, ν. This probability density can be expressed as the product, $p_{\delta,\nu}(\delta,\nu) = p_\delta(\delta) p_{\nu|\delta}(\nu)$, where $p_\delta(\delta)$ is the marginal density on the displacements and $p_{\nu|\delta}(\nu)$ is the conditional density on the exponent coefficients. As noted above, the displacement for a given spectrum is simply the average level for that spectrum. Consequently, the marginal probability density $p_\delta(\delta)$ is identical to the probability density for the displacements of the optimum displacement model, provided in Fig. 16, translated by the ensemble averaged spectrum level. Thus, it suffices to estimate the conditional density $p_{\nu|\delta}(\nu)$. Except for the special case where δ and ν are statistically independent, the number of ship spectra required to estimate $p_{\nu|\delta}(\nu)$ is comparable to the number of ship spectra required to estimate $p_{\delta,\nu}(\delta,\nu)$. This number is well in excess of the number of ships in the current analy-

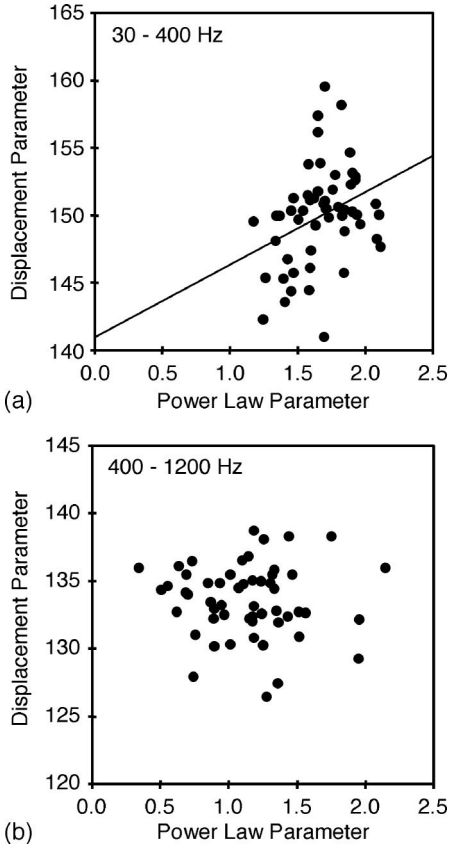


FIG. 17. Scatter plot of level 1 rational spectra parameters (displacement parameter and the power-law parameter): (a) 30–400-Hz band; (b) 400–1200 Hz. The line in (a) represents the best fit described in the text.

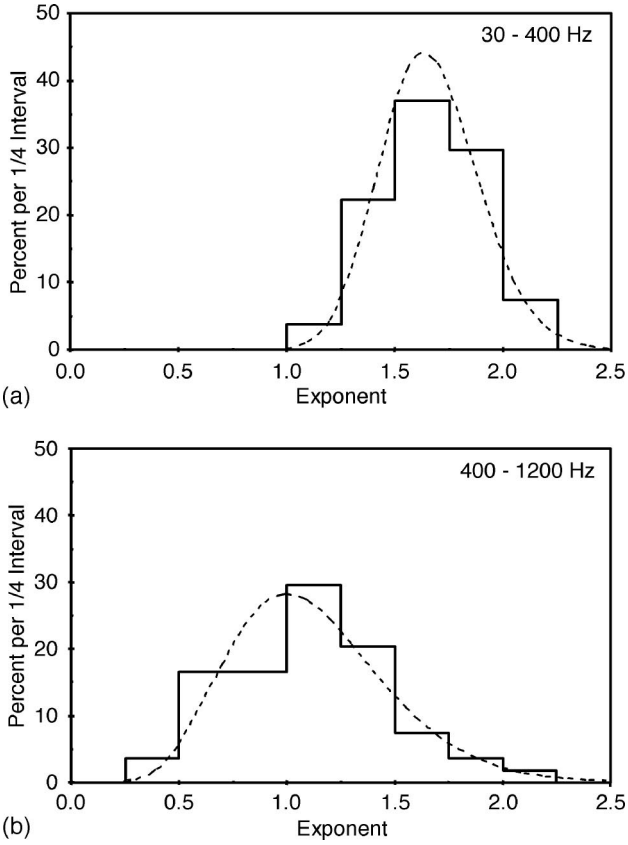


FIG. 18. Histograms of the exponent coefficient (solid lines) and an approximating gamma function (dashed curves): (a) 30–400-Hz band; (b) 400–1200 Hz.

sis. To provide a preliminary estimate of the joint density, we assume that δ and ν are close enough to being statistically independent that the conditional density $p_{\nu|\delta}(\nu)$ can be approximated by the marginal density $p_{\nu}(\nu)$. With this approximation, the joint density is then given by $p_{\delta,\nu}(\delta,\nu) = p_{\delta}(\delta)p_{\nu}(\nu)$.

To provide insight into the limitations of this approximation, we have plotted scatter diagrams of δ and ν for the two frequency bands in Fig. 17. In the scatter diagram for the low-frequency band [Fig. 17(a)], the straight-line segment represents a least-mean-squares fit to δ and ν. The correlation coefficient corresponding to this fit is 0.33. The positive slope of this line (5.3/dB) indicates that large values of the displacements tend to occur for large magnitudes of the exponent coefficients. Stated in other terms, large average spectrum levels tend to be associated with stronger (negative) slopes in the power-law approximation to the spectrum. However, the comparatively small value of the correlation coefficient indicates that this is at most a weak correspondence. Nevertheless, the statistically independent approximation neglects this correspondence. The high-frequency scatter plot [Fig. 17(b)] shows no relationship between δ and ν. Indeed, the small value of the correlation coefficient for this band (-0.056) indicates that δ and ν are essentially uncorrelated. Although it does not follow that δ and ν are statistically independent, the distribution of the points in the scatter plot of Fig. 17(b) suggests that $p_{\delta,\nu}(\delta,\nu)$ might be well approximated by the product $p_{\delta}(\delta)p_{\nu}(\nu)$.

Figure 18 shows plots of the histograms of the magnitude of the exponent coefficient for the RS-1 model along with gamma probability densities determined by the mean and standard deviation of the coefficients. An inspection of these plots indicates that in both frequency bands, the measured histograms are reasonably well approximated by the gamma probability density. The mean and standard deviation of the exponent coefficients are 1.67 and 0.3 in the low-frequency band, respectively, and 1.12 and 0.37 in the high-frequency band.

VI. SUMMARY AND DISCUSSION

This paper has presented an evaluation of the classical model for determining ensembles of the source spectra that describe ship-radiated noise and proposed an alternate model to overcome deficiencies in the classical model.

In the first section, we presented an overview of the methodology used to acquire the ship-radiated noise data and to derive the estimates of the source spectra from that data. In this section, we asserted that neither the surface monopole source representation assumed by Ross nor the finite dipole source representation assumed by Wright and Cybulski and by Scrimger and Heitmeyer are adequate to describe the sound generated by the individual ships at the short ranges where the propagation is dominated by the continuous mode spectrum. The former cannot describe the surface interaction propagation effects observed in the radiated noise spectrograms for some of the larger ships; the latter overestimates

those effects in the source spectrogram. The major difference between the two source distribution models is that the finite dipole model can introduce an L^{-2} into the source spectrum at the very low frequencies. In this paper, we have assumed that the source consists of a vertical distribution of incoherent point sources. The point sources were weighted according to a Gaussian with a standard deviation equal to one-quarter of the propeller diameter. This distribution represents a compromise between the surface monopole representation and the finite dipole representation. Clearly, none of these source distributions can represent either the azimuthal dependence of the source directionality relevant for all source ranges or the distributed nature of the source potentially relevant for very short ranges. The data required to support such a source distribution were not available to the authors.

In Sec. III, we described properties of the source spectra ensemble relevant to the analysis. In this section, we identified two partitions of the 30–1200-Hz frequency band over which the analysis was conducted. The first partition was based on a consideration of the form of the individual source spectra. In particular, it was asserted that for frequencies greater than about 400 Hz most of the individual spectra showed a simple power-law dependence in frequency with exponents that were concentrated around a mean value of about 2.0. Conversely, for frequencies less than 400 Hz many of the source spectra exhibited a more complex frequency dependence and there was a much greater variability in the spectra across the ensemble. This partition into a low-frequency band (30–400 Hz) and a high-frequency band (400–1200 Hz) provided the rationale for developing separate sets of model parameters for each band and for comparing the rms errors for the individual spectra and for the ensemble of spectra in each band. The second frequency partition formed the basis for characterizing the relative source level in terms of histograms and cumulative distribution functions. This partition was based on the frequency dependence of the ensemble standard deviation, which varied around a nominal value of about 5.3 dB for frequencies below 150 Hz and about 3.1 dB for frequencies greater than 400 Hz. Consequently, a probability density on the relative source level based on the measured histograms for each of these frequency bands should be applicable to any frequency in that band, provided that the underlying probability density is completely determined by its moments up to order two. For simplicity, we also took the high-frequency band to be 400 Hz in the estimates of the relative source level histograms. The form of these histograms for both frequency bands was reasonably well approximated by a Gaussian probability density. Finally, we presented a brief description of the spectral line properties in terms of estimates of the frequency-dependent line occurrence probabilities and the average ratios of the line energy to the total energy. It was seen that although there are a significant number of lines below about 60 Hz (line occurrence probabilities of about 0.15), these lines do not significantly alter the ensemble standard deviation, and hence do not significantly alter the mean-square-error calculations used to determine the model parameters.

In Sec. IV, we presented an evaluation of the power-law model conducted on the current ensemble of source spectra. The analysis was presented in several stages. First, we considered the power-law model with the power-law exponents proposed by Ross (sixth power of the ship speed, second power of the ship length). This model, referred to as the historical Ross model, yielded rms errors for the individual ship spectra in excess of 10 dB for more than 25 percent of the ships in both frequency bands. Furthermore, the cumulative distribution function of the relative source level predicted by the historical Ross model significantly overestimates the variability in the source level in both frequency bands. Next, we considered an optimum power-law model where the power-law exponents were derived from the current source spectrum ensemble itself. Although this model yielded an ensemble error of about half that of the historical Ross model in both frequency bands, these errors were only slightly less than those obtained by approximating each spectrum in the ensemble by the mean spectrum. Furthermore, the ensemble errors for the optimum power-law model were significantly larger than those obtained from an optimum displacement model, where the displacements to the baseline spectrum were determined to minimize the mean-square error. This discrepancy was seen to be a consequence of the fact that correlation coefficients between both the source level and the ship speed and the source level and the ship length were not significant. Finally, we recomputed these coefficients using source spectra derived from the surface monopole assumption to test the hypothesis that the deficiencies in the power-law model were due to the distributed source distribution assumed in the current analysis. No significant increase in the correlation coefficients was observed.

Finally, in Sec. V, we presented an analysis of the error performance for the rational spectrum model proposed in this paper and an approximation to the joint probability density on the parameters of this model. It was seen that the one-term rational spectrum model (RS-1) achieved slightly smaller rms errors than the optimum displacement model by accounting for the variability in the form of the spectra across the ensemble (about one-quarter of a decibel in each band). Furthermore, it was seen that increasing the number of terms in the rational spectrum model to three resulted in a significant reduction in the error for the low-frequency band (an additional 0.7 dB) and a smaller reduction of only 0.15 dB in the low-frequency band. The negligible reduction in the high-frequency band was consistent with the observation that the frequency dependence of the source spectra in this band is well approximated by a simple power-law dependence. The larger reduction in the low-frequency band was attributed to the fact that that there are a number of source spectra with a complex frequency dependence in the low-frequency band which are better handled by the three-term rational spectrum model. It was also seen that the cumulative distribution function of the relative source level predicted by the RS-3 model provided a better approximation to the measured CDF in the low-frequency band. The CDF for the optimum displacement model differed significantly from the measured CDF in the low-frequency band. Finally, the joint probability density on the displacement and the exponent coefficient for the RS-1 model was based on a statistical

independence approximation and the measured histograms of these parameters. Support for the independence approximation was presented in the form of a scatter diagram for the two parameters. This diagram showed a negligible correlation between the parameters in the high-frequency band and only a modest correlation in the low-frequency band. The measured parameter histograms suggested that the exponent parameter was well approximated by a gamma probability density and the displacement parameter was reasonably well approximated by a normal probability density.

ACKNOWLEDGMENTS

This work was supported by ONR base funding at NRL. The data were obtained under the auspices of an international data exchange agreement, Navy International Programs Office, and provided to us by the MITRE Corporation. J. Scott Martin of MITRE identified the ships in Lloyd's ship registry and provided us with the results. The authors would also like to thank Dr. Orest Diachok (NRL) for his suggestions and Paul Arveson (NSWC-Carderock) for providing us with a draft of his paper. Finally, we would like to thank the reviewer for the insightful comments and helpful suggestions made.

APPENDIX: SUPPLEMENTAL MODEL DETAILS

1. The baseline model solutions

The solutions for the minimum mean-square error and the corresponding optimum parameters for the three baseline spectrum models can be obtained by representing the ensemble spectra, the predicted spectra, and the baseline spectrum as vectors, $\mathbf{s}_i$, $\hat{\mathbf{s}}_i$, and $\mathbf{s}_o$ and then representing the model by a matrix equation of the form, $\hat{\mathbf{s}}_i = \mathbf{s}_o + \mathbf{A}_i \mathbf{q}_i$, where $\mathbf{q}_i$ is the vector of unknown constants and $\mathbf{A}_i$ is the matrix that specifies the relationship of those constants to the spectrum estimate. For the source-specific, optimum power-law model [Eq. (3)], $\mathbf{q}_i$ is the P-dimensional vector with elements $q_p = c_p$ and $\mathbf{A}_i$ is the N_F by P matrix with elements $a_{j,p} = \alpha_p$. For the generic optimum displacement model [Eq. (2)], $\mathbf{q}_i$ is the N-dimensional vector with elements $q_p = \delta_p$ and $\mathbf{A}_i$ is the N_F by N matrix whose entries are all zero except for the ith column, whose entries are all ones. For the degenerate mean spectrum model, $\mathbf{q}_i$ and $\mathbf{A}_i$ are zero. The solutions to the linear least-squares problem for these models are presented in Sec. IV C.

2. Subband mean-square-error relationships

Let B_1 and B_2 be the bandwidths of a partition of a frequency band of width $B = B_1 + B_2$. Further, let $\varepsilon^2_{B''}(B')$ be the minimum mean-square error computed over a band of width B' using model parameters derived on a band of width B''. It follows from Eq. (1a) that

$$\varepsilon_B^2(B) = \left(\frac{B_1}{B}\right)\varepsilon_B^2(B_1) + \left[1 - \left(\frac{B_1}{B}\right)\right]\varepsilon_B^2(B_2). \tag{A1}$$

Furthermore, $\varepsilon^2_{B_i}(B_i) \leqslant \varepsilon^2_B(B_i)$, since the parameters for the subband, B_i, were derived to minimize $\varepsilon^2_{B_i}(B_i)$, whereas, the parameters for total band, B, were derived to minimize $\varepsilon_B^2(B)$ and not $\varepsilon_B^2(B_i)$. It follows from these inequalities and Eq. (A1) that

$$\varepsilon_o^2(B) = \left(\frac{B_1}{B}\right)\varepsilon^2_{B_1}(B_1) + \left[1 - \left(\frac{B_1}{B}\right)\right]\varepsilon^2_{B_2}(B_2) \leqslant \varepsilon_B^2(B), \tag{A2}$$

with equality if and only if $\varepsilon^2_{B_1}(B_1) = \varepsilon^2_B(B_1)$ and $\varepsilon^2_{B_2}(B_2) = \varepsilon^2_B(B_2)$, or equivalently, if and only if there is no improvement in the mean-square error obtained through subband partitioning. Thus, the difference between $\varepsilon_B^2(B)$ and $\varepsilon_o^2(B)$ is a measure of that improvement. We also note that although Eq. (8) implies that $\min[\varepsilon_B^2(B_1), \varepsilon_B^2(B_2)] \leqslant \varepsilon_B^2(B) \leqslant \max[\varepsilon_B^2(B_1), \varepsilon_B^2(B_2)]$ for the subband models evaluated on the full band, we can only conclude that $\min[\varepsilon^2_{B_1}(B_1), \varepsilon^2_{B_2}(B_2)] \leqslant \varepsilon_B^2(B)$ for the subband models evaluated on the subbands themselves. Finally, we note that $\varepsilon_B^2(B) < \varepsilon^2_{B_i}(B_i)$ does not imply that it is better to use the full-band model when approximating a spectrum the subband, B_i, since the inequality $\varepsilon^2_{B_i}(B_i) \leqslant \varepsilon^2_B(B_i)$ must always hold.

3. The rational spectrum model solutions to the linear problem

The rational spectra results were obtained using the linear theory to solve for the linear parameters in terms of the unknown frequency parameters and then minimizing the mean-square error over the frequency parameters. According to the linear theory, the displacement parameter is simply the average spectrum level, $\delta_i = S_i^b$ which depends neither on the unknown frequency parameters nor the order of the model. Furthermore, the linear theory completely determines the order-one model, since the RS-1 model does not depend on the frequency parameters. For the RS-1 model, the exponent coefficient is given by

$$\tilde{\nu}_1 = \frac{\gamma_1}{r_{1,1}}, \tag{A3}$$

where $\gamma_1 = \langle (S_i(f) - S_i^b)(\beta_1(f) - \overline{\beta}_1(f)) \rangle_F$ and $r_{1,1} = \langle (\beta_1(f) - \beta_1(f))^2 \rangle_F$. For the higher-order models, the expressions for the exponent coefficients and the mean-square error take the usual form, with γ_p and $r_{p,n}$ representing frequency correlations of the data with the approximating functions and the approximating functions with themselves, respectively.

[1] R. J. Urick, *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York, 1983), Chap. 7.

[2] I. A. Aleksandrov, "Physical nature of the 'rotation noise' of ship propellers in the presence of cavitation," Sov. Phys. Acoust. **8**(1), 23–28 (1962).

[3] L. M. Gray and D. S. Greeley, "Source level model for propeller blade rate radiation for the world's merchant fleet," J. Acoust. Soc. Am. **67**, 516–522 (1980).

[4] D. Ross, *Mechanics of Underwater Noise* (Pergamon, New York, 1976).

[5] W. K. Blake, *Aero-Hydroacoustics for Ships* (David W. Taylor Naval Ship Research and Development Center, Bethesda, MD, DTNSRDC-84-7010, June, 1984), Vols. I and II.

[6] J. P. Breslin and P. Anderson, *Hydrodynamics of Ship Propellers* (Cambridge U.P., Cambridge, 1994).

[7] J. P. Breslin and P. Anderson, *Hydrodynamics of Ship Propellers* (Cambridge U.P., Cambridge, 1994), Chaps. 20–22, and references therein.

[8] Y.-Z. Kehr, C.-Y. Hsin, and Y.-C. Sun, "Calculations of Pressure Fluctua-

tions on the Ship Hull Induced by Intermittently Cavitating Propellers,'' in *Twenty-First Symposium on Naval Hydrodynamics* (National Academy, Washington, DC, 1997), pp. 882–897.
[9] A. A. Pudovkin, ''Noise emission by the cavitation zone of a marine propeller,'' Sov. Phys. Acoust. **22**(2), 151–154 (1976).
[10] Lloyd's Registry of Shipping. Ship Particulars Database, Lloyd's Maritime Information Services, Stamford, CT.
[11] H. W. Kutschale, ''Rapid Computation by Wave Theory of Propagation Loss in the Arctic Ocean,'' Lamont-Doherty Geological Observatory of Columbia University, Technical Report No. 8, CU-8-73 (1973).
[12] F. R. DiNapoli and R. L. Deavenport, ''Theoretical and numerical Green's function solution in a plane multilayered medium,'' J. Acoust. Soc. Am. **67**, 92–105 (1980).
[13] L. M. Gray and D. S. Greeley, ''Source level model for propeller blade rate radiation for the world's merchant fleet,'' J. Acoust. Soc. Am. **67**, 516–522 (1980).
[14] Gray and Greeley's formula has been modified to account for the fact that we have not assumed that the bottom of the propeller extends to the bottom of the ship. Thus, where they used, Draft—0.85 Diameter, we have used, 0.875 Draft—0.85 Diameter, further adjusted for the reported loading.
[15] Time in the spectrogram progresses from positive beam aspect angles to stern aspect angles. Figures 1–4 span the angular interval 58° to −56° and are thus not expected to be effected by hull shielding (forward) or propagation wake absorption (stern).
[16] H. Takahashi and T. Ueda, ''An experimental investigation into the effect of cavitation on fluctuating pressures around a marine propeller,'' in *Proceedings of the Twelfth International Towing Tank Conference* (Consiglio Nazionale dele Ricerche, Rome, 1969), pp. 315–317.
[17] This distribution was cut off at the surface and the keel.
[18] P. T. Arveson and D. J. Vendittis, ''Radiated noise characteristics of a modern cargo ship,'' J. Acoust. Soc. Am. **107**, 118–129 (2000).
[19] The maximum aspect angle for the 2-minute period used for each ship varied from 14° to 54° with the majority of ships less than 25°.
[20] Unless otherwise stated, all averages discussed in this paper are conducted on a decibel scale, rather than a power scale.
[21] J. J. Wolcin, ''On the Statistical Properties of Noise Background Equalization Schemes,'' NUSC Tech. Memo. No. 781159, Naval Underwater Systems Center, New London, CT (31 July 1978).
[22] W. A. Struzinski and E. D. Lowe, ''A performance comparison of four background normalization schemes proposed for signal detection systems,'' J. Acoust. Soc. Am. **76**, 1738–42 (1984).
[23] P. Scrimger and R. M. Heitmeyer, ''Acoustic source-level measurements for a variety of merchant ships,'' J. Acoust. Soc. Am. **89**, 691–699 (1991).
[24] D. Ross, in Ref. 4, Chap. 8.
[25] Ross, Donald, and F. F. Alvarez, unpublished work (1964).
[26] In the usual expression of this model, the factor of 10 is absorbed into the coefficients.
[27] R. M. Hamson and R. A. Wagstaff, ''An Ambient-Noise Model That Includes Coherent Hydrophone Summation for Sonar System Performance in Shallow Water,'' SACLANT ASW Research Centre Report, SR-70 (15 June 1983).
[28] J. E. Breeding, Jr., L. A. Pflug, M. Bradley, M. H. Walrod, and W. McBride, ''Research Ambient Noise Directionality (RANDI) 3.1 Physics Description,'' NRL Report, NRL/FR/7176-95-9628, Naval Research Laboratory, Stennis Space Center, MS (August 1996).
[29] The difference between the two at 1200 Hz is only 0.2 dB from a common reference point at 400 Hz.
[30] E. B. Wright and J. Cybulski, ''Low-Frequency Acoustic Source Levels of Large Merchant Ships,'' NRL Report 8677, Naval Research Laboratory, Washington, DC (14 March 1983).

Identification of complex stiffness tensor from waveform reconstruction

N. Leymarie[a)]
Université Bordeaux 1, UMR CNRS n°5469, Laboratoire de Mécanique Physique, École MATMECA, 351, Cours de la Libération, 33405 Talence Cedex, France

C. Aristégui[b)]
INSERM Unité 556, 151, Cours Albert Thomas, 69424 Lyon Cedex 03, France and
Université Claude Bernard Lyon 1, 43, bd du 11 Novembre 1918, 69622 Villeurbanne Cedex, France

B. Audoin and S. Baste
Université Bordeaux 1, UMR CNRS n°5469, Laboratoire de Mécanique Physique, École MATMECA, 351, Cours de la Libération, 33405 Talence Cedex, France

(Received 22 May 2001; revised 2 November 2001; accepted 20 November 2001)

An inverse method is proposed in order to determine the viscoelastic properties of composite-material plates from the plane-wave transmitted acoustic field. Analytical formulations of both the plate transmission coefficient and its first and second derivatives are established, and included in a two-step inversion scheme. Two objective functions to be minimized are then designed by considering the well-known maximum-likelihood principle and by using an analytic signal formulation. Through these innovative objective functions, the robustness of the inversion process against high level of noise in waveforms is improved and the method can be applied to a very thin specimen. The suitability of the inversion process for viscoelastic property identification is demonstrated using simulated data for composite materials with different anisotropy and damping degrees. A study of the effect of the rheologic model choice on the elastic property identification emphasizes the relevance of using a phenomenological description considering viscosity. Experimental characterizations show then the good reliability of the proposed approach. Difficulties arise experimentally for particular anisotropic media. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1451072]

PACS numbers: 43.35.Cg [SGK]

I. INTRODUCTION

Nondestructive evaluation of material viscoelastic properties opens promising prospects for material knowledge and structure behavior. In this paper, the notion of viscoelasticity acoustically related to wave energy loss in solids covers both the viscosity and all other damping mechanisms[1] (scattering, frictional damping, structural damping,...).

Well-established characterization ultrasonic methods use bulk waves obliquely incident from an outer medium (water or air) onto a plate. They are usually based on the modeling of either plate[2,3] or bulk[4–9] waves. Thus, the associated ultrasonic devices are, respectively, designed to measure either the leakage of the waves propagating within the plates into the surrounding medium, or the waves propagating through the plate thickness. The main disadvantage of the first mentioned approach using leaky Lamb waves lies in the difficulties of the inversion processes for identifying the material viscosity properties. Studies using bulk waves propagating through plates have shown that the wave speed measurements coupled with an energy (or amplitude) criterion allow the determination of the complex stiffnesses by means of simple inversion processes. These complex constants are usually introduced to model the viscoelastic mechanisms. The tensor real part is related to the elasticity constants while the imaginary part is related to the viscosity constants. The main advantages of this well-known technique concern the simplicity of the experimental setup design and the rapidity of identification from a suitable set of waveforms. The elasticity constants are extracted from wave speed measurements, and the viscosity parameters are then determined from both the amplitude measurements and the elastic properties already recovered. However, questions still remain regarding the reliability of these inverse methods for some materials or samples. Of particular concerns is the time resolution of overlapped waves when dispersive effects are present or when the wavelength-to-thickness ratio is large.[10,11] A solution consists of increasing the working frequency while staying below the upper frequency boundary associated with the homogeneous medium assumption. Due to the difficulties to meet at the same time these two frequency requirements, the signal processing limits can lead to a characterization failure.[12]

In order to extend the limits of these methods, inversion methods using the plate transfer function calculation have been suggested. First, mathematical approaches used an objective function defined from the shift between the Euclidian norms of the measured and predicted moduli of the plate

[a)]Electronic mail: n.leymarie@lmp.u-bordeaux.fr
[b)]Electronic mail: christophe.aristegui@univ-lyonl.fr

reflection[13] or transmission[14] coefficient. Recently, an improvement has been proposed[15] by considering the signal spectrum modulus instead of the transmission coefficient and introducing in this way the "frequency weight of the source signal." However, because these techniques do not take into account explicitly the signal phase features, characterization difficulties arise both with the decreasing plate thickness and with increasing the medium dispersion or noise in signals. Another approach deals with the minimization of the Euclidian norm of the shift between the measured and predicted moduli of the plate reflection.[16] This method considers explicitly the phase spectrum, but its formulation still induces a non-negligible sensitivity to noise in data.

In this paper we outline a new inverse method with a very reduced noise sensitivity that is suited to both very thin plates and strongly dispersive media. As for the previous works, the proposed method is based on the use of the plate transfer function in the inverse problem. However, a first improvement concerns a mathematical development in which the triclinic plate transfer function in transmission and its first and second derivatives with respect to the complex stiffnesses are analytically established. The second improvement lies in the optimization strategy defined by introducing two innovative objective functions inferred from the maximum-likelihood principle[17,18] applied to analytic signals. The reliability and the accuracy of the inversion method for the complex stiffness reconstruction are investigated first by means of numerical simulations. For this purpose, several tests are carried out on calculated signal sets for different plate thicknesses and material dampings. Next, attention is focused on the effect of an erroneous choice of the rheologic model on the elastic property recovery. Finally, two experiments are performed on one isotropic and one anisotropic thin plate in order to examine the suitability of the inverse process for the experimental identification of viscoelastic properties.

II. TRICLINIC PLATE TRANSFER FUNCTION IN TRANSMISSION

In this section we describe the different steps to establish the transfer function H of a viscoelastic triclinic plate immersed in a fluid, its gradient vector ∇H and its Hessian matrix $\nabla^2 H$ with respect to the complex stiffness components. These last two developments lead to an exact formulation of the iterative scheme used in the minimization processes presented further in Sec. III. Establishing the transfer function of a plate requires the study of the problem of wave reflection and refraction at liquid–anisotropic solid–liquid interfaces for a homogeneous, infinite and harmonic plane wave.

A. Wave propagation in triclinic plates

Let $\mathbf{k}$ be the complex wave number, or wave bivector, and $\mathbf{S}$ the slowness bivector related to $\mathbf{k}$ by $\mathbf{k}=\omega\mathbf{S}$, where ω is the angular frequency. The use of slownesses removes the frequency dependence in numerous calculations. According to the Snell–Descartes laws, the projections of the slowness bivector on the **2** and **3** axes, denoted S_2 and S_3, respectively, are identical in each medium for every partial waves m generated in the plate from the known incident mode I (Fig. 1). Then one has

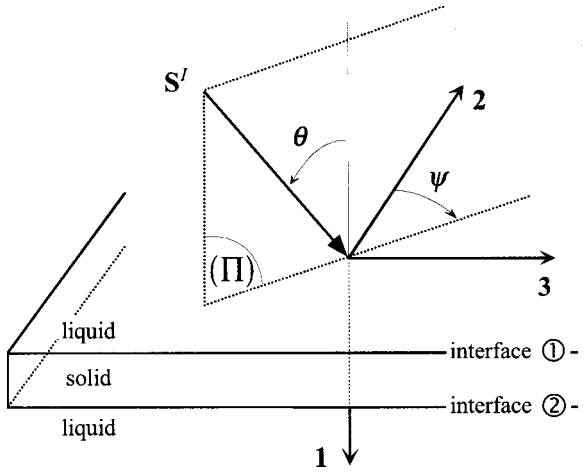


FIG. 1. Coordinate system (**1**,**2**,**3**) associated with the incident plane wave at the liquid–solid–liquid interfaces. The azimuthal propagation plane (Π) is defined by the angle ψ, and the angle between the normal to the interface and the direction of the propagation of the incident wave I corresponds to the angle of incidence θ.

$$S_2^I = S_2^{(m)} \quad \text{and} \quad S_3^I = S_3^{(m)}, \tag{1}$$

where the superscript in italics represents the incident wave I or the partial wave m.

As usual, the slowness component $S_1^{(m)}$ along the **1** axis and the polarization vector $\mathbf{P}^{(m)}$ both associated with the partial wave m in the solid medium are deduced from the Christoffel equation,[19]

$$(\mathbf{S.C.S}-\rho\mathbf{I}_3)\cdot\mathbf{P}=\mathbf{0}, \tag{2}$$

where $\mathbf{C}$ is the fourth-rank complex stiffness tensor of the solid, ρ its density, and $\mathbf{I}_n$ the $(n\times n)$ unit matrix. Using the abbreviated subscript notation, the tensor $\mathbf{C}$ is expressed by the 21 independent complex components of a symmetric (6×6) matrix (C_{ij}).[19] The slowness components $S_1^{(m)}$ are the roots of the following sixth-degree polynomial $\mathcal{P}$,

$$\mathcal{P}(S_1;C_{ij},\rho,S_2^I,S_3^I)=\det(\mathbf{S.C.S}-\rho\mathbf{I}_3), \tag{3}$$

expressed in terms of the slowness component S_1. For each slowness solution $S_1^{(m)}$, the associated normalized polarization vector $\mathbf{P}^{(m)}$ is then deduced by solving Eq. (2). Thus, the six pairs of solutions $(S_1^{(m)},\mathbf{P}^{(m)})$ define the six partial waves that propagate in the plate. The partial waves propagating forward through the plate thickness along the direction **1** are noted QL^+, $QT1^+$, and $QT2^+$, whereas the three others propagating backward are denoted QL^-, $QT1^-$, and $QT2^-$, as shown in Fig. 2. The superscripts QL, QT1, and QT2 indicate the features of the quasilongitudinal and of the two quasitransverse waves, respectively. The waves generated in the fluid, denoted by the superscript R and T, correspond to the reflected and transmitted plane waves, respectively.

The formulation used for the displacement field associated with the partial wave m at any space point $\mathbf{M}$ and at time t is defined by

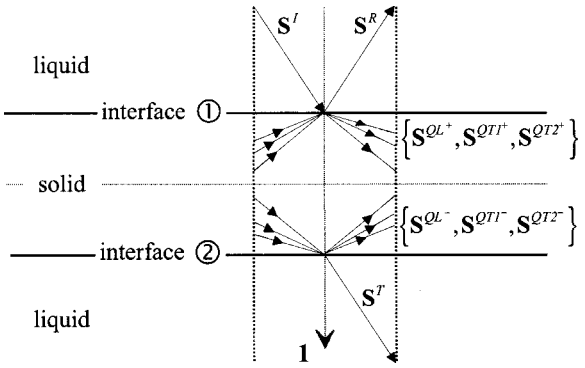


FIG. 2. Waves propagating in the azimuthal plane (Π), defined in Fig. 1, at the liquid–anisotropic solid–liquid interfaces.

$$\mathbf{u}^{(m)} = \mathrm{Re}\{A^{(m)}\mathbf{P}^{(m)} \exp[i\omega(t-\mathbf{S}^{(m)}\cdot\mathbf{M})]\}, \tag{4}$$

where $A^{(m)}$ is the complex amplitude.

B. Formulation of the transmission coefficient for a triclinic plate immersed in a fluid

The anisotropic plate transmission coefficient is classically determined from continuity of normal displacements and mechanical tension vectors at each interface. The calculation method used has previously been presented by Deschamps and Hosten[20] for an orthotropic plate. Although the formulations of the plate transfer functions in both transmission and reflection are accessible by this method, we focus only on the transmission coefficient. By defining the (1×4) ''continuity vector'' $\mathbf{l}^{(m)}$ of the partial wave m such as

$$\mathrm{Re}\{A^{(m)} \exp[i\omega(t-\mathbf{S}^{(m)}\cdot\mathbf{M})]\times\mathbf{l}^{(m)}\} = (u_1^{(m)}, \sigma_{11}^{(m)}, \sigma_{12}^{(m)}, \sigma_{13}^{(m)})^t, \tag{5}$$

with $\sigma^{(m)}$ the stress tensor associated with the displacement field $\mathbf{u}^{(m)}$, the four equations of continuity at each plate interface can be replaced by the following (8×8) linear system:

$$\mathcal{A}\cdot\mathbf{X} = \mathbf{b}, \tag{6}$$

where the (1×8) vector $\mathbf{X}$ is defined by

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_① \\ \mathbf{X}_② \end{pmatrix}, \tag{7}$$

with $\mathbf{X}_① = (A^R, A^{\mathrm{QL}^+}, A^{\mathrm{QT1}^+}, A^{\mathrm{QT2}^+})^t$ and $\mathbf{X}_② = (A^T, A^{\mathrm{QL}^-}, A^{\mathrm{QT1}^-}, A^{\mathrm{QT2}^-})^t$. Note that the subscripts ① and ② refer to the plate interfaces, as shown in Fig. 1. In Eq. (6), $\mathbf{b}$ represents the (1×8) known source vector, such that

$$\mathbf{b} = \begin{pmatrix} A^I e^{i\alpha S_1^I}\mathbf{l}^I \\ \mathbf{0} \end{pmatrix}, \quad \text{with} \quad \alpha = \omega\frac{d}{2}, \tag{8}$$

where d is the plate thickness and $\mathcal{A}$ denotes the continuity matrix corresponding to the continuity equations on both plate interfaces. This (8×8) matrix $\mathcal{A}$ can be rewritten in terms of (4×4) block matrices, such as

$$\begin{aligned}\mathcal{A} &= \begin{pmatrix} \mathbf{L}_①\Phi_① & \mathbf{L}_②\mathbf{U}\Phi_②^{-1} \\ \mathbf{L}_①\mathbf{U}\Phi_①^{-1} & \mathbf{L}_②\Phi_② \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{M}_① & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_② \end{pmatrix} + \begin{pmatrix} \mathbf{0} & \mathbf{B}_① \\ \mathbf{B}_② & \mathbf{0} \end{pmatrix} \\ &= \mathcal{M} + \mathcal{B}, \end{aligned} \tag{9}$$

where

$$\mathbf{L}_① = (-\mathbf{l}^R, \mathbf{l}^{\mathrm{QL}^+}, \mathbf{l}^{\mathrm{QT1}^+}, \mathbf{l}^{\mathrm{QT2}^+}),$$

$$\mathbf{L}_② = (-\mathbf{l}^T, \mathbf{l}^{\mathrm{QL}^-}, \mathbf{l}^{\mathrm{QT1}^-}, \mathbf{l}^{\mathrm{QT2}^-}),$$

$$\Phi_① = \mathrm{diag}(e^{-i\alpha S_1^I}, e^{i\alpha S_1^{\mathrm{QL}^+}}, e^{i\alpha S_1^{\mathrm{QT1}^+}}, e^{i\alpha S_1^{\mathrm{QT2}^+}}),$$

$$\Phi_② = \mathrm{diag}(e^{-i\alpha S_1^I}, e^{i\alpha S_1^{\mathrm{QL}^-}}, e^{i\alpha S_1^{\mathrm{QT1}^-}}, e^{i\alpha S_1^{\mathrm{QT2}^-}}),$$

$$\mathbf{U} = \mathrm{diag}(0,1,1,1).$$

It is important to note that the expressions (8) and (9) have been established by fixing the phase origin at the plate middle corresponding to the zero abscissa along the direction **1**. In addition, the term $\exp[i\omega(t-S_2^{(m)}x_2-S_3^{(m)}x_3)]$, common to all the partial waves, has been factorized and omitted. The advantage of this block decomposition is to separate both interfaces in the calculation. Then rewriting Eq. (9) as

$$\mathcal{A} = \mathcal{M}(\mathbf{I}_8 - \mathcal{R}), \tag{10}$$

allows us to define the reflection–refraction matrix $\mathcal{R}$ as

$$\mathcal{R} = \begin{pmatrix} \mathbf{0} & \mathbf{R}_① \\ \mathbf{R}_② & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & -\mathbf{M}_①^{-1}\mathbf{B}_① \\ -\mathbf{M}_②^{-1}\mathbf{B}_② & \mathbf{0} \end{pmatrix}. \tag{11}$$

Finally, the first component of the vector $\mathbf{X}_②$, noted $(\mathbf{X}_②)_1$, corresponding to the transmission coefficient A^T, Eq. (7), is determined by solving Eq. (6), where the substitution (10) for $\mathcal{A}$ has been integrated. Hence the analytical formulation of the transfer function H in transmission for a triclinic plate immersed in a fluid is established as

$$\begin{aligned} H &= A^T/A^I \\ &= (\mathbf{X}_②)_1/A^I \\ &= e^{i\alpha S_1^I}\left(\mathbf{R}_② \sum_{n=0}^{N_e} (\mathbf{R}_①\mathbf{R}_②)^n \mathbf{M}_①^{-1}\cdot\mathbf{l}^1\right)_1, \end{aligned} \tag{12}$$

by considering the simple transmitted wave corresponding to $n=0$, and the N_e echoes propagating through the plate (transient assumption), or as

$$\begin{aligned} H &= A^T/A^I \\ &= (\mathbf{X}_②)_1/A^I \\ &= e^{i\alpha S_1^I}\left(\mathbf{R}_②(\mathbf{I}_4 - \mathbf{R}_①\mathbf{R}_②)^{-1}\mathbf{M}_①^{-1}\cdot\mathbf{l}^I\right)_1, \end{aligned} \tag{13}$$

by considering the simple transmitted wave and all the multireflected echoes in the plate ($N_e=\infty$, harmonic assumption).[21] Note that the transfer function depends implicitly on the complex stiffnesses by way of both the slownesses $S_1^{(m)}$ and their associated polarization vectors $\mathbf{P}^{(m)}$. Hence, the matrix formulations established in Eqs. (12) and

(13) make easier the use of the plate transfer function in the further calculations.

By assuming plane wave propagation, the expressions (12) and (13) can then be used to simulate the temporal waveform y_n by

$$y_n = \mathcal{F}^{-1}[e^{-i\alpha S_1^I} E(\omega) H_n(\omega;\rho,d,C_{ij})], \tag{14}$$

where E is the spectrum of the experimental reference signal e propagating in water without the plate, H_n is the plate transfer function and $\mathcal{F}^{-1}$ is the inverse Fourier transform operator. Note that the subscript n refers to a given couple of angles (θ_n, ψ_n).

C. Derivatives of the transfer function

The analytical expression of the transmission coefficient of an anisotropic plate has been established as a function of frequency, density, thickness, incident, and azimuthal angles and obviously complex stiffnesses. Let us now describe the analytical development of the first and second derivatives of the transfer function H, Eq. (12) or (13), with respect to the complex stiffnesses. The function H is expressed in terms of the two matrixes $\mathcal{R}$ and $\mathcal{A}$, which depend implicitly on the slowness components $S_1^{(m)}$ and the stiffness components C_{ij}, via the polarization vectors $\mathbf{P}^{(m)}$, displacements and stresses. The problem is then equivalent to assess the $S_1^{(m)}$ component variations with respect to each component C_{ij}, knowing that Eq. (3) has to be verified. Thus, applying the implicit function theorem to Eq. (3) yields

$$\frac{\partial \varphi^{(m)}}{\partial C_{ij}} = -\frac{\partial \mathcal{P}/\partial C_{ij}}{\partial \mathcal{P}/\partial S_1} \tag{15}$$

and

$$\frac{\partial^2 \varphi^{(m)}}{\partial C_{ij}\,\partial C_{kl}} = -\left(\frac{\partial^2 \mathcal{P}}{\partial C_{ij}\,\partial C_{kl}} + \frac{\partial \varphi^{(m)}}{\partial C_{ij}} \frac{\partial^2 \mathcal{P}}{\partial S_1\,\partial C_{kl}} + \frac{\partial \varphi^{(m)}}{\partial C_{kl}} \frac{\partial^2 \mathcal{P}}{\partial C_{ij}\,\partial S_1} + \frac{\partial \varphi^{(m)}}{\partial C_{ij}} \frac{\partial \varphi^{(m)}}{\partial C_{kl}} \frac{\partial^2 \mathcal{P}}{\partial S_1^2} \right) \Bigg/ \frac{\partial \mathcal{P}}{\partial S_1}, \tag{16}$$

where $\varphi^{(m)}$ is the local function defined in the vicinity of $S_1^{(m)}$ such as $S_1 = \varphi^{(m)}(C_{ij})$. Then for any polarization, displacement, or stress component associated with a partial wave m and denoted by $q^{(m)}$, their first and second derivatives with respect to all the complex stiffnesses can be easily deduced from

$$dq^{(m)} = \left(\frac{\partial q^{(m)}}{\partial C_{ij}} + \frac{\partial \varphi^{(m)}}{\partial C_{ij}} \frac{\partial q^{(m)}}{\partial S_1} \right) dC_{ij} \tag{17}$$

and

$$d^2 q^{(m)} = \left(\frac{\partial^2 q^{(m)}}{\partial C_{ij}\,\partial C_{kl}} + \frac{\partial \varphi^{(m)}}{\partial C_{ij}} \frac{\partial^2 q^{(m)}}{\partial C_{kl}\,\partial S_1} + \frac{\partial \varphi^{(m)}}{\partial C_{kl}} \frac{\partial^2 q^{(m)}}{\partial C_{ij}\,\partial S_1} + \frac{\partial \varphi^{(m)}}{\partial C_{ij}} \frac{\partial \varphi^{(m)}}{\partial C_{kl}} \frac{\partial^2 q^{(m)}}{\partial S_1^2} + \frac{\partial q^{(m)}}{\partial S_1} \frac{\partial^2 \varphi^{(m)}}{\partial C_{ij}\,\partial C_{kl}} \right) dC_{ij}\,dC_{kl}. \tag{18}$$

Therefore, an application of the differentiation rules on the algebraic linear expressions (12) and (13) allows one to establish the analytical forms of both the gradient vector ∇H and the Hessian matrix $\nabla^2 H$ as a function of the components C_{ij}. The vector ∇H and the matrix $\nabla^2 H$ will be introduced in the optimization process described in Sec. III B.

III. INVERSE PROBLEM

The establishment of the suggested method has been guided by our desire to take into account explicitly all the features of the transmitted waveforms. For this purpose, two objective functions based on a physical approach are constructed by considering information on both the signal phase and the signal energy.

Our work is directly inferred from the maximum-likelihood principle and the expression of ambiguity functions.[17,18] These tools were first used in radar and sonar studies.[22] The use of the analytic formulation of signals, introduced by Gabor[23] and Ville,[24] and developed during the last decade for signal processing and time–frequency analysis,[25] appears to be suited to this kind of assessment problem.

A. Objective function based on the maximum-likelihood principle

A well-known investigation tool of ambiguity functions is the scalar product defined in the square integrable function space L_2. The scalar product between the analytic signals $y_0(t)$ and $y_1(t,K)$ with finite energy can then be written as follows:

$$\langle y_1(\cdot,K), y_0 \rangle = \int_{-\infty}^{+\infty} y_1(t,K)\overline{y_0(t)}\,dt, \tag{19}$$

where K is an unknown parameter and $\overline{y_0}$ denotes complex conjugation of the analytic signal y_0. Note that the Parseval identity $\langle y_1, y_2 \rangle = \langle Y_1, Y_2 \rangle$, with Y_i the Fourier transform of y_i $(i=1,2)$, allows the calculation of the scalar product in either the time or the frequency domain. In contrast with the "real scalar product" between two real signals, we will call "analytic scalar product" the scalar product between the two analytic signals deduced from these real signals. The analytic scalar product provides more information than the real scalar product.

By definition, the spectrum of an analytic signal deduced from a real signal contains no negative-frequency components. Hence, for the analytic scalar product calculation in the frequency domain, only the positive frequency range, denoted D_f, is considered. Thus, the frequency formulation based on the Parseval identity reduces the computation time of the scalar product calculation.

Investigations are then carried out in order to establish a minimization process based on the maximum-likelihood principle between the experimental and predicted signals with C_{ij} as parameters. Let us introduce three functions useful for a clear presentation of the objective functions. The signal interaction energy ϵ_n^{int}, and the predicted and experimental signal energies, ϵ_n^{pre} and ϵ_n^{exp} are defined, respectively, as

$$\epsilon_n^{\text{int}}(C_{ij}) = \langle y_n(\cdot, C_{ij}), y_n^{\text{exp}} \rangle = \int_{D_f} Y_n(\omega, C_{ij}) \overline{Y_n^{\text{exp}}(\omega)} d\omega,$$

$$\epsilon_n^{\text{pre}}(C_{ij}) = \langle y_n(\cdot, C_{ij}), y_n(\cdot, C_{ij}) \rangle = \int_{D_f} |Y_n(\omega, C_{ij})|^2 \, d\omega, \tag{20}$$

$$\epsilon_n^{\text{exp}} = \langle y_n^{\text{exp}}, y_n^{\text{exp}} \rangle = \int_{D_f} |Y_n^{\text{exp}}(\omega)|^2 \, d\omega,$$

where $y_n^{\text{exp}}(t)$ and $y_n(t, C_{ij})$ are the experimental and predicted analytic signals associated with the pair of angles (θ_n, ψ_n). The symbol $|\cdot|$ represents here the modulus of a complex value. Introducing the transfer function H_n in Eq. (20), we find that the functions ϵ_n^{int} and ϵ_n^{pre} take the form

$$\epsilon_n^{\text{int}}(C_{ij}) = \int_{D_f} H_n(\omega, C_{ij}) \gamma_n(\omega) d\omega,$$

$$\epsilon_n^{\text{pre}}(C_{ij}) = \int_{D_f} |H_n(\omega, C_{ij})|^2 \xi_n(\omega) d\omega, \tag{21}$$

with $\gamma_n(\omega) = E(\omega) \overline{Y_n^{\text{exp}}(\omega)}$ and $\xi_n(\omega) = |E(\omega)|^2$.

The introduction of the previous energetic quantities bridges a gap between the maximum-likelihood principle and the objective function construction. A multistep estimation scheme is proposed. Let N_S be the chosen number of experimental signals to be recorded. The following objective functions F^{init} and F^{final} are considered:

$$F^{\text{init}}(C_{ij}) = \sum_{n=1}^{N_S} \left(1 - \frac{|\epsilon_n^{\text{int}}(C_{ij})|^2}{\epsilon_n^{\text{pre}}(C_{ij}) \cdot \epsilon_n^{\text{exp}}} \right) \tag{22}$$

and

$$F^{\text{final}}(C_{ij}) = \sum_{n=1}^{N_S} |f_n(C_{ij})|^2, \tag{23}$$

with

$$f_n(C_{ij}) = 1 - \frac{\epsilon_n^{\text{int}}(C_{ij})}{\epsilon_n^{\text{exp}}}, \tag{24}$$

and successively minimized in order to identify the optimal set of parameters (C_{ij}^{opt}), such that

$$(C_{ij}^{\text{opt}}) = \min_{(C_{ij})} \{F^{\text{final}}(C_{ij})\}. \tag{25}$$

The two objective functions, Eqs. (22) and (23), have been chosen to be positive and locally convex around their *minimum minimorum*, which is identical for both of them and equals zero. In addition, these objective functions have been associated because of their complementarity. As a first approach, the function F^{init} based on an "envelope recognition" principle presents the advantage to be regular away from the global minimum (C_{ij}^{opt}), while the second F^{final} is sharp and possesses more local minima due to its greater phase sensitivity. Thus, the signal recognition is improved near the optimal solution. Hence a multiple step estimation scheme based on the successive minimization of these two functions enables the accurate identification of the material viscoelastic properties from signal recovery.

B. Solving method

Since the function, F, representing either F^{init} or F^{final}, is a real function of complex parameters, its minimization is not straightforward. Let us separate the real and imaginary terms of each component C_{ij}, denoted by the subscripts C'_{ij} and C''_{ij}, respectively. This leads us to introduce the real function $\hat{F}$, defined as

$$\hat{F}(C'_{ij}, C''_{ij}) = F(C_{ij}). \tag{26}$$

The minimization of the objective function $\hat{F}$ is then carried out by using the generalized Newton method associated with the steepest descent technique. The function H being *a priori* analytic, the computations of both the gradient vector and the Hessian matrix of the function $\hat{F}$ are performed by using the Cauchy–Riemann relationships. Additional details are given in Appendix A. Finally, the analytical formulation of the descent direction, which allows a faster and more reliable identification than in the case of numerical assessments of derivatives, can be easily established.

In addition, a confidence interval $I(C_{ij})$ is associated with each identified complex stiffness. A discussion of the limit of the recovery reliability from a set of recorded waveforms can be based on the analysis of these statistic quantities. Details of the confidence interval calculation for the complex stiffnesses, $I(C'_{ij})$ and $I(C''_{ij})$, respectively, are given in Appendix B. A study of the confidence interval behavior with respect to the level of noise in waveforms will be carried out in the following section.

IV. MAIN FEATURES OF THE INVERSION METHOD: NUMERICAL SIMULATION

The different features of the proposed optimization process are investigated by simulating a typical experiment. The calculation of the simulated signals $(y_n)_{1 \leqslant n \leqslant N_S}$ is performed by assuming plane wave propagation. The harmonic assumption has been made for the signal modeling in order to avoid some problems related to the divergence of the Debye's series[21] used for the transfer function computation under transient assumption, Eq. (12). The influence of the parameters, such as the signal-to-noise ratio (SNR) or plate thickness, is evaluated by comparing the set of recovered complex stiffnesses (C_{ij}^{opt}) with the corresponding values (C_{ij}^{true}) used for the signal simulation.

A. Test description

The signals are simulated in different azimuthal plane (Π) for an absorbing orthotropic material. Its viscoelastic properties are then expressed by nine complex constants. According to the working frequency range and the material

TABLE I. Viscoelastic properties of the composite material used for simulation.

Stiffnesses in GPa	C_{11}	C_{22}	C_{33}	C_{12}	C_{13}	C_{23}	C_{44}	C_{55}	C_{66}
C'_{ij}	12	12	136	6.8	5.4	5.4	6.2	6.2	3.4
$C''_{ij,0}$	0.65	0.65	1.1	0.12	0.23	0.23	0.22	0.22	0.24

properties, a unique rheologic model cannot represent all the damping phenomena. In most cases, the complex stiffnesses have the form

$$C_{ij}=C'_{ij}+iC''_{ij}=C'_{ij}+i\left(\frac{\omega}{\omega_0}\right)^p C''_{ij,0}, \tag{27}$$

where p is equal[20] to 0, 1, or 2. The parameters C'_{ij} are independent of the frequency and represent the elasticity constants of the material. The set of values $C''_{ij,0}$, representative of the damping property, are related to the frequency-dependent parameters C''_{ij} through the rheologic model (27).

The complex constants used for simulation are similar to those of a unidirectional carbon-epoxy composite, as seen in Table I. Its density is 1560 kg/m^3. The effects of anisotropy and damping on the optimization process are examined by considering the three cases summarized in Table II. For these simulations the p coefficient, defined in Eq. (27), is fixed and equal to 1, and the center frequency of the reference signal e measured experimentally, is 2 MHz. These three tests differ from each other by their own degrees of anisotropy and damping. From Table I, it can be seen that the plane (**1**,**2**) is quasi-isotropic while the plane (**1**,**3**) is anisotropic. The damping degree varies by considering that the sct of constants $C''_{ij,0}$ given in Table I corresponds to the intrinsic damping constants at different frequencies, 2 MHz for the two first cases, and 0.5 MHz for the last one. Hence, at a given frequency, the constants C''_{ij} are larger for the third case than for the first two ones. Finally, the N_c complex stiffnesses to be recovered by the inversion process ($N_c=4$), the number N_s of considered signals, and the range of incident angles used for the waveform simulation are listed for each implemented test in the fifth, sixth, and seventh columns of Table II, respectively. Note that the value reported below each arrow represents the incident angle increment. In addition, the influence of the plate thickness is examined by considering 0.5, 1, 2, 4, and 8 mm thick samples. We note that the wavelength at 2 MHz of the longitudinal wave at normal incidence is 1.4 mm. Thus, the set of implemented simulations should allow us to study the behavior of the proposed inversion process as well in the case of thin plates as in the case of dispersive materials.

B. Optimization process efficiency without noise in the simulated signals: Stability on initial guess

The process efficiency is studied with respect to the parameter initialization. Numerous factors going from the wavelength-to-thickness ratio to the material properties can affect the shape of the used objective functions by inducing several minima. Therefore the implemented optimization process does not lead to the global minimum for any initial set of unknowns. However, this problem can be solved both by relying on the laboratory knowledge and by constructing an optimization method with the widest convergence domain. This last point has been worked out by designing a two-step inversion scheme. As a first approach the function F^{init}, Eq. (22), based on an ''envelope recognition,'' is minimized. Then the minimization of the second function F^{final}, Eq. (23), is performed in order to improve the signal recognition by increasing the phase sensitivity.

To test the stability on an initial guess, a random perturbation is imposed on the original set of variables. For the set of simulation tests presented in Sec. IV A, the original complex stiffnesses are recovered independently of the initial values selected in a range of $\pm 50\%$ from the original values. The errors on the viscoelastic property reconstruction are always lower than 0.1% for plates of thickness between 0.5 and 8 mm. That shows the excellent efficiency and accuracy of the implemented process even for very thin plates.

C. Optimization reliability with additional noise in the simulated signals

One of the main advantages of the chosen objective functions concerns their slight sensitivity to noise. Let us consider a zero-mean Gaussian white noise [denoted $b(t)$] added to a signal. Since this kind of noise is decorrelated from the modeled signal [i.e., $\langle y_n(\cdot,C_{ij}),b)\rangle=0$], the noise only has an effect on the objective function F^{final} via the denominator of the function (24). The optimal set of parameters (25) is therefore weakly affected by the noise level in waveforms. Thus, in contrast with the optimization processes

TABLE II. Main aspects and characteristics of each performed simulation. The complex stiffnesses associated with the azimuthal planes (**1**,**2**) and (**1**,**3**) are shown in Table I.

Simulation test ($p=1$)	$\omega_0/2\pi$	Anisotropy degree	Damping degree	Sensitive stiffnesses	Range of incident angles	Number of simulated signals
First plane (**1**,**2**)	2 MHz	low	low	$(C_{11},C_{22},C_{12},C_{66})$	$\left[\begin{smallmatrix}step\\ 0°\rightarrow 60°\\ 4°\end{smallmatrix}\right]$	$N_s=16$
Second plane (**1**,**3**)	2 MHz	high	low	$(C_{11},C_{33},C_{13},C_{55})$	$\left[\begin{smallmatrix}step & step & step\\ 0°\rightarrow 6°\rightarrow 10°\rightarrow 24°\\ 1° & 0.5° & 2°\end{smallmatrix}\right]$	$N_s=22$
Third plane (**1**,**3**)	0.5 MHz	high	high	$(C_{11},C_{33},C_{13},C_{55})$	$\left[\begin{smallmatrix}step & step & step\\ 0°\rightarrow 6°\rightarrow 10°\rightarrow 24°\\ 1° & 0.5° & 2°\end{smallmatrix}\right]$	$N_s=22$

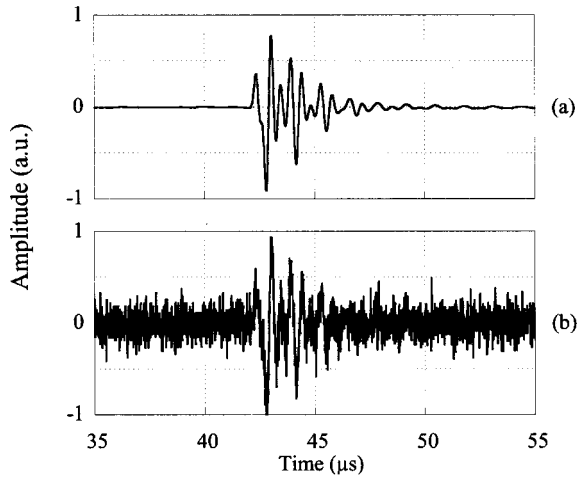


FIG. 3. Signal simulated for a 2 mm thick plate in water with $\theta_n = 7°$ and $\psi_n = 90°$: (a) without noise; (b) with additional noise corresponding to a SNR of 0 dB. Both simulations have been performed for the third simulation test given in Table II.

already proposed,[13–16] the successive use of the objective functions (22) or (23) reduces significantly the noise sensitivity of the inversion process.

The noise sensitivity is quantified by comparing the reconstructed results for four different values (40, 24, 12, and 0 dB) of the SNR defined as the energy ratio. Figure 3(b) shows an example of a scattered signal with a SNR of 0 dB, simulated by adding noise to the original signal reported in Fig. 3(a). Thus, the noisy waveform is randomized within an interval defined by the SNR level.

In order to illustrate the noise and thickness influences, quadratic errors are defined as

$$\Delta \mathbf{C}' = \frac{1}{N_c} \sqrt{\sum_{i \leqslant j} \left(\frac{C'^{\,\mathrm{opt}}_{ij}}{C'^{\,\mathrm{true}}_{ij}} - 1 \right)^2}$$

and (28)

$$\Delta \mathbf{C}'' = \frac{1}{N_c} \sqrt{\sum_{i \leqslant j} \left(\frac{C''^{\,\mathrm{opt}}_{ij,0}}{C''^{\,\mathrm{true}}_{ij,0}} - 1 \right)^2},$$

for the real and imaginary unknowns of the N_c sensitive complex stiffnesses reported in the fifth column of Table II.

In an attempt to get rid of the randomization of a noisy signal amplitude, 16 runs were made for each value of the plate thickness and SNR level. Then the quadratic errors $\Delta \mathbf{C}'$ and $\Delta \mathbf{C}''$ for reconstructed elasticity and damping constants were computed for each of the 16 runs. Finally, the means over the 16 runs are performed for the three simulation tests.

For the sake of clarity, only the results for the second simulation test are displayed in Fig. 4, with the results obtained from the sets of waveforms without random noise. It is interesting to note that, the errors $\Delta \mathbf{C}'$ and $\Delta \mathbf{C}''$ for no additional noise in the waveforms are nearly zero for all processed cases. That points out the accuracy of the suggested method. It can also be observed that, although the errors on the damping constants $C''^{\,\mathrm{opt}}_{ij,0}$ become significant with the increasing noise level, the errors on the real parts are negligible, even for both large noise levels and very thin plates. As expected from the theory, the shifts $\Delta \mathbf{C}'$ and $\Delta \mathbf{C}''$ between the optimal parameters and the original values are very small and increase with the noise level. The slight values of the error $\Delta \mathbf{C}'$ on the identification from noisy signals point out the excellent determination of the elastic properties. The results for the damping property identification are also very satisfying for the noise levels of 40 and 20 dB. For the two other higher noise levels, the errors become nonnegligible, even if the greatest error is always lower than 10%. However, it is important to note that these noise levels are only reached for the case of nonaveraged signals propagating in plates immersed in air.

From the comparison between the three abovementioned simulation tests, the effects of both the anisotropy and the damping on the reconstruction accuracy can also be discussed. No conclusion can be drawn for the elasticity constant identification due to the negligible errors, $\Delta \mathbf{C}'$, always lower than 1%. The comparison between the identifications performed for the first and second tests shows that the plate anisotropy does not really affect the identification of the damping constants $C''^{\,\mathrm{opt}}_{ij,0}$. In addition, from the second and third tests, we have noted that the errors occurring on the damping constants decrease with the increasing damping. This can be explained by the increasing influence of phase dispersion, which makes the final objective functions more sensitive to the damping constants.

The reconstruction reliability and robustness can also be examined by comparing the sets of original and reconstructed signals for the largest deviations $\Delta \mathbf{C}'$ and $\Delta \mathbf{C}''$ reported in Fig. 4. Thus, the signals corresponding to the least accurate signal reconstruction ($\theta = 40°$) for the first simulation test with a SNR of 0 dB and a 0.5 mm thick plate are displayed in Fig. 5. Figure 5(c) shows the signal recovered

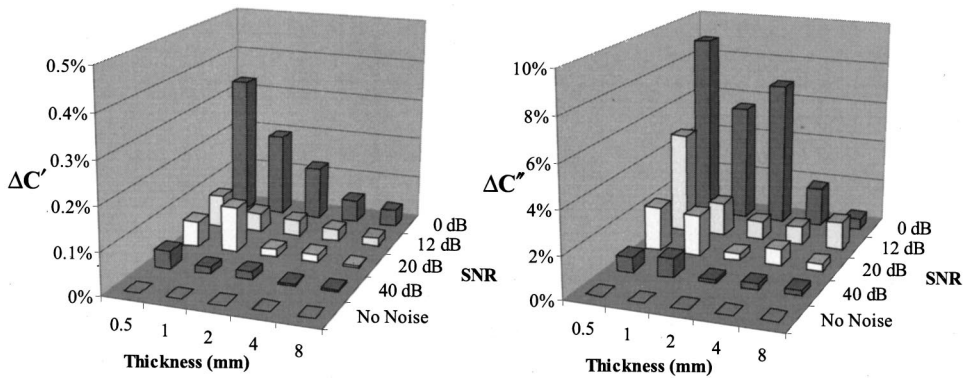


FIG. 4. Errors on the real and imaginary recovered stiffnesses, Eq. (28), for the second simulation test defined in Table II.

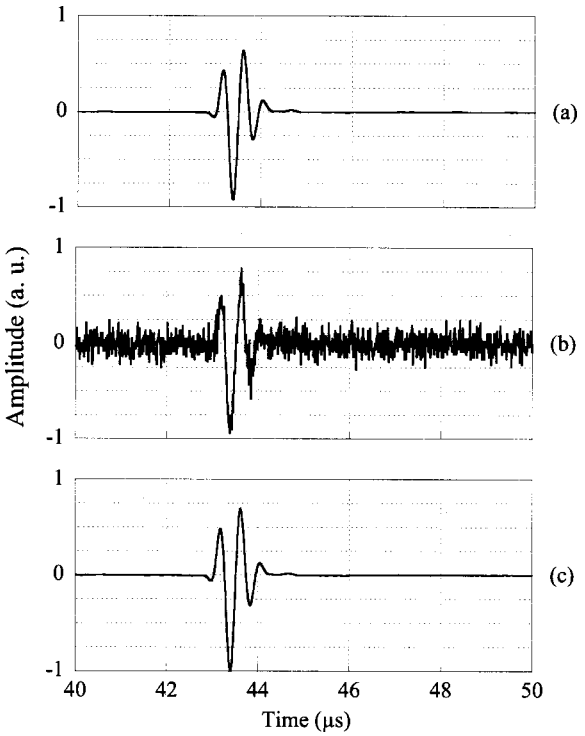


FIG. 5. Signal reconstruction for the first simulation test (with a SNR of 0 dB and a 0.5 mm thick plate) corresponding to the largest deviation $\Delta\mathbf{C}''$. For the 40° incident angle: (a) the original signal simulated without noise, (b) the signal simulated with additional noise, and (c) the signal reconstructed from (b).

from the noisy signal, Fig. 5(b), which has previously been simulated from the original signal represented in Fig. 5(a). The excellent agreement between the original and recovered signals, points out the robustness of the inversion process.

The tensor deviation norms (28) do not allow us to illustrate the variation of each identified stiffness with respect to the signal noise and the plate thickness. We did not choose to represent all the corresponding curves because of the interesting property of our inversion process associating a confidence interval with each identified parameter: all the complex stiffnesses recovered for all implemented simulations coincide, to the nearest confidence intervals, with the original constants used.

However, the uncertainties can be quantified by defining the norms,

$$\Delta I(\mathbf{C}') = \frac{1}{N_c}\sqrt{\sum_{i\leqslant j}\left(\frac{I(C_{ij}'^{\,\mathrm{opt}})}{C_{ij}'^{\,\mathrm{true}}}\right)^2}$$

and (29)

$$\Delta I(\mathbf{C}'') = \frac{1}{N_c}\sqrt{\sum_{i\leqslant j}\left(\frac{I(C_{ij,0}''^{\,\mathrm{opt}})}{C_{ij,0}''^{\,\mathrm{true}}}\right)^2}.$$

The $\Delta I(\mathbf{C}')$ and $\Delta I(\mathbf{C}'')$ results regarding the noise level and the plate thickness are similar for all processed cases. Thus, an illustration of their behaviors is given in Fig. 6 for the second simulation test. The norms $\Delta I(\mathbf{C}')$ and $\Delta I(\mathbf{C}'')$ increase as the noise level in the simulated signals whatever the plate thickness is. Subsequently, it can be concluded that, as expected, the higher the noise level, the more uncertain the optimal parameters. That demonstrates that the confidence intervals quantify the reliability of the inversion algorithm for identifying stiffnesses. Thus, the reconstruction results become uncertain for a 0.5 mm thick plate when noise is introduced in signals.

The slight values of $\Delta I(\mathbf{C}')$ in Fig. 6 show the great sensitivity, and therefore the strong reliability, of the inversion algorithm for identifying the elasticity constants C_{ij}'. It can also be seen that the norm $\Delta I(\mathbf{C}'')$ on the damping constants $C_{ij,0}''$ becomes important with the increasing noise level and the decreasing plate thickness, while the errors $\Delta\mathbf{C}''$ remain fairly slight, as shown in Fig. 6(b). The large values of $\Delta I(\mathbf{C}'')$ are mainly due to the nonzero values of the function F^{final} minimum increasing with the noise level.

Minimizations of the objective functions already introduced by previous authors[13–16] have also been carried out. As expected from the theory, noise induces a more important shift in the recovered stiffnesses. We have noted an increase of the error on the real part of stiffnesses with respect to our approach that takes into account explicitly the signal phase, Eqs. (22) and (23).

D. Influence of the rheologic model on the real part of the stiffness tensor

As already mentioned, an *a priori* choice of the rheologic model (27) needs to be made for the identification of viscoelastic properties. Damping mechanisms can be due to a combination of different energy dissipation phenomena: diffraction, friction, or viscoelastic features of the material. Thus, one must keep in mind that the damping phenomena may be unknown or inexactly modeled. Hence, it becomes

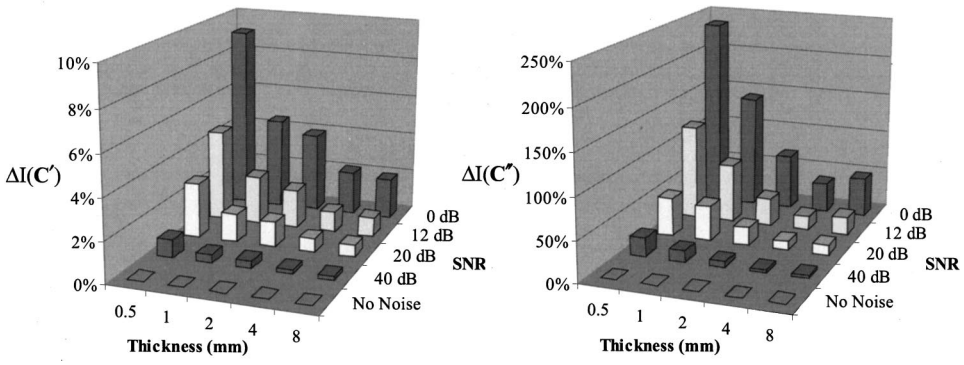


FIG. 6. Real and imaginary relative confidence intervals, Eq. (29), for the second simulation test defined in Table II.

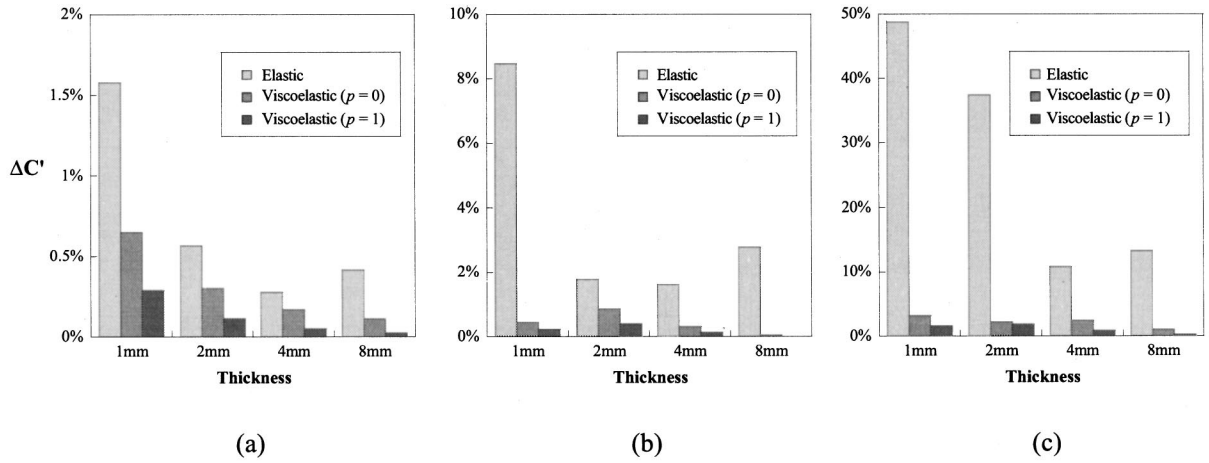


FIG. 7. Errors, Eq. (28), on the real stiffnesses recovered under different rheologic model assumptions: the imaginary stiffnesses C''_{ij} are successively assumed to be zero (i.e., elastic material), frequency independent ($p=0$) or even linear-frequency dependent ($p=1$). Signals have been simulated assuming a quadratic-frequency dependence of C''_{ij} ($p=2$) for the first (a), the second (b), and the third (c) simulation tests given in Table II.

necessary to study the consequences of this choice on the identification of the real part of the stiffness tensor.

For the three configurations given in Table II, a quadratic frequency dependence of the imaginary constants C''_{ij} [i.e., p is fixed to 2 in Eq. (27)] is assumed and signal sets are simulated without additional noise. Complex stiffnesses are then identified under different rheologic model assumptions, and compared with the original stiffnesses used for signal simulation, by assessing the norm $\Delta\mathbf{C}'$, Eq. (28). The imaginary stiffnesses C''_{ij} are successively assumed to be 0, frequency independent ($p=0$), linear-frequency dependent ($p=1$), and quadratic-frequency dependent ($p=2$). Thus, the medium to be identified is considered elastic, and then viscoelastic with different models ($p=0$, 1, or 2). Results for $\Delta\mathbf{C}'$ are presented by Fig. 7.

The noise level in the simulated signals being zero, the results of the quadratic-damped medium characterization are identical to the original stiffnesses ($\Delta\mathbf{C}'=0$) and therefore are not reported in Fig. 7. Whatever the plate thickness and the processed case, the error on the real stiffness identification is much greater for the assumed-elastic medium than for the assumed-viscoelastic media. As expected, the higher the deviation between the simulated and assumed rheologic models, the higher the error on the real part stiffness identification. However, the slight values of $\Delta\mathbf{C}'$ shown in Fig. 7 for any of the three viscoelastic models, point out the need to consider a viscosity parameter in our modeling. In addition, Figs. 7(a) and (b) show that the effect of the material anisotropy on the wrong recognition of the intrinsic rheologic model is slight, whereas by comparing Figs. 7(b) and (c), the effect of material damping appears to be significant.

The effect of these four rheologic laws on complex stiffness identification can also be illustrated by displaying the temporal waveforms recovered assuming the medium to be successively elastic and viscoelastic ($p=0,1,2$), as seen in Fig. 8, for the second simulation test with a 2 mm thick plate and $\theta=7°$. The original signal used in the inversion process is similar to the one reported in Fig. 8(d). The echoes multi-

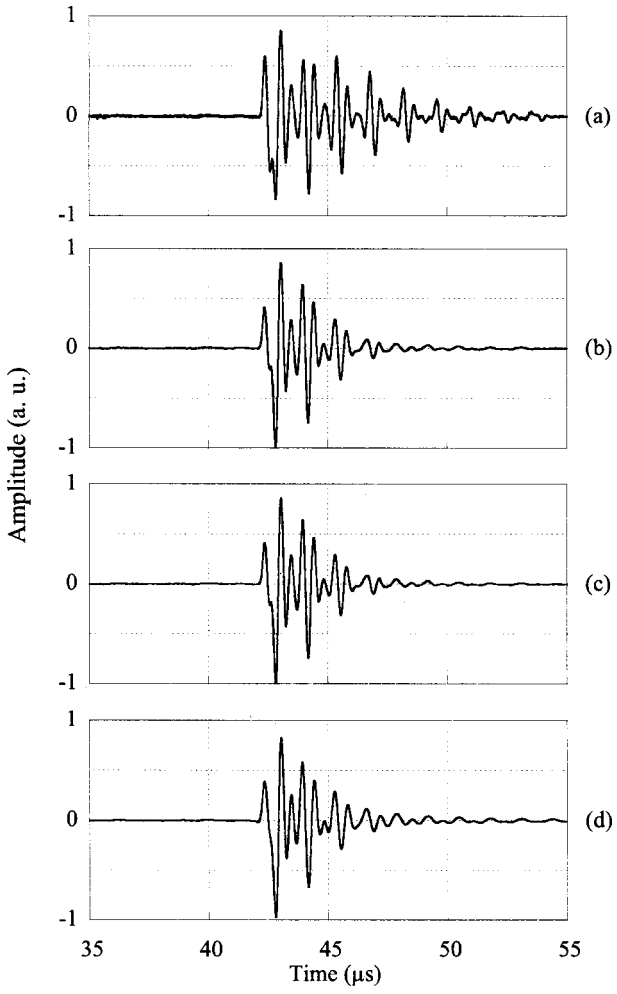


FIG. 8. Signal reconstructions for the second simulation with a 2 mm thick plate under different rheologic model assumptions: the imaginary stiffnesses C''_{ij} are successively (a) fixed to 0 (i.e., elastic material), (b) frequency independent ($p=0$), (c) linear-frequency dependent ($p=1$), and (d) quadratic-frequency dependent ($p=2$). The original signal used in the inversion process has been simulated assuming ($p=2$) for the 7° incident angle and is similar to the one reported in (d).

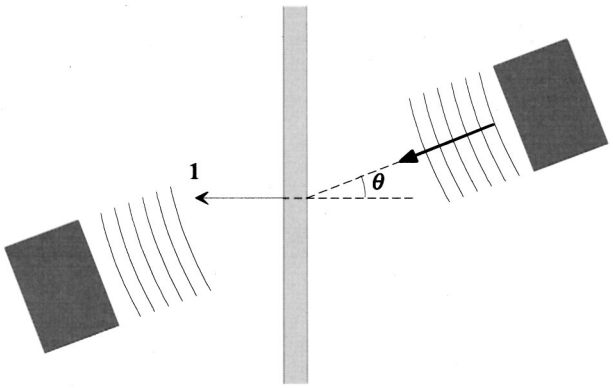


FIG. 9. A schematic diagram of the experimental setup: the sample is immersed in water and placed between two transducers. An ultrasonic pulse is generated by the transmitter, propagates through the plate, and is then recorded by the receiver.

reflected in the plate that are characteristic of elastic media can be seen in Fig. 8(a). Figures 8(b) and (c) show that the taking into account of damping in the rheologic model improves the reconstruction by inducing a phase dispersion and a decrease of echo amplitude. However, despite the bad correlation between the signals displayed in Figs. 8(a) and (d), the inversion with the elastic modeling provides an order of magnitude for the elasticity constants.

Thus, even if the damping model is unknown, the optimization process leads to a fairly good evaluation of the elasticity constants if damping is considered through the rheologic model. That is an interesting advantage of our characterization technique, which is still valid for strongly dispersive media.

V. IDENTIFICATION FROM EXPERIMENTAL DATA

An ultrasound is generated and sensed by a pair of identical underwater piezoelectric transducers between which the sample is placed. They are supported by two goniometers that allow one to select any propagation direction of incident wave. During the angular inspection, the transmitter and receiver are always facing each other.

The experimental setup presented in Fig. 9 does not take into account the lateral shift of both the simple transmitted beam and the echoes multireflected in the plate. Thus, since the transducer size is finite, the maximum energies of all signals propagating in the plate may not reach the receiver. However, the plate considered in this paper being thin, the effect of these lateral shifts deduced from the Snell–Descartes law tends to be negligible. Let us recall that these deviations are also small for the waves slowly propagating in the plate, and so for media with a weak rigidity-to-density ratio.

Two experiments are conducted to examine the suitability of the established inverse method based on the plane wave assumption for viscoelastic property characterization. The experimental characterization conditions are chosen in order to check the validity of the inverse method for large wavelength-to-thickness ratios.

TABLE III. PMMA viscoelastic properties. (a) Reconstruction performed by using the proposed inverse process from a set of signals propagating through a 2.4 mm thick plate. The numbers in the parentheses indicate the confidence interval (in GPa) associated with each identified constant. (b) Results extracted from Ref. 3.

Stiffnesses in GPa with $p=0$		C_{11}	C_{22}	C_{12}	C_{66}
(a)	C'_{ij}	8.49(0.06)	8.32(0.47)	4.21(0.26)	2.08(0.16)
	$C''_{ij,0}$	0.11(0.03)	0.34(0.30)	0.13(0.15)	0.08(0.05)
(b)	C'_{ij}	8.61	8.61	4.34	2.14
	$C''_{ij,0}$	0.096	0.096	0.000	0.048

A. Isotropic viscoelastic plate

The experimental characterization of a 2.4 mm thick isotropic viscoelastic plate of polymethylmethacrylate (PMMA) whose density is 1185 kg/m^3, is conducted at low frequency. The center frequency of the reference pulse is around 0.4 MHz and the diameter of both transducers is 29 mm. Signals are caught for incident angles varying from 0° to 46° by a step of 2° in the symmetry plane (**1**,**2**) defined by $\psi=0°$. Due to the properties of PMMA (a weak rigidity-to-density ratio), the harmonic assumption remains valid over the large incident angle range considered. The recovered properties are given in Table III, with the ones extracted from Ref. 3.

The reliability of this characterization can be investigated *a posteriori* by comparing the experimental and reconstructed signals. The experimental reference pulse e, and the experimental and reconstructed signals for $\theta=30°$, are repre-

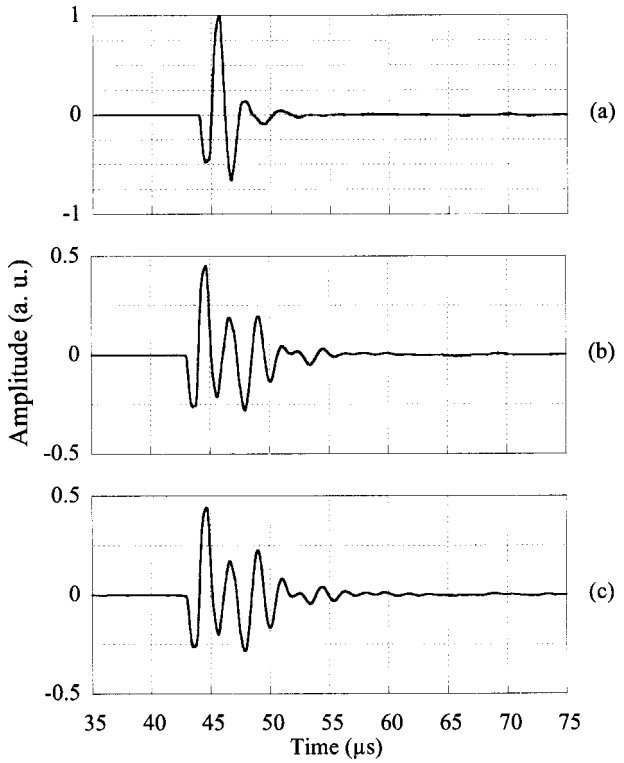


FIG. 10. (a) The 0.4 MHz center frequency reference signal propagating in water without the sample. Experimental (b) and reconstructed (c) signals propagating through a 2.4 mm thick plate of PMMA for the 30° incident angle. This angle corresponds to the least accurate signal reconstruction.

TABLE IV. Material properties (in GPa) and associated confidence intervals of a 2.2 mm thick plate of a carbon–SiC composite. (a) Viscoelastic properties recovered by using the proposed inversion process with $p=1$. (b) Elasticity constants identified from time-of-flight measurements. These two characterizations have been performed at two different working frequencies: 2 MHz (a) and 5 MHz (b).

	Stiffnesses in GPa with $p=1$	C_{11}	C_{22}	C_{12}	C_{66}
(a)	C'_{ij}	17.8(0.5)	98.6(4.2)	7.3(1.0)	10.0(0.4)
	$C''_{ij,0}$	3.4(0.2)	7.4(3.7)	1.8(0.6)	0.2(0.3)
(b)	C'_{ij}	16.7(0.2)	88.5(5)	5.3(0.5)	8.5(0.1)

sented in Figs. 10(a)–(c), respectively. Note that this incident angle corresponds to the largest value of f_n, Eq. (24), and therefore to the least accurate reconstruction of experimental signals. It can be observed from Figs. 10(c) and (b) that the waveform of the reconstructed signal matches well with the shape of the measured signal. However, slight differences appear in the signal tails. These deviations arise from the spatial shifts of the maximum energy of the echoes multireflected in the plate, which are not taken into account by the plane wave theory. The maximum energy in measured signals depends indeed on the receiver location in the phase plane.[26] Thus, as long as the incident angle increases, the plane wave assumption tends to be unrealistic. That affects essentially the identification of the imaginary part of variables strongly sensitive to the signals measured for large incident angles, as $C''_{22,0}$. In addition, the comparison of the wavelength around 0.4 MHz with the transducer diameter shows that the beam divergence is about 15°. Hence, the plane wave assumption may be unrealistic and can explain the inaccuracy of the imaginary stiffness identification.

However, the results summarized in Table III show, to the nearest confidence intervals, a fairly good agreement between the identified viscoelastic properties and the ones extracted from Ref. 3. From Table III(b), it is also interesting to note that the isotropic material symmetry has been experimentally recovered. Finally, the recovery of the well-known viscoelastic properties of PMMA and the good coincidence between the sets of measured and recovered signals validates the identification.

B. Thin anisotropic damping plate

In an attempt to study the anisotropy effect, a second test is performed on a 2.2 mm thick plate of a two-dimensional (2-D) carbon–SiC composite whose density is 2025 kg/m^3. This specimen is a high damping material due to porosities. The center frequency of the reference pulse is around 2 MHz and the diameter of both transducers is 29 mm. Signals are measured for incident angles varying from 0° to 18° by step of 0.5° in the anisotropic symmetry plane (**1**,**2**). The fairly large rigidity-to-density ratio of carbon–SiC has imposed the choice of the incident angle range width. In order to minimize the influence of the echoes that do not impinge upon the receiver faced toward the transmitter, the optimization process is performed with transient assumption ($N_e=2$). The recovered complex stiffnesses are given in Table IV,

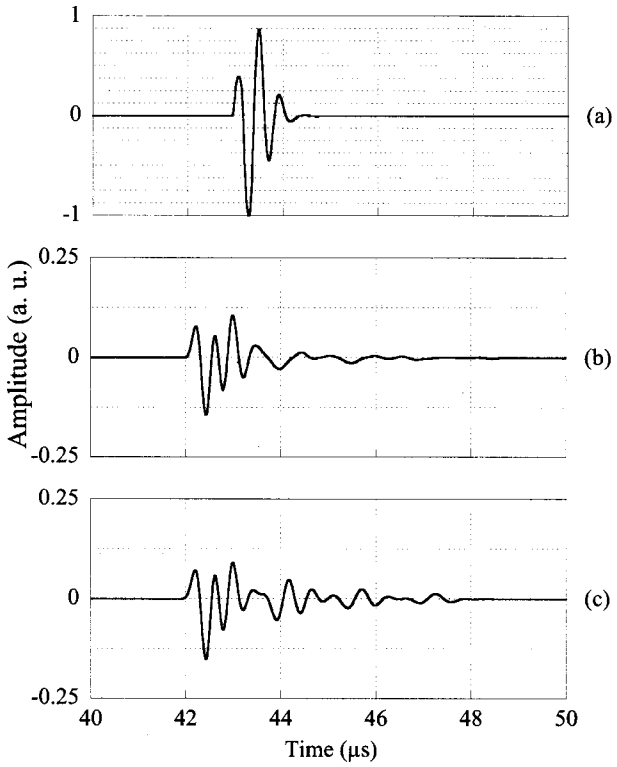


FIG. 11. (a) The 2 MHz center frequency reference signal propagating in water without the sample. Experimental (b) and reconstructed (c) signals propagating through a 2.2 mm thick plate of a carbon–SiC composite for the 7.5° incident angle. This angle corresponds to the least accurate signal reconstruction.

with the elasticity constants identified from time-of-flight measurements.

Note that the elasticity constants, Table IV(b), had to be determined from wave speed measurements performed at 5 MHz, the lower working frequency at which the mode overlapping does not affect the experimental characterization. The good agreement between the real parts of the stiffnesses identified at 2 MHz, Table IV(a), and the elasticity constants determined from the wave speed measurements show the accuracy of the proposed inversion method for elastic property reconstruction. In addition, since the presented results are obtained at a lower working frequency (2 instead of 5 MHz), i.e., for a larger wavelength-to-thickness ratio, the limits of methods based on wave speed measurements are extended.

Only the superimposition of the measured and reconstructed signal sets can supply information about the accuracy of the $C''_{ij,0}$ identification. The experimental reference pulse *e*, and the experimental and reconstructed signals for $\theta=7.5°$, are represented in Figs. 11(a)–(c), respectively. Note that this incident angle corresponds to the least accurate reconstruction of experimental signals. From the comparison between Figs. 11(c) and (b), it can be seen that the coincidence between the waveforms is excellent for the simple transmitted wave. Although the discrepancies observed for the following echoes can be explained by the spatial shift of echo energy discussed in Sec. V A, it is crucial to mention a second phenomenon related to material anisotropy. In anisotropic media the energy flux direction may differ from the

phase plane direction.[27,28] Therefore, the uniform distribution of energy on the phase plane assumed by the plane wave theory is not experimentally verified. Thus, for high anisotropic materials, the assumption of plane wave propagation becomes more approximate because of the experimental setup. Finally, since the results of the identification of the unknown constants $C''_{ij,0}$ are closely related to the used modeling (plane wave theory, transient assumption), uncertainties remain concerning the values given in Table IV(a) for $C''_{ij,0}$.

VI. CONCLUSION

An analytical development of the transfer function in transmission of an anisotropic plate immersed in a fluid has been outlined in harmonic and transient assumptions. By applying the implicit function theorem to the Christoffel equation, the analytical formulations of the first and second derivatives of this transmission coefficient can be obtained. Hence, the direct problem being fully analytically developed, an original inversion scheme based on the maximum-likelihood principle has been proposed. Two complementary objective functions have been constructed. Simulation tests have shown the efficiency and accuracy with thin isotropic and anisotropic plates, regardless of the noise level and the damping properties. It has also been demonstrated that the accurate elastic property identification can be achieved only by taking into account the damping mechanisms in the modeling. The suitability of the established inversion method for experimental characterization has been examined by conducting experiments on one isotropic (PMMA) and one anisotropic (carbon–SiC composite material) solids. The success of the PMMA viscoelastic properties identification at very low working frequency has demonstrated the strong reliability of the process for large wavelength-to-thickness ratios. The characterization of the carbon–SiC plate has shown the satisfactory reconstruction of the elasticity constants and the limits of the plane wave assumption. This modeling difficulty can be overcome by an already proposed[15,29] experimental adjustment, where large transducers are used to improve transmission signal measurement with plane wave assumptions.

APPENDIX A: DERIVATIVES OF THE OBJECTIVE FUNCTIONS

The analytical formulations of the first and second derivatives of the objective functions (22) or (23) require the examination of the derivative properties of the squared modulus of a complex function f. Let $z_i = x_i + iy_i$ be a element of a complex variable set. Then splitting f into two real functions P and Q such as

$$f(z) = P(x,y) + iQ(x,y), \tag{A1}$$

the relationship

$$\frac{\partial f(z)}{\partial z_i} = f_{,i} = \frac{\partial P(x,y)}{\partial x_i} + i\,\frac{\partial Q(x,y)}{\partial x_i} \tag{A2}$$

satisfies the Cauchy–Riemann equations,

$$\frac{\partial P(x,y)}{\partial x_i} = \frac{\partial Q(x,y)}{\partial y_i} \quad \text{and} \quad \frac{\partial Q(x,y)}{\partial x_i} = -\frac{\partial P(x,y)}{\partial y_i}. \tag{A3}$$

Defining the function $\hat{F}: \mathbb{R}^{2n} \to \mathbb{R}$ such as

$$\hat{F}(x,y) = |f(z)|^2, \tag{A4}$$

the following relationships can then be established:

$$\frac{\partial \hat{F}}{\partial x_i} = 2\ \mathrm{Re}\{f \cdot \overline{f_{,i}}\}$$

$$\frac{\partial \hat{F}}{\partial y_i} = 2\ \mathrm{Im}\{f \cdot \overline{f_{,i}}\}$$

and (A5)

$$\frac{\partial^2 \hat{F}}{\partial x_i\,\partial x_j} = 2\ \mathrm{Re}\{f_{,i} \cdot \overline{f_{,j}} + f \cdot \overline{f_{,ij}}\},$$

$$\frac{\partial^2 \hat{F}}{\partial y_i\,\partial y_j} = 2\ \mathrm{Re}\{f_{,i} \cdot \overline{f_{,j}} - f \cdot \overline{f_{,ij}}\},$$

$$\frac{\partial^2 \hat{F}}{\partial x_i\,\partial y_j} = 2\ \mathrm{Im}\{f_{,i} \cdot \overline{f_{,j}} + f \cdot \overline{f_{,ij}}\}.$$

APPENDIX B: CONFIDENCE INTERVAL CALCULATION

The confidence intervals $I(C_{ij})$ associated with each identified complex stiffness C_{ij} are determined by processing separately the real and imaginary parts of the analytic function f_n, Eq. (24). That led us to examine independently the real and imaginary unknowns of each complex stiffness, C'_{ij} and $C''_{ij,0}$, respectively. Consequently, we associate with each of them a real confidence interval, $I(C'_{ij})$ and $I(C''_{ij,0})$, respectively.

These two intervals have been determined from the $(1 \times 2N_S)$ error vector $\boldsymbol{\epsilon}$

$$\begin{aligned} \epsilon_{2n+1} &= \mathrm{Re}\{f_n(C_{ij}^{\mathrm{opt}}) - f_n(C_{ij}^{\mathrm{true}})\} = \mathrm{Re}\{f_n(C_{ij}^{\mathrm{opt}})\}, \\ \epsilon_{2n} &= \mathrm{Im}\{f_n(C_{ij}^{\mathrm{opt}}) - f_n(C_{ij}^{\mathrm{true}})\} = \mathrm{Im}\{f_n(C_{ij}^{\mathrm{opt}})\}, \\ &\text{for } n \in [1, N_S], \end{aligned} \tag{B1}$$

where (C_{ij}^{opt}) are the estimated stiffnesses, (C_{ij}^{true}) the intrinsic stiffnesses, and N_S the number of recorded signals. On assuming that the set of variables $\boldsymbol{\epsilon}_n$ follows a normal distribution with a zero mean and a constant variance, the covariance matrix Ω of the assessed stiffnesses (C_{ij}^{opt}) can then be expressed as

$$\Omega = \frac{\boldsymbol{\epsilon}^t \cdot \boldsymbol{\epsilon}}{2(N_S - N_c)} \left[\left(\frac{\partial g_n}{\partial \mathbf{c}} \right)^t \cdot \left(\frac{\partial g_n}{\partial \mathbf{c}} \right) \right]$$

$$\text{with } g_n = \begin{cases} \mathrm{Re}\{f_n\}, & \text{if } n \text{ is odd}, \\ \mathrm{Im}\{f_n\}, & \text{if } n \text{ is even}, \end{cases} \tag{B2}$$

where the $(1 \times 2N_c)$ vector $\mathbf{c}$, formed by the N_c variables to be identified, C'_{ij} or $C''_{ij,0}$. Since the distribution of the ratios $(c_i^{\mathrm{opt}} - c_i^{\mathrm{true}})/\sqrt{\Omega_{ii}}$ satisfies a student's t distribution with $2(N_S - N_c)$ degrees of freedom, a confidence interval $I(c_i)$ can then be associated with the variable c_i^{opt} such as

$$|c_i^{\mathrm{opt}} - c_i^{\mathrm{true}}| < I(c_i), \tag{B3}$$

with, for example, a 95% level of confidence.

[1] R. Chandra, S. P. Singh, and K. Gupta, ''Damping studies in fiber-reinforced composites—a review,'' Compos. Struct. **46**, 41–51 (1999).
[2] M. R. Karim, A. K. Mal, and Y. Bar-Cohen, ''Inversion of leaky lamb wave data by simplex algorithm,'' J. Acoust. Soc. Am. **88**, 482–491 (1990).
[3] A. Safaeinili, O. I. Lobkis, and D. E. Chimenti, ''Quantitative materials characterization using air-coupled leaky Lamb waves,'' Ultrasonics **34**, 393–396 (1996).
[4] M. F. Markham, ''Measurement of the elastic constants of fibre composites by ultrasonics,'' Composites **1**, 145–149 (1970).
[5] R. E. Smith, ''Ultrasonic elastic constants of carbon fibers and their composites,'' J. Appl. Phys. **43**, 2555–2561 (1972).
[6] B. Hosten, M. Deschamps, and B. R. Tittmann, ''Inhomogeneous wave generation and propagation in lossy anisotropic solids. Application to the characterization of viscoelastic composite materials,'' J. Acoust. Soc. Am. **82**, 1763–1770 (1987).
[7] B. Castagnède, J. T. Jenkins, W. Sachse, and S. Baste, ''Optimal determination of the elastic constants of composite materials from ultrasonic wavespeed measurements,'' J. Appl. Phys. **67**, 2753–2761 (1990).
[8] S. I. Rokhlin and W. Wang, ''Double through-transmission bulk wave method for ultrasonic phase velocity measurement and determination of elastic constants of composite materials,'' J. Acoust. Soc. Am. **91**, 3303–3312 (1992).
[9] C. Aristégui and S. Baste, ''Optimal recovery of the elasticity tensor of general anisotropic materials from ultrasonic velocity data,'' J. Acoust. Soc. Am. **101**, 813–833 (1997).
[10] B. Castagnède, J. Roux, and B. Hosten, ''Correlation method for normal mode tracking in anisotropic media using an ultrasonic immersion system,'' Ultrasonics **27**, 280–287 (1989).
[11] B. Audoin and J. Roux, ''An innovative application of the Hilbert transform to time delay estimation of overlapped ultrasonic echoes,'' Ultrasonics **34**, 25–33 (1996).
[12] N. Leymarie and S. Baste, ''Guided waves and ultrasonic characterization of three-dimensional composites,'' in *Review of Progress in Quantitative Nondestructive Evaluation* (AIP, New York, 2000), Vol. 19B, pp. 1175–1181.
[13] A. I. Lavrentyev and S. I. Rokhlin, ''Determination of elastic moduli, density, attenuation, and thickness of a layer using ultrasonic spectroscopy at two angles,'' J. Acoust. Soc. Am. **102**, 3467–3477 (1997).
[14] B. Hosten, M. Castaings, and T. Kundu, ''Identification of viscoelastic moduli of composite materials from the plate transmission coefficient,'' in Ref. 12, Vol. 17B (Plenum Press, New York, 1998), pp. 1117–1124.
[15] M. Castaing, B. Hosten, and T. Kundu, ''Inversion of ultrasonic, plane-wave transmission data in composite plates to infer viscoelastic material properties,'' NDT & E Int. **33**, 377–392 (2000).
[16] V. K. Kinra and V. R. Iyer, ''Ultrasonic measurement of the thickness, phase velocity, density or attenuation of a thin-viscoelastic plate. Part II: the inverse problem,'' Ultrasonics **33**, 111–122 (1995).
[17] H. L. Van Trees, *Detection, Estimation and Modulation Theory* (Wiley, New York, 1968).
[18] A. D. Whalen, *Detection of Signals in Noise* (Academic, New York, 1971).
[19] B. A. Auld, *Acoustic Fields And Waves in Solids* (Krieger, Stanford, 1990).
[20] M. Deschamps and B. Hosten, ''The effects of viscoelasticity on the reflection and transmission of ultrasonic waves by an orthotropic plate,'' J. Acoust. Soc. Am. **91**, 2007–2015 (1992).
[21] M. Deschamps and C. Cao, ''Reflection/refraction of a solid layer by Debye's series expansion,'' Ultrasonics **29**, 288–293 (1991).
[22] P. M. Woodward, *Probability and Information Theory with Application to Radar* (Pergamon, London, 1953).
[23] D. Gabor, ''Theory of communication,'' J. Inst. Electr. Eng. **93**, 429–457 (1946).
[24] J. Ville, ''Theorie et applications de la notion de signal analytique,'' Cables Transm. **1**, 61–74 (1947).
[25] S. Mallat, *A Wavelet Tour of Signal Processing* (Academic, New York, 1999).
[26] M. Deschamps and P. Chevée, ''Reflection and transmission of a transient bounded beam by a dispersive and orthotropic plate,'' Acta Acust. (Paris) **1**, 3–14 (1993).
[27] M. Deschamps and F. Assouline, ''Attenuation along the Poynting vector direction of inhomogeneous plane waves in absorbing and anisotropic solids,'' Acustica-acta acustica **86**, 295–302 (2000).
[28] M. J. P. Musgrave, *Crystal Acoustics* (Holden-Day, San Francisco, 1970).
[29] P. Cawley and B. Hosten, ''The use of large ultrasonic transducers to improve transmission coefficient measurements on viscoelastic anisotropic plates,'' J. Acoust. Soc. Am. **101**, 1373–1379 (1997).

Determination of the elastic constants of a composite plate using wavelet transforms and neural networks

Jing Yang and Jianchun Cheng[a)]
State Key Laboratory of Modern Acoustics and Institute of Acoustics, Nanjing University, Nanjing 210093, People's Republic of China

Yves H. Berthelot
Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0405

(Received 12 June 2001; revised 20 October 2001; accepted 12 December 2001)

An inverse method based on a combination of the wavelet transform and artificial neural networks is presented. The method is used to recover the elastic constants of a fiber-reinforced composite plate from experimental measurements of ultrasonic Lamb waves generated and detected with lasers. In this method, the elastic constants are not recovered from the dispersion curves but rather directly from the measured waveforms. Transient waveforms obtained by numerical simulations for different elastic constants are used as input to train the neural network. The wavelet transform is used to extract the eigenvectors from the Lamb wave signals to simplify the structure of the neutral network. The eigenvectors are then introduced into a multilayer internally recurrent neural network with a back-propagation algorithm. Finally, experimental waveforms recoded on a titanium-graphite composite plate are used as input to recover the elastic constants of the material. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1451071]

PACS numbers: 43.35.Cg, 43.35.Yb [DEC]

I. INTRODUCTION

Ultrasound can be used to characterize the mechanical properties of a medium. Resonant Ultrasound Spectroscopy,[1–4] for instance, is an accurate method for measuring the elastic constants of a material provided that the medium is nondispersive. Laser ultrasonics[5,6] is also a powerful tool because it enables generation and detection of elastic waves in solids with minimum contact between the optical transducer and the surface thus providing a means to perform noncontact testing of materials and structures. Laser ultrasonics is particularly well suited to analyze the elastic behavior of new composite materials intended to be used at elevated temperatures. Because of the inherent point-source/point-receiver nature of laser ultrasonics, an important application is the inverse determination of the elastic constants from either the group or the phase velocity of ultrasonic propagation.[7–9] However, the inversion of elastic constants for fiber-reinforced composite materials, which are anisotropic in nature, usually requires knowledge of group or phase velocities of ultrasonic wave propagating in a particular plane of symmetry for the measurements of the pure bulk waves.[10,11] For a thin fiber-reinforced composite plate, Wu and Liu[12] presented a hybrid inverse method to determine the anisotropic elastic constants by using ultrasonic bulk wave and laser-generated lamb wave.

Many recent studies have dealt with the quantitative determination of the elastic constants and the thickness of isotropic sheet materials[13–16] from laser-generated Lamb waves. The measurement is based on the fact that, for a plate with thickness much less than a wavelength, the transient waveform is characterized by the lowest Lamb wave modes. The first arrival is the lowest nondispersive symmetric mode (S_0) and it is followed by the dispersive lowest antisymmetric mode (A_0). This feature is used to measure the thickness and the elastic constants of isotropic materials without any prior knowledge of other acoustic parameters. However, the transient waveforms depend strongly on the longitudinal acoustic wavelength and the plate thickness. For plates with thickness on the order of millimeters, the transient waveforms excited by a laser pulse with typical duration of the order of 10 ns are much more complicated because of contributions from the higher-order Lamb wave modes. As each frequency component propagates with a different phase velocity, the initial shape of the transient waveform is temporally distorted. The wavelet transform (WT) has been applied to analyze such signals with time-varying spectra to extract characteristics of transient Lamb waveforms.[17–19] Han *et al.*[20] have used the wavelet transform to study anisotropic behaviors and mode distribution of the laser-generated Lamb waves in fiber-reinforced composite plates. They found that it is often very difficult to invert directly the elastic properties of composite materials from dispersion curves or time-frequency representations. Thus, it is necessary to develop a method for extracting the elastic constants of composite plates directly from experimental laser-generated Lamb waveforms. This is the objective of this paper.

Applications of artificial neural network (ANN) to quantitative nondestructive evaluation (QNDE) have been developed in recent years. Thomsen and Lund[21] investigated the classification of ultrasonic power spectra from composite laminates. Udpa and Udpa[22] used neural networks to classify the results of QNDE using electromagnetic eddy current techniques. Windsor *et al.*[23] investigated welded structures using neural networks. Challis *et al.*[24] and Todd and

[a)]Author to whom correspondence should be addressed. Electronic mail: jccheng@nju.edu.cn

Challis[25] presented applications of neural networks to the study bonded structures using Lamb waves. For the most part, previous work focused on the classification capability of neural networks to identify whether or not cracks or defects were present in the materials or the structure. A major practical difficulty is that a very large number of waveforms are needed to train the network.

In this paper, we present a new inverse method to determine the elastic constants of a fiber-reinforced composite plate from transient Lamb waveforms generated and detected by lasers, using a combination of neural networks and wavelet transforms. The samples for training the network are obtained from theoretical simulations based on the method of the normal mode expansion[26,27] for analyzing laser-generated ultrasonic Lamb waves propagating along the principal directions of a fiber-reinforced composite plate, which is here modeled as a transversely isotropic medium. On the basis of this theory, the transient displacements of Lamb waves can be rapidly obtained for many sets of different elastic constants. These theoretical transient waveforms are then taken as the samples to train the artificial neural network. In order to simplify the structure of the neural network, the raw Lamb waves signals are first transferred to a pre-processor using the wavelet packet analysis for extracting the eigenvectors from the signals. The pre-processor output is then input into a multilayer internally recurrent neural network (IRNN) with back-propagating algorithm for training and learning purposes. Finally, the experimental waveforms are used as input in the complete system to recover the elastic constants of the material used in the experiment. Inversion results show the effectiveness of the method. Overall, this work provides a quantitative method for ultrasonic nondestructive evaluation of the elastic stiffness properties of the composite plate by the laser-generate Lamb wave technique.

II. EXPERIMENTAL WAVEFORMS

The sample under consideration is a titanium graphite composite plate: SCS-6/Timetal 21S with a half-thickness $h=0.723$ mm. It is modeled as a transversely isotropic material with five independent elastic constants. The ultrasonic Lamb wave is generated by a Nd: YAG pulsed laser (as shown in Fig. 1) with an approximate duration 10 ns and an energy of 5.7 mJ. The normal surface velocity of the ultrasonic Lamb waves at the rear surface is detected by a Mach–Zehnder interferometer. Measurements are recorded at different locations away from the pulsed laser spot. In order to increase efficiency of the ultrasonic excitation by a laser pulse, a very thin coating of oil at the generation spot has been used in experiments. Figure 2 shows the transient Lamb waves propagating along the fiber direction at two different distances $D=12$ mm and $D=20$ mm between the source and the detecting laser. The waveforms are quite complex because of the large number of modes present in the signal. Nevertheless, the waveforms depend upon the five elastic constants of the sample and it is shown below that one can use wavelet transforms and neural networks to extract quantitative information about the elastic constants of the material.

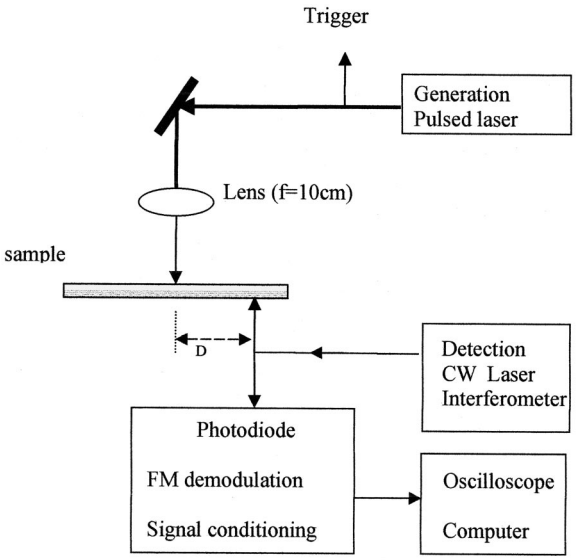


FIG. 1. Experimental configuration.

III. THEORETICAL WAVEFORMS

The fiber-reinforced composite is modeled as an orthotropic plate bounded by two parallel surfaces with free of traction planes at $x_3=\pm h$, $x_3=0$ being the mid-plane of the plate. The coordinate axes x_1, x_2, and x_3 corresponding to the fiber, cross fiber and thickness directions, respectively, are the three principal directions. Lamb waves propagates along the x_1-direction. The total displacement field $\mathbf{U}=(U_1,U_2,U_3)$ is independent of x_2, so that the Lamb wave motion in the plane (x_1-x_3) is decoupled from the displacement component U_2 in the x_2 direction. The elastic equations can be written as:[26]

$$\rho\frac{\partial^2 U_1}{\partial t^2}=C_{11}U_{1,1}+C_{13}U_{3,1}+C_{55}(U_{1,3}+U_{3,1})+f_1,$$
$$\rho\frac{\partial^2 U_3}{\partial t^2}=C_{33}U_{3,3}+C_{13}U_{1,1}+C_{55}(U_{1,1}+U_{3,1})+f_3, \tag{1}$$

with the boundary conditions at $x_3=\pm h$ given by

$$C_{55}(U_{1,3}+U_{3,1})=s_1,\ C_{13}U_{1,1}+C_{33}U_{3,3}=s_3, \tag{2}$$

where ρ is the volume density of the plate, C_{ij} are elastic constants, $\mathbf{f}=(f_1,f_3)$, and $\mathbf{s}=(s_1,s_3)$ are the body and surface force densities, respectively, induced by the laser pulse. The displacement field U_i can be expanded by the generalized Rayleigh–Lamb modes for the x_3 variable and by the Fourier transform for the x_1 variable as

$$U_i(x_1,x_3,t)=\int_{-\infty}^{+\infty}\sum_n \Phi_n(\omega_n,k,t)\psi_{in}(\omega_n,k,x_3)\times\exp(ikx_1)dk,\quad i=1\ \text{and}\ 3, \tag{3}$$

where Φ_n and $\psi_n=(\psi_{1n},\psi_{3n})^t$ are the generalized Fourier coefficients and Rayleigh–Lamb modes, respectively.[26] The normal displacement U_3 has been calculated numerically for certain elastic constants in the fiber direction, as shown in Fig. 3. Compared to Fig. 2, the theoretical results show gen-

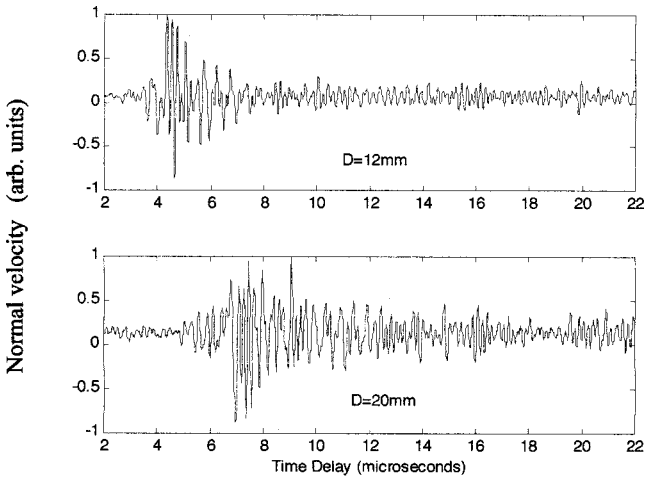


FIG. 2. Experimental waveforms propagating in the fiber direction of a plate. The waveforms are recorded at two receiver positions ($D=12$ and $D=20$ mm).

eral features that correspond to those of the experimental waveforms.

IV. INVERSION STRATEGY

A. Preprocessing techniques

It is difficult to invert directly the elastic constants from the waveforms because of their inherent complexity. Therefore, as shown in Fig. 4, the raw Lamb waveform signal is first transferred to a pre-processor. The chosen preprocessing technique is the wavelet packet analysis[28] which is used here to extract the eigenvector of the transient waveform so as to simplify the structure of the neural network. The pre-processor output is then input into a multilayer internally recurrent neural network (IRNN) with backpropagation algorithm for training and learning. The wavelet packet analysis method is a generalization of the wavelet decomposition that offers a richer range of possibilities for signal analysis. In wavelet analysis, a signal is split into an approximation and a detail. The approximation then itself is

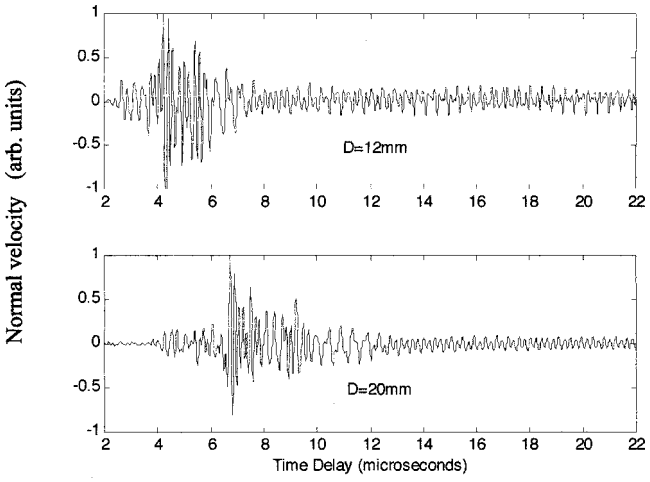


FIG. 3. Predicted waveforms propagating in the fiber direction of a transversely isotropic plate. The waveforms are calculated for two receiver positions ($D=12$ mm and $D=20$ mm).

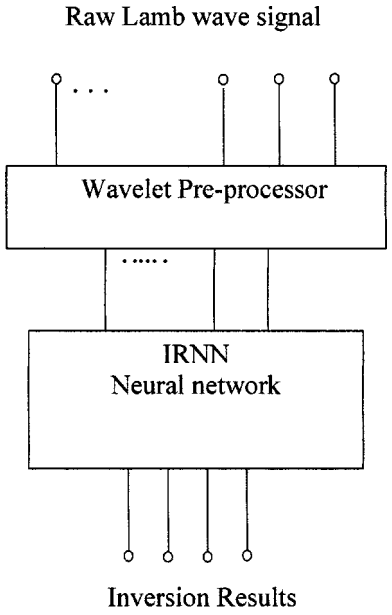


FIG. 4. Overall inversion strategy to recover the elastic constants.

split into a second-level approximation and another detail, and the process is repeated. In the wavelet packet analysis, the details as well as the approximations can be split, so it is a more complex and flexible analysis. The procedure of the wavelet packet analysis consists of three steps.

First, the transient Lamb wave is decomposed for a given wavelet by a three-layer wavelet packet, and then the wavelet coefficients of the signal are computed separately from the third layer of the tree, which describes the eight frequency components from low to high frequency. The corresponding wavelet packet decomposition tree is shown in Fig. 5, where *S* denotes the signal to be analyzed, *A* denotes low frequency, *D* denotes high frequency, and the number that follows denotes the layer number of wavelet packet decomposition. The choice of a particular wavelet function used for wavelet packet decomposition depends strongly on the transient characteristics and the information content of the signal being analyzed. In view of the matching of signals, the wavelet function used for wavelet packet decomposition here is Daubechies wavelet ($db1$), i.e., the Haar wavelet function. (It is even possible that a smoother wavelet might work even better.) The second step is to reconstruct the wavelet packet decomposition coefficients to extract the signal in each frequency band. The third and final step is to construct the eigenvector by computing the energy of the signal in each frequency band. The eigenvector is subsequently used as an input vector into the neural network.

In this paper, the raw signal (Lamb wave) is pre-

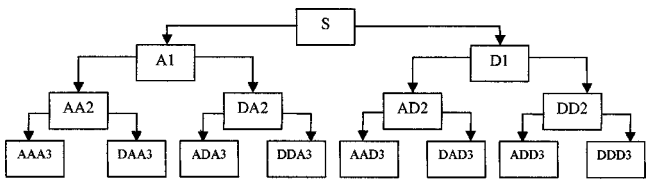


FIG. 5. Wavelet packet three-layer decomposition tree. *S* denotes the signal. *A* and *D* denote low and high frequencies, respectively.

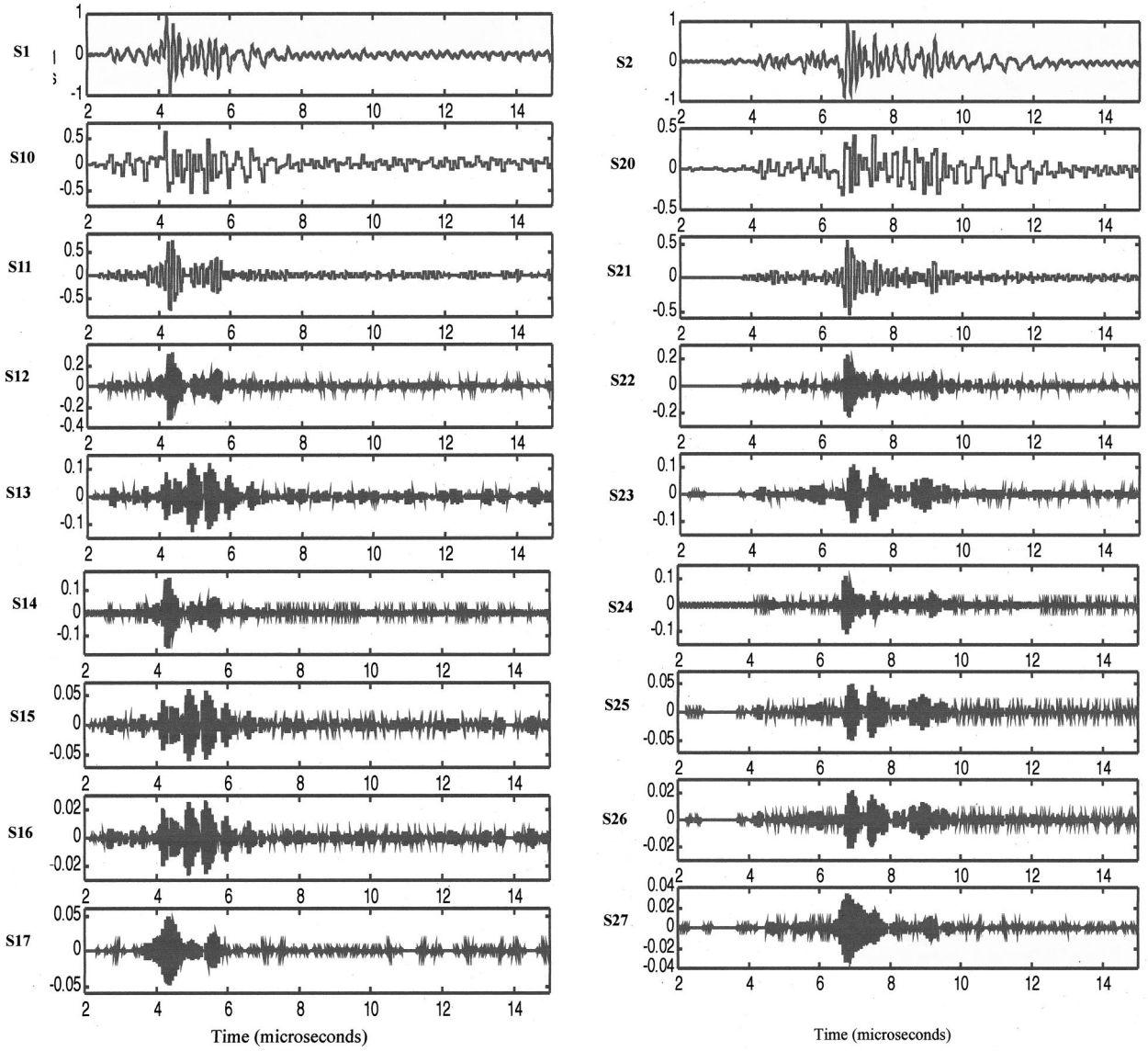


FIG. 6. Original Lamb waveforms and their eight eigen-components extracted by the wavelet decomposition. (a) For the waveform ($S1$) recorded at $D=12$ mm, and (b) for the waveform ($S2$) recorded at $D=20$ mm.

processed using the method described above so that an eigenvector is constructed. For example, the theoretical Lamb waveform signals s_1 and s_2 in Fig. 3 are decomposed in each frequency band using a three-level wavelet packet decomposition, as shown in Figs. 6(a) and (b). The signals denoted by $s_{10}\sim s_{17}$ and $s_{20}\sim s_{27}$ represent reconstructed signals within each frequency band.

B. Neural network architecture

The multilayered feed-forward neural network with back propagation algorithm has been widely used in many fields because it can be trained to realize any arbitrary mapping from an input to output by a gradient-descent based learning algorithm. However, it also has some limitations, notably a slower convergence speed, which affects practical application to a large extent. In this work, a somewhat different neural network architecture is used. It combines the multilayered feed-forward neural network and the recurrent neural network, and it is referred to as an internally recurrent neural network (IRNN).[29–31] As shown in Fig. 7, the topology of the IRNN with bias units used here is composed of an input layer of nodes, a hidden layer and an output layer of sigmoid neurons. Unlike the ordinary feed-forward neural network, it adds an additional unit to the input layer of the above network called a context or state layer. Two bias nodes are separately added to the hidden and output layer. In this architecture, the context units remember the previous internal state. Thus, the hidden units have the task of accepting both an output signal from the input layer, and also the previous internal output state of its own. Each individual node sums all of its inputs, which are formed from the products of incoming weights and the preceding node outputs. If $y(t)$ denotes the output vector at step t, $I_j(t)$ denotes the net input of node j at step t, and $x_j(t)$ denotes the output of the hidden node j at step t. The IRNN with bias units can be described as:

$$y(t)=\sum_{j=1}^{Nh} WO_j x_j(t)+WO_{\text{bias}}, \tag{4}$$

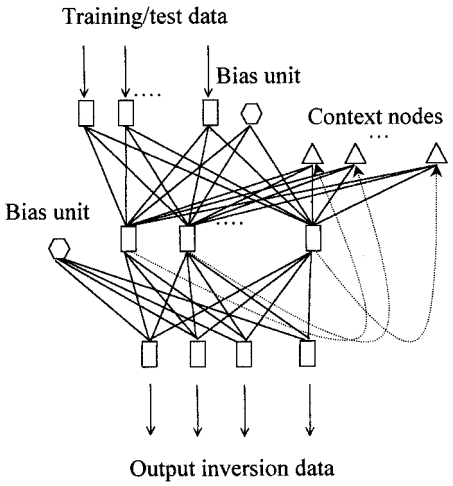


FIG. 7. Internally recurrent neural network (IRNN) with bias units.

$$x_j(k)=\sigma(S_j(t)), \tag{5}$$

$$S_j(k)=\sum_{i=1}^{NH} WR_{ij}x_i(t-1)+\sum_{i=1}^{NI} WI_{ij}I_i(k)+WI_{i\text{bias}}, \tag{6}$$

where WI is the weight from the input to the hidden layer, WR is the weight from the hidden layer to itself, WO is the weight from the hidden layer to the output layer, WI_{bias} is the weight of the bias unit added to the hidden layer, WO_{bias} is the weight of the bias unit added to the output layer, and $\sigma(\cdot)$ is an nonlinear activation function of the hidden nodes. The binary sigmoid activation function is commonly used.

The addition of the context layer and the bias nodes is convenient for the introduction of experience knowledge during the learning course, which largely improves the learning speed and overcomes the lack of a pure feed-forward neural network. Because of the similarity with the multilayered feed-forward neural network, the propagation error and the weight adjustments employed in the IRNN follow the standard back-propagation algorithm.[31] During training, a column of numerical patterns is applied to the input nodes, and fed forward through the network to the output nodes. In a supervised learning scheme, the output node values are compared with those expected from the input pattern, and any difference is interpreted as an error that is then used to adjust the weights and biases of the network using the error back propagation and gradient descent algorithm. This algorithm uses a gradient search technique for minimizing the squared error between the actual output and the desired output by iterative adapting the interconnection weights. It cycles through the training data until the error drops below a specified value or the maximum time is reached. The learning rule adopted for this work is the generalized delta rule.

V. SIMULATION RESULTS

The overall inversion method including the wavelet preprocessing and the IRNN training was evaluated using a transient Lamb waveform database developed under the normal mode expansion method proposed by Cheng *et al.*[26,27] In order to verify the feasibility and accuracy of the proposed method, the waveforms were numerically simulated for different elastic constants and first used for wavelet feature extraction and IRNN inversion. A total of 116 samples were obtained by numerical simulations for different elastic constants, where 100 samples were used for training and the remaining 16 samples were used for testing. The IRNN architecture for the implementation consisted of 8 input nodes, 12 hidden nodes, and 4 output nodes. Some inversion results are shown in Table I. The error of the inversion algorithm is within $\pm 5\%$. Finally, the experimental Lamb waveforms propagating in the fiber direction for two detecting distances (16 mm and 20 mm) are used as input into the whole system to determine the elastic constants of the material. The results are shown in Table II. The two sets of inverted constants are in reasonable agreement with each other and provide a good fit to the values for the material used in the experiment.

TABLE I. Input data and corresponding inverted values.

	C_{11} (GPa)	C_{33} (GPa)	C_{13} (GPa)	C_{55} (GPa)
Input	254.7	191.5	84.9	49.0
	251.7	196.2	87.7	50.4
Inverted	(1.2%)	(2.5%)	(2.9%)	(2.0%)
Input	254.7	191.5	101.8	49.0
	252.7	198.8	98.3	50.0
Inverted	(1.1%)	(3.7%)	(3.4%)	(2.0%)
Input	254.0	183.0	89.90	57.6
	249.3	180.0	92.5	56.0
Inverted	(1.8%)	(1.7%)	(2.8%)	(2.7%)

VI. CONCLUSIONS

An inverse method has been used to determine the elastic constants of a fiber-reinforced composite plate from laser-based ultrasonic Lamb waveforms. The method combines wavelet decomposition and neural network analysis. The salient features of the method are as follows.

(1) The elastic constants of the materials are directly recovered from the transient Lamb waveforms by combining neural networks and wavelet transforms. Neither the phase or group velocity of the Lamb wave modes are needed, as in the conventional method, and no direct least-square fitting of the experimental waveforms is necessary. The latter is not only computationally too costly, but it also involves problems with the stability of the algorithm.

(2) The structure of the neural network is simplified by pre-processing with the wavelet packet analysis to extract the eigenvectors from the signals, which consists of the energy of the signal in each frequency band. The wavelet preprocessing technique shows great potential for analyzing

TABLE II. Inverted elastic constants using experimental data measured at two source–receiver distances along the titanium-graphite composite plate.

D	C_{11} (GPa)	C_{33} (GPa)	C_{13} (GPa)	C_{55} (GPa)
16 mm	249.3	195.9	97.3	52.2
20 mm	248.5	194.2	99.6	53.6

transient Lamb wave signals. The major advantage of this technique is the ability to perform multi-resolution analysis of the signal. The localization of the signal in both time and frequency domains allows extraction of features carrying useful information about the elastic constants. Once the eigenvector is derived from the signal, the neural network offers an effective tool to recover the elastic constants with both accuracy and speed.

(3) Theoretical simulations based on the model of propagation of Lamb waves in the fiber-reinforced composite plate are used to train the neural network. This method can be extended for recovering the elastic constants from more complex experimental waveforms, as long as the theoretical elastic waveforms can be obtained in an anisotropic material.

This work provides a quantitative analysis for noncontact and nondestructive evaluation of elastic stiffness properties of the composite plate by the laser-generated Lamb wave technique.

ACKNOWLEDGMENT

The authors (J. Yang and J. C. Cheng) thank the financial supports by the National Science Foundation of China (Grant No. 19974017).

[1] J. Maynard, ''Resonant ultrasonic spectroscopy,'' Phys. Today **49**, 26–31 (1996).
[2] P. Heyliger *et al.*, ''Elastic constants of isotropic cylinders using resonant ultrasound,'' J. Acoust. Soc. Am. **94**, 1482–1487 (1993).
[3] A. Migliori and J. L. Sarao, *Resonant Ultrasound Spectroscopy* (Wiley, New York, 1997).
[4] C. Remken *et al.*, ''An experimental evaluation of the resonant ultrasonic spectroscopy method, in *Review of Progress in Quantitative NDE*, Vol. 17B, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York, 1998), pp. 2029–2044.
[5] C. B. Scruby, R. J. Dewhurst, and S. B. Palmer, ''Quantitative studies of thermally generated elastic waves in laser-irradiated metals,'' J. Appl. Phys. **51**, 6210–6216 (1980).
[6] C. B. Scruby and L. E. Drain, *Laser Ultrasonics Techniques and Applications* (Hilger, Bristol, 1990).
[7] S. Guilbaud and B. Audoin, ''Measurement of the stiffness coefficients of a viscoelastic composite material with laser-generated and detected ultrasound,'' J. Acoust. Soc. Am. **105**, 2226–2235 (1999).
[8] B. Audoin and C. Bescond, ''Measurement by laser-generated ultrasound of four stiffness coefficients of an anisotropic material at elevated temperatures,'' J. Nondestruct. Eval. **16**, 91–100 (1997).
[9] B. Audoin, C. Bescond, and M. Deschamps, ''Measurement of stiffness coefficients of anisotropic materials from pointlike generation and detection of acoustic waves,'' J. Appl. Phys. **80**, 3760–3771 (1996).
[10] Y. C. Chu and S. I. Rokhlin, ''A method for determination of elastic constants of a unidirectional lamina from ultrasonic bulk velocity measurements on [0/90] cross-ply composites,'' J. Acoust. Soc. Am. **96**, 342–352 (1994).
[11] A. G. Every and W. Sachse, ''Determination of the elastic constants of an anisotropic solid from acoustic-wave group-velocity measurements,'' Phys. Rev. B **42**, 8196–8205 (1990).
[12] T. T. Wu and Y. H. Liu, ''On the measurement of anisotropic elastic constants of fiber-reinforced composite plate using ultrasonic bulk wave and laser generated Lamb wave,'' Ultrasonics **37**, 405–412 (1999).
[13] H. Sontag and A. C. Tam, ''Optical monitoring of photoacoustic pulse propagation in silicon wafers,'' Appl. Phys. Lett. **46**, 725–727 (1985).
[14] R. J. Dewhurst, C. Edwards, A. D. W. Mckie, and S. B. Palmer, ''Estimation of the thickness of thin metal sheet using laser generated ultrasound,'' Appl. Phys. Lett. **51**, 1066–1068 (1987).
[15] D. A. Hutchins, K. Lundgren, and S. B. Palmer, ''A laser study of transient Lamb waves in thin materials,'' J. Acoust. Soc. Am. **85**, 1441–1448 (1989).
[16] L. Noui and R. J. Dewhurst, ''Two quantitative optical detection techniques for photoacoustic Lamb waves,'' Appl. Phys. Lett. **57**, 551–553 (1990).
[17] J. Chao, L. Udpa, and S. S. Udpa, ''Ultrasonic signal analysis using wavelet transform,'' Rev. Prog. Quant. Nondestr. Eval. **12**, 735–742 (1996).
[18] T. Onsay and A. G. Haddow, ''Wavelet transform analysis of transient wave propagation in a dispersive medium,'' J. Acoust. Soc. Am. **95**, 1441–1448 (1994).
[19] A. Abbate, J. Frankel, and P. Das, ''Wavelet transform signal processing applied to ultrasonics,'' Rev. Prog. Quant. Nondestr. Eval. **15**, 741–748 (1996).
[20] J. B. Han, J. C. Cheng, T. H. Wang, and Y. Berthelot, ''Mode analyses of laser-generated transient ultrasonic Lamb waveforms in a composite plate by wavelet transform,'' Mater. Eval. **57**, 837–840 (1999).
[21] J. J. Thomsen and K. Lund, ''Quality control of composite materials by neural network analysis of ultrasonic power spectra,'' Mater. Eval. **49**, 594–600 (1991).
[22] L. Udpa and S. S. Udpa, ''Neural networks for the classification of nondestructive evaluation signals,'' IEEE Proc. F **138**, 41–45 (1991).
[23] C. G. Windsor, F. Anselme, L. Capineri, and J. P. Mason, ''The classification of weld defects from ultrasonic images: A neural network approach,'' Br. J. Non-Destr. Test. **35**, 15–22 (1993).
[24] R. E. Challis, U. Bork, and C. P. D. Todd, ''Ultrasonic NDE of adhered T-joints using Lamb waves and intelligent signal processing,'' Ultrasonics **34**, 455–459 (1996).
[25] C. P. D. Todd and R. E. Challis, ''Quantitative classification of adhesive bondline dimensions using Lamb waves and artificial neural networks,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 167–181 (1999).
[26] J. C. Cheng and Y. Berthelot, ''Theory of laser-generated transient Lamb waves in orthotropic plates,'' J. Phys. D **29**, 1857–1864 (1996).
[27] J. C. Cheng, S. Y. Zhang, and Y. Berthelot, ''Experimental and theoretical analyses on laser-generated Lamb waves in an orthotropic plate,'' Rev. Prog. Quant. Nondestr. Eval. **19**, 1151–1158 (2000).
[28] B. Carnero and A. Drygajlo, ''Fast short-time orthogonal wavelet packet transform algorithms,'' Proc. ICASSP, 1161–1164 (1995).
[29] P. J. Angeline, G. M. Saunders, and J. B. Pollack, ''An evolutionary algorithm that constructs recurrent neural networks,'' IEEE Trans. Neural Netw. **5**, 54–65 (1994).
[30] R. J. Schalkoff, *Artificial Neural Networks* (McGraw-Hill, New York, 1997).
[31] C. M. Bishop, *Neural Networks for Pattern Recognition* (Clarendon, Oxford, UK, 1995).

Adaptive tuning of an electrodynamically driven thermoacoustic cooler

Yaoyu Li, Brian L. Minner, George T.-C. Chiu, Luc Mongeau, and James E. Braun
Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, West Lafayette, Indiana 47907

(Received 16 August 2000; revised 16 April 2001; accepted 31 August 2001)

The commercial development of thermoacoustic coolers has been hampered in part by their low efficiencies compared to vapor compression systems. A key component of electrodynamically driven coolers is the electromechanical transducer, or driver. The driver's electroacoustic transduction efficiency, defined as the ratio of the acoustic power delivered to the working gas by the moving piston and the electrical power supplied, must be maintained near its maximum value if a high overall system efficiency is to be achieved. Modeling and experiments have shown that the electroacoustic efficiency peaks sharply near the resonance frequency of the electro-mechano-acoustic system. The optimal operating frequency changes as the loading condition changes, and as the properties of the working gas vary. The driver efficiency may thus drop significantly during continuous operation at a fixed frequency. In this study, an on-line driver efficiency measurement scheme was implemented. It was found that the frequency for maximum electroacoustic efficiency does not precisely match any particular resonance frequency, and that the efficiency at resonance can be significantly lower than the highest achievable efficiency. Therefore, a direct efficiency measurement scheme was implemented and validated using a functional thermoacoustic cooler. An adaptive frequency-tuning scheme was then implemented. Experiments were performed to investigate the effectiveness of the control scheme to maintain the maximum achievable driver efficiency for varying operating conditions. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1413755]

PACS numbers: 43.35.Ud, 43.58.Fm [SGK]

I. INTRODUCTION

Research efforts have been made over the past 20 years to develop thermoacoustic heat pumping systems for air-conditioning, refrigeration, and other cooling applications.[1–3] These systems do not require refrigerants. They use environmentally benign, inert gases as the working fluid. They are mechanically simple, requiring no sliding seals or lubrication. At this stage of their development, the performance of thermoacoustic systems has been poor relative to traditional vapor compression systems. However, existing models based on linear theory suggest that better performance is achievable. One key to improving performance relates to the maximization of efficiency.

The overall efficiency of an electrodynamically driven thermoacoustic heat pumping device is a function of heat pumping effectiveness and electroacoustic transduction efficiency. In this study, methods to maintain optimal electroacoustic efficiency for varying operating conditions were investigated. This efficiency is defined as the ratio between the acoustic power delivered and the electrical power supplied. The acoustic power is defined as the time-averaged product of the acoustic pressure over the driving piston and the piston volume velocity. The electrical input power is the time-averaged product of the voltage fed to the coil and the corresponding coil current. Both the acoustic and the electric power are very sensitive to the phase difference between the pressure and velocity signals, or between the voltage and current signals. The acoustic power, and thus the electroacoustic transduction efficiency, is highly frequency dependent. In practical applications as well as in the laboratory, it is imperative to maintain peak efficiency in order to minimize power consumption, and thus operating costs. In operation, variations in refrigeration loads cause the temperature of the working gas and the heat exchangers to vary, which alters the acoustic load impedance and thus the input electrical impedance. This in turn changes the frequency of optimal electroacoustic efficiency of the driver, referred to here as the ''tuned frequency.''

The design of thermoacoustic coolers has been the object of many previous studies. Minner *et al.*[4] performed systematic design optimization for a few different applications. Grant[5] investigated the use of a passive mass element to replace the gas-filled half-wavelength resonance tube. The selection of electrodynamic drivers for thermoacoustic refrigerators was discussed recently in a study by Wakeland,[6] which investigated the effects of key parameters and frequency on driver efficiency. The optimization processes utilized in these[4–6] and other studies may be designated as *a priori* or ''off-line'' tuning since they do not require the knowledge (in real-time) of measured quantities. In practice, thermoacoustic phenomena often involve nonlinear, time-varying processes affected by many operating and design factors. Even carefully designed and fabricated devices are subject to uncertainty, and variability which cannot be accounted for using existing linear models.[7] From a control system viewpoint, the system needs to adapt to transient variations of the plant. It must be able to tune itself without user intervention, *a posteriori* or after the initial tuning pro-

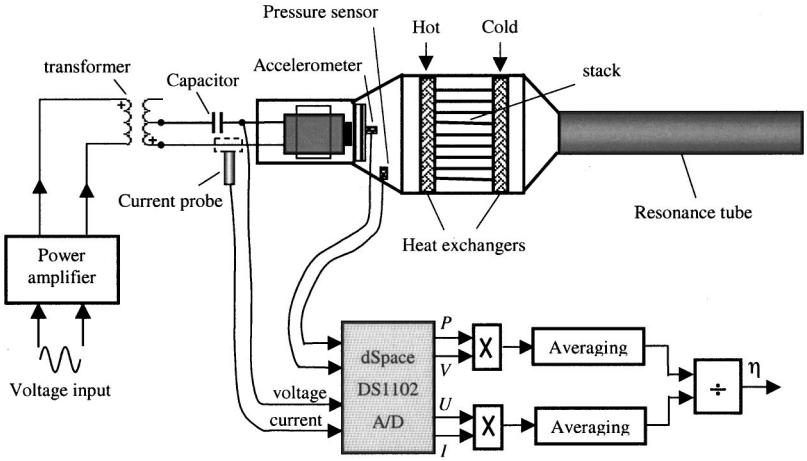


FIG. 1. Sketch of the thermoacoustic cooler and the on-line efficiency measurement scheme.

cess, utilizing dynamic signals fed to and processed by a digital processor. This is referred to as ''on-line'' tuning.

On-line monitoring and control of electroacoustic efficiency to maintain peak efficiency for varying operating conditions is possible. Garrett[8] developed a passive method of frequency compensation for changes in the temperature of the working gas by placing a gas mixture and an adsorbent within the resonator. Hofler[9] used a method for real-time acoustic power measurements. In Hofler's work, a voltage-controlled-frequency (VCF) signal generator was used along with a proportional-plus-integral (PI) controller to tune the driving frequency for acoustic resonance, i.e., for pressure and velocity to be in phase on the piston. Another PI controller was used to maintain a constant output pressure amplitude. Issues such as filtering and overload protection were considered. The system was implemented using analog circuits.

In the present study, a method for acoustic power measurement similar to that in Ref. 9 was used. Real-time efficiency monitoring and control schemes were implemented using a DSP (digital signal processor). The electroacoustic efficiency is not a directly measurable quantity. It is usually calculated ''off-line'' from measured pressure, piston velocity, coil voltage, and current data. Either time or frequency domain methods can be used to calculate the acoustical and electrical power. The approach here utilized digital signal processing techniques to calculate the instantaneous acoustic and electrical power of the driver for a prototype thermoacoustic cooler on-line. The instantaneous driver efficiency was supplied to a real-time adaptive frequency tuning algorithm to ensure maximum driver efficiency for changing and uncertain operating conditions.

The remainder of the paper is organized as follows. A brief description of the prototype thermoacoustic cooler is given in the next section, followed by the description and the analysis of the on-line driver efficiency measurement technique. It is then shown that there can be a significant efficiency loss if the system is operated at resonance, and that a direct measurement of efficiency is required. The real-time adaptive algorithm for tuning the electromechanical driver is finally described. Some implementation details are included in the hope that they may be of some assistance to other workers in the field.

II. DESCRIPTION OF THE THERMOACOUSTIC COOLER

The thermoacoustic cooler is shown schematically in Fig. 1. The system was driven by a moving-magnet CFIC Model B-300 electromechanical transducer. The driver was designed to deliver 300 W acoustic power at 33 Hz, with an electroacoustic transduction efficiency of 70% and a maximum displacement of 6 mm. The alternator was comprised of radially extending neodymium magnet arms mounted to a central plunger (the moving element) and copper wound laminated iron pole pieces (the stationary element), which extend inward from the outer periphery. The outward-directed magnet fingers (each comprised of two magnets in contact) were thus separated by gaps, which were filled by the inward-directed laminated iron fingers. The magnetic flux pathlines in such devices go circumferentially through a magnet finger and a pole piece, radially up through the coil to the outer periphery, circumferentially to the other pole piece adjacent to the magnet finger, radially down through the coil and back to the magnet finger. Each pole piece then supports two flux paths, one for each adjacent magnet finger. The alternating current causes a fluctuated force that moves the magnet-mounted plunger axially back and forth. The plunger-mounted piston was mounted within a bore, with extremely tight clearance seals extending over a length many times the piston displacement. This minimized ''blow down,'' or gas leakage between the rear and front cavities. The plunger was balanced and self-centering, and supported by a single flexure. Leaf springs were added to increase the resonance frequency of the driver in-vacuo to 154 Hz. The driver was fully enclosed, and the static pressure equalized across the piston.

The measured specifications of the driver are listed in Table I. The equivalent linear parameters of the driver, described for example in Beranek's text,[10] were measured. They were found to depend on amplitude and frequency. The values listed in Table I are thus only approximations over a narrow range of piston displacements. More information on the driver characteristics is available in Refs. 11 and 12.

The other components of the system included (1) a stack, made of rolled up polyester film, (2) finned-tube copper heat exchangers, and (3) a pressure vessel. The working

TABLE I. Specifications of CFIC-B300 driver.

Parameter	Notation	Value	Unit
Force constant	Bl	9	N/A
Stiffness	K	74	kN/m
Mech. resistance	b	25	N s/m
Mass	M	1.63	kg
Piston area	S	0.0068	m^2
Coil resistance (DC)	R	0.11	Ω
Coil inductance	L	0.9	mH

fluid was a 55% helium–argon mixture with a mean pressure up to 2 MPa. A single-frequency sinusoidal signal was fed to a power amplifier (TECHRON Model 5530) using a signal generator. Since the power amplifier itself could not supply the required current (up to 20 A-rms), a transformer was installed between the amplifier and the coil. Two TALEMA UR0500 500-VA transformers connected in parallel were used, each with a 1:5 winding ratio. The resulting maximum loading condition was 21 A-rms at 24 V-rms. A capacitor was connected in series with the coil to improve the power factor of the device. A PCB U353B16 piezoelectric accelerometer was attached to the moving part of the driver, and a PCB dynamic pressure sensor was installed in the duct near the piston. The coil voltage was measured with a voltage divider, and the coil current was measured using a HP 1146A current probe.

III. ONLINE EFFICIENCY MEASUREMENT PROCEDURES

Recall that the acoustic power and electrical power are defined as

$$\mathrm{AP}=\frac{1}{T}\int_0^T \vec{P}\cdot\vec{V}\,dt, \tag{1}$$

$$\mathrm{EP}=\frac{1}{T}\int_0^T \vec{U}\cdot\vec{I}\,dt, \tag{2}$$

where AP and EP denote acoustic power and electrical power, respectively. Driver pressure, piston velocity, input voltage, and current are denoted by P, V, U, and I, respectively. The arrows indicate that these are complex variables. The averaging time interval, T, was chosen to be an integral number of periods. The discrete form of these equations is

$$\mathrm{AP}=\frac{1}{N}\sum_{i=1}^{N} P(i)\,V(i), \tag{3}$$

$$\mathrm{EP}=\frac{1}{N}\sum_{i=1}^{N} U(i)\,I(i), \tag{4}$$

where N is the number of averaging time steps. The basic on-line efficiency measurement scheme is shown in Fig. 1. The signals for the four measured quantities were sent to the *A*/*D* ports of a dSPACE DSP controller board. The piston velocity was obtained by integrating the signal from the accelerometer. All real-time computations and the filtering processes were performed using a TI TMSC31 DSP. MATLAB® Real-Time Workshop and SIMULINK were used for developing the embedded system software.

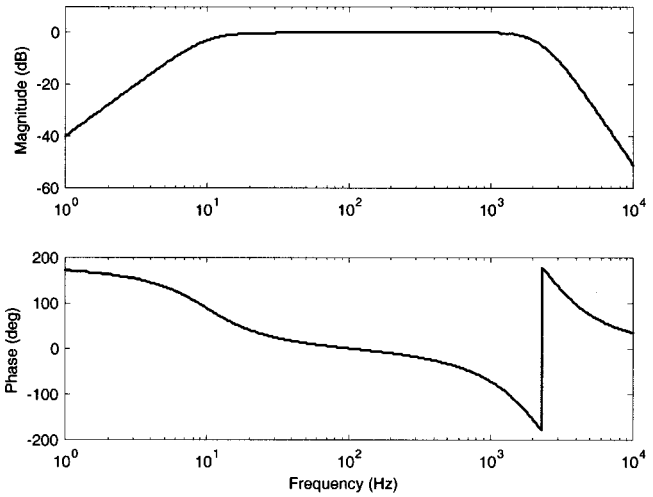


FIG. 2. Frequency response of the band-pass filter for the velocity measurements.

To compensate for the DC drift of the accelerometer signal, a high-pass filter was used to eliminate DC offsets in the integrated signal. High frequency noise was minimized through the application of a low-pass filter. The combined band-pass filter frequency response was

$$F_{\mathrm{bp}}(s) = \frac{1.578\,64\times10^{8}(s^2+8.796\times10^{-3}s+3.9478\times10^{-5})}{(s^2+879.6s+63\,165)(s^2+17\,592.9s+1.579\,16\times10^{8})}, \tag{5}$$

where $s=j\omega$ is the Laplacian operator, with ω the angular frequency (rad/s). A Bode plot of the filter's frequency response is shown in Fig. 2. The -3 dB criterion was used for a 10 Hz to 2000 Hz pass band. Since the thermoacoustic cooler driving frequency was between 100 and 200 Hz, the pass band response for the driving frequency range was within 0.002 dB. To compensate for the filter's phase response, the same band-pass filter was applied to the pressure signal. Note that the band-pass filter was implemented in the Simulink program for the efficiency measurement.

Averaging a data sequence over a time interval is equivalent to a low pass FIR (Finite Impulse Response) filtering process with a very low cutoff frequency and a large roll-off. The computational requirements of averaging over a large number of data samples to obtain acceptable variance in the estimates may limit the bandwidth of the efficiency measurement. To solve this problem, an IIR (Infinite Impulse Response) filter was employed instead. It is known that

$$F_n(z^{-1})=\frac{1-2z^{-1}+z^{-2}}{1-2rz^{-1}+r^2z^{-2}} \quad (r<1) \tag{6}$$

is a notch filter with a notch at DC,[13] where the discrete-time operator is $z=e^{sT_s}$, with s the Laplacian operator and T_s the sampling period. So the complementary filter

$$F_{cn}(z^{-1})=1-F_n(z^{-1})=\frac{2(1-r)z^{-1}+(r^2-1)z^{-2}}{1-2rz^{-1}+r^2z^{-2}} \tag{7}$$

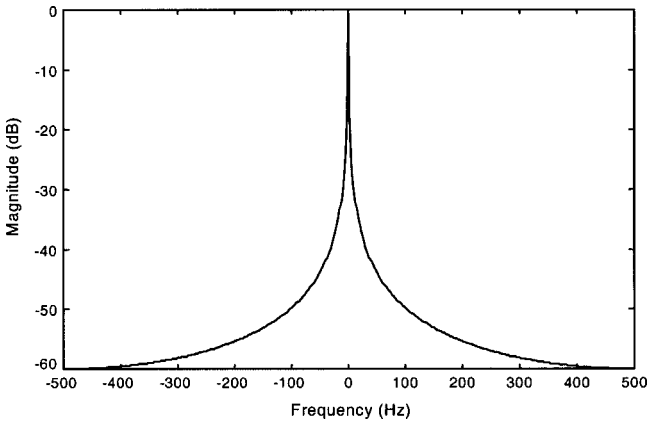


FIG. 3. Frequency response of the IIR filter used to compute the average power.

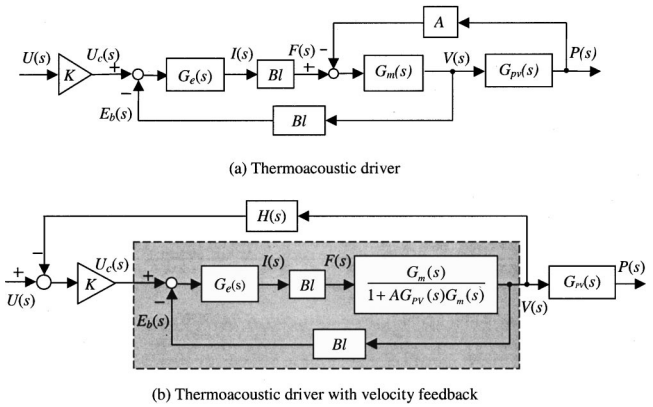


FIG. 5. Block diagrams of thermoacoustic driver.

acts as a low-pass filter. When r gets very close to 1, a steeper notch is obtained for (6) and thus a sharper roll-off is obtained for (7). The frequency response of such a filter for $r = 0.99$ and a sampling rate of 1 kHz is plotted in Fig. 3. A 40 dB roll-off is obtained at about 30 Hz. The computation time for averaging 100 samples was about 220 μs. Use of an IIR filter reduced the computational time to 15 μs. Both digital filters were implemented in the Simulink program.

A comparison was made between the on-line approach just described and an off-line approach. The off-line measurement was performed using a Brüel & Kjær PULSE system to acquire the four sensor signals. The acoustical and electrical power were obtained from the cross spectra of the signals using 1 second long records and a 2 kHz sampling rate. The results in the vicinity of the best efficiency frequency are shown in Fig. 4. Excellent agreement was obtained, except around the frequency of maximum efficiency where the largest relative error in efficiency was about 4%, which was deemed acceptable. Note that the error on tuned frequency is very small.

IV. EFFICIENCY ANALYSIS

Analytical expressions were derived for the electrodynamic efficiency of the driver in terms of the acoustic input impedance of the thermoacoustic cooler at the piston. The effects of feedback control and those of adding electrical capacitors for power factor improvement were evaluated.

Figure 5(a) shows a block diagram of the electrodynamic driver with piston velocity feedback. A transfer-function formulation is used in the following, which is consistent with the impedance-based circuit formulation used elsewhere.[10] Transfer function models are convenient for feedback control. In this figure, $U(s)$ is the driving voltage, K is the power amplification factor, $U_c(s)$ is the voltage on the output side of the amplifier, and $G_e(s)=1/[Ls+R+(1/Cs)]$ is the electrical impedance of the coil in series with a capacitor. The added capacitor created a second order pole in the system frequency response. This reduced the phase difference between delivered voltage and current, thereby reducing the power required by the amplifier for the same delivered power. This enabled a higher electrical power output around the driving frequency. Bl is the electromechanical transduction coefficient (N/A). The mechanical impedance of the driver itself is $G_m(s)=s/ms^2+bs+k$, where m, b, and k are the mass, the mechanical resistance, and the stiffness of the driver, respectively. The transfer function $G_{pv}(s)$ is the equivalent mechanical impedance of the acoustic load at the piston, denoting the relationship between particle velocity and sound pressure at that point. A is the piston area. The sound pressure exerts a force on the driver, represented in the diagram as a negative feedback at the input of

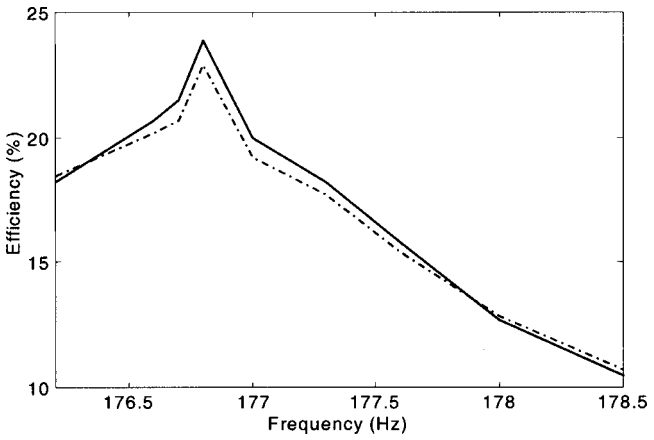


FIG. 4. Efficiency versus frequency.

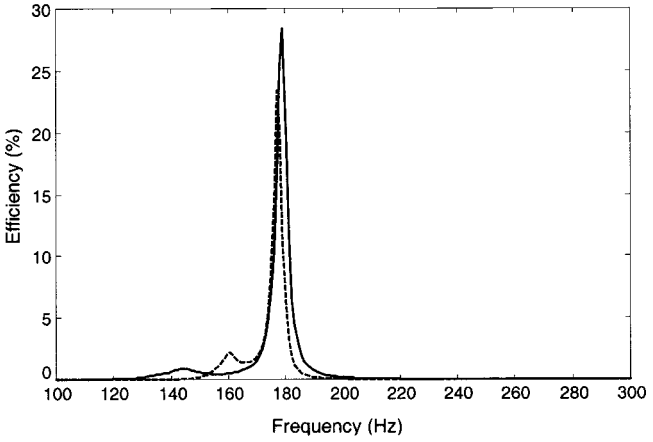


FIG. 6. Electrodynamic efficiency versus frequency.

the driver element. The driver velocity results in a ''back-emf'' $E_b(s)$ fed back to the input of the electrical element.

An equivalent block transformation including an admissible feedback controller, $H(s)$, is shown in Fig. 5(b). The frequency response of the electroacoustic efficiency of the driver was obtained from the block diagram. The transfer functions from the input voltage to the piston velocity and to the acoustic pressure are given by

$$\frac{V(s)}{U(s)}=\frac{(Bl)G_m(s)Cs}{[LCs^2+RCs+1][1+AG_{PV}(s)G_m(s)]+(Bl)[Bl+KH(s)]G_m(s)Cs} \tag{8}$$

and

$$\frac{P(s)}{U(s)}=\frac{G_m(s)G_{PV}(s)Cs}{[LCs^2+RCs+1][1+AG_{PV}(s)G_m(s)]+(Bl)[Bl+KH(s)]G_m(s)Cs}. \tag{9}$$

The acoustic power delivered by the piston is given by $\mathrm{AP}(\omega)=\frac{1}{2}A|P(j\omega)||V(j\omega)|\cos(\Phi_{PV})$, where A is the piston area and Φ_{PV} is the phase difference between the pressure and velocity. The transfer functions from input voltage to the coil current and to the coil voltage are

$$\frac{I(s)}{U(s)}=\frac{[1+AG_{PV}(s)G_m(s)]Cs}{[LCs^2+RCs+1][1+AG_{PV}(s)G_m(s)]+(Bl)[Bl+KH(s)]G_m(s)Cs} \tag{10}$$

and

$$\frac{U_c(s)}{U(s)}=\frac{Cs\{[Ls+R][1+AG_{PV}(s)G_m(s)]+(Bl)[Bl+KH(s)]G_m(s)\}}{[LCs^2+RCs+1][1+AG_{PV}(s)G_m(s)]+(Bl)[Bl+KH(s)]G_m(s)Cs}. \tag{11}$$

The electrical power is given by $\mathrm{EP}(\omega)=\frac{1}{2}|U_C(j\omega)||I(j\omega)|\cos(\Phi_{UcI})$, where Φ_{UcI} is the phase angle between the coil voltage and current. The electroacoustic efficiency is then

$$\eta(\omega)=\frac{\mathrm{AP}(\omega)}{\mathrm{EP}(\omega)}=\frac{A|G_{PV}(j\omega)|\left|\frac{V(j\omega)}{U(j\omega)}\right|^2\cos(\Phi_{PV})}{\left|\frac{U_C(j\omega)}{U(j\omega)}\right|\left|\frac{I(j\omega)}{U(j\omega)}\right|\cos(\Phi_{UcI})}. \tag{12}$$

Substituting Eqs. (8), (10), and (11) into Eq. (12) we have

$$\eta(\omega)=\frac{A|G_{PV}(s)|(Bl)^2|G_m(s)|^2\cos(\Phi_{PV})}{|1+AG_{PV}(s)G_m(s)||[Ls+R][1+AG_{PV}(s)G_m(s)]+(Bl)^2|\cos(\Phi_{UcI})}. \tag{13}$$

Equation (13) shows that the electroacoustic efficiency is related to driver parameters and to the acoustic impedance of the tube. Note that the capacitance C does not appear in Eq. (13). This shows that the addition of a capacitor in series with the coil does not affect efficiency.

The predicted and measured electroacoustic efficiency responses of the thermoacoustic system operating with a 55% helium–argon mixture at 2 MPa are shown in Fig. 6. The maximum efficiency is found within a narrow frequency range near the overall system resonance frequency, at which the electrical input reactance is zero. This resonance varies for different static pressures and different refrigeration loading conditions. For example, the resonance frequency increases from about 173 Hz to 178 Hz for an increase in mean pressure from 0.5 to 2 MPa.

To determine the impact of velocity feedback on efficiency, the numerator and denominator of Eq. (12) may be divided by $|U_c(j\omega)|^2$. The efficiency response becomes

$$\eta(\omega)=\frac{\mathrm{AP}(\omega)}{\mathrm{EP}(\omega)}=\frac{A|G_{PV}(j\omega)|\left|\frac{V(j\omega)}{U_c(j\omega)}\right|^2\cos(\Phi_{PV})}{\left|\frac{I(j\omega)}{U_c(j\omega)}\right|\cos(\Phi_{UcI})}. \tag{14}$$

The terms on the right-hand side of Eq. (14) involve only the quantities within the shaded block in Fig. 5(b), i.e., inside the feedback loop. None of them is influenced by velocity feedback. Therefore velocity feedback does not affect the efficiency for this particular problem. A similar conclusion was also obtained for current feedback.

Hofler[9] used a phased-lock loop to enforce operation of the electrodynamic driver at the acoustic resonance frequency. This was considered an ''indirect efficiency control.'' The acoustic resonance may not always correspond to

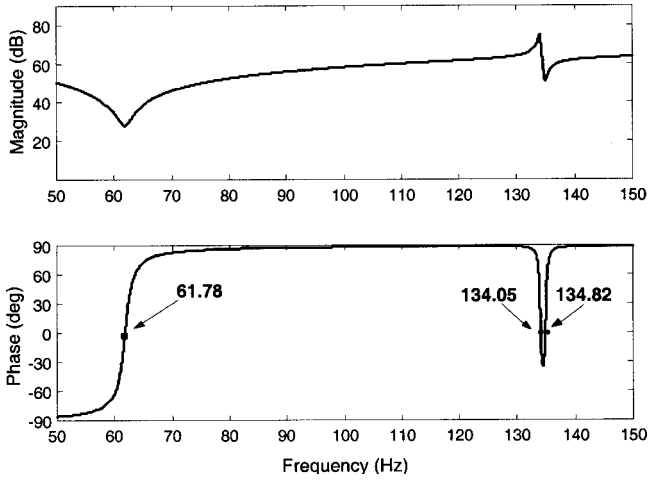


FIG. 7. Frequency response of total mechanical impedance F/V.

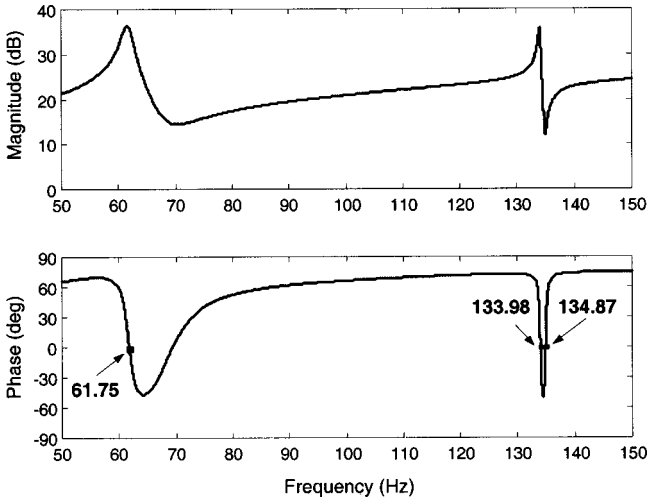


FIG. 8. Frequency response of system impedance U_cI.

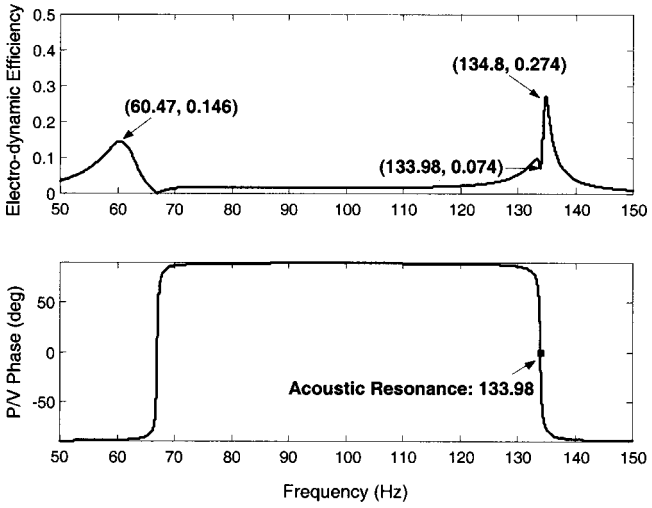


FIG. 9. Efficiency and phase of acoustic impedance versus frequency.

TABLE II. Variable of interest in the frequency tuning scheme.

f_0	Previous driving frequency	f	Next driving frequency
Δf_0	Previous frequency shift	Δf	Next frequency shift
e_0	Previous efficiency	e	Currently measured efficiency

the optimal operating frequency. To show this, the models described earlier were exercised for a mechanical stiffness of 300 KN/m, a mechanical resistance coefficient 25 N s/m (these values are representative of the driver), and assuming the working gas was air at mean pressure of 1 MPa. Other parameters are shown in Table I. All computation was performed over a frequency range 50–150 Hz with a frequency interval of 0.01 Hz.

Here the specific acoustic impedance is defined as the ratio of sound pressure and particle velocity at the driver. The driver mechanical impedance is the ratio of force and piston velocity for the driver in vacuo. The total mechanical impedance is the ratio of force and piston velocity in presence of the acoustic load. Referring to Fig. 5(a), the total impedance is the sum of the driver and the acoustic impedance, i.e.,

$$\frac{F}{V}=\frac{1}{G_m}+G_{PV}A=\left(\frac{ms^2+bs+k}{s}\right)+G_{PV}A. \tag{15}$$

The system impedance is defined as the coil voltage (after the capacitor) over current, U_cI (Fig. 5). Figure 7 shows the frequency response of the total mechanical impedance. The anti-resonance at 61.78 Hz is related to the driver response.

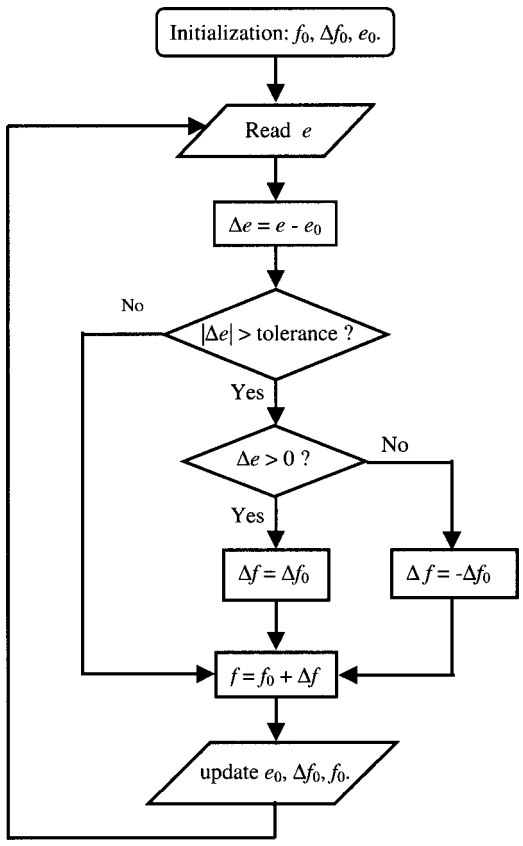
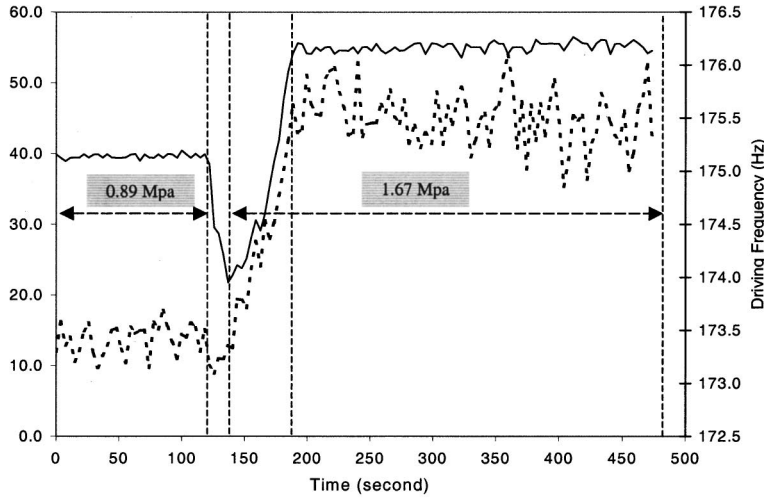
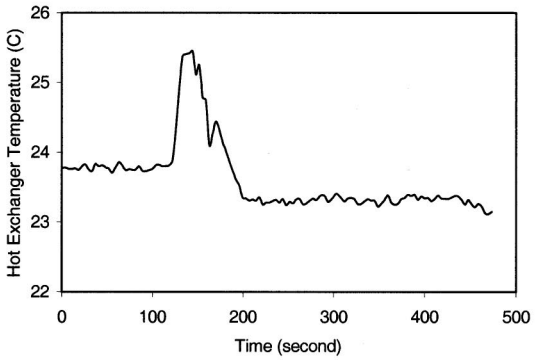


FIG. 10. The algorithm for adaptive frequency tuning.



(a) Efficiency and driving frequency vs. time. ——: Efficiency (%); ----: Frequency (Hz)

FIG. 11. Experimental results of the "pressure jump" test.



(b) Hot heat exchanger temperature vs. time.

The resonance at 134.05 Hz and the anti-resonance at 134.82 Hz are due to the acoustic load. Figure 8 shows the frequency response of the system impedance. Similarly, the resonance at 61.75 Hz is due to the driver. The resonance at 133.98 Hz and the antiresonance at 134.87 Hz are due to the acoustic load.

Figure 9 shows the driver efficiency and the phase response of the acoustic impedance of the tube. There are two peaks (local maxima) for efficiency, related to the antiresonance of the mechanical driver and one acoustic duct resonance. This is an agreement with the results in Ref. 6. In this example, the overall maximum efficiency is achieved at 134.8 Hz, close to the acoustic resonance at 133.98 Hz. The other local maximum efficiency at 60.47 Hz is close to the antiresonance of the mechanical driver at 61.78 Hz. Increasing stiffness or reducing the damping of the driver may change the relative amplitudes of these maxima.

The global maximum efficiency at 134.8 Hz is close to but not the same as the acoustic resonance at 133.98 Hz, and also different from the neighboring resonance/antiresonances in Figs. 7 and 8. Similarly, the local maximum efficiency at 60.47 Hz is close to the anti-resonance at 61.78 Hz in Fig. 7 and the resonance at 61.75 Hz in Fig. 8. This discrepancy is due to the complexity of the interplay among different parameters of the system, as evidenced in Eqs. (12), (13), and (14). Peak efficiencies (either local or global) do not perfectly coincide with any resonance or antiresonance condition. For example, in Fig. 9, the efficiency at the acoustic resonance frequency (133.98 Hz) is 7.4% which is about 4 times less than the optimal value of 27.4%. This occurs for a mere 0.8 Hz change in frequency. Similar observations were reported by Hunt:[14] "*While it is customary to assume that the maximum efficiency always occurs at 'resonance', a closer examination of Eqs (4.19) and (4.20) reveals that a still higher efficiency can be obtained for some other value of frequency or reactance than the one for which* $p=0=X_M+X_L$*.*"

V. ADAPTIVE FREQUENCY TUNING PROCEDURES

The adaptive frequency tuning scheme was implemented using a dSPACE DS1102 controller board (Refs. 15 and 16). Table II lists the variables of interest. A flow diagram for the frequency tuning algorithm is shown in Fig. 10. The frequency, frequency shift, and efficiency are first initialized. After reading the efficiency value, a new target efficiency is calculated. If the efficiency change does not exceed a predefined threshold, the frequency change is kept the same as for the previous step. If the efficiency change is significant, a decision on the next frequency adjustment is made depending on if the change is positive of negative. If the efficiency is increasing, the sign of the adjustment is kept; otherwise, the sign is reversed.

To verify the effectiveness of this adaptive tuning algorithm, a test was designed to expose the thermoacoustic cooler to a ''sudden'' change in operating conditions (not necessarily representative of typical transients during normal operation). The mean pressure of the cooler was increased from 0.89 MPa to 1.67 MPa over a time period of 20 seconds. The measured electrodynamic efficiency and the corresponding driving frequency are shown as functions of time in Fig. 11(a). The test started at the instant $t=120$ s. Initially, the driver efficiency dropped from 40% to about 21% due to the change in resonance frequency. The performance degraded dramatically while the driving frequency was kept constant. The adaptive tuning, within about 45 seconds, allowed the system performance to reach a new peak efficiency of 55% at a new driving frequency of 176.2 Hz. During the pressurization process, the temperature within the tube also experienced a transient evolution. The temperature profile near the center of the heat exchanger on the hot side is shown in Fig. 11(b). A temperature fluctuation around 4 F (2 C) was observed.

VI. CONCLUSIONS

An on-line efficiency measurement scheme was implemented for a functional electrodynamically driven thermoacoustic cooler. It was found by analysis that regular feedback should not change the optimal electrodynamic efficiency of the driver. The acoustic resonance frequency does not match the frequency for optimal efficiency, and a significant performance drop can result from operation at acoustic resonance. An adaptive frequency tuning algorithm was proposed and implemented to maintain the driver at its maximum efficiency for varying operating conditions. A ''pressure jump'' test was performed to illustrate the effectiveness of the controller in dealing with transients in the tuned frequency. The approach can deal with most changes in operating conditions during normal operation.

[1] G. W. Swift, ''Thermoacoustic engines,'' J. Acoust. Soc. Am. **84**, 1145–1180 (1988).
[2] G. W. Swift, ''Thermoacoustic engines and refrigerators,'' Phys. Today **48**, 22–28 (1995).
[3] S. L. Garrett, and S. Backhaus, ''The power of sound,'' Am. Sci. **88**, 516–525 (2000).
[4] B. L. Minner, J. E. Braun, and L. G. Mongeau, ''Theoretical evaluation of the optimal performance of a thermoacoustic refrigerator,'' ASHRAE Trans. **103**, 873–887 (1997). ASHRAE, Atlanta, GA, USA.
[5] L. A. Grant, Master thesis, Naval Postgraduate School, 1992.
[6] R. S. Wakeland, ''Use of electrodynamic drivers in thermoacoustic refrigerators,'' J. Acoust. Soc. Am. **107**, 827–832 (2000).
[7] B. Ward and G. Swift, Tutorial and User's Guide for DELTAE: Design Environment for Low-Amplitude Thermoacoustic Engines, Version 4.6, 1998. Available at no charge from the Los Alamos National Laboratory web site: www.lanl.gov/thermoacoustics.
[8] S. L. Garrett, Passive Frequency Stabilization in an Acoustic Resonator, U.S. Patent No. 5857340, 1999.
[9] T. Hofler, ''Accurate acoustic power measurements with a high-intensity driver,'' J. Acoust. Soc. Am. **83**, 777–786 (1988).
[10] L. L. Beranek, *Acoustics* (The Acoustical Society of America, Washington, DC, 1996).
[11] G. A. Yarr and J. A. Corey, Linear Electrodynamic Machine, U.S. Patent No. 5389844, 1995.
[12] CFIC Inc., www.cficinc.com, Troy, NY.
[13] J. G. Proakis and D. G. Manolakis, *Introduction to Digital Signal Processing* (Macmillan, New York, 1988).
[14] Hunt, *Electroacoustics* (Harvard University Press, Cambridge, MA, 1954).
[15] DSPACE GmbH, User's Guide: Floating-Point Controller Board, 1996.
[16] DSPACE GmbH, User's Guide: Real-Time Interface to Simulink 2, 1996.

Single half-wavelength ultrasonic particle filter: Predictions of the transfer matrix multilayer resonator model and experimental filtration results

Jeremy J. Hawkes[a] and W. Terence Coakley
School of Biosciences, Cardiff University, Park Place, Cardiff CF10 3TL, United Kingdom

Martin Gröschl and Ewald Benes
Institut für Allgemeine Physik, Technische Universität Wien, Wiedner Hauptstrasse 8-10/134, A-1040 Wien, Austria

Sian Armstrong and Paul J. Tasker
Centre for High Frequency Engineering, School of Engineering, Cardiff University, Newport Road, Cardiff CF24 3TF, United Kingdom

Helmut Nowotny
Institut für Theoretische Physik, Technische Universität Wien, Wiedner Hauptstrasse 8-10/136, A-1040 Wien, Austria

(Received 26 June 2001; revised 6 November 2001; accepted 26 November 2001)

The quantitative performance of a ''single half-wavelength'' acoustic resonator operated at frequencies around 3 MHz as a continuous flow microparticle filter has been investigated. Standing wave acoustic radiation pressure on suspended particles (5-μm latex) drives them towards the center of the half-wavelength separation channel. Clarified suspending phase from the region closest to the filter wall is drawn away through a downstream outlet. The filtration efficiency of the device was established from continuous turbidity measurements at the filter outlet. The frequency dependence of the acoustic energy density in the aqueous particle suspension layer of the filter system was obtained by application of the transfer matrix model [H. Nowotny and E. Benes, J. Acoust. Soc. Am. **82**, 513–521 (1987)]. Both the measured clearances and the calculated energy density distributions showed a maximum at the fundamental of the piezoceramic transducer and a second, significantly larger, maximum at another system's resonance *not* coinciding with any of the transducer or empty chamber resonances. The calculated frequency of this principal energy density maximum was in excellent agreement with the optimal clearance frequency for the four tested channel widths. The high-resolution measurements of filter performance provide, for the first time, direct verification of the matrix model predictions of the frequency dependence of acoustic energy density in the water layer. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1448341]

PACS numbers: 43.35.Yb, 43.35.Bf, 43.20.Ks, 43.25.Qp [RW]

I. INTRODUCTION

Particles suspended in a fluid exposed to an ultrasonic standing wave experience an acoustic radiation pressure that drives them towards the (usually pressure) nodes of the acoustic field, where they form ''bands'' of particles bounded by clarified medium.[1–8] The particles can also then experience a transverse radiation force that causes them to aggregate in localized clumps within the plane of the pressure node.[7,9–12] The discrete clumps thus sediment much more rapidly than individual particles, to give a clarified supernatant phase if the sound-field pressure envelope is amplitude modulated, e.g., by pulsing.[11] Particle clumps are also easily entrained and concentrated in frequency-modulated ''drifting''[13,14] or quasistationary fields[11,15] and can be concentrated in one preferred location in a multiwavelength standing wave.[11]

Significant developments of continuous flow ultrasonic particle filters have occurred in recent years.[16–22] A highly efficient filter based on ultrasound-enhanced sedimentation has been commercially developed for perfusion culture of mammalian cell fermentors where the bioproduct is of high value.[16–18,22] Essentially, cells in a suspension flowing out of the fermentor in the vertical direction are trapped in the standing wave so that the aggregates sediment back, against the outflow, into the fermentor. Ultrasonic filters that are not dependent on sedimentation, and are thus particle-concentration-independent, have also been developed.[23,24] These narrow channel filters are formed between plane transducer and reflector plates that are typically only one-half-wavelength of the acoustic wave in the fluid channel apart. Flow parallel to the surfaces in this narrow chamber is laminar, with a parabolic velocity profile.[23] Particles in that flow are rapidly concentrated close to the axial plane of the chamber as they pass across the transducer. The concentrated particles can be separated downstream from the clarified liquid that is closer to the walls. Proof of principle experiments showed that large (100-μm diameter) latex particles flowing in a 250-kHz single half-wavelength (3 mm) separation channel were very significantly more concentrated in the out-

[a]Electronic mail: HawkesJJ@Cardiff.ac.uk

flow from the axial region compared to that from the wall region.[23] Hawkes and Coakley[24] recently described a high-frequency (3-MHz, 0.25-mm channel) high radiation pressure system that can achieve 1000-fold concentration reduction of 5-μm yeast cells at a flow rate of 6 ml min^{-1}. These authors found good agreement between experimental results on particle concentration and predictions of a numerical model of the behavior of particles exposed to the standing wave radiation pressure and drag in the parabolic-profile flow.

In the work presented here an acoustic particle filter, where the thickness of the suspension layer is of the order of a single half-wavelength, is investigated with respect to filter performance (particle clearance). Filter properties are measured and calculated, whereby calculations are based on a transfer matrix model of piezoelectric multilayer resonators developed by Nowotny and Benes.[8,25,26] The calculated frequency dependence of the stored acoustic energy density given by the transfer matrix model in the different layers of the multilayered filter system is presented. These results are compared with the experimentally determined frequency dependence of the filtration efficiency of the system. It is shown that this frequency dependence is well described by the model. The approach described provides the first correlation between experimental ultrasonic suspension clarification and energy density simulations over a wide frequency range.

II. METHODS

A. Experiment

1. Particle suspension

Test suspensions were prepared by diluting a suspension of 4.89-μm diameter polystyrene latex (Polymer Labs., Church Stretton, UK) in degassed water to a concentration of $7.6 \cdot 10^6$ ml^{-1}.

2. Standing wave chamber

The construction of the stainless-steel filter chamber is shown schematically in Fig. 1. A 30×30-mm^2 PZ26 piezoceramic (Ferroperm, Krisgard, Denmark) was used as the ultrasonic transducer. Both electrodes of the transducer were etched to central dimensions of 20-mm length in the direction of flow and 10 mm in the orthogonal direction. The separation channel had a cross section of 10 mm by *l*, with *l* denoting the channel width in the direction of sound propagation. For the experiments described here, the channel width *l* was chosen as 0.18, 0.24, 0.26, or 0.30 mm. The thicknesses and other relevant material parameters of the different layers of the system are given in Table I. A particle suspension is pumped upwards through the chamber as previously described.[24] Laminar flow develops as the fluid passes 40 mm along the channel before entering the 20-mm-long ultrasound field. Acoustic radiation pressure drives the particles towards the single pressure node in the center of the channel, forming a band of particles along the chamber axis. The band remains intact in the laminar flow on leaving the sound field. Continuing through the chamber, 10 mm beyond the sound field, the clarified liquid on the left side of the band is drawn off through a side arm at an angle of 9°. The main channel extends for a further 45 mm and the side arm for 40 mm in order to maintain laminar flow conditions around the divider before entering circular cross-section tubes. The fluid circuit has been described previously.[24] The flow rates were 0.144 ml s^{-1} through the ultrasound field and 0.028 ml s^{-1} for extraction of cleared medium.

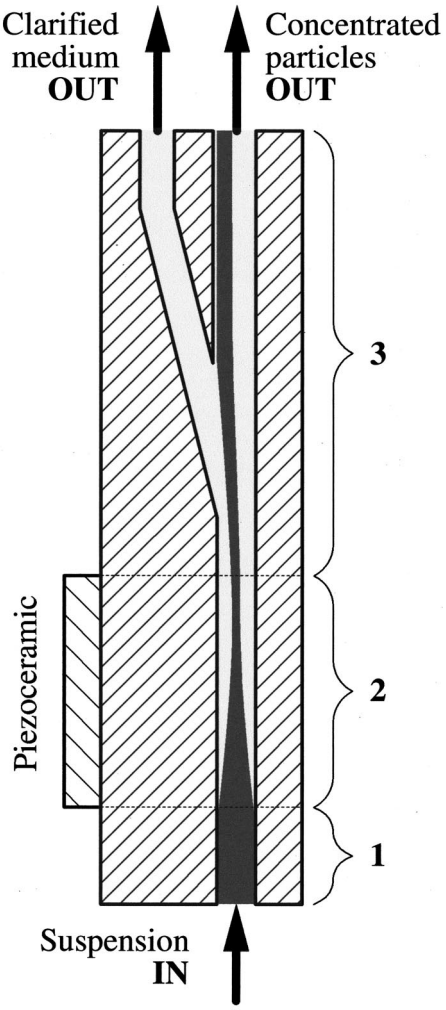


FIG. 1. Diagrammatic representation of the acoustic half-wavelength filter showing particles concentrated by the sound field forming a thin band and withdrawal, from a side outlet, of clear medium from the left side of the band. (1) Laminar flow stabilization; (2) Ultrasound standing waves concentrate particles into a single band; (3) Laminar flow splitting.

3. Ultrasound generation

The piezoceramic transducer was driven by an amplifier (model 240L RF amplifier, ENI, Rochester, NY). The acoustic filter was attached directly to the amplifier output through a BNC connector. The sinusoidal amplifier input came from a function generator (3326A, Hewlett-Packard). The frequency of the function generator was scanned from 2 to 4 MHz in 660 steps of 7-s duration. For the filtration experiments, the amplifier was operated at a constant input voltage level set for an open-circuit output voltage of 16 V peak. The amplifier's output resistance was 50 Ω.

TABLE I. Physical properties of filter layers.

Layer	Piezoceramic A	Glue B	Transmission C	Water D	Reflector E
Material	Pz26	Super glue	Stavax stainless steel	Water	Stavax stainless steel
Thickness (mm)	0.68[a]	0.005[b]	3.1[a]	Varied[c]	1.61[a]
Sound speed (m/s)	4600[a]	2400[b]	6100[b]	1464[b]	6100[b]
Density (kgm^{-3})	7700[d]	1100[b]	7800[a]	1000	7800[a]
Elastic stiffness constant (GPa)	125	6.33	290	2.25	290
Eff. acoustic quality factor	350[b]	5[b]	1000[b]	150[b]	1000[b]

[a]Measured value.
[b]Fitted value.
[c]Four thicknesses of the water layer were used, 0.18, 0.24, 0.26, and 0.30 mm.
[d]Ferroperm data sheet.

4. Determination of filtration efficiency

Scattered light measurements were made with a particle detector based on an LED/photodiode placed 100 mm downstream from the ''cleared medium'' filter outlet. The LED light beam passed through the sample and was detected by the photodiode after scattering through a right angle as the suspension flowed through a 30-mm-long borosilicate glass tube (Camlab, Cambridge, UK) of $3\times3\text{-mm}^2$ inner square section and 0.5-mm wall thickness. The photodiode output voltage (PDV) was linearly proportional to particle numbers at concentrations of yeast (particles of ca. 5-μm diameter) up to $2.5\cdot10^7\ \text{ml}^{-1}$.[24] The threshold concentration for detection of cells was $5\cdot10^5\ \text{ml}^{-1}$.

The filter's performance was characterized in terms of clearance. Clearance is defined as [in]/[clear] where [in] is the concentration of particles entering the filter, and [clear] is the concentration of particles exiting from the side arm (clarified outlet). For practical reasons, both concentrations were measured with the same detector placed in the side arm, with the sound field switched off for determination of [in]. For the experimental results measurements of both [in] and [clear] fell within the established linear region of the PDV dependence on particle concentration. Clearance was therefore expressed as PDV(sound off)/PDV(sound on). The output from the photodiode was interfaced to a computer.

5. Admittance measurements

Electric admittance of the acoustic filter was measured using an Agilent Technologies Network Analyzer (HP 8753E). The instrument was calibrated up to the acoustic filter's housing connector to ensure signal integrity and to eliminate any influence of the screened RF connecting cable.

B. The transfer matrix multilayer resonator model

The transfer matrix model of piezoelectric multilayer resonators used in this work for calculation of filter properties was developed by Nowotny and Benes.[25,26] Application of the model to piezoelectric resonators for particle separation was described by Gröschl.[8] The model is based on the fundamental equations of piezoelectricity[27] that relate the coupled electroacoustic field quantities, acoustic displacement u, mechanical stress T, electric potential φ (quasistatic approximation), and dielectric displacement D, whereby the medium under consideration is described by its elastic stiffness constant c, piezoelectric constant e, and dielectric constant ε. The model is generally restricted to harmonic time dependence and to the one-dimensional case, that is, all considered quantities are assumed to show space dependence in only one direction (direction of sound propagation, thickness direction of the layers). Furthermore, in some situations, the displacement of the sound wave may be restricted to this direction as well (longitudinal waves only). This treatment is justified here, because the piezoceramic plate transducers used essentially permit electrical excitation of longitudinal waves only. As a consequence of this restriction to a single displacement direction, all material constants, which are tensor quantities in the general case, are reduced to scalars.

According to the model, a homogeneous material layer is represented by a transfer matrix relating the boundary values of u, T, φ, and D at adjacent surfaces. The transfer matrix elements depend only on the material constants ρ (mass density), c, e, ε, the layer thickness d, and the angular frequency ω. The explicit expressions for the matrix elements were given by Nowotny.[25] For multiple layers, the total transfer matrix of the layered structure can be obtained by multiplication of the transfer matrices of each single layer. This follows from the continuity conditions that apply at interfacing surfaces. The electrodes of the piezoelectric layer have to be treated as separate films, but may be regarded as massless if they are sufficiently thin. On the outer free surfaces of the total sandwich arrangement, stress and dielectric displacement must be zero. This general boundary condition establishes an expression for the electrical admittance Y between the electrodes. The explicit result for Y as a function of angular frequency ω and electrode area A was given by Nowotny.[25]

A general way of considering losses (viscoelastic damping, dielectric and piezoelectric losses) is the introduction of complex material constants.[28] Consequently, the transfer matrix elements and the acoustic and electric field quantities become complex, too. Specifically, viscoelastic and dielectric

TABLE II. Additional physical properties of active layer (piezoceramic).

Material	Pz26	
Dielectric constant (F/m)	$8 \cdot 10^{-9}$	(a)
Tangent of dielectric loss angle	$3 \cdot 10^{-3}$	(b)
Electromechanical coupling factor	0.48	(b)
Piezoelectric constant (N/Vm)	17.3	
Electrode area (mm^2)	200	(a)

[a]Measured value.
[b]Ferroperm data sheet.

losses can be expressed by an acoustic quality factor Q and a dielectric loss angle δ, that form the imaginary parts of the material constants c and ε, respectively. (Piezoelectric losses can be neglected in most practical cases.) Other loss mechanisms relevant for acoustic particle filters, like sound attenuation due to bubbles or dispersed particles, losses due to divergence of the sound beam, etc., can be accounted for in a global way by using reduced Q values (effective quality factors) for each material layer, as discussed by Gröschl.[8]

For a given multilayer resonator, typically comprising an active piezoelectric layer followed by several passive solid and/or liquid layers, in a first step, the transfer matrix model allows calculation of the electrical admittance (real and imaginary parts) between the electrodes of the active layer as a function of frequency. In a second step, the spatial progress of the primary electroacoustic field quantities u, T, φ, D within each layer, can be determined for a fixed frequency and a given voltage amplitude applied across the electrodes. This is accomplished by stepwise evaluation of the transfer matrix of the corresponding layer, whereby the layer thickness d is replaced by a variable thickness running from one boundary to the next. From the primary field quantities the spatial progress of other quantities of interest, like displacement velocity or acoustic energy density in the considered layer, can be derived easily.[8]

Practical application of the model requires knowledge of all material parameters involved, these are for each layer: thickness d, mass density ρ, elastic stiffness constant c, effective quality factor Q_{eff}; furthermore, if the considered layer is located between electrodes (within an electric field): dielectric constant ε and dielectric loss angle δ; and, in addition for piezoelectric layers: piezoelectric constant e. The parameters used for the calculations presented in this work are listed in Tables I and II. The mathematical expressions used were given by Gröschl[8] and Nowotny.[25]

III. RESULTS AND DISCUSSION

Figure 2 shows the electrical admittance spectrum (absolute values) of the filter chamber with 0.26-mm channel width, measured with a Network Analyzer type HP 8753E, compared to the values predicted by the transfer matrix model described in the previous section. The calculation is based on the layered structure of the filter chamber, comprising a piezoceramic layer A, followed by a glue layer B, a steel part C, the water layer D, and another steel part E, as shown in the upper part of Fig. 4. Sound propagates along the axis in the horizontal direction. All parameters of the investigated multilayer resonator are listed in Tables I and II. Each peak shown in the graph represents a resonance of the

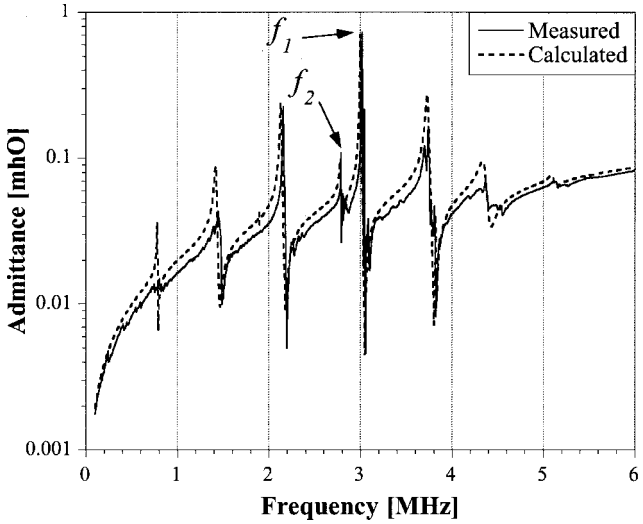


FIG. 2. Measured and calculated electrical admittance of the water-filled filter chamber (0.26-mm channel width).

multilayer structure (composite resonator). Two characteristic frequencies are marked: f_1 coinciding with the fundamental resonance frequency of the piezoceramic (this coincidence is a consequence of chosen dimensions of layers A and C), and f_2, which will be identified as the frequency of optimum filter performance (see below). The resonance f_2 occurs due to the liquid (water) layer of the structure; the admittance spectra of the empty chamber (not shown here) look essentially the same as the admittance curves of Fig. 2, with the resonance f_2 vanishing.

Each resonance of the multilayer structure is characterized by a total phase shift of the acoustic wave between the outer boundaries equal to a multiple of π. For the fundamental resonance frequency f_1 (3 MHz) the thickness of layer A corresponds to one half-wavelength ($\lambda/2$) of the acoustic wave in the piezoceramic and the thickness of layer C corresponds to $3\lambda/2$ in the steel. (The thin glue layer B can be neglected in this consideration.) Thus, at 3 MHz the phase shift of the acoustic wave is π between the boundaries of layer A and 3π between the boundaries of layer C. For the empty chamber (layers A and C only) the resonance at $f_1 = 3$ MHz therefore corresponds to a total phase shift between the outer boundaries of 4π. Additional resonances can be seen clearly in Fig. 2, corresponding to a total phase shift of π ($f = 0.75$ MHz), 2π ($f = 1.5$ MHz), 3π ($f = 2.25$ MHz), and 5π ($f = 3.75$ MHz). These resonances remain nearly unchanged if the chamber is filled with water, although, in principle, layers D and E have to be considered then. This small influence of the water layer arises from the significantly lower acoustic impedance of water compared to that of steel (about 1:30). However, an additional resonance develops in the water-filled structure if the width of the water layer D is in the order of (a multiple of) a half-wavelength of the acoustic wave in water. For the chamber represented in Fig. 2 the width of water layer D is 0.26 mm. With a sound speed value of 1464 m/s (Table I) this results in a resonance frequency f_2 of approximately 2.8 MHz. This resonance, which is only present for the water-filled chamber, is influenced strongly by the properties of the water layer (thickness, elastic con-

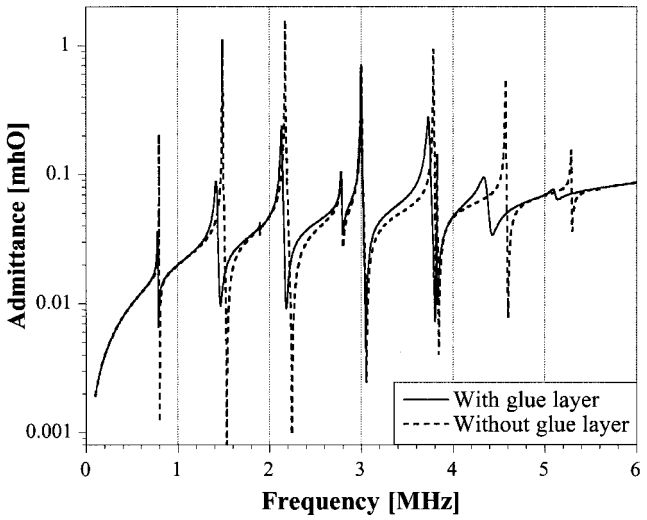


FIG. 3. Calculated admittance of the water-filled filter chamber for situations where the influence of a glue layer between piezoceramic and contacting steel is included or ignored (0.26-mm chamber).

stant or sound speed) and is strongly dependent on temperature (because of the temperature effect on sound speed in water).

Correct modeling of chamber characteristics with respect to resonance peak heights and quality factors requires consideration of a glue layer between the piezoceramic and the contacting steel part, as can be seen from Fig. 3. Modeling without a glue layer results in an over-representation of resonance peaks neighboring the fundamental at 3 MHz. Introduction of a glue layer reduces peak heights of the neighbored resonances, but not of the 3-MHz fundamental, thus leading to excellent agreement with the measured admittance curve (Fig. 2). The effect of the glue can be explained as follows: Around the piezoceramic's fundamental at 3 MHz the displacement amplitude of the acoustic wave is a maximum at the ceramic surfaces. (Exactly one half-wavelength fits into the ceramic layer.) Therefore, in this frequency range the glue layer lies in a region of maximum amplitude and vanishing stress (pressure node) and has almost no effect on the resonator. At other resonances the glue layer is not in a pressure node, thus being stressed significantly and because of its low Q value ($Q_{\text{eff}}=5$ was assumed for calculations) causing strong resonance degradation.

Figure 4 shows in its upper part the layered structure of the filter chamber, as described previously. Below, two graphs show the calculated spatial progress of displacement velocity amplitude and stored acoustic energy density, respectively, within these layers for the 0.26-mm chamber. Each graph gives the results for two selected frequencies f_1 and f_2 (see Fig. 2). For both frequencies, amplitudes and energy densities are higher in the water layer *D* than in the other layers; however, the maxima (in the water layer) do not appear at the piezoceramic's fundamental resonance f_1 but at the system's resonance f_2. With respect to filtration efficiency, for a given electric energy supplied, maximum stored acoustic energy in the water (liquid) layer is desired. Thus, optimum system performance is achieved at resonance frequencies *not* coinciding with eigenfrequencies of the piezo-

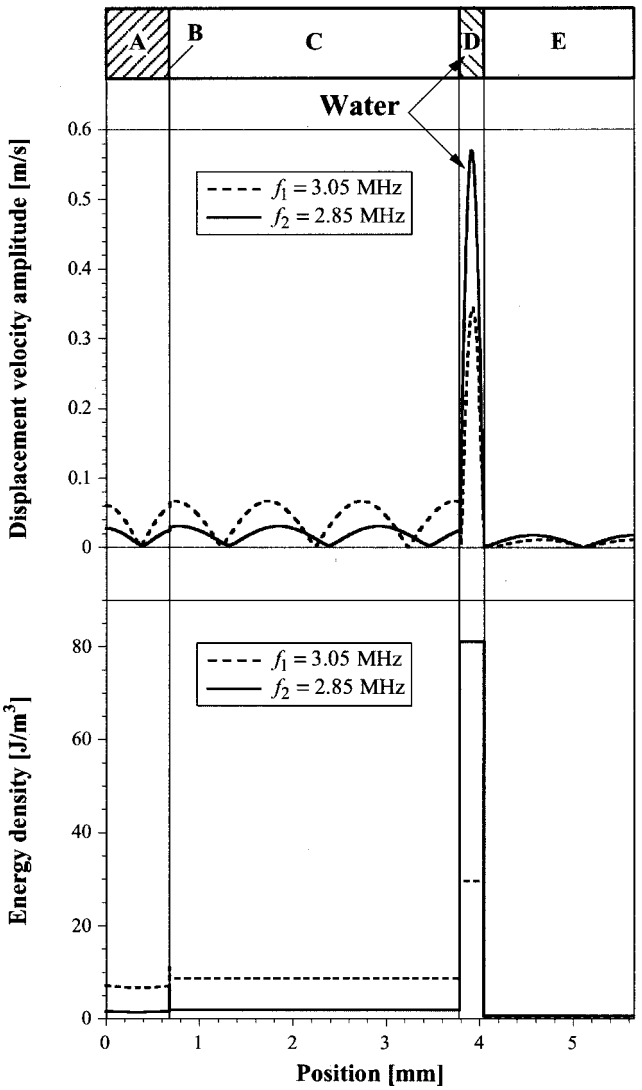


FIG. 4. One-dimensional structure of filter chamber, together with corresponding displacement velocity amplitude and stored acoustic energy density distributions (0.26-mm chamber), each calculated for two characteristic frequencies (as marked in Fig. 2); f_1: System's resonance coinciding with fundamental resonance of piezoceramic; f_2: Resonance with maximum stored acoustic energy density in the liquid layer *D*.

ceramic (layer *A*) or the ceramic–steel compound (empty chamber, layers *A*, *B*, *C*), as already shown by Gröschl for ultrasonic filters developed for biotechnology applications.[8] For the calculations shown in Fig. 4, a driving voltage of 16 V peak (open-circuit) and a source resistance of 50 Ω have been assumed, according to the experimental conditions applied.

Measured filter performance (particle clearance) was compared to calculated acoustic energy density in the water layer of the 0.26-mm chamber, over a frequency range from 2 to 4 MHz (Fig. 5). In the preferred operation frequency range near 3 MHz a distinct double-peak structure appears in the clearance as well as in the energy density spectrum. Measured and calculated peak frequencies are in excellent agreement. The remaining slight frequency shifts are primarily due to the strong influence of changes in the elastic constant or sound speed, respectively, of the water layer (temperature dependence) and could even be eliminated. The peak height

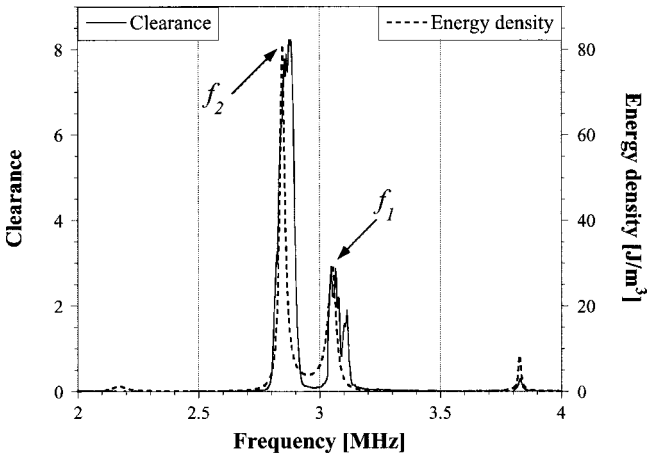


FIG. 5. Measured clearance of latex particles compared to calculated acoustic energy density spectra for the 0.26-mm chamber.

ratio corresponds to the findings of Fig. 4; that is, best performance (maximum clearance) is obtained at the resonance frequency f_2, at which the stored acoustic energy density in the water layer is a maximum.

Figure 6 corresponds to Fig. 5 but shows the respective results for a filter with 0.30-mm channel width. All parameters other than channel width remained unchanged. However, the spectra of clearance and energy density look remarkably different. The peak of optimum performance is shifted to a far lower frequency (approximately 2.5 MHz). The peak at the ceramic's fundamental almost vanishes. This strong influence of channel width is a consequence of the specific geometric conditions applying to the multilayer structure. As discussed previously, the resonances already present for the empty chamber are only slightly influenced by the water layer. A change of the water layer thickness (here, from 0.26 to 0.30 mm) produces a significant shift only of that resonance that is specific to the water-filled structure. Thus, the resonance frequency f_2 around 2.8 MHz for the 0.26-mm chamber (Figs. 2 and 4) is shifted to approx. 2.5 MHz for the 0.30-mm chamber (Fig. 6). The value of 2.5 MHz corresponds to one half-wavelength of the acoustic wave in the 0.30-mm water channel. The observed strong

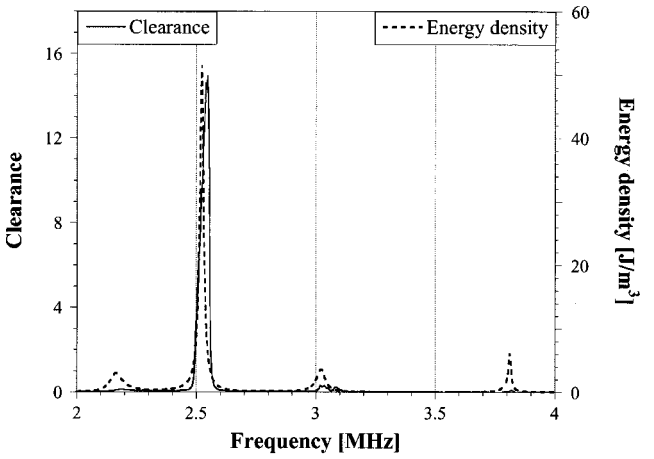


FIG. 6. Measured clearance of latex particles compared to calculated acoustic energy density spectra for the 0.30-mm chamber.

dependence of filter performance on channel width has been verified experimentally for another two channel widths (0.18, 0.24 mm, data not shown) and was well represented by the model in all cases.

In Figs. 5 and 6 we have compared experimental data (clearance) to calculated results (energy density in water layer) both representing different physical quantities, although they are connected strongly. In particular, both properties show essentially the same frequency dependence. The transfer matrix model in its current state is perfectly suited to predict peak frequencies of filter clearance. The ratios of the peak heights for energy density and particle clearance are also in good agreement, as can be seen from Fig. 5. Nevertheless, the model cannot deliver absolute values of clearance. A model for predicting filter clearance in an ideal half-wavelength chamber has already been presented.[24] Development of a model for predicting filter clearance from energy density data over a wide frequency range and for different channel widths is in progress.

In our measurements the frequencies associated with the calculated energy density distribution peaks appear to be slightly lower than those of the clearance measurements. Part of this difference arises because, in the experimental scanning process, frequency increases with time. Since the detector is downstream from the ultrasound active volume, the frequency associated with a measured clearance, i.e., the frequency being applied to the transducer when the clearance is measured, is higher than the frequency at which that increment of sample was cleared. The volume between the ultrasound field and the detector was 0.55 ml. Since the clarified medium flow rate was 0.028 ml s^{-1}, a time delay in the region of 20 s occurs between sample clarification and clearance measurement. Since the frequency scan rate was 3 kHz in 7-s steps, the above delay corresponds to a 10-kHz overestimate of the frequency associated with a measured clearance. The data presented in Figs. 5 and 6 are not corrected for this overestimate since the difference in frequency is small in terms of these figures' frequency range.

In practical filter operation the output resistance of the electric driving source used is of great influence. Due to the resonant behavior of the filters considered, their electric impedance (admittance) varies strongly with frequency (Fig. 2). Consequently, maximum energy transfer from the source to the resonator (liquid layer) occurs at certain frequencies depending on the source resistance. This is illustrated in Fig. 7, showing in expanded view the region around 3 MHz of the calculated admittance spectrum of Fig. 2, together with the calculated distribution of energy density in the 0.26-mm channel. Calculations were carried out for an open-circuit source voltage of 16 V peak and a source resistance of 50 Ω (the experimental conditions applied). The peaks of energy density, already identified in Fig. 5, almost coincide with the frequencies of the admittance *minima*, the so-called parallel resonance frequencies. Thus, under the given electric conditions, the optimum operation frequency of the filter is close to a *parallel* resonance rather than a series resonance (admittance maximum). This is because of the low resonator impedance (less than 10 Ω at series resonances) compared to the source, leading to electric drive conditions close to a

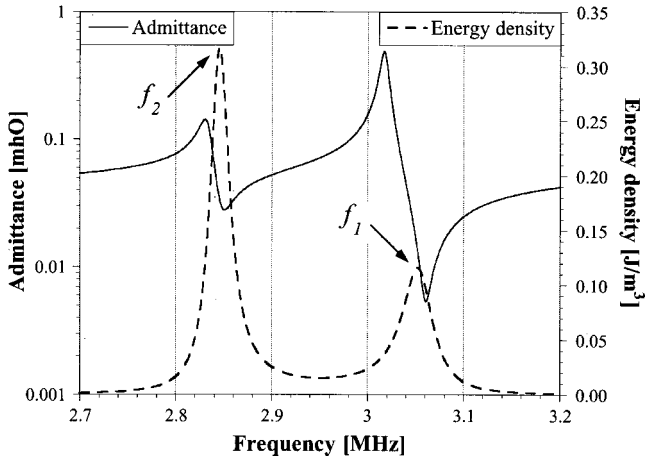


FIG. 7. Calculated electrical admittance and acoustic energy density in the frequency range around the resonances f_1, f_2.

constant current source configuration. For an ideal voltage source (zero source impedance) energy density peaks would coincide with the series resonances, while for an ideal current source (infinite source impedance) the optimum frequencies would exactly match the parallel resonances. A method of identifying and maintaining the optimum operation frequency for acoustic particle filters by controlling electric true power input into the resonator has been described by Gröschl.[29]

IV. CONCLUSIONS

Application of the transfer matrix model to a multilayered filter system consisting of a piezoceramic transducer, glue, steel transmission plate, fluid channel, and steel reflector showed excellent agreement between measured and predicted electrical admittance spectra when the channel contained water or air. Calculation of admittance showed that neglect of the thin glue layer attaching the transducer to the steel transmission plate gave good agreement with the glue-included case and with admittance measurements close to the piezoceramic's fundamental resonance, but showed significant differences away from this resonance frequency. The model also showed that the frequency dependence of stored acoustic energy density in the water layer was in excellent agreement with the experimentally determined frequency distribution of particle filtration efficiency. Comparisons of results for four channel widths showed that the frequency dependence for the energy density and clearance efficiency had a small maximum at the system's resonance corresponding to the fundamental of the piezoceramic (resonance of the empty chamber) and a larger maximum (optimum clearance) at another specific system's resonance *not* coinciding with any of the empty chamber's resonances. Gröschl[8] previously applied the matrix model to a long, aqueous pathlength ultrasonic filter. Quantitative comparison of experimental and model results was restricted to electric properties of the filter (admittance spectra) in that work. He did conclude qualitatively that the filter worked better if operated at resonances not coinciding with piezoceramic or empty chamber resonances, but there was no determination of separation performance as a function of frequency. The high level of agreement between the frequency dependence of the measured filter performance in this work and the model predictions of acoustic properties such as acoustic energy density in the water layer provide the first detailed experimental verification of matrix model predictions of filter performance.

Further comparisons showed that the peaks of energy density were close to the frequencies of minima in the admittance spectrum, indicating that the filter in combination with the connected amplifier performs optimally close to the parallel resonance condition (operation in constant current mode).

The successful test of the model against experimentally determined admittance and clearance data identifies it as a powerful design tool for further optimization of multilayer ultrasound systems.

ACKNOWLEDGMENTS

Jeremy J. Hawkes was supported by a DERA Contract No. CU013-0000010785. Work supported in part by the European Commission's TMR Programme, Contract No. ERBFMRXCT97-0156, EuroUltraSonoSep.

[1] L. V. King, "On the acoustic radiation pressure on spheres," Proc. R. Soc. London, Ser. A **147**, 212–240 (1934).
[2] K. Yosioka and Y. Kawasima, "Acoustic radiation pressure on a compressible sphere," Acustica **5**, 167–173 (1955).
[3] L. P. Gor'kov, "On the forces acting on a small particle in an acoustical field in an ideal fluid," Sov. Phys. Dokl. **6**, 773–775 (1962).
[4] W. L. Nyborg, "Radiation pressure on a small rigid sphere," J. Acoust. Soc. Am. **42**, 947–952 (1967).
[5] L. A. Crum, "Acoustic force on a liquid droplet in an acoustic stationary wave," J. Acoust. Soc. Am. **50**, 157–163 (1971).
[6] R. K. Gould and W. T. Coakley, "The effects of acoustic forces on small particles in suspension," in *Proceedings of the 1973 Symposium on Finite-Amplitude Wave Effects in Fluids* (Pergamon, Guildford, UK, 1974), pp. 252–257.
[7] G. Whitworth and W. T. Coakley, "Particle column formation in a stationary ultrasonic field," J. Acoust. Soc. Am. **91**, 79–85 (1992).
[8] M. Gröschl, "Ultrasonic separation of suspended particles-Part I: Fundamentals," Acust. Acta Acust. **84**, 432–447 (1998).
[9] M. A. H. Weiser and R. E. Apfel, "Interparticle forces on red cells in a standing wave field," Acustica **56**, 114–119 (1984).
[10] D. G. Kilburn, D. J. Clarke, W. T. Coakley, and D. W. Bardsley, "Enhanced sedimentation of mammalian cells following acoustic aggregation," Biotechnol. Bioeng. **34**, 559–562 (1989).
[11] G. Whitworth, M. A. Grundy, and W. T. Coakley, "Transport and harvesting of suspended particles using modulated ultrasound," Ultrasonics **29**, 439–444 (1991).
[12] S. M. Woodside, J. M. Piret, M. Gröschl, E. Benes, and B. D. Bowen, "Acoustic force distribution in resonators for ultrasonic particle separation," AIChE J. **44**, 1976–1984 (1998).
[13] E. Benes, F. Hager, W. Bolek, and M. Gröschl, "Separation of dispersed particles by drifting ultrasonic resonance fields," in *Ultrasonics International '91, Conference Proceedings* (Butterworth-Heinemann, Oxford, UK, 1992), pp. 167–170.
[14] B. Handl, M. Gröschl, F. Trampler, E. Benes, S. M. Woodside, and J. M. Piret, "Particle trajectories in a drifting resonance field separation device," in *Proceedings of the 16th International Congress on Acoustics and 135th Meeting Acoustical Society of America* (ISBN 1-56396-817-7, Acoustical Society of America, New York, 1998), Vol. III, pp. 1957–1958.
[15] S. Peterson, G. Perkins, and C. Baker, "Development of an ultrasonic blood cell separator," in *Proceedings of the 8th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (IEEE, New York 1986), pp. 154–156.
[16] F. Trampler, S. A. Sonderhoff, P. W. S. Pui, D. G. Kilburn, and J. M. Piret, "Acoustic cell filter for high density perfusion culture of hybridoma cells," BioTechnology (NY) **12**, 281–284 (1994).
[17] O. Doblhoff-Dier, T. Gaida, H. Katinger, W. Burger, M. Gröschl, and E. Benes, "A novel ultrasonic resonance field device for the retention of

animal cells,'' Biotechnol. Prog. **10**, 428–432 (1994).
[18] T. Gaida, O. Doblhoff-Dier, H. Katinger, W. Burger, M. Gröschl, and E. Benes, ''Scale-up of ultrasonic resonance field cell separation devices used in animal cell technology,'' in *Animal Cell Technology: Developments towards the 21st Century*, edited by E. C. Beuvery, J. B. Griffiths, and W. P. Zeijlemaker (Kluwer, Dordrecht, 1995), pp. 699–703.
[19] P. W. S. Pui, F. Trampler, S. A. Sonderhoff, M. Gröschl, D. G. Kilburn, and J. M. Piret, ''Batch and semicontinuous aggregation and sedimentation of hybridoma cells by acoustic resonance fields,'' Biotechnol. Prog. **11**, 146–152 (1995).
[20] T. Gaida, O. Doblhoff-Dier, K. Strutzenberger, H. Katinger, W. Burger, M. Gröschl, B. Handl, and E. Benes, ''Selective retention of viable cells in ultrasonic resonance field devices,'' Biotechnol. Prog. **12**, 73–76 (1996).
[21] J. J. Hawkes and W. T. Coakley, ''A continuous flow ultrasonic cell-filtering method,'' Enzyme Microb. Technol. **19**, 57–62 (1996).
[22] M. Gröschl, W. Burger, B. Handl, O. Doblhoff-Dier, T. Gaida, and C. Schmatz, ''Ultrasonic separation of suspended particles-Part III: Application in biotechnology,'' Acust. Acta Acust. **84**, 815–822 (1998).
[23] D. A. Johnson and D. L. Feke, ''Methodology for fractionating suspended particles using ultrasonic standing wave and divided flow fields,'' Separations Technol. **5**, 251–258 (1995).
[24] J. J. Hawkes and W. T. Coakley, ''Force field particle filter, combining ultrasound standing waves and laminar flow,'' Sens. Actuators B **75**, 213–222 (2001).
[25] H. Nowotny and E. Benes, ''General one-dimensional treatment of the layered piezoelectric resonator with two electrodes,'' J. Acoust. Soc. Am. **82**, 513–521 (1987).
[26] H. Nowotny, E. Benes, and M. Schmid, ''Layered piezoelectric resonators with an arbitrary number of electrodes (general one-dimensional treatment),'' J. Acoust. Soc. Am. **90**, 1238–1245 (1991).
[27] B. A. Auld, *Acoustic Fields and Waves in Solids* (Wiley, New York, 1973), Vol. I.
[28] R. Holland, ''Representation of dielectric, elastic, and piezoelectric losses by complex coefficients,'' IEEE Trans. Sonics Ultrason. **14**, 18–20 (1967).
[29] M. Gröschl, ''Ultrasonic separation of suspended particles-Part II: Design and operation of separation devices,'' Acust. Acta Acust. **84**, 632–642 (1998).

Measurement of an aeroacoustic dipole using a linear microphone array

Peter Jordan and John A. Fitzpatrick[a)]
Department of Mechanical Engineering, Trinity College, Dublin 2, Ireland

Jean-Christophe Valière
LEA/CNRS UMR-6609, University of Poitiers, Poitiers, France

(Received 4 October 2001; accepted for publication 27 November 2001)

It is shown that the standard beamformer technique is inadequate for both the source location and the measurement of a simple dipole and that this is due to the assumption of monopole propagation in the calculation of the phase weights used to steer the focus of the array. A numerical simulation is used to illustrate the problem and to develop a correction to the signal processing algorithm to account for the dipole propagation characteristic. This is then applied to array measurements for an aeroacoustic dipole produced by a cylinder in a cross flow. The resulting source map and the beamformed spectrum are shown to give a true representation of the source energy and frequency content. A secondary effect of this correction is that the array becomes insensitive to other source types so that in addition to acting as a spatial filter, the array can perform as a source filter. This work also demonstrates how an array measurement can be misinterpreted if applied without consideration of the source mechanism. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1446052]

PACS numbers: 43.38.Hz, 43.28.Ra, 43.28.Tc [MSH]

I. INTRODUCTION

The use of acoustic beamforming has become increasingly commonplace in experimental aeroacoustics, and microphone arrays have proved a useful tool for source localization in a number of applications. Billingsley and Kinns[1] used a microphone array based on the beamforming technique to measure jet noise whereas a polar correlation technique was used by Fisher *et al.*[2] also to measure jet noise. Marcolini and Brooks,[3] examined helicopter rotor noise by using an array placed above the rotor and Brooks *et al.*[4] reported on the parameters affecting the response characteristics of an array. The advantages and disadvantages of a phased array have been given by Siller *et al.*[5] who implemented the system for monitoring engine core noise. The performance of an array is usually characterized by the directivity pattern as this defines its spatial sensitivity. The directivity pattern can result in the contamination of measurements from extraneous noise sources, a form of spatial aliasing, and this is a function of both source frequency and microphone array geometry. This has resulted in the development of very large arrays with random spacing so that these effects can be minimized over a range of frequencies.[6]

The implementation of beamforming for a microphone array normally assumes monopole propagation characteristics in the calculation of the phase weights used to steer the focus of the array. As a consequence, the capacity of the system to identify different source types and provide a measure of the contribution of that source in the far-field is limited in certain circumstances. Since the far-field characteristics of an aeroacoustic system depend on the aerodynamic source mechanisms and the manner by which they interact, it is important to identify these mechanisms so that their contribution to the acoustic far-field can be assessed.

The noise source for many aeroacoustic systems is generated by fluctuating lift forces and these are usually dipole type mechanisms. In this work it is shown numerically and experimentally that the standard beamformer method fails when used to measure a simple dipole source represented by an Aeolian tone generated by cross flow over a circular cylinder. For this, the dipole is generated principally by the fluctuating lift force at the vorticity shedding frequency with very much smaller unsteady drag at twice this frequency as described by Phillips.[7] Experimental results reported by Leehey and Hanson[8] investigated the effect of Reynolds number and spanwise correlation length on the noise intensity. Their results, which agreed well with predictions based on the analysis of Phillips,[7] indicated that the oscillating lift coefficient was the controlling factor in the radiated sound intensity and that this increased sharply with increasing Reynolds number. Thus, identification of coherent sound sources typical of dipole mechanisms is important for practical applications in aeroacoustics.

A linear microphone array is used to make measurements of the noise generated by a cylinder in a cross-flow. The failure of the beamformer is manifest in both the location of the source and in the derived spectrum which shows almost no energy above the background levels in a frequency range known to contain a discrete frequency source. This is shown to be due to the assumption of monopole propagation and a correction which accounts for dipole propagation characteristics is developed and applied so that the beamformer can be modified to accurately measure the source. After the correction is applied, the spectrum shows, very clearly, a peak at the source frequency, the level of which is a measure

[a)]Author to whom correspondence should be addressed. Electronic mail: john.fitzpatrick@tcd.ie

of the source energy. It is further shown how, due to the source-specific nature of the beamformer modification, higher and lower frequency duct, fan, nozzle and jet noise is eliminated from the spectrum. This work demonstrates how array measurements can be misinterpreted if some effort is not made to account for the physical nature of the source mechanisms. It also demonstrates how, through application of corrections to account for different source types, the array can be used, not only to discriminate spatially between sources, but also to discriminate between different source types.

II. AEROACOUSTIC SOURCES

The far-field pressure of monopole, dipole and quadrupole sources have been given by, for example, Powell[9] respectively as

$$p_M(t;x)=\frac{\rho_0 a_m^2}{x}\frac{dU_m^*}{dt},$$

$$p_D(t;x)=\frac{\rho_0 a_m^2}{x}\left(\frac{y\cos\vartheta}{c}\right)\frac{d^2U_m^*}{dt^2}, \tag{1}$$

$$p_Q(t;x)=\frac{\rho_0 a_q^2}{x}\left(\frac{a_q}{3c}\right)^2\frac{d^3U_{q,x}}{dt^3},$$

where ρ_0 is the air density, a_m is the radius of an imaginary sphere of fluid containing the source, and y is the distance between the two poles of the dipole (assumed small compared to an acoustic wavelength). U_m^* is the radial velocity of the monopole surface, x is the distance from the source, and c the speed of sound. $U_{q,x}$ is the x component of velocity of the quadrupole surface.

Monopole sources are associated with volumetric fluctuations of the local medium and the resultant sound power is proportional to first power of the local Mach number. For aeroacoustics, they arise in propeller or fan noise as described by Ffowcs-Williams and Hawkings.[10] Dipoles are associated with fluctuating forces, the resultant sound power varying with the third power of the local Mach number, and are dominant sources for many aeroacoustic systems at moderate Mach numbers. Finally, quadrupole sources, associated with the turbulent distortion of the fluid, generate sound power proportional to the fifth power of the local Mach number and arise mainly in high speed jet noise systems. In terms of far-field propagation, the monopole is the most efficient radiator of sound, followed by the dipole, which is less efficient because of the extra degree of freedom of motion associated with the source mechanism and, then, by the quadrupole which is less efficient again, having yet another degree of freedom of motion. This work concentrates on the dipole source, which tends to generate narrow-band acoustic energy due to unsteady aerodynamic loading in fluid–structure interactions.

A. The dipole source

From Eqs. (1) it can be seen that the acoustic field of a dipole is dependent on θ, where θ is the azimuthal angle in the dipole plane. The resulting unsteady pressure fluctuations propagate spherically into space as shown in Fig. 1 and there exists a planar region between the two poles, and perpendicular to the dipole axis, where the sound waves combine to interfere destructively as the waves are perfectly out of phase. As an observer moves around the dipole, the cancellation becomes less efficient, until a point is reached (on the dipole axis) where there is no cancellation and the sound pressure observed is a maximum.

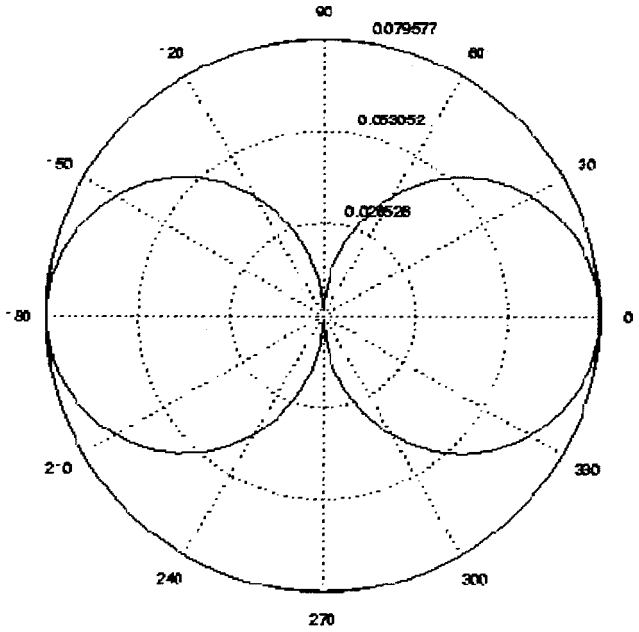


FIG. 1. Polar pressure distribution of a point dipole.

III. BEAMFORMING AND THE DIPOLE

A. Modeling the dipole

The expression for the acoustic field of a monochromatic dipole can be written as

$$p(x,\vartheta,k)=\frac{B}{x}\cos\vartheta e^{jkx}, \tag{2a}$$

where $k=\omega/c$ is the wavenumber, B is the sound pressure magnitude and ω the sound frequency in radians, or, alternatively as

$$p(x,\vartheta,t)=\frac{B}{x}\cos\vartheta e^{-j\omega t}, \tag{2b}$$

i.e., a source with time dependence proportional to $e^{-j\omega t}$, where $j=\sqrt{-1}$. Using this expression, a far-field radiation pattern for a dipole was generated as given in Fig. 1. Using this acoustic field, the response of a linear array of 30 microphones arranged parallel to the dipole axis as shown in Fig. 2 was determined. The frequency of the source was 800 Hz, the array was located 1 m from the source and the microphone spacing was 30 mm. Figure 3 shows ten fully reconstructed temporal signals, where it can be seen that there are variations in both magnitude and phase across the array as would be expected.

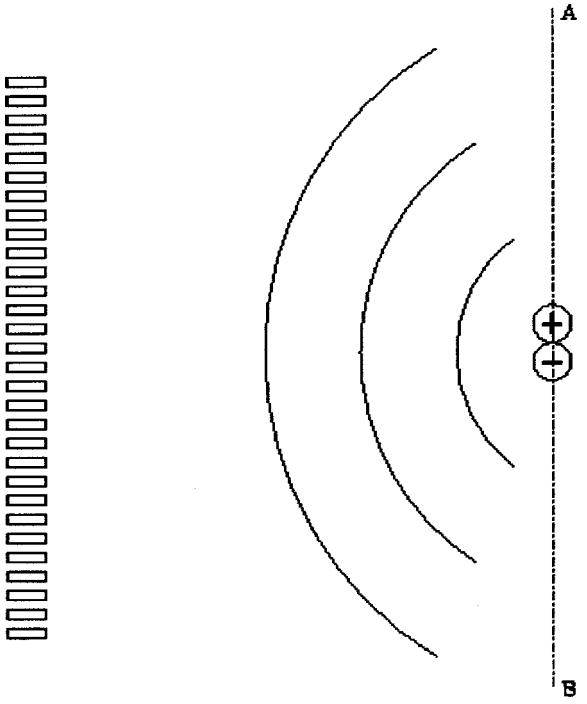


FIG. 2. Setup for numerical simulation.

B. Beamforming algorithm

The beamforming operation is defined by[11]

$$X(t,\omega)=\sum_{m=1}^{M}\frac{W_m}{r_m}Y_m(t,\omega)e^{-j\omega\Delta_m}, \tag{3}$$

which, in the frequency domain, is

$$X(f)=\sum_{m=1}^{M}\frac{W_m}{r_m}Y_m(f)e^{-j(2\pi f/c)r_m}, \tag{4}$$

where W_m is the element weighting (nominally equal to 1), r_m is the distance from the mth sensor to the focus position, Y_m is the Fourier transform for the mth sensor, f is the frequency and c is the speed of sound in air. It can be seen from Eqs. (3) and (4) that the output of the array, as well as being a function of the source frequency, is also a function of the microphone-source vectors, r_m. The influence of these parameters on the resolution power of an array is significant and has been well documented in the literature.[1,2,4,11] Higher frequency sources are more effectively resolved by the array due to a narrowing of the mainlobe with increasing frequency whereas low frequency sources cause a broadening of the mainlobe and are less well resolved. An increase in the overall aperture of an array, defined as the angle subtended between its extremities and the source, results in an improvement in the resolution capacity of the array. Thus, for a fixed aperture, the addition of more microphones (i.e., reducing the microphone separation) results in both an improvement in the dynamic gain of the array, which is defined as the difference between the mainlobe peak and the next highest sidelobe, and its resolution power.

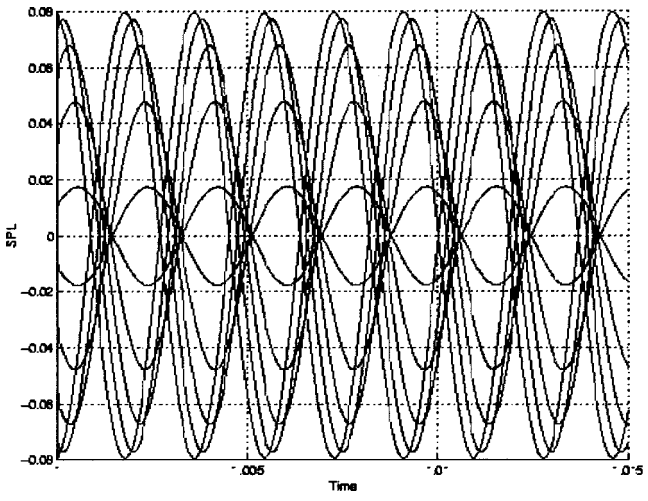


FIG. 3. Time domain array data from the simulations.

C. Results of numerical simulation

Using the numerically generated acoustic field, the simulated microphone signals were weighted so as to move the focus position over the linear region AB containing the source as shown in Fig. 2. It is clear from the beamformer response shown in Fig. 4(a) that the measured source distribution is in error, with a minimum sound pressure at the source location. By comparison, Fig. 4(b) shows the results of a simulation for a point monopole source with the same frequency as the dipole, and it is here readily observed that the beamformer correctly identifies the source.

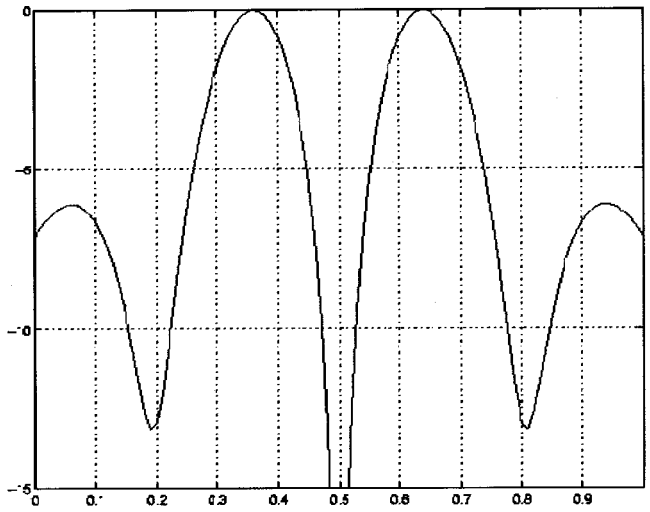
The mechanism responsible for the null measurement at the source location for the dipole is illustrated in Fig. 5, where the phase weighted signals for that focus position are shown. This shows that the signals prior to summation are such that destructive interference will occur, as the first 15 microphone signals are all in phase, with the second 15 out of phase. Thus, the out of phase characteristic of the microphones on either side of the zero-axis results in perfect destructive interference when the entire array of signals is combined. It is clear that, physically, this is due to the spherical wave-fronts being 180 degrees out of phase. This is the distinctive phase signature of the dipole and a modified procedure is required so that beamforming can be used to identify dipole sources which can result in coherent noise generation in aeroacoustic systems.

In order to implement such a procedure, the steps required will be as follows :

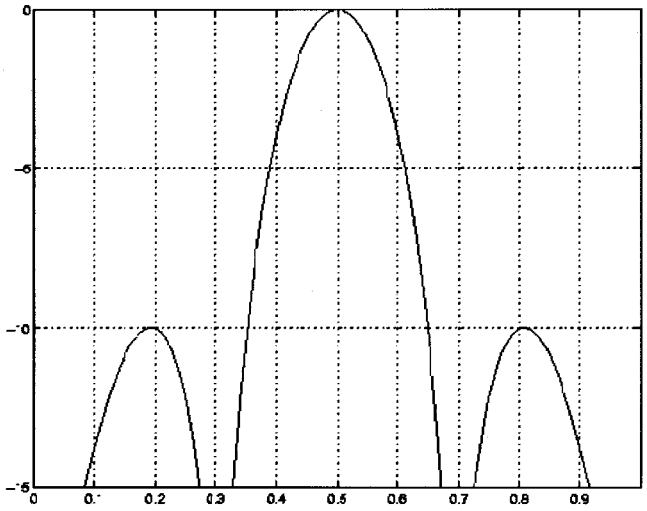
(1) Examine the phase characteristics of the signals from the array for each focus position of interest.
(2) Locate a potential dipole source as the position where the phase clearly moves through 180 degrees across the array response.
(3) Realign the data to correct for the phase difference.
(4) Proceed with the beamforming procedure as normal.

IV. EXPERIMENTAL SETUP AND ANALYSIS

A series of experiments was conducted to implement the procedures as described above. A cylinder was placed in an open jet flow so that an Aeolian tone was generated representing a realistic aeroacoustic dipole source. The experimental facility of Fig. 6 shows the semi-anechoic environ-



(a) dipole source



(b) Monopole source

FIG. 4. Beamformer response: (a) dipole source and (b) monopole source.

ment (effective down to a frequency of 500 Hz), the fan and nozzle arrangement and the microphone array. A support structure for the 4-mm-diam cylinder was mounted at the exit from the nozzle which had a maximum velocity of 50 m/s. A linear array consisting of 30 microphones was used to perform measurements with microphone spacing of 35 mm so that frequencies up to 4.9 kHz could be resolved. The array consisted of 30 KE4 Sensheiser electret microphones with a 20–20 000 Hz range and integrated preamplifiers. Data acquisition was performed using two 16-channel Kinetic Systems V200 acquisition cards, mounted on a National Instruments VXI chassis with each card capable of acquiring 16 channels of data simultaneously up to 200 kHz. A controller provided local control of the system and was connected to a PCI card on a PC via an MXI interface.

A single acquisition consisted of 150 000 data points acquired at 12.5 kHz. The signals were broken down into blocks of 4096 data points, and each block Fourier transformed. Phase weights were applied to the transforms according to Eq. (4), in order to steer the array focus over a

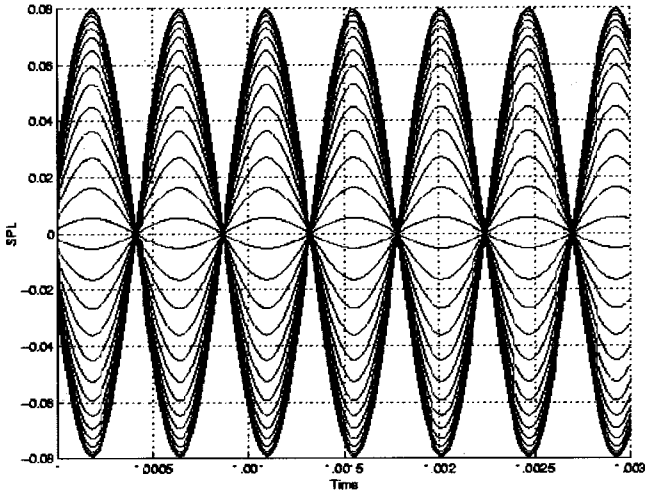


FIG. 5. Array response weighted for focus on source.

range of locations. This range was a 1-m linear region, parallel to the array axis and centered at the source. The weighted frequency domain signals were then summed, for each focus position, in order to give a beamformed FFT. This process was repeated for each consecutive block of 4096 data points, which for a total of 150 000 points results in 36 beamformed data sets. The beamformed spectrum was determined from these in the usual way with 36 ensemble averages. This then represents an estimation of the acoustic energy generated by a source at a given focus position.

V. RESULTS

The 4-mm cylinder in a cross-flow of 50 m/s, generated strong Aeolian tones at a frequency of 2560 Hz, and these were clearly audible above all other noise sources. Figure 7

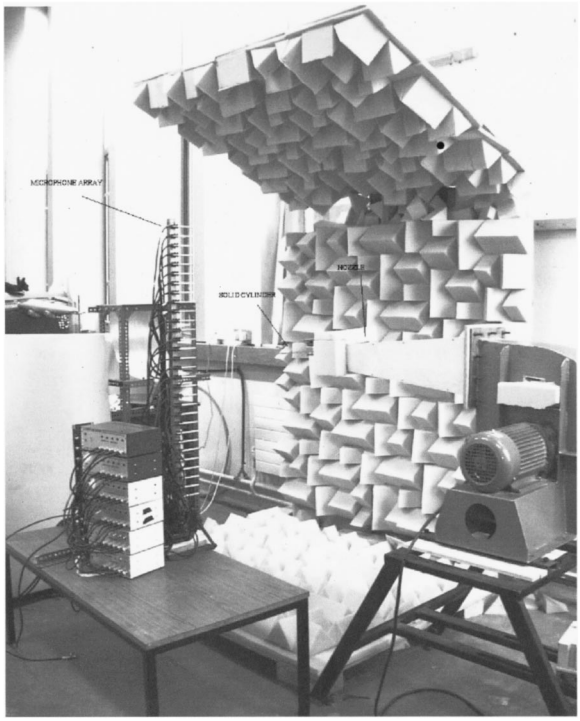


FIG. 6. Experimental setup.

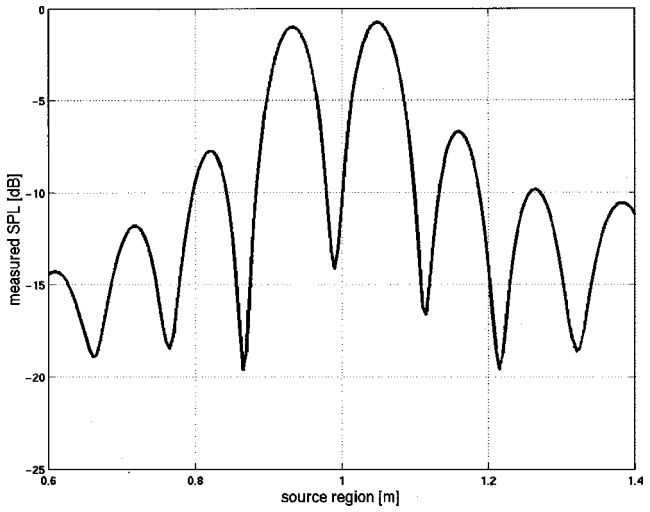


FIG. 7. Normal beamformer source map.

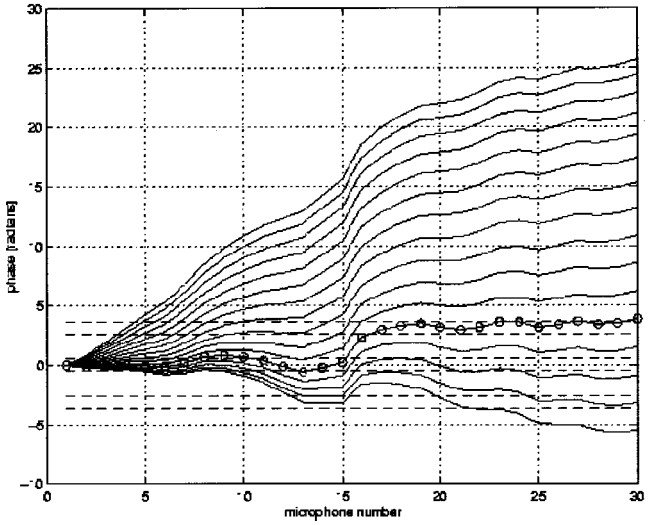


FIG. 9. Phase distributions for beamformer focus positions.

shows the beamformed spectral levels at the source frequency as a function of focus position. The source is located at 1 m and the results demonstrate clearly the inability of the array using the standard beamformer method either to locate the source or to estimate its amplitude. Even more striking is the result of the beamformed spectrum shown in Fig. 8. Although a strong peak was expected at a frequency of 2560 Hz, the spectrum is dominated by lower frequency fan, duct and nozzle noise, showing only a very weak response over the background level at the source frequency. Figure 9 shows phase distributions across the array, for a range of focus positions. A dipole signature can be identified where the phase moves through 180 degrees across the array and this is marked as a series of circles. This is the dipole source location, and to measure the source using the array, the beamformer must be corrected to allow for the distinctive phase characteristic of the dipole source.

A. Beamformer correction

Measurements taken with a beamformer give an average measure of sound power radiated from the source region in the direction of the array. However, this is only true when the signals are weighted such that those components of the signals representing energy which originated at the focus position are perfectly in phase. Even in the case of a monopole, standard phase weighting will not give perfect phase alignment, due to the nonideal nature of real acoustic sources and their propagation characteristics. For measurement of a dipole source the correction procedure proposed earlier in the article is applied to compensate for the phase misalignment shown in Fig. 9. Once this correction has been applied to give the phase distribution of Fig. 10, summation of the signals will result in constructive interference between components of the signal representative of the dipole. The effect of this phase correction on the source map is shown in Fig. 11 together with the original result from the standard beamformer. This source map now gives a maximum at the source where the ordinary beamformer showed a large attenuation. More dramatic, however, is the change in the beamformed spectrum shown in Fig. 12(a) where a peak at 2560 Hz is now clearly evident. To demonstrate the effectiveness of the method in identifying a dipole type source, results were obtained at two other flow velocities and the corrected beam-

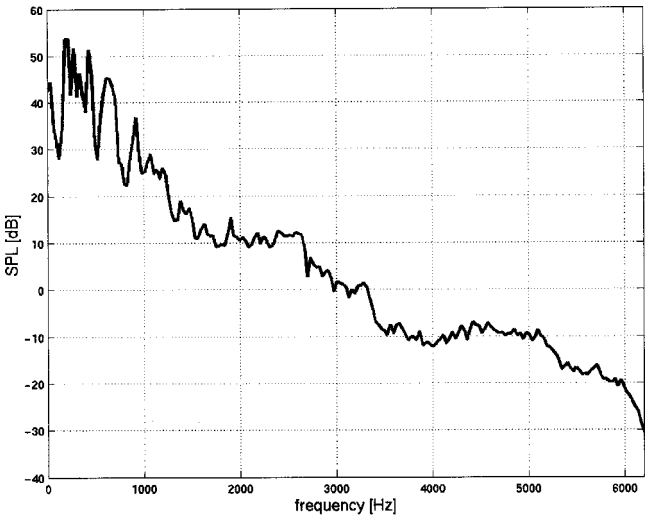


FIG. 8. Normal beamformer spectrum.

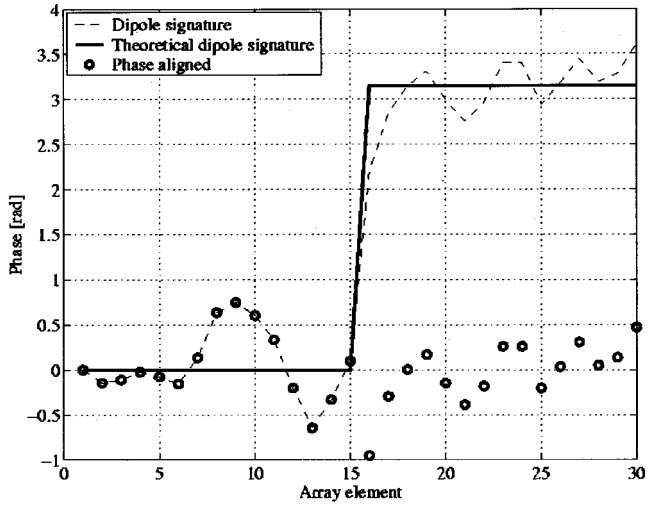


FIG. 10. Beamformer phase correction.

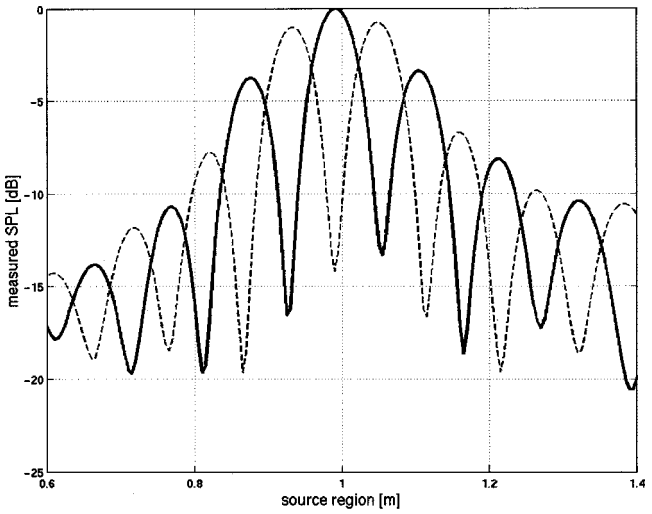


FIG. 11. Source map (- - : normal; ---: with dipole correction).

formed spectra are shown in Figs. 12(b) and (c). For these, clear peaks at 2200 and 1850 Hz can be observed whereas the spectra from the uncorrected beamforming approach show the same characteristics for all three velocities.

VI. CONCLUSIONS

The use of the standard beamforming method fails to locate a simple aeroacoustic dipole source and results in a spectrum which does not represent the true character of the source. The spectrum generated by the standard beamformer method demonstrates the inability of this technique to identify a discrete frequency dipole source in an aeroacoustic system. The source location and the spectra resulting from the phase correction for measurement of a dipole show that the proposed method works well with experimental data. The results of the corrected spectra are indicative of a number of phenomena. The first is the capacity of the beamformer to identify discrete frequency dipole energy sources in the frequency range 1800–2570 Hz. Secondary effects are the reductions in energy shown in the frequency ranges 0–1500 Hz and at higher frequencies than that of vortex shedding. By phase correction of the beamformer to increase its sensitivity to the dipole source, the array appears to have become less sensitive to other source types, and, so, sound energy which was present in the system as a result of other aeroacoustic phenomena has been attenuated. This indicates that the spectrum generated by the standard beamformer represents the sound energy of sources such as the fan, duct, nozzle and jet (all of these lie on a horizontal axis containing the dipole, perpendicular to the array axis, and, thus, in the same look direction), whereas the modified spectrum represents the dipole energy. Thus, there is potential for the technique to be developed so that it can be used to discriminate between different source types as well as acting as a spatial filter.

The results demonstrate the degree to which an array measurement can be in error. Aeroacoustic systems are generally directional, with a large variety of different source mechanisms, all combining to produce the resultant acoustic field. This work constitutes a worst case scenario, where a very obvious source is completely missed by the standard technique. This is clearly a consequence of the linear array being aligned normal to the dipole axis. Obviously, if the array was aligned parallel to the dipole axis, it would see an

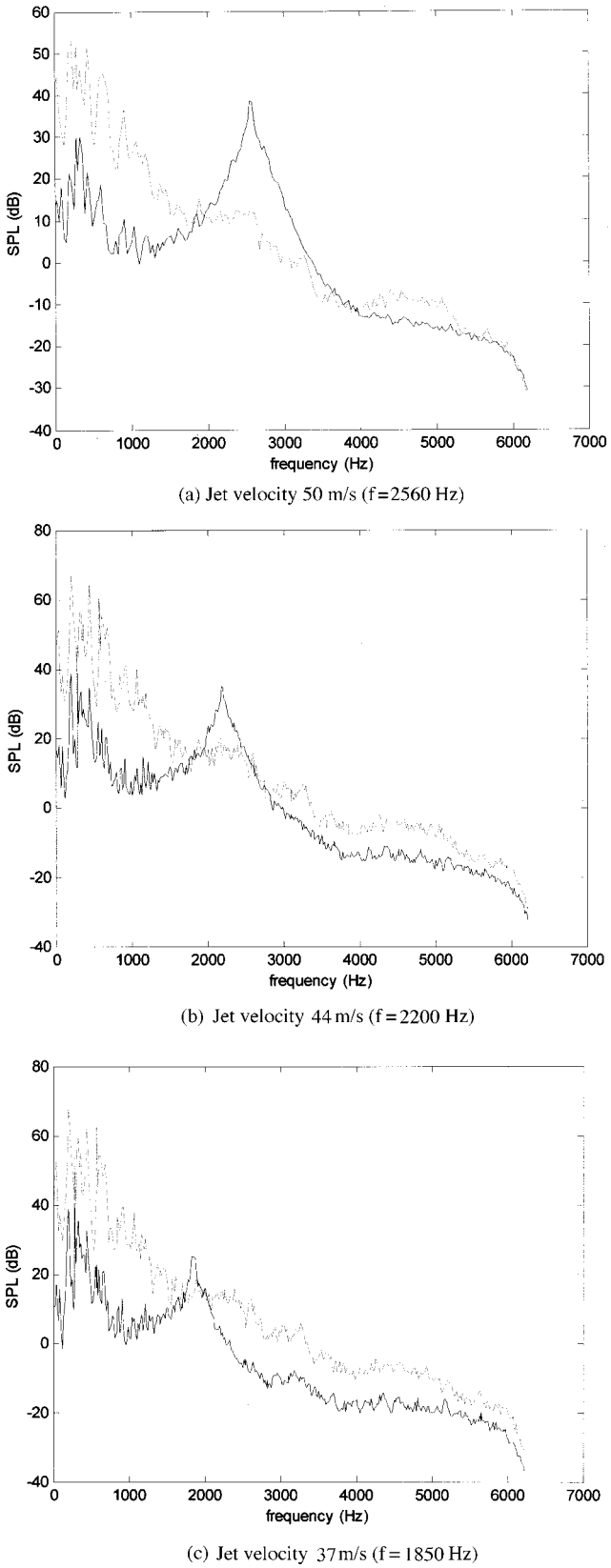


FIG. 12. Computed spectra (- - : normal; --- : with dipole correction). (a) Jet velocity 50 m/s (f=2560 Hz). (b) Jet velocity 44 m/s (f=2200 Hz). (c) Jet velocity 37 m/s (f=1850 Hz).

apparent monopole and the source map would identify this. In practical cases, it is likely that the array axis will be oblique and, therefore, the likely orientation of potential aeroacoustic sources should be considered before experiments are performed using arrays. For many airframe applications, the aeroacoustic sources arise from fluctuating loading forces and these are principally of the dipole type. Further work is required to investigate the application of the proposed method to more complex systems.

[1] J. Billingsley and R. Kinns, ''The acoustic telescope,'' J. Sound Vib. **48**, 485–510 (1976).

[2] M. J. Fisher, M. Harper-Bourne, and S. A. L. Glegg, ''Jet Engine Noise Source Location: The Polar Correlation Technique,'' J. Sound Vib. **51**, 23–54 (1977).

[3] M. A. Marcolini and T. F. Brooks, ''Rotor Noise Measurement Using a Directional Microphone Array,'' J. Am. Helicopter Soc. **37**, 11–22 (1992).

[4] T. F. Brooks and W. M. Humphreys, Jr., ''Effect Of Directional Array Size On The Measurement Of Airframe Noise Components,'' AIAA Paper No. 99-1958, 5th AIAA/CEAS Aeroacoustics Conference, Bellevue, Washington, May 1999.

[5] H. A. Siller, F. Arnold, and U. Michel, ''Investigation of Aero-Engine Core Noise using a Phased Microphone Array,'' AIAA Paper No. 2001-2269, 7th AIAA/CEAS Aeroacoustics Conference, Maastricht, Netherlands, May 2001.

[6] H. Holthusen and H. Smit, ''A New Data Acquisition System for Microphone Array Measurements in Wind Tunnels,'' AIAA Paper No. 2001-2169, 7th AIAA/CEAS Aeroacoustics Conference, Maastricht, Netherlands, May 2001.

[7] O. M. Phillips, ''The Intensity of Aeolian Tones,'' J. Fluid Mech. **1**, 607–624 (1956).

[8] P. Leehey and C. E. Hanson, ''Aeolian Tones Associated with Resonant Vibrations,'' J. Sound Vib. **13**, 465–483 (1971).

[9] A. Powell, ''Some Aspects of Aeroacoustics: From Rayleigh Until Today,'' J. Vibr. Acoust. **112**, 145–159 (1990).

[10] J. E. Ffowcs-Williams and D. L. Hawkings, ''Sound Generation by Turbulence and Surfaces in Arbitrary Motion,'' Phil. Trans. R. Soc. A, **264**, 321–343 (1969).

[11] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing* (Prentice Hall, Englewood Cliffs, NJ, 1993).

Estimation of broadband acoustic power radiated from a turbulent boundary layer-driven reinforced finite plate section due to rib and boundary forces

M. L. Rumerman
Naval Surface Warfare Center, Carderock Division, Signatures Directorate (Code 7204), 9500 MacArthur Boulevard, West Bethesda, Maryland 20817-5700

(Received 30 March 2001; accepted for publication 31 October 2001)

Previous papers considered an infinite fluid-loaded plate with parallel line attachments, driven by a wave-number-white pressure excitation invariant in the direction of an attachment, and established the conditions and procedure for estimating the broadband radiated power by assuming the ribs to radiate independently. This paper applies those results to a finite rectangular ribbed plate, and extends the methodology to include the contribution of the plate's boundary support forces to the radiation and the consideration of excitation that varies in the direction parallel to the ribs. The approach is relevant to problems of sound radiation by underwater stiffened steel plates driven by turbulent boundary layer (TBL) pressures, and is also applicable to stiffened circular cylindrical shells when the response is dominated by bending. Comparisons of sample calculations with results of rigorous models validate the approximation. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1446049]

PACS numbers: 43.40.At, 43.40.Hb, 43.40.Rj, 43.30.Jx [PJR]

I. INTRODUCTION

It was shown by Ffowcs Williams[1,2] that the sound power radiated by a turbulent boundary layer (TBL) on an extended surface is due to the sum of two principal effects. There is direct radiation from the multipole structure of the turbulence whose magnitude is not significantly affected by the nature of the surface. There is also radiation due to the interaction of the surface vibration with structural discontinuities, and this component usually dominates the first.

Outboard surfaces of underwater structures are often extended panels having discontinuities in the form of reinforcing parallel ribs, and the mutual forces exerted between the panel and each rib, in response to the TBL-generated panel vibration, generate radiated sound power. If there are competing loss mechanisms, a mathematical model must be able to properly apportion power radiated with that lost to structural damping and other mechanisms in order to determine the panel response and the resulting radiated power. A rigorous calculation requires a complete solution for the system because the phase relationships among the various rib forces are important in determining the net radiation. Such a solution is difficult to formulate and can require a very time-consuming execution for random excitation and response of a realistic panel. It is inherently a narrow-band calculation and delivers a precision that may not be justified by the ability to describe the structure.

References 3 and 4 developed an approach for directly estimating the rib-related radiated power in broad (e.g., one-third-octave) frequency bands. An idealized one-dimensional model of a string, loaded at discrete points with identical attachments (masses or springs) and driven by a wave-number-white mechanical excitation, was considered in Ref. 3. It was demonstrated that, when a frequency band contains at least two resonances of the finite string, and the attachment forces can be considered to radiate independently and similarly, then the mean-squared force within the band can be estimated without considering interactions among the attachments. The radiated power is then equal to the product of the number of attachments and the mean-squared force and the power radiated per unit mean-squared force. The assumption that the attachments could be considered to radiate independently and similarly was examined in Ref. 4 using the model of a membrane containing equally spaced parallel line masses, with excitation and response invariant in the direction parallel to an attachment. It was shown to be valid when at least one Bloch wave number of the periodic structure falls within the band of radiating wave numbers. This always results when the acoustic wavelength is smaller than two times the rib spacing, and generally occurs within the next lower frequency octave, in which the acoustic wavelength is smaller than four times the rib spacing.

The approximate solution was tested in Ref. 4 by using it to estimate the power radiated per attachment for infinite membranes and plates containing periodically spaced line masses. The structure was assumed to be driven by a random driving pressure on the surface, invariant in the direction parallel to an attachment and wave-number-white for the direction transverse to the attachments. This was intended to represent the spectral density of a blocked TBL pressure in the subconvective range. The parameters represented steel panels in water, and underwater flow speeds are low enough, compared to the plating bending wave speed, so that hydrodynamic coincidence is not a factor. An exact solution was obtained by applying the formalism of Ref. 5 to a wave-number-white random surface excitation. The results validated the approximation.

The analysis of Ref. 3 ignored radiation from the ends of the finite string, and because a string was used as the structural model, the effects of finite width in the direction paral-

 0001-4966/2002/111(3)/1274/6/$19.00 © 2002 Acoustical Society of America

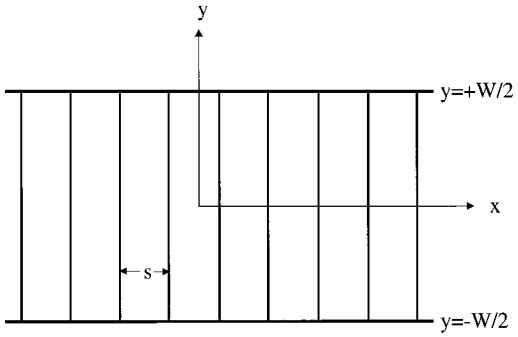


FIG. 1. Ribbed plate strip of finite width.

lel to an attachment and the variation of excitation and response in that direction were also ignored. This paper considers these issues by first extending the formulation to plates having finite width parallel to a rib, and then suggesting an approximation for the contribution of radiation from panel edges that are parallel to the ribs.

II. FINITE WIDTH PANELS

Consider a flat plate lying in the x–y plane, as shown in Fig. 1. The plate is of infinite length in the x direction and of finite width, W, in the y direction, occupying the strip $|y| < W/2$. The plate is reinforced by identical ribs that are parallel to the y axis and have average spacing, s, along the x axis. It is assumed to be rigidly baffled in the x–y plane for $|y| > W/2$ with a semi-infinite acoustic fluid in the space above the plate.

The plate is assumed to respond in bending only, and is described by thin-plate theory. Letting α and γ be the wave numbers corresponding to the x- and y directions, respectively, and with ω equal to 2π times the frequency, the wave number-frequency impedance of the unstiffened plate, $Z_p(\alpha,\gamma;\omega)$, is given by Eq. (1). In this equation, M is the mass per unit area of the plate, D is its bending rigidity, η is the associated loss factor for bending, and k_p is the complex bending wave number of the plate at the particular frequency

$$Z_p(\alpha,\gamma;\omega) = -i\omega M\left[1-\frac{(\alpha^2+\gamma^2)^2}{k_p^4}\right]; \quad k_p^4 = \frac{M\omega^2}{D(1-i\eta)}. \tag{1}$$

The wave-number-frequency impedance of the acoustic fluid above the plate is given by Eq. (2), in which ρ is the mass density of the fluid, c is its sound speed, and k is the acoustic wave number at the frequency

$$Z_f(\alpha,\gamma;\omega) = \frac{\rho\omega}{\sqrt{(k^2-\alpha^2-\gamma^2)}}; \quad k = \frac{\omega}{c}. \tag{2}$$

The wave-number-frequency impedance of a rib, modeled as a beam in bending, is given by Eq. (3a), in which τ is the mass per unit length of the beam, EI is the equivalent bending rigidity of the attached beam's cross section, η_r is the associated bending loss factor, and k_r is the beam's complex bending wave number

$$Z_{\text{rib}}(\gamma;\omega) = -i\omega\tau\left[1-\frac{\gamma^4}{k_r^4}\right]; \quad k_r^4 = \frac{\tau\omega^2}{EI(1-i\eta_r)}. \tag{3a}$$

The admittance of the beam is defined as the reciprocal of the impedance

$$Y_{\text{rib}}(\gamma;\omega) = Z_{\text{rib}}^{-1}(\gamma;\omega). \tag{3b}$$

A time-dependence $\exp(-i\omega t)$ has been assumed.

The basis of the approximate solution is that the structural dissipation in the ribs and their radiation losses (when the rib forces can be considered to radiate independently) can be spatially averaged to produce loss factors that can be added to the loss factor of the homogeneous plate. Sections III and IV of Ref. 4 summarize the approximate solution for membranes and plates infinite in the y direction and with no y dependence of excitation and response (i.e., $\gamma=0$). However, any value of γ could have been assumed in the analyses leading to those results. The steps are summarized here for the plate infinite in the y direction, with arbitrary γ.

(1) Compute $Y_\infty(\gamma;\omega)$, the input admittance for a line force acting on the infinite, uniform fluid-loaded plate, and $Y_{\text{rad}}(\gamma;\omega)$, the acoustic power radiated per unit line force

$$Y_\infty(\gamma;\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}[Z_p(\alpha,\gamma;\omega)+Z_f(\alpha,\gamma;\omega)]^{-1}d\alpha, \tag{4}$$

$$Y_{\text{rad}}(\gamma;\omega) = \frac{1}{2\pi}\int_{-k}^{+k} Z_f(\alpha,\gamma;\omega)\left|Z_p(\alpha,\gamma;\omega) + Z_f(\alpha,\gamma;\omega)\right|^{-2}d\alpha. \tag{5}$$

(2) Compute β_{diss} and β_{rad}, the spatially averaged equivalent damping coefficients for the rib dissipation and radiation, respectively. The dissipation coefficient for the rib is given in explicit form, and represents the power lost in bending deformation

$$\beta_{\text{diss}}(\gamma;\omega) = \frac{\omega(\eta_r D\gamma^4/\omega^2)|Y_{\text{rib}}^2(\gamma;\omega)|}{s|Y_\infty(\gamma;\omega)+Y_{\text{rib}}(\gamma;\omega)|^2} \tag{6a}$$

$$\beta_{\text{rad}}(\gamma;\omega) = \frac{Y_{\text{rad}}(\gamma;\omega)}{s|Y_\infty(\gamma;\omega)+Y_{\text{rib}}(\gamma;\omega)|^2}. \tag{6b}$$

(3) Compare the mean-squared velocity of the equivalent plate, assuming a wave-number-white excitation spectral density, $S(\omega)$

$$\langle|v(\gamma;\omega)|^2\rangle = S(\omega)\int_{-\infty}^{+\infty}|Z_p(\alpha,\gamma;\omega)+Z_f(\alpha,\gamma;\omega) + [\beta_{\text{diss}}(\gamma;\omega)+\beta_{\text{rad}}(\gamma;\omega)]|^{-2}d\alpha. \tag{7}$$

(4) Compute the mean-squared attachment force

$$\langle|F^2(\gamma;\omega)|\rangle = \frac{\langle|\nu^2(\gamma;\omega)|\rangle}{|Y_\infty(\gamma;\omega)+Y_{\text{rib}}(\gamma;\omega)|^2}. \tag{8}$$

(5) Compare the acoustic power radiated per rib in a frequency band $\Delta\omega$ by considering the rib attachment force in isolation. When evaluating the quantities in the previous equations, ω is to correspond to the band's center frequency

$$\langle P_{\mathrm{rad}}\rangle = \langle |F^2(\gamma;\omega)| \rangle Y_{\mathrm{rad}}(\gamma;\omega)\Delta\omega. \tag{9}$$

A. Expansion in cross modes

If the excitation and responses of the plate illustrated in Fig. 1 vary in the y direction, their y dependence may be represented in terms of a full-range Fourier series in exponential form. For an excitation that is deterministic in the y direction, the normal force that a rib exerts on the plate can be written in the forms of Eq. (10)

$$F(y) = \frac{1}{W}\sum_{n=-\infty}^{+\infty} A_n \exp\left[\frac{+2\pi i n y}{W}\right] \tag{10a}$$

$$A_n = \int_{-w/2}^{+w/2} F(y)\exp\left[\frac{-2\pi i n y}{W}\right] dy. \tag{10b}$$

Let $Y_{\mathrm{rad}}(2\pi n/W)$ be the real part of the modal acoustic radiation admittance of a rib force applied along a finite length line segment. If W is large enough in comparison to the acoustic wavelength, then $Y_{\mathrm{rad}}(2\pi n/W)$ may be approximated by Eq. (5) with $\gamma = 2\pi n/W$. When $|n| > kW/(2\pi)$, Y_{rad} vanishes. Below the coincidence frequency, this is equivalent to ignoring radiation from corner modes in comparison to edge modes along the rib. Above the coincidence frequency, edge modes along the rib are ignored in comparison to surface modes. When coupling among these ''cross modes'' is ignored, the acoustic power radiated by the line force is given by Eq. (11), in which the asterisk denotes the complex conjugate and J is the largest positive integer less than $kW/(2\pi)$

$$\begin{aligned} P_{\mathrm{rad}} &= \int_{-w/2}^{+w/2} dy \sum_{n=-J}^{+J} \frac{A_n^*}{W}\exp\left[\frac{-2\pi i n y}{W}\right] \sum_{m=-J}^{+J} \frac{A_m}{W} \\ &\quad \times \exp\left[\frac{+2\pi i m y}{W}\right] Y_{\mathrm{rad}}\left[\frac{2\pi m}{W}\right] \\ &= \sum_{n=-J}^{+J} \frac{|A_n^2|}{W} Y_{\mathrm{rad}}\left[\frac{2\pi n}{W}\right]. \end{aligned} \tag{11}$$

Because the excitation and response are random in the y direction, the expected value of the power is required, which is defined by Eq. (12)

$$\langle P_{\mathrm{rad}}\rangle = \frac{1}{W}\sum_{n=-J}^{+J} \langle A_n A_n^*\rangle Y_{\mathrm{rad}}\left[\frac{2\pi n}{W}\right], \tag{12a}$$

$$\begin{aligned} \langle A_n A_n^*\rangle &= \int_{-w/2}^{+w/2} dy_1 \int_{-w/2}^{+w/2} dy_2 \langle F(y_1)F^*(y_2)\rangle \\ &\quad \times \exp\left[\frac{-2\pi i n y_1}{W}\right]\exp\left[\frac{+2\pi i n y_2}{W}\right]. \end{aligned} \tag{12b}$$

If the random response in the y direction is treated as spatially homogeneous, even though the width is finite, then

$$\langle F(y_1)F^*(y_2)\rangle = R_y(y_1 - y_2), \tag{13}$$

in which $R_F(\varepsilon)$ is the autocorrelation function of F at interval ε. Equation (13) may be substituted into Eq. (12b)

$$\langle A_n A_n^*\rangle = \int_{-w/2}^{+w/2} dy_2 \int_{-y_2-w/2}^{-y_2+w/2} R_f(\varepsilon)\exp\left[\frac{-2\pi i n \varepsilon}{W}\right] d\varepsilon. \tag{14a}$$

Assuming that W is much larger than the correlation length in the y direction, Eq. (14a) may be replaced by Eq. (14b), in which $\Phi_F[2\pi n/W]$ is the mean-squared spectral density of F in the y direction

$$\begin{aligned} \langle A_n A_n^*\rangle &= \int_{-w/2}^{+w/2} dy_2 \int_{-\infty}^{+\infty} R_f(\varepsilon)\exp\left[\frac{-2\pi i n \varepsilon}{W}\right] d\varepsilon \\ &= 2\pi W \Phi_F\left[\frac{2\pi n}{W}\right]. \end{aligned} \tag{14b}$$

The spectral density Φ_F is equivalent to the left-hand side of Eq. (8) with $\gamma = [2\pi n/W]$; therefore, upon substituting Eq. (14b) into Eq. (12a), Eq. (15) is obtained

$$\langle P_{\mathrm{rad}}\rangle = 2\pi \sum_{n=-J}^{+J} \left\langle \left| F\left(\frac{2\pi n}{W}\right)\right|^2 \right\rangle Y_{\mathrm{rad}}\left[\frac{2\pi n}{W}\right]\Delta W. \tag{15}$$

B. Comparison with exact results

1. Exact solution

As described in Sec. III B of Ref. 4, an exact solution can be generated from the results of Ref. 5 with the addition of fluid loading and the assumption of a wave-number-white random excitation. For the two-dimensional problem, the resulting mean-squared spectral density of the plate velocity has the general form of Eq. (16)

$$S_\nu(\alpha,\gamma;\omega) = S_\nu^0(\alpha,\gamma;\omega) + S_\nu^{0r}(\alpha,\gamma;\omega) + S_\nu^r(\alpha,\gamma;\omega). \tag{16}$$

The first term on the right-hand side represents the velocity spectral density that the excitation would produce on the uniform plate without ribs. The third term represents the velocity spectral density produced on the uniform plate by the induced rib attachment forces. The second term represents an interaction between the other two velocity fields. The acoustic power radiated per ''bay'' (i.e., portion of the plate between two adjacent attachments) within the frequency band by that part of the velocity field due to the rib forces alone, and equivalent to the power radiated per rib, is given by Eq. (17)

$$\begin{aligned} P_{\mathrm{rad}}^r &= 2\pi s \int_{\Delta\omega} d\omega \sum_{n=-J}^{+J} \int_{-k}^{+k} S_\nu^r\left[\alpha,\left(\frac{2\pi n}{W}\right);\omega\right] \\ &\quad \times Z_f\left[\alpha,\left(\frac{2\pi n}{W}\right);\omega\right] d\alpha. \end{aligned} \tag{17}$$

The results of Eq. (17) will be compared to those of Eq. (15).

2. Description of plates

Comparisons of calculations will be presented for the same two infinitely extended steel plates in water that were considered in Ref. 4. Each plate is reinforced by an infinite number of equally spaced steel ribs having a cross-sectional area of 75 cm^2 and an effective moment of inertia for bending of 7500 cm^4. The mass density and sound speed for

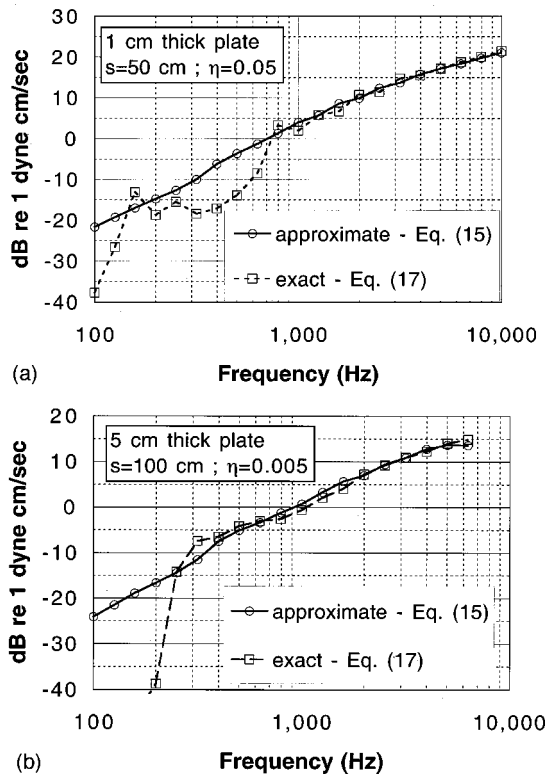


FIG. 2. (a) Comparison of calculated power radiated per rib for 1-cm plate. (b) Comparison of calculated power radiated per rib for 5-cm plate.

water are taken as 1.0 g/cm^3 and 150 000 cm/s, respectively. The mass density and plate modulus for steel are taken as 7.8 g/cm^3 and 2.0×10^{12} dyne/cm^2, respectively. No mechanical damping is associated with the rib. The first plate is 1 cm thick with a bending loss factor of 0.05 and a rib spacing of 50 cm. The second plate is 5 cm thick with a loss factor of 0.005 and a rib spacing of 100 cm. For both plates, $W=500$ cm. The driving spectral density is again assumed to be wave-number-white and frequency independent

$$S(\alpha,\gamma;\omega)=S_0(\omega)=1(\text{dyne/cm})^2\text{s}. \tag{18}$$

3. Results

The development in Ref. 4 indicates that the approximate calculation of power radiated per rib (or per bay) should be within 3 dB of the exact calculation when $ks>\pi/2$. Accordingly, the minimum frequency of applicability is 750 Hz for the 1-cm-thick plate having a rib spacing of 50 cm and 375 Hz for the 5-cm-thick plate having a rib spacing of 100 cm. Figures 2(a) and (b) show agreement of the approximate and exact calculations above these frequencies.

The mean-squared velocity of the panel, $|V^2|$, is also of interest. Analogous to the result for radiated power, the approximation for the panel of width W is given by Eq. (19). Here, the summation is over all cross modes

$$\langle|V^2|\rangle=\frac{2\pi}{W}\sum_{n=-\infty}^{+\infty}\left\langle\left|\nu^2\left(\frac{2\pi n}{W};\omega\right)\right|\right\rangle\Delta\omega. \tag{19}$$

There is a corresponding expression for the exact solution. The approximate and exact results for the two panels are compared in Fig. 3.

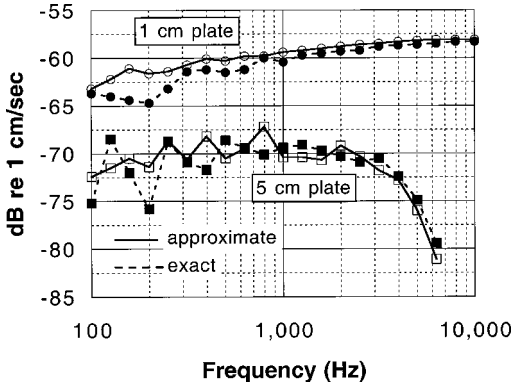


FIG. 3. Comparison of calculated mean-squared plate velocities.

III. FINITE PANEL SECTIONS

When the edge of a panel is explicitly accounted for in structural acoustic problems, the panel is usually considered to be in free space or bounded by an infinite impedance coplanar baffle. For most practical problems of steel structures in water, neither the free space nor baffle idealizations represents the actual situation, because the panel section of interest is connected to the rest of the structure, often at a large impedance discontinuity. The coupling between the panel and remaining structure should be accounted for. This will be done through an idealized representation, illustrated in Fig. 4, in which the panel of interest is composed of $N+1$ bays separated by N parallel ribs with average spacing s and terminated on each end by an infinite impedance line force support. The length of the section in the x direction (transverse to the ribs) is $L=(N+1)s$. The rest of the structure to which this section is attached is taken as its periodic extension, including the infinite impedance supports. The width of the panel in the y direction (parallel to a rib) is W.

A. Extension of approximation

As discussed above, the ribs can always be considered to radiate independently when $ks>\pi$ because, at these frequencies, there must be at least one radiating Bloch wave. As a practical matter, the assumption can be extended down to $ks>\pi/2$ because that Bloch wave is unlikely to suddenly drop out of the radiation band as the frequency decreases below that cutoff. For the doubly periodic system, there is a second set of Bloch waves based on the length L. At least

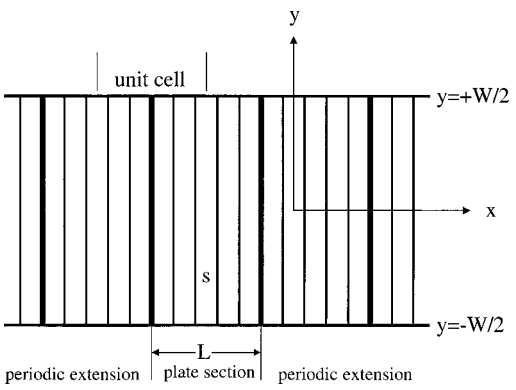


FIG. 4. Finite ribbed-plate section with boundary supports.

one of these will always radiate when $kL > \pi$, and the supports can be assumed to radiate independently when $kL > \pi/2$.

Just as a spatially averaged equivalent radiation damping coefficient for a cross mode was defined for the ribs based on the assumption of independent rib radiation, a cross-mode radiation damping coefficient, β_b, can be defined for the infinite impedance supports. This is given by Eq. (20)

$$\beta_b(\gamma;\omega) = \frac{Y_{\text{rad}}(\gamma;\omega)}{L|Y_\infty(\gamma;\omega)|^2}. \tag{20}$$

The mean-squared plate velocity for the cross mode is then given by Eq. (21), in which β_{diss} and β_{rad} are given by Eq. (6)

$$\langle|\nu(\gamma;\omega)|^2\rangle = S(\omega)\int_{-\infty}^{+\infty}|Z_p(\alpha,\gamma;\omega)+Z_f(\alpha,\gamma;\omega) + [\beta_{\text{diss}}(\gamma;\omega)+\beta_{\text{rad}}(\gamma;\omega) + \beta_b(\gamma;\omega)]|^{-2}d\alpha. \tag{21}$$

The corresponding mean-squared boundary support force is given by Eq. (22)

$$\langle|F_b^2(\gamma;\omega)|\rangle = \frac{\langle|\nu^2(\gamma;\omega)|\rangle}{|Y_\infty(\gamma;\omega)|^2}. \tag{22}$$

The power radiated by the panel section of length L and width W is then represented by Eq. (23), in which Y_{rad} and F_r are given by Eqs. (5) and (8), respectively, and it is assumed that $kL > \pi/2$

$$\langle P_{\text{rad}}\rangle = 2\pi\sum_{n=-J}^{+J}\left[N\left\langle\left|F_r\left(\frac{2\pi n}{W}\right)\right|^2\right\rangle + \left\langle\left|F_b\left(\frac{2\pi n}{W}\right)\right|^2\right\rangle\right]Y_{\text{rad}}\left[\frac{2\pi n}{W}\right]\Delta\omega; \quad ks > \pi/2, \tag{23a}$$

$$\langle P_{\text{rad}}\rangle = 2\pi\sum_{n=-J}^{+J}\left\langle\left|F_b\left(\frac{2\pi n}{W}\right)\right|^2\right\rangle Y_{\text{rad}}\left[\frac{2\pi n}{W}\right]\Delta\omega; \quad ks < \pi/2. \tag{23b}$$

In Eq. (23), explicit use has been made of the fact that the ribs cannot be *expected* to radiate at frequencies below $ks < \pi/2$. In this frequency range, β_{rad} is set equal to zero in Eq. (21). Note that only one boundary support is allocated to the panel section. This can be justified by defining a unit cell of the infinite periodic extension so that the support is at its center.

B. Comparison with exact results

Mace[6] has presented an exact analysis of an infinite plate reinforced by a doubly periodic stiffener-support system. This can be used to check the approximation of radiation by the finite section in the same manner as for one bay of the singly periodic ribbed plate. The two plates previously considered will again be used for illustration. Each will be given nine ribs between the major supports so that $L = 500$ cm for the 1-cm-thick plate and $L = 1000$ cm for the 5-cm-thick plate. Results will be given only for the uniform cross mode (i.e., $n = 0$), so that the width, W, does not enter. Figures 5(a) and (b) show comparisons in relative dB.

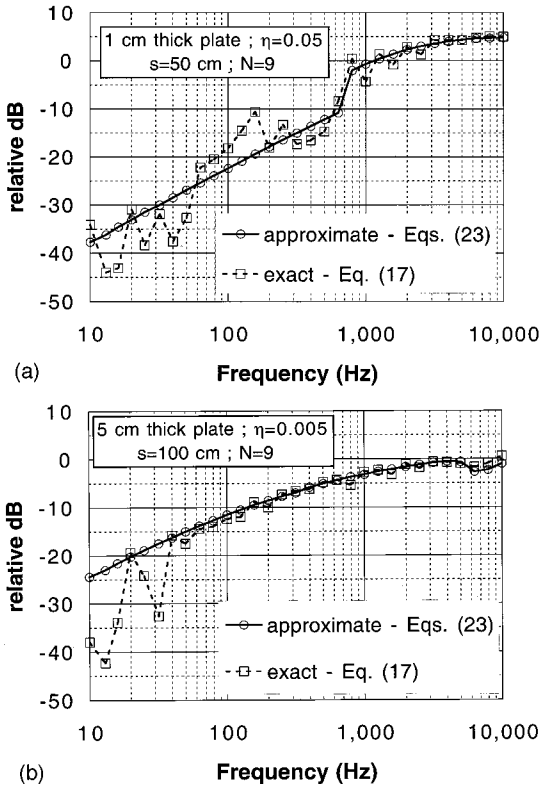


FIG. 5. (a) Comparison of calculated power radiated by 1-cm plate section. (b) Comparison of calculated power radiated by 5-cm plate section.

The theory indicates that, based on the length L, the approximations for the thinner and thicker plate are valid above 75 and 37.5 Hz, respectively. The results indicate very good agreement between the approximate and exact formulations above 40 Hz for the thicker plate. Similar agreement is found above 80 Hz for the thinner plate, except in the 100–200-Hz band. This discrepancy is due to a rib-related Bloch wave that has passed back into the radiating band, but cannot be accounted for by the approximate theory when $ks < \pi/2$.

IV. CONCLUSION

This paper has considered sound radiation due to TBL excitation of a rectangular plate stiffened by a set of ribs parallel to two of the sides. Results of previous papers were used to establish a procedure for directly estimating broadband levels of radiated power due to the interaction forces between the ribs/parallel boundary supports and the plate. Because the intended application is to steel (or aluminum) plates in water, for which the flow speeds are much less than typical plate-bending wave speeds, there are no effects due to hydrodynamic coincidence. The plate was assumed to be driven by the subconvective portion of the frequency-wavenumber spectral density of the blocked TBL pressure in the neighborhood of the plate's bending wave number in the fluid. For typical steel plates and frequencies of interest, the spectral density in this region is considered to be slowly

varying and, for purposes of development and illustrative calculations, the spectral density was assumed to be wave-number-white. Because the results of the derivations in Refs. 3 and 4 displayed themselves in terms of robust responses such as mean-squared plate velocity, the approximate formalism appears to be also applicable to the slowly varying TBL pressure spectral density. Results for a wave-number-dependent form can be used by replacing $S(\omega)$ appearing outside the integrals of Eqs. (7) and (21), with $S(\alpha,\gamma;\omega)$ in the integrands.

An infinite periodic extension of a series of ribs between major supports was used to evaluate the approach to including radiation due to boundaries parallel to the ribs. Although an idealization, it allows an approximation of the coupling between connected sections that are not too dissimilar, and is better than a rigid baffle extension. Because the distance between the plate's edge supports is $(N+1)s$, the frequency at which the supports can be considered to radiate independently is much lower than the cutoff frequency for the ribs. By including the contribution of the supports, the rib radiation can be shut off when $ks<\pi/2$, and a good estimate of the panel's radiation is still achieved. The exception is when a rib-related Bloch wave number has moved back into the radiation band. The approximate theory cannot account for this. With this exception, the comparison of exact and approximate powers shows very good agreement, typically within 3 dB.

The coplanar extension of the plate in the y direction was taken to be a rigid baffle. This assumption does not significantly affect the radiation due to the supersonic edge-mode components of the rib forces, but it does omit the effects of coupling to the rest of the structure.

In addition to radiation due to rib and boundary forces, there is direct radiation from the TBL multipoles. This power, which corresponds to radiation from the supersonic wave-number range of the TBL pressure spectral density, is not significantly dependent upon panel properties unless the panel is very compliant.[1] Unfortunately, estimates of spectral density levels in the supersonic wave-number range are less well-established than those for higher subconvective wave numbers, and corresponding predictions of directly radiated power are more problematic. This means that estimates of rib- and boundary-related power above the coincidence frequency are also questionable, because the bending wave number is, by definition, in the supersonic range.

Although thin-plate theory was used in the calculations, a thick-plate bending model can also be used. Additionally, the approach can be applied to circular cylindrical shells, whose response is dominated by bending, by using structural and acoustic impedances appropriate to that geometry, and setting the width, W, equal to 2π times the shell radius.

ACKNOWLEDGMENT

This work was supported by the Office of Naval Research Hydroacoustics Program: L. Patrick Purtell, Program Manager.

APPENDIX: CLARIFICATIONS

It can be seen from Eq. (7) that the quantities designated β have the dimensions of wave-number impedance, and perhaps should have been called equivalent damping *impedances* rather than *coefficients*. In addition, due to a manuscript error in Ref. 3, Eqs. (33a) and (33b) of that paper are given incorrectly. They should read as follows:

$$\nu(x)=\nu_u(x) + i\frac{F}{mc_0}\frac{\sin[k(L/2+\xi)]\sin[k(L/2-x)]}{\sin(kL)}; \quad x>\xi,$$

$$\nu(x)=\nu_u(x) + i\frac{F}{mc_0}\frac{\sin[k(L/2-\xi)]\sin[k(L/2+x)]}{\sin(kL)}; \quad x<\xi.$$

[1] J. E. Ffowcs Williams, "Sound radiation from turbulent boundary layers formed on compliant surfaces," J. Fluid Mech. **22**, 347–358 (1965).
[2] J. E. Ffowcs Williams, "The influence of simple supports on the radiation from turbulent flow near a plane compliant surface," J. Fluid Mech. **26**, 641–649 (1966).
[3] M. L. Rumerman, "Estimation of broadband acoustic power due to rib forces on a reinforced panel under turbulent boundary layer-like pressure excitation. I. Derivations using strong model," J. Acoust. Soc. Am. **109**, 563–575 (2001).
[4] M. L. Rumerman, "Estimation of broadband acoustic power due to rib forces on a reinforced panel under turbulent boundary layer-like pressure excitation. II. Applicability and validation," J. Acoust. Soc. Am. **109**, 576–582 (2001).
[5] M. L. Rumerman, "Vibration and wave propagation in ribbed plates," J. Acoust. Soc. Am. **57**, 370–373 (1975).
[6] B. R. Mace, "Sound radiation from a plate reinforced by two sets of parallel stiffeners," J. Sound Vib. **71**, 435–441 (1980).

Effect of phase on discomfort caused by vertical whole-body vibration and shock—Experimental investigation

Yasunao Matsumoto[a] and Michael J. Griffin[b]
Human Factors Research Unit, Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, England

(Received 23 February 2001; accepted for publication 31 October 2001)

An experimental study has investigated the effect of ''phase'' on the subjective responses of human subjects exposed to vertical whole-body vibration and shock. The stimuli were formed from two frequency components: 3 and 9 Hz for continuous vibrations and 3 and 12 Hz for shocks. The two frequency components, each having $1.0\ \mathrm{ms}^{-2}$ peak acceleration, were combined to form various waveforms. The effects of the vibration magnitude on the discomfort caused by the input stimuli were also investigated with both the continuous vibrations and the shocks. Various objective measurements of acceleration and force at the seat surface, the effects of different frequency weightings and second and fourth power evaluations were compared with judgments of the discomfort of the stimuli. It was found that a 6% to 12% increase in magnitude produced a statistically significant increase in discomfort with both the continuous vibrations and the shocks. Judgments of discomfort caused by changes in vibration magnitude were highly correlated with all of the objective measurements used in the study. The effects on discomfort of the phase between components in the continuous vibrations were not statistically significant, as predicted using evaluation methods with a power of 2. However, small changes in discomfort were correlated with the vibration dose value (VDV) of the W_b frequency-weighted acceleration. The effect of phase between frequency components within the shocks was statistically significant, although no objective measurement method used in the study was correlated with the subjective judgments. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1446051]

PACS numbers: 43.40.Ng [PJR]

I. INTRODUCTION

Subjective responses, including feelings of discomfort, are among the principal reactions to whole-body vibration and shock. Studies of the quantitative relation between motions and subjective responses led to the frequency weightings defined in British Standard 6841 (1987) [e.g., Miwa (1967), Shoenberger and Harris (1971), Jones and Saunders (1972), Griffin *et al.* (1982a, b), Corbridge and Griffin (1986), reviewed in Griffin (1990)]. International Standard 2631-1 (1997) includes frequency weightings modified slightly from those in BS 6841, without explanation (Griffin, 1998). The frequency weightings are applied to the acceleration time history on the assumption that they represent the dependency of human response on vibration frequency. The mathematical definition of the frequency weightings presented in the standards gives the modulus and the phase of the weightings. Although the moduli of the weightings were based on the results of psychophysical studies, the phases were determined arbitrary so as to achieve the desired moduli using convenient filters. There has been little study of the effect of ''phase'' on subjective responses to whole-body vibration.

The subjective responses to simple vibratory stimuli have been explored analytically in a separate article (Matsumoto and Griffin, 2002). That study investigated continuous vibrations and shocks formed from two frequency components so as to investigate the effects of phase when using alternative methods of evaluating the stimuli. The study suggested suitable conditions for an experimental study. The frequency-weighted root-mean-square (rms) acceleration and the frequency-weighted vibration dose value (VDV) were assumed to represent the subjective reactions as defined in ISO 2631-1 (1997) and BS 6841 (1987). The predicted effects on subjective reactions of both the phase between two frequency components and the phase response of the body were investigated. The phase response of the body was assumed to be represented by frequency weightings defined in the standards (i.e., the W_k frequency weighting in ISO 2631-1 or the W_b frequency weighting in BS 6841) or by a filter representing the apparent mass of seated subjects (Matsumoto and Griffin, 1998), although there is no scientific evidence for these assumptions.

The analytical investigation showed that, for continuous vibration, the VDV was dependent on the phase between two frequency components in the input stimulus only when the ratio between the frequencies of the two components was 3, while the rms value was not altered by changing the phase between the components in the input stimulus. For the shocks used in the investigation, both the rms value and the VDV were dependent on the phase between components in the input stimulus, irrespective of the frequency ratio between the two components. An effect of the phase response of the body, represented by the assumed frequency weight-

[a]Now at Department of Civil and Environmental Engineering, Saitama University, 255 Shimo-Ohkubo, Saitama 338-8570, Japan.
[b]Author to whom correspondence should be addressed. Electronic mail: M.J.Griffin@soton.ac.uk

 © 2002 Acoustical Society of America

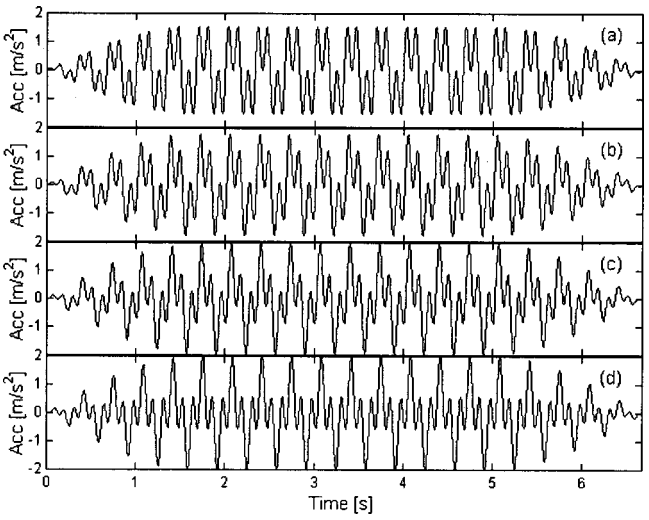


FIG. 1. "Stationary" vibrations used in the experiment. Phases between base and added vibrations are (a) 0 degrees (Stationary 1), (b) −60 degrees (Stationary 2), (c) −120 degrees (Stationary 3), and (d) −180 degrees (Stationary 4), respectively.

ings, was observed for most input stimuli and resulted in varying VDVs for the continuous vibrations when the frequency ratio was 3, and in varying rms values and VDVs for the shocks due to different phases within the input stimulus. The differences in the rms and VDV calculated for continuous vibrations and shocks were sufficient to be detected by human subjects, according to previous studies of difference thresholds for whole-body vibration by Morioka and Griffin (2000) and Mansfield and Griffin (2000).

An experimental study conducted by Clarke *et al.* (1965) indicates a possible dependence of subjective reactions on the phase between two frequency components. However, their experiment was conducted with very limited conditions based on "vibration representative of low altitude high speed flight buffeting" (i.e., at frequencies of 1 and 2 Hz at magnitudes of about 5 and 10 ms^{-2} rms with only two phases between two frequency components). There has been no experimental evidence showing an effect of phase within stimuli having short durations, for which the effects were predicted to be more significant than with continuous stimuli in the analytical study conducted by the current authors. The predictions of the analytical investigation were tested in an experimental study described in this article. In this study, the subjective responses (i.e., judgments of discomfort) produced by both continuous vibration and shocks were investigated experimentally to determine the maximum effect of phase on discomfort. The effects of the magnitude of the stimuli on discomfort were also investigated so as to assess the practical significance of any effects of phase.

II. METHOD

A. Input stimuli

If the alternative evaluation methods defined in ISO 2631-1 (1997) and BS 6841 (1987) (i.e., the frequency-weighted rms acceleration and the frequency-weighted VDV) represent human response to vibration and shock, the analytical investigation described in Matsumoto and Griffin (2002) implies that (i) human response to continuous vibration consisting of two sinusoidal components is generally unaffected by the phase between the two components; (ii) however, when the frequency ratio between two components is 3, human response to continuous vibration will be affected by the phase, according to the VDV, but not according to the rms; and (iii) the phase between two frequency components in shock type accelerations may influence human response. These hypotheses were investigated in an experiment in which two types of vertical oscillatory stimuli were used: continuous "statistically stationary" vibration and shock type stimuli.

The "statistically stationary" stimuli were based on two sinusoidal vibrations superimposed with different phases. The base vibration was 20 cycles of sinusoidal acceleration at 3 Hz having a peak amplitude of 1.0 ms^{-2}, with the first five cycles and the last five cycles tapered by a quarter of a sinusoidal function to avoid residual velocity and displacement. For the vibration added to the base vibration, a frequency ratio of 3 was used (i.e., the frequency of the second vibration was 9 Hz) so as to investigate the different predictions for the rms method and the VDV method. The amplitude of the second acceleration was 1.0 ms^{-2}, the same as that of the base acceleration. For the frequency of the base vibration at 3 Hz, the effects of phase of the W_k and the W_b frequency weightings on the VDV were greatest among the conditions investigated in the analytical investigation (Matsumoto and Griffin, 2002). Four different phases of the second vibration with respect to the base vibration were used: 0°, −60°, −120° and −180°. The waveforms of the four input accelerations used in the experiment, referred to in this article as "stationary 1" to "stationary 4," are presented in Fig. 1. The rms and the VDV of each input stimulus were calculated analytically without frequency weighting and with the W_k frequency weighting defined in ISO 2631-1 (1997),

TABLE I. Unweighted and weighted rms and VDV obtained analytically for "stationary" accelerations (stationary 1–4) used in the experiment. (Unwght: unweighted; W_k: W_k frequency weighting in ISO 2631-1 (1997); W_b: W_b frequency weighting in BS 6841 (1987); A.M.: apparent mass filter defined in Matsumoto and Griffin (2002).

Stimulus no.	Phase (degrees)	rms (ms^{-2})				VDV ($ms^{-1.75}$)			
		Unwght	W_k	W_b	A.M.	Unwght	W_k	W_b	A.M.
1	0	0.866	0.778	0.721	0.882	1.68	1.64	1.50	1.69
2	−60	0.866	0.778	0.721	0.882	1.74	1.66	1.50	1.83
3	−120	0.866	0.778	0.721	0.882	1.84	1.62	1.46	1.90
4	−180	0.866	0.778	0.721	0.882	1.88	1.56	1.42	1.85

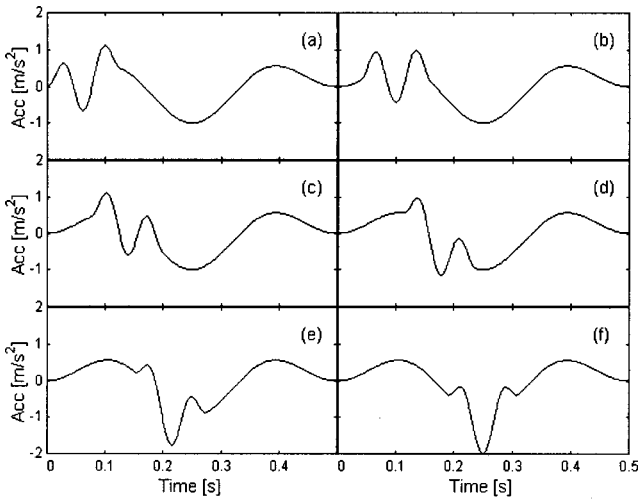


FIG. 2. Shock stimuli used in the experiment. Normalized delays of added shock with respect to base shock are (a) 0 (Shock 1), (b) 0.1 (Shock 2), (c) 0.2 (Shock 3), (d) 0.3 (Shock 4), (e) 0.4 (Shock 5), and (f) 0.5 (Shock 6), respectively.

with the W_b frequency weighting defined in BS 6841 (1987), and with the ''apparent mass filter'' defined in Matsumoto and Griffin (2002). These values are tabulated in Table I and were used as the target values for the input stimuli in the experiment. The rms values were calculated over the full duration of each exposure. It can be seen in the table that the frequency weightings changed the order of the VDV among the four stimuli, due to the phase responses of the weightings, assumed to represent the phase responses of human subjects.

The shock stimuli consisted of two acceleration components based on tapered one-and-half-sine waveforms, as shown in Fig. 2. The base shock was generated from a 3-Hz sinusoidal vibration at an amplitude of 1.0 ms^{-2}. The added shock was based on a 12-Hz sinusoidal acceleration (i.e., a frequency ratio of 4) at an amplitude of 1.0 ms^{-2}. For this combination of frequency, the effects of phase of the W_k and the W_b frequency weightings on the VDV were greatest among the conditions investigated in the analytical investigation (Matsumoto and Griffin, 2002). Six different phases, or time delays, were used to combine the two acceleration components: normalized delays of 0, 0.1, 0.2, 0.3, 0.4 and 0.5 [i.e., 0 corresponded to the two components beginning simultaneously, whereas 0.5 corresponded to the peaks of the two components occurring simultaneously; see Matsumoto and Griffin (2002)]. The waveforms of the six shocks used in the experiment, referred to in this article as ''shock 1'' to ''shock 6,'' are shown in Fig. 2. The unweighted and weighted rms and VDV for each shock stimulus obtained analytically are tabulated in Table II. The rms values were calculated over a duration of 0.5 s. It can be seen in the table that the phase response of the W_k and W_b frequency weightings caused the weighted VDV of shock 5 to be greater than the weighted VDV of shock 6.

B. Design, procedure and analysis

For each subject, a single session included four experiments. Two experiments were designed to investigate the effect of phase between two frequency components with either the stationary vibration or the shocks (i.e., the waveform effect). In the other two experiments, judgments of the relative discomfort caused by stimuli having a common waveform but different magnitudes were obtained for both the stationary vibrations and the shocks (i.e., the magnitude effect), so as to obtain some quantitative indication of the practical significance of the effect of waveform on discomfort in terms of the change in discomfort with a change in the magnitude of vibration or shock.

A modified Scheffe's method of paired comparison was employed to determine the relative discomfort caused by the different input stimuli (Noro *et al.*, 1973). When a group of subjects, $S_1, S_2, \ldots, S_N$, evaluate a group of stimuli, $A_1, A_2, \ldots, A_M$, a discomfort score, x_{qrs}, assigned to the pair of stimuli having the order of $[A_q, A_r]$ by the subject S_s was assumed to be the following structure:

$$x_{qrs} = (\alpha_q - \alpha_r) + (\alpha_{qs} - \alpha_{rs}) + \gamma_{qr} + (\delta + \delta_s) + \varepsilon_{qrs}. \quad (1)$$

Here, α_q is the average effect of the stimulus A_q, or the average discomfort score assigned to the stimulus A_q, with the assumption that $\Sigma_q \alpha_q = 0$; α_{qs} is the individual difference in the discomfort score, that is, the difference between the discomfort score assigned to the stimulus A_q by the subject S_s and the average discomfort score α_q, with the assumption that $\Sigma_q \alpha_{qs} = 0$ and $\Sigma_s \alpha_{qs} = 0$. γ_{qr} is the effect of the combination with the assumption that $\Sigma_r \gamma_{qr} = 0$ and $\gamma_{qr} = -\gamma_{rq}$. δ is the average effect of the order within a pair and δ_s is the individual difference in the order effect with the assumption that $\Sigma_s \delta_s = 0$. ε_{qrs} is the error that has an equal

TABLE II. Unweighted and weighted rms and VDV obtained analytically for shock accelerations (shocks 1–6) used in the experiment. (Nrmlz delay: normalized delay; Unwght: unweighted; W_k: W_k frequency weighting in ISO 2631-1 (1997); W_b: W_b frequency weighting in BS 6841 (1987); A.M.: apparent mass filter defined in Matsumoto and Griffin (2002).

Stimulus no.	Nrmlz delay	rms (ms^{-2})				VDV (ms$^{-1.75}$)			
		Unwght	W_k	W_b	A.M.	Unwght	W_k	W_b	A.M.
1	0	0.555	0.449	0.408	0.681	0.562	0.443	0.407	0.709
2	0.1	0.542	0.431	0.390	0.652	0.553	0.423	0.382	0.680
3	0.2	0.542	0.428	0.386	0.642	0.558	0.432	0.397	0.653
4	0.3	0.556	0.442	0.399	0.667	0.568	0.510	0.490	0.655
5	0.4	0.575	0.462	0.418	0.704	0.685	0.611	0.567	0.779
6	0.5	0.583	0.471	0.426	0.721	0.770	0.568	0.510	0.831

TABLE III. Scale used by subjects in paired comparison to judge relative discomfort of the second stimulus to the first stimulus.

	DISCOMFORT caused by stimulus
+3	2nd VERY MUCH MORE than 1st
+2	2nd DEFINITELY MORE than 1st
+1	2nd SLIGHTLY MORE than 1st
0	2nd THE SAME as 1st
−1	2nd SLIGHTLY LESS than 1st
−2	2nd DEFINITELY LESS than 1st
−3	2nd VERY MUCH LESS than 1st

average value for each pair of stimuli having an order and is assumed to have the normal distribution.

In the two experiments investigating the magnitude effect, the waveform of the stationary vibration shown in Fig. 1(d) and the waveform of the shock shown in Fig. 2(f) were used. Five magnitudes were employed, with each magnitude 6% greater than the magnitude at one level lower. These fractional changes in the stimulus magnitude were based on previous studies of difference thresholds by Mansfield and Griffin (2000) and Morioka and Griffin (2000); these investigations found that difference thresholds for seated subjects exposed to whole-body vertical vibration were consistent with Weber's law. The middle level of magnitude used for the two experiments investigating the effect of input magnitude was set to be comparable to the median magnitude of the input stimuli used to investigate the waveform effect. The target unweighted VDVs used in the two experiments were 1.60, 1.70, 1.80, 1.91, and 2.02 $ms^{-1.75}$ for the stationary vibrations and 0.600, 0.636, 0.674, 0.715, and 0.757 $ms^{-1.75}$ for the shocks.

In the two paired comparison experiments investigating the phase effect, there were four stationary vibration stimuli and six shock stimuli, giving 12 pairs of stationary vibration stimuli and 30 pairs of shock stimuli, including pairs with both orders of presentation. In the two studies of the magnitude effect, there were 20 pairs for both the stationary vibration and the shock stimuli. A total of 82 paired comparisons were judged by each subject in a single session lasting about 75 min, including a break for about 5 min. For each paired comparison, a series of two different stimuli, separated by an interval of 2 s, were presented. A subject judged the discomfort caused by the second stimulus relative to the discomfort caused by the first stimulus in terms of category numbers, or words, as presented in Table III. If the subject responded by a number, the experimenter confirmed the response by repeating the words corresponding to the category. In a preliminary experiment conducted prior to the experiment described in this article, subjects tended to be biased in favor of judging the second stimulus more uncomfortable than the first stimulus. A series of two pairs of stimuli was, therefore, presented before obtaining a judgment so as to reduce this order effect. The orders of the four experiments, and the order of presentation of pairs within each experiment, were randomized between subjects. Analysis of variance was conducted on the results of the paired comparison experiments for the subjects considered in three groups: all 20 subjects, a subgroup of ten male subjects and a subgroup of ten female subjects. Average discomfort scores of the input stimuli were also calculated. The analysis procedure is explained by Noro *et al.* (1973) in the Japanese language and also summarized by Ebe (1998) in the English language. A lower score means more discomfort if the analysis described by Noro *et al.* (1973) is applied to the results from the scale presented in Table III. The scores presented in this article were obtained by multiplying original scores by −1 so that a higher score corresponds to more discomfort. This does not distort any conclusions derived from the analysis presented in this article since only linear calculations are used.

C. Subjects and apparatus

Twenty subjects, ten males and ten females, participated in the experiment. The ages, heights and weights of the subjects are summarized in Table IV. The subjects were seated on a flat rigid seat secured to the platform of a 1-m stroke vertical electro-hydraulic shaker in the Institute of Sound and Vibration at the University of Southampton. The subjects were asked to sit in a comfortable upright posture with their hands on their laps and avoid any unnecessary movements. There was no backrest or footrest (i.e., the feet were allowed to hang freely) so that subjects would judge discomfort caused by input stimuli transmitted from the seat surface only. The subjects were asked to keep their eyes open and look straight ahead. The subjects were required to wear a loose fitting lap belt for safety reason. The shaker platform, $1.5\times0.9\ m^2$, was surrounded by stationary walls such that the subjects did not have a distant external view.

The acceleration of each stimulus was measured by a piezo-resistive accelerometer, Entran EGCSY-240D-10, attached to the shaker platform beneath the center of the seat surface. The force at the seat surface was measured by a force plate, Kistler 9281B, with mass cancellation applied to the measured values so as to compensate for the moving mass of the force plate. The accelerations and forces were used to investigate the relation between the subjective response obtained in the paired comparison method and the dynamic response of subjects in relation to the apparent mass filter mentioned above. The root-mean-square value and the accumulated fourth power (similar to the VDV of the accel-

TABLE IV. Characteristics of subjects.

	Male			Female		
	Mean	Minimum	Maximum	Mean	Minimum	Maximum
Age (yr)	27	18	34	25	20	34
Height (m)	1.78	1.68	1.88	1.66	1.58	1.82
Weight (kg)	74	61	87	58	49	76

TABLE V. Summary of analysis of variance on paired comparison experiments with 20 subjects obtained by modified Scheffe's method.

Effect	Sums of squares	Degrees of freedom	Variance	F-ratio	Significance
Stationary, magnitude					
Primary	231	4	57.9	86.1	$p<0.01$
Combination	2.02	6	0.337	0.500	not significant
Order	69.2	10	6.92	10.3	$p<0.01$
Error	255	380	0.672		
Stationary, phase					
Primary	4.21	3	1.40	1.78	not significant
Combination	2.19	3	0.730	0.925	not significant
Order	39.7	6	6.62	8.39	$p<0.01$
Error	179.9	228	0.789		
Shock, magnitude					
Primary	118	4	29.4	47.1	$p<0.01$
Combination	2.11	6	0.352	0.563	not significant
Order	13.0	10	1.30	2.08	$p<0.05$
Error	237	380	0.624		
Shock, phase					
Primary	126	5	25.3	29.9	$p<0.01$
Combination	5.49	10	0.55	0.650	not significant
Order	22.5	15	1.50	1.77	$p<0.05$
Error	482	570	0.845		

eration) were calculated for the measured force, without frequency weighting.

III. RESULTS

A summary of the analyses of variance conducted on the results of the four experiments with paired comparisons using 20 subjects are provided in Table V. The primary effect (i.e., the differences between the input stimuli) was statistically significant in all experiments ($p<0.01$), except for the effect of phase on subjective responses to the statistically stationary vibration: judgments of discomfort caused by the four vibration stimuli shown in Fig. 1 were not significantly different from each other. However, for ten male subjects, there was a statistically significant difference between judgments of discomfort caused by the four vibrations shown in Fig. 1 ($p<0.05$), although the data are not presented here. There were no significant effects caused by the combination of stimuli, but the order of presentation of stimuli was significant for both the stationary vibration ($p<0.01$) and for the shocks ($p<0.05$). The subjects tended to judge the second stimulus to cause more discomfort than the first stimulus, as in the preliminary study, even though a series of two pairs of stimuli was presented so as to minimize this order effect.

The averages of the subjective judgments for each input stimulus are shown in Fig. 3, where a higher score corresponds to more discomfort. The statistical significance of differences between scores shown in Fig. 3 is presented for each experiment in Table VI. The statistical significance was obtained from the differences in scores required for statistical significance levels of 0.01 and 0.05 [i.e., by yardsticks, see Noro *et al.* (1973)], which are also shown in Table VI. It can be observed that the subjects gave significantly different judgments for pairs of stimuli having a common waveform but a magnitude difference of 6% [$p<0.01$, Tables VI(a) and (c), and Figs. 3(a) and (c)]. This is observed more clearly when the subjects were exposed to the stationary vibrations than when exposed to the shocks.

For the stationary vibrations with different phases between two frequency components, there was no statistically significant difference in discomfort caused by the four stimuli [Table VI(b) and Fig. 3(b)]. This is predicted by the rms values of the stimuli as shown in Table I. For the shock stimuli, the phase between two frequency components influenced subjective judgments: shocks 1 and 2 (i.e., shocks with normalized delays of 0 and 0.1) caused more discomfort than shocks 3–6 (i.e., shocks with normalized delays of 0.2–0.5) [$p<0.01$, Table VI(d) and Fig. 3(d)]. Shock 6 caused less discomfort than shocks 1 and 2, although both the rms and VDV of shock 6 were greater than those of shocks 1 and 2, as shown in Table II.

The average scores representing discomfort caused by the stimuli were normalized by dividing the scores by the corresponding yardstick for a significance level of 0.01 for each experimental result. A difference in the normalized average scores of unity was taken to indicate that the difference between two scores was statistically significant at a level of 0.01. Results from different experiments have been compared using the normalized scores. Figure 4 compares the normalized average discomfort scores for stimuli having different waveforms with those for stimuli of the same waveform but with different magnitudes. For the stationary vibrations, the differences in discomfort caused by the different waveforms employed here (i.e., with different phases between two components) were less than the differences in discomfort caused by two input stimuli having a magnitude difference of 6% [Fig. 4(a)]. For the shock type inputs, the differences between the discomfort caused by shock 1 and

TABLE VI. Statistical significance of differences between scores for each experiment. Yardsticks for significance levels of 0.01 and 0.05 (i.e., $Y_{0.01}$ and $Y_{0.05}$) are also shown. (**: $p<0.01$; *: $p>0.05$).

(a) Stationary, magnitude						
$Y_{0.01}$:0.273	1	2	3	4	5	
$Y_{0.05}$:0.227	−0.690	−0.335	0.000	0.365	0.660	
1	−0.690		−0.355**	−0.690**	−1.055**	−1.350**
2	−0.335			−0.335**	−0.700**	−0.995**
3	0.000				−0.365**	−0.660**
4	0.365					−0.295**
5	0.660					
(b) Stationary, phase						
$Y_{0.01}$:0.316	1	2	3	4		
$Y_{0.05}$:0.259	0.075	0.069	−0.019	−0.125		
1	0.075		0.006	0.094	0.200	
2	0.069			0.088	0.194	
3	−0.019				0.106	
4	−0.125					
(c) Shock, magnitude						
$Y_{0.01}$:0.263	1	2	3	4	5	
$Y_{0.05}$:0.219	−0.500	−0.150	−0.050	0.165	0.535	
1	−0.500		−0.350**	−0.450**	−0.665**	−1.035**
2	−0.150			−0.100	−0.315**	−0.685**
3	−0.050				−0.215	−0.585**
4	0.165					−0.370**
5	0.535					
(d) Shock, phase							
$Y_{0.01}$:0.289	1	2	3	4	5	6
$Y_{0.05}$:0.243	0.521	0.271	−0.092	−0.225	−0.167	−0.308
1	0.521		0.250*	0.613**	0.746**	0.688**	0.829**
2	0.271			0.363**	0.496**	0.438**	0.579**
3	−0.092				0.133	0.075	0.217
4	−0.225					−0.058	0.083
5	−0.167						0.142
6	−0.308						

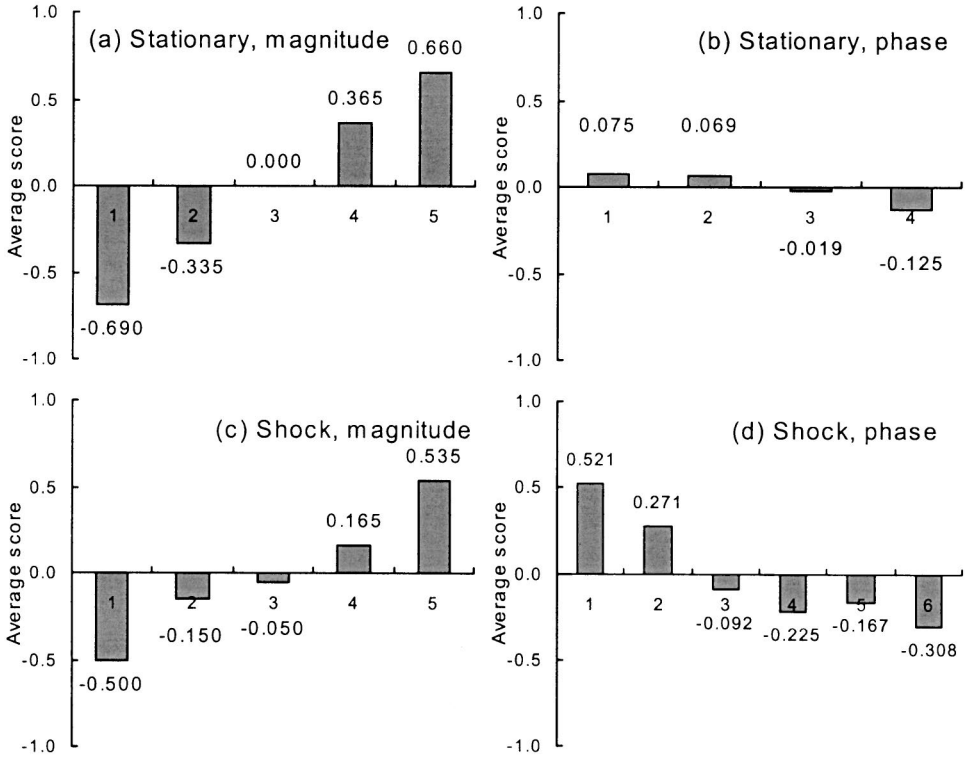


FIG. 3. Average score of each stimulus in the four experiments. A higher score means more discomfort. (a) Effect of magnitude of stationary vibration, (b) effect of phase of stationary vibration, (c) effect of magnitude of shock, and (d) effect of phase of shock. In (a) and (c), a smaller number in the label of the horizontal axis corresponds to the stimulus at less physical magnitude. In (b) and (d), stimulus numbers shown in Tables I and II are used in the label of the horizontal axis.

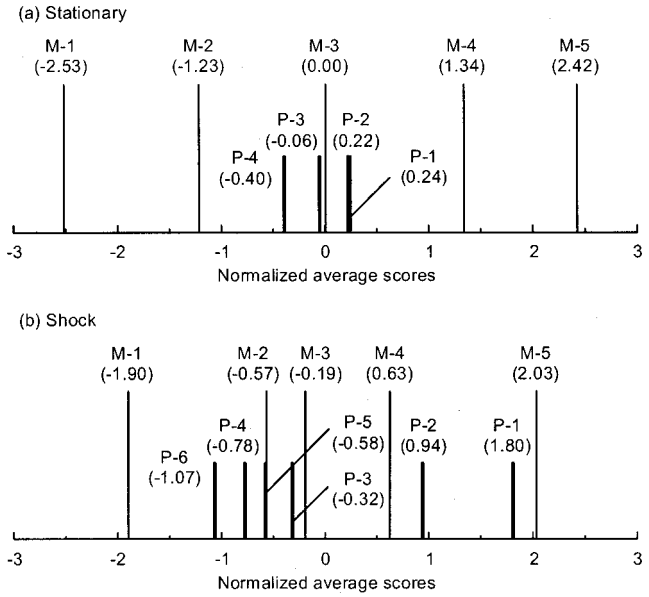


FIG. 4. Normalized average scores for the stationary vibrations and shocks. Comparison between magnitude effect and phase effect (M: magnitude; P: phase).

the discomfort caused by shocks 3–6 correspond to the differences in discomfort caused by two shocks with a magnitude difference of about 12% to 19% [Fig. 4(b)].

Several objective measurements mentioned in Sec. II are compared with the average discomfort scores in Fig. 5 (for the stationary vibrations) and in Fig. 6 (for the shocks). The objective measurements presented in Figs. 5 and 6 are the rms and the VDV of unweighted acceleration, the rms and the VDV of the W_k frequency-weighted acceleration, the rms and the VDV of the W_b frequency-weighted acceleration, the rms and the VDV of the acceleration frequency-weighted by

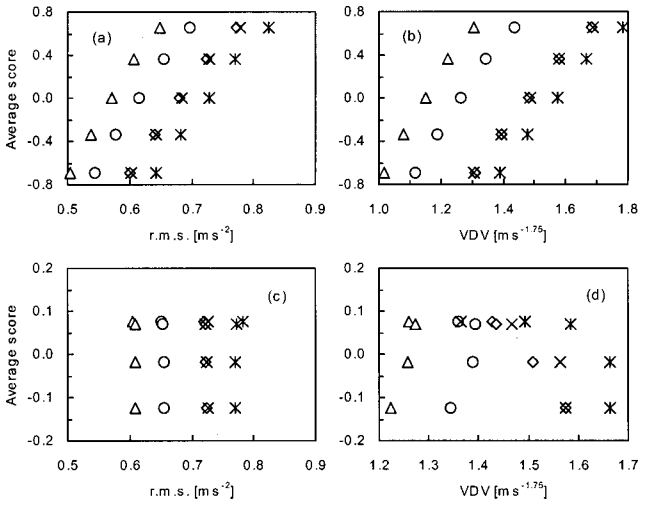


FIG. 5. Comparison between average discomfort scores and objective measurements for stationary vibrations. (a), (b) magnitude effect; (c), (d) phase effect. Keys in (a) and (c): ◇, unweighted rms acceleration; ○, W_k weighted rms acceleration; △, W_b weighted rms acceleration; ×, apparent mass filter weighted rms acceleration; *, rms of normalized force. Keys in (b) and (d), ◇, VDV of unweighted acceleration; ○, VDV of W_k weighted acceleration; △, VDV of W_b weighted acceleration; ×, VDV of apparent mass filter weighted acceleration; *, VDV of normalized force. (Normalized force was obtained by dividing the measured force by subject mass.)

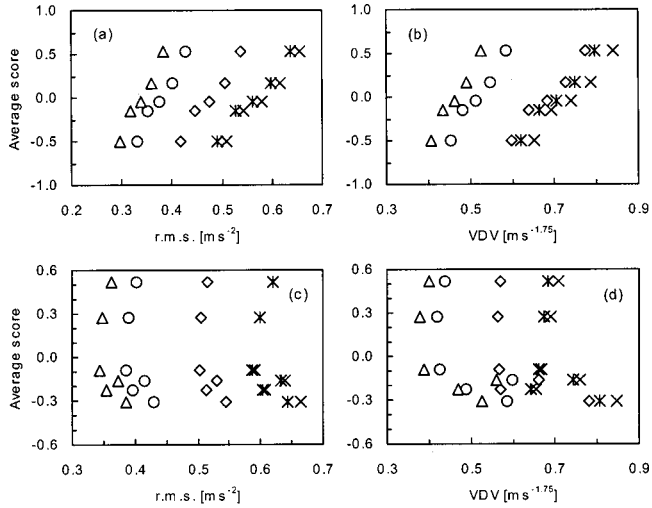


FIG. 6. Comparison between average discomfort scores and objective measurements for shocks. (a), (b) magnitude effect; (c), (d) phase effect. Keys are the same as those in Fig. 5.

the apparent mass filter, and the rms and the accumulated fourth power (referred to as VDV) of the force measured at the seat surface. For the continuous vibrations, the actual acceleration levels were some 15% less than the target values. However, the differences in magnitudes were systematic and do not affect the findings. The force measurements were divided by the mass of each subject before obtaining the rms and the VDV of the force, so that the units of the resulting values were the same as those for the acceleration. It can be observed in Figs. 5(a) and (b) and 6(a) and (b) that all objective measurements used here might be considered to represent the subjective judgment reasonably when the magnitude of an input stimuli having a common waveform was changed, for both the stationary vibrations and the shocks: the discomfort caused by an input stimulus increased with value for all objective measurements. However, such simple relations were not found when different waveforms were used, either with the stationary vibrations or the shocks [Figs. 5(c) and (d) and 6(c) and (d)]: an increased discomfort score did not necessarily correspond to higher objective measures.

The associations between subjective responses and the objective measures shown in Figs. 5 and 6 were investigated by calculating Kendall's τ_b correlation coefficients. A one-tailed correlation was used because it was predicted that discomfort would increase with increased vibration magnitude. For the experiments with stationary vibrations and shocks having a common waveform at different magnitudes, the Kendall's τ_b correlation coefficients between the average subjective scores and all ten objective measures were 1.0 at $p<0.01$ (i.e., the rank order of subjective judgments was correctly predicted by all objective measures). For the stationary vibrations with different waveforms, the Kendall's τ_b correlation coefficient was positive and marginally significant at $p<0.1$ for correlations between the average discomfort score and two objective measures: the W_b weighted VDV ($\tau_b=0.667$, $p=0.087$) and the rms obtained by normalized force at the seat surface ($\tau_b=0.667$, $p=0.087$). For the shocks with different waveforms, there were no positive

and statistically significant correlations between the average discomfort scores for any of the objective measures.

IV. DISCUSSION

The judgments of discomfort caused by stimuli having a common waveform were significantly increased by 6% increases in the magnitudes of the stimuli ($p<0.01$). This was found with each 6% increase for the stationary vibrations, and with only two of the four 6% increases with the shock-type stimuli [see Tables VI(a) and (c), and Fig. 4]. This implies that the group of subjects perceived a 6% difference between stimuli when the waveforms were the same. It also implies that the subjects more easily detected a difference in the magnitude of the continuous motions than the short transient motions.

For the statistically stationary stimuli having the waveforms shown in Fig. 1(d), the average discomfort scores for different magnitudes were well correlated with all objective measures, including the objective measures defined in the standards and the measures based on the force at the seat surface. All objective measures employed here might, therefore, be considered to represent the relative discomfort caused by such vibrations having a common waveform but different magnitudes.

For the stationary vibrations having different waveforms, arising from different phase relationships between two fixed components, the differences in discomfort were not statistically significant. This is as predicted when using the rms accelerations of the waveforms, which were the same (see Table I). However, none of the rms measures showed a statistically significant correlation with the average discomfort scores [Fig. 5(c)]. The VDVs of stationary vibrations having different waveforms, with and without frequency weightings, show a difference of between 5.6% and 12% (see Table I). If differences in VDV greater than 6% produce significant differences in discomfort, as observed for the stationary vibrations of different magnitude having a common waveform, differences in the unweighted VDV of about 12% would be expected to be reflected in subject judgments. The absence of differences in the discomfort scores seems inconsistent with these changes in the unweighted VDVs. However, differences of about 6% in the W_k weighted VDV and the W_b weighted VDV might have been too small to cause statistically significant differences in the discomfort scores. It was interesting that although the differences between the four stimuli were not statistically significant, the W_k frequency-weighted VDV and the W_b frequency-weighted VDV were lowest for ''stationary 4,'' which was judged by the subjects to cause the least discomfort [see Fig. 5(d)]. The W_b frequency-weighted VDV was almost the same for ''stationary 1'' and ''stationary 2,'' to which similar average scores were given by the subjects [Figs. 3(b) and 5(d)]. A statistically significant correlation was found between the average discomfort score and the W_b frequency-weighted VDV for the stationary vibrations with different waveforms [$p<0.1$, Fig. 5(d)]. Therefore, although the unweighted VDV calculated from the input acceleration did not represent the discomfort caused by the stationary vibrations with different waveforms, the W_b frequency-weighted VDV did represent the feeling of subjects.

Clarke *et al.* (1965) showed that, for continuous vibration consisting of a base frequency component at frequencies of 1 or 2 Hz and its third harmonic superimposed with a phase of 0 or 180 degrees, the subjects reported greater severity with the phase between two components at 180 degrees (i.e., with a greater peak acceleration). This finding is inconsistent with the results in this article: the discomfort score for the continuous vibration with the phase at 0 degrees (''stationary 1'') tended to be greater than the discomfort score for the vibration with the phase at 180 degrees ''stationary 4'' [Fig. 3(b)], although the difference was not statistically significant. Clarke *et al.* (1965) used input stimuli with five or ten times greater magnitudes (i.e., about 5 and 10 ms^{-2} rms) than those in this study and the frequencies of their input stimuli were far lower (i.e., 1 and 2 Hz). There are a variety of possible reasons for their results differing from those of the present study with lower vibration magnitudes, higher vibration frequencies and different seating and restraint conditions.

The average discomfort scores obtained with different magnitudes of the shock were highly correlated with all objective measures [Figs. 6(a) and (b)]. This is not inconsistent with previous findings that the discomfort caused by impulsive whole-body vibration is more highly correlated with measures based on a power of 4 (e.g., VDV) than a power of 2 (e.g., rms) (Griffin and Whitham, 1980; Howarth and Griffin, 1991). These previous studies explored stimuli having different durations as well as different magnitudes whereas the stimuli used in this study had variations in magnitude but the same duration.

For the shocks having different waveforms (produced by different phases between the two frequency components), none of the objective measures employed were correlated with the average discomfort scores [Figs. 6(c) and (d)]. The subjects judged there was more discomfort caused by shocks 1 and 2 [Figs. 2(a) and (b)] than by shocks 3–6 [Figs. 2(c)–(f)] as shown in Table VI(d) and Fig. 3(d), even though shocks 1 and 2 had relatively low rms and VDV compared to the others (Table II). Some subjects mentioned they experienced more discomfort because they felt two consecutive bumps with shocks 1 and 2, compared to feeling only one shock for the other stimuli. The number of shocks perceived by the subject may have been a significant factor when the discomfort caused by the shocks used in this experiment was judged.

Whether a subject felt two shocks or one will have depended on the time interval between the occurrences of the peak accelerations of two frequency components. It will also have depended on the ''phase response'' of the subject, which might depend on various characteristics of the human response to such mechanical stimuli. On the assumption that this ''phase response'' exists, suppose the peak accelerations of two frequency components [$f=f_1, f_2$(Hz)($f_1<f_2$)] occur at $t=t_1$(s) and $t=t_2$(s) ($t_1>t_2$ in the experiment). The time lag between the peak acceleration that the subject perceives, t_1, could be expressed as

$$t_l = t_1 - t_2 - t_p, \tag{2}$$

$$t_p = (\phi_1/f_1 - \phi_2/f_2)/(2\pi). \tag{3}$$

Here, the phase response of a subject at f_1(Hz) and f_2(Hz) are represented as ϕ_1 and ϕ_2 (in radians), respectively. t_p in Eqs. (2) and (3) corresponds to the effect of the phase response of a subject on the time lag t_l. For example, assuming the phase of the W_k frequency weighting in ISO 2631-1 (1997) represents the phase response of the human subjects and substituting $[f_1 \quad f_2] = [3\ 12]$, as used in the experiment, and the corresponding phases of the W_k frequency weighting $[\phi_1 \quad \phi_2] = [0.6058 \ -0.7787]$ into Eq. (2), then $t_p = 0.0425$(s) will be obtained. This t_p corresponds to more than half of the period of the 12-Hz frequency component, 0.0833 s. The value of $t_p = 0.0425$(s) in Eq. (2) means that the phase response of a subject reduces the time lag between two frequency components by 0.0425 s, which may make it more difficult for a subject to perceive two shock components separately. This effect might be observed in the experimental results here for shock 3 shown in Fig. 2(c) for which most subjects might have perceived only one shock although two separate acceleration peaks can be clearly seen in the acceleration waveform. This may illustrate the importance of the phase response of the human subjects in the evaluation of shocks as used in this study. The above illustration assumes that the relative phase of the W_k frequency weighting between 3 and 12 Hz represents the relative phase response of the human subjects, even though it is recognized that the phase of W_k was selected arbitrarily without reference to human responses.

V. CONCLUSIONS

For input stimuli having a common waveform, increases in the discomfort of seated subjects have been observed with a 6% increases in the magnitude of continuous vibrations and with 6% to 12% increases in the magnitude of shocks. These increases were correlated with all objective measures involving acceleration or force at the seat surface investigated in the study. For continuous vibrations, although changes in the phase between two frequency components did not cause statistically significant overall changes in discomfort (as predicted by the rms measures used in this study), the VDV of the W_b frequency-weighted acceleration showed a statistically significant correlation with ratings of discomfort. For shocks, the phase between two frequency components in the shocks affected judgments of discomfort, although there was no correlation between the discomfort caused by the shocks and the evaluations obtained by any of the measures used in the study, including the evaluation methods defined in current standards. Subjects thought that the number of shocks they perceived might be one of the factors affecting their reactions to the shocks. It is shown that the phase response of subjects could influence the perception of the number of shocks.

ACKNOWLEDGMENTS

This research was supported by the Korea Research Institute of Standards and Science, Taejon, Korea. The support of Dr. Wan-Sup Cheung is appreciated.

British Standards Institution (**1987**). ''Guide to measurement and evaluation of human exposure to whole-body mechanical vibration and repeated shock,'' BS 6841, London.

Clarke, N. P., Mohr, G. C., Brinkley, J. W., Martin, P. J., Wooding, H. C., Henzel, J. H., and von Gierke, H. E. (**1965**). ''Evaluation of peak vs RMS acceleration in periodic low frequency vibration exposures,'' Aerosp. Med. **36**, 1083–1089.

Corbridge, C., and Griffin, M. J. (**1986**). ''Vibration and comfort: vertical and lateral motion in the range 0.5 to 5.0 Hz,'' Ergonomics **29**, 249–272.

Ebe, K. (**1998**). ''Predicting overall seat discomfort from their static and dynamic characteristics,'' Ph.D. thesis, University of Southampton, England.

Griffin, M. J. (**1990**). *Handbook of Human Vibration* (Academic, London), Chap. 3, pp. 53–71.

Griffin, M. J. (**1998**). ''A comparison of standardized methods for predicting the hazards of whole-body vibration and repeated shocks,'' J. Sound Vib. **215**, 883–914.

Griffin, M. J., and Whitham, E. M. (**1980**). ''Discomfort produced by impulsive whole-body vibration,'' J. Acoust. Soc. Am. **68**, 1277–1284.

Griffin, M. J., Parsons, K. C., and Whitham, E. M. (**1982a**). ''Vibration and comfort: IV. Application of experimental results,'' Ergonomics **25**, 721–739.

Griffin, M. J., Whitham, E. M., and Parsons, K. C. (**1982b**). ''Vibration and comfort: I. Translational seat vibration,'' Ergonomics **25**, 603–630.

Howarth, H. V. C., and Griffin, M. J. (**1991**). ''Subjective reaction to vertical mechanical shocks of various waveforms,'' J. Sound Vib. **147**, 395–408.

International Organization for Standardization (**1997**). ''Mechanical vibration and shock—evaluation of human exposure to whole-body vibration—Part 1: General requirements,'' ISO 2631-1, Geneva.

Jones, A. J., and Saunder, D. J. (**1972**). ''Equal comfort contours for whole-body vertical pulsed sinusoidal vibration,'' J. Sound Vib. **23**, 1–14.

Mansfield, N. J., and Griffin, M. J. (**2000**). ''Difference thresholds for automotive seat vibration,'' Appl. Ergon **31**, 255–261.

Matsumoto, Y., and Griffin, M. J. (**1998**). ''Movement of the upper-body of seated subjects exposed to vertical whole-body vibration at the principal resonance frequency,'' J. Sound Vib. **215**, 743–762.

Matsumoto, Y., and Griffin, M. J. (**2002**). ''Effect of phase on discomfort caused by vertical whole-body vibration and shock–analytical investigation,'' J. Sound Vib. (awaiting publication).

Miwa, T. (**1967**). ''Evaluation methods for vibration effect. 1. Measurements of threshold and equal sensation contours of whole-body for vertical and horizontal vibrations,'' Ind. Health **5**, 183–205.

Morioka, M., and Griffin, M. J. (**2000**). ''Difference thresholds for intensity perception of whole-body vertical vibration: Effect of frequency and magnitude,'' J. Acoust. Soc. Am. **107**, 620–624.

Noro, K., Ura, S., and Sato, S. (**1973**). ''Paired comparison,'' in *Sensory Evaluation Handbook*, Union of Japanese Scientists and Engineers (JUSE, Tokyo), Chap. 10, pp. 349–393 (in Japanese).

Shoenberger, R., and Harris, C. S. (**1971**). ''Psychophysical assessment of whole-body vibration,'' Hum. Factors **13**, 41–50.

Backscattering of transients by tilted truncated cylindrical shells: Time-frequency identification of ray contributions from measurements

Scot F. Morse[a] and Philip L. Marston
Department of Physics, Washington State University, Pullman, Washington 99164-2814

(Received 10 September 2001; accepted for publication 11 December 2001)

Impulse backscattering measurements by a thick-walled finite cylindrical shell are examined in the time-frequency domain to identify and characterize individual ray contributions from generalized Lamb waves excited on the shell. Previous experiments and analysis in the frequency-aspect angle domain [S. F. Morse *et al.*, J. Acoust. Soc. Am. **103**, 785–794 (1998)] indicate that large backscattering enhancements occur in the midfrequency region for the shell tilted at large angles. Presently this experimental data is examined in the time-frequency domain for selected angles of incidence. Individual ray contributions are evident and their evolution over aspect angle is discussed. The most prominent contribution is due to the meridional ray of the a_0 leaky Lamb wave. This feature distinctly highlights the truncation of the shell and is found over a range of aspect angles spanning 200° for the frequencies examined. Also observed are periodic features corresponding to end-reflected helical waves of the a_{0-}. These scattering features are significantly different from those reported for thin-walled finite cylinders at low frequencies. The present results may be useful for target identification and localization and as a comparison tool for high-frequency computational scattering models. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1448518]

PACS numbers: 43.40.Rj, 43.30.Gv, 43.20.Fn [ANN]

I. INTRODUCTION

The acoustic backscattering from simple finite cylindrical shells has been shown to be enhanced over regions of aspect-angle frequency space.[1–7] The enhancements reported[4,5] occurred near and above the coincidence frequency for thick, hollow bluntly-truncated steel cylindrical shells submerged in water. The mechanism responsible for the enhancements is the excitation and subsequent radiation of a meridional ray of the supersonic a_0 (quasi-flexural mode) leaky Lamb wave. The meridional ray propagates in the axial direction along the meridian of the cylinder (see Fig. 1, paths EGHBI and FHBI) and is launched when the aspect angle of the cylinder is near the leaky wave coupling angle, given by $\gamma \approx \theta_l \equiv \sin^{-1}(c/c_l)$, which is dependent on the phase velocity c_l of the leaky wave. The backscattered levels are larger than one may expect from the scattering by a curved shell because the wavefront radiated by the leaky wave has a vanishing Gaussian curvature even at oblique incidences. A ray model has been shown to describe this enhancement for infinite cylindrical shells[5,8] and for Rayleigh waves on infinite and finite solid cylinders.[3,9]

In this paper an analysis of the scattering response for the thick-walled cylinder in Ref. 4 is made in the time and time-frequency domain for the full range of scattering aspect angles. The time-frequency analysis is used to identify and characterize the a_0 meridional enhancement with respect to other features which have already been examined by other authors. A brief analysis is made of the contribution to the scattering by helical surface waves identified as the subsonic a_{0-} (Scholte–Stoneley type).

Time and joint time-frequency domain analysis can be useful in understanding the relevant mechanisms which contribute to the overall scattering. This is especially true for the case of a finite cylindrical shell, where the nature of the scattering is highly dependent on the tilt angle of the cylinder, on the cylinder end conditions and the frequency. For a thick-walled finite cylindrical shell Bao[10] identified the scattering due to various guided waves in the time domain and studied their progression as the cylinder tilt angle was increased from broadside aspect to 24 degrees tilt. More recently Touraine *et al.*[11] have examined the time domain response for extra thin shells having hemispherical endcaps. Haumesser *et al.*[12] studied the response of the lowest order guided waves for a bluntly truncated thin-walled shell over a broad range of frequencies and for various cylinder lengths. In both Ref. 11 and Ref. 12 when the scattering response due to the endcaps is neglected, the examination is limited to the range of aspect angles within 30 degrees of broadside incidence. Within this range of angles and frequencies the surface waves excited on the shell correspond to helical waves of the lowest order quasicompressional, s_0, quasishear, T_0, and Scholte–Stoneley type, a_{0-} waves (although the a_{0-} is not present for the extra thin shell in Ref. 11). For the aspect-angle frequency domain the scattering response of the s_0 and T_0 is discussed by Rumerman.[13] In the time domain Refs. 11 and 12 find good agreement between experiment and predicted echo arrival times for the s_0 and T_0 helical waves.

In this paper the scattering due to the a_0 and a_{0-} waves are investigated in the joint time-frequency domain. This analysis enables the identification of individual ray compo-

[a]Present address: Division of Computer Science, Western Oregon University, 345 North Monmouth Avenue, Monmouth, OR 97361.

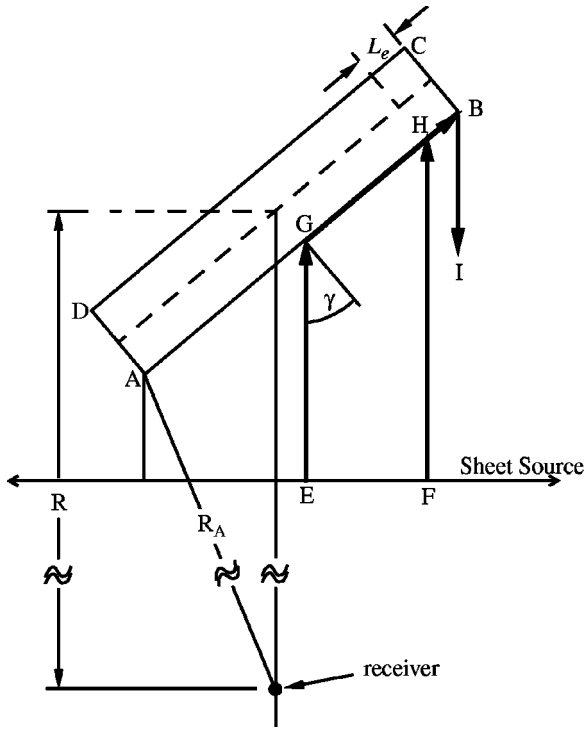


FIG. 1. Diagram of experimental setup showing meridional ray path and notation for echo return expressions.

nents and allows for the separation of the scattering response for different wave types. Time-frequency methods have proven useful in understanding the complex scattering response from submerged elastic targets.[14,15] Additional analysis of the time-frequency response has been used to determine the velocity dispersion of various Lamb waves.[16,17] The time-frequency response can also be valuable in target classification algorithms.[18,19] It may also be noted that the present identification of midfrequency scattering mechanisms for the noncanonical finite cylindrical shell may be of use in validating high frequency computational methods such as is found in Ref. 20.

II. EXPERIMENT

Impulse backscattering measurements were carried out in a large water tank with the configuration detailed in Fig. 1. A detailed description of the experiment may be found in Ref. 4, with two modifications noted below. The shell is constructed of stainless steel (type 304) with the following dimensions, and has flat plexiglas endcaps: $b/a=0.924$, $a=19.05\,\mathrm{mm}$, $L/a=12.0$, where a and b are the outer and inner radii, respectively, and L is the cylinder length. In the present experiment the source, a large planar sheet of PVDF, is driven with a short unipolar electrical pulse. Previously a step voltage was used. Appendix A gives a brief discussion of the source in the present configuration. The center of the cylinder is positioned 1.55 m from the receiver.

The unprocessed time domain scattering response is shown in Fig. 2 as a function of the angle of the incident sound. Broadside incidence corresponds to 0 degrees. The arrows at left indicate the tilt angles selected for time-frequency analysis in Fig. 3. The overlaid solid black curves labeled t_A-t_D indicate the expected return times for direct propagation from the source to the four furthest extents of the cylinder in its plane of rotation (henceforth referred to as corners) and back to the receiver. They are calculated with respect to the specular reflection measured at broadside incidence; expressions are given in Appendix B. These curves give a convenient reference for comparing actual scattering returns and also highlight the simple diffraction by the cylinder and earliest portions of scattering by the endcaps. The dashed magenta curves are the predicted echo arrival times for helical waves of the a_{0-} (see below and Appendix B for a brief discussion of this response).

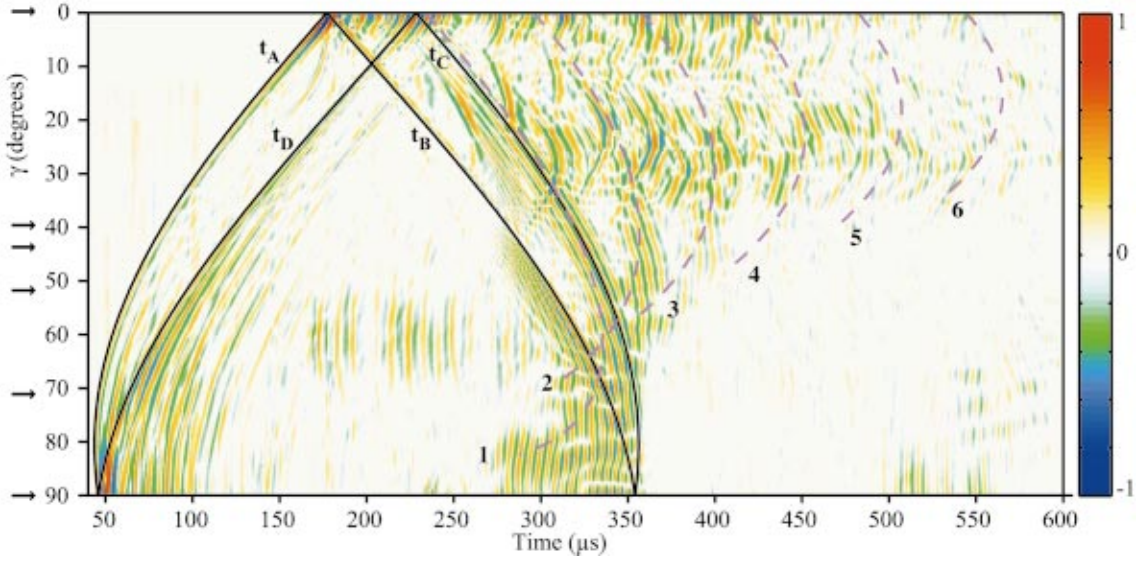


FIG. 2. Measured impulse response time series after clutter subtraction. Broadside incidence corresponds to $\gamma=0°$. The linear color scale represents a logarithmic scale where red is positive high voltage [1→(+)0 dB], blue is negative high voltage [−1→(−)0 dB], and white corresponds to low levels (−50 dB). The overall scale is relative to the peak signal observed. The solid black lines indicate the expected arrival times at the receiver of pulses travelling from the source to each of the four ''corners'' of the cylindrical shell directly through the water. Dashed magenta lines are the calculated arrival times for the first six helical waves of the subsonic a_{0-}. Arrows at left indicate the angles selected for time-frequency analysis.

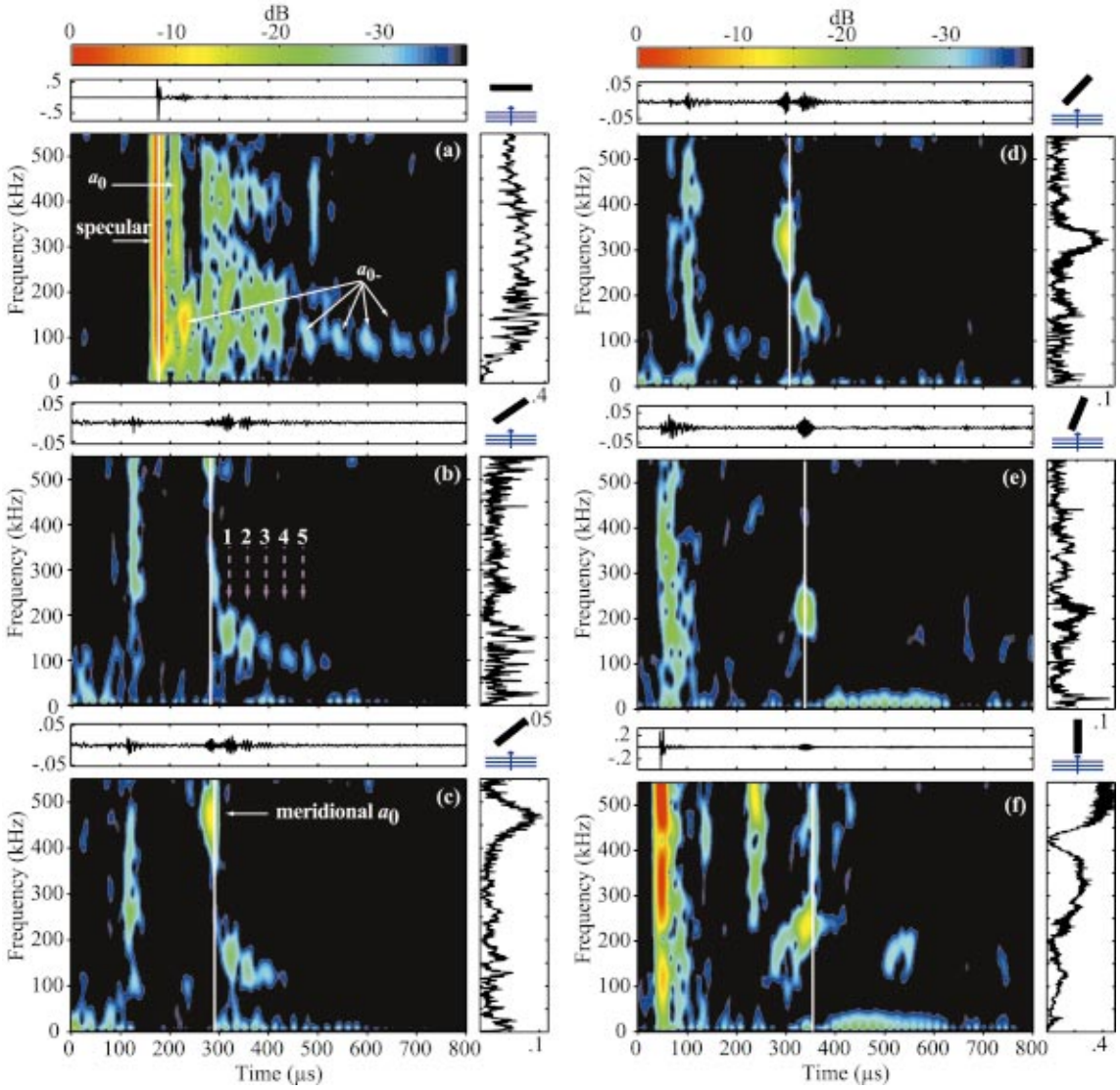


FIG. 3. Magnitude of the short time Fourier transform of the measured scattering response at selected angles. (a) $\gamma=0°$, (b) $\gamma=40°$, (c) $\gamma=44°$; (d) $\gamma=52°$, (e) $\gamma=71.2°$, (f) $\gamma=90°$. For each spectrogram the raw time series data (top) and its associated un-windowed FFT (right) are shown as well. The vertical white line corresponds to the time for a pulse travelling directly through the water from the source to the nearest rear corner of the cylinder (labeled **B** and t_B in Figs. 1 and 2, respectively) to reach the receiver. The magenta lines in (b) correspond to the a_{0-} helical wave echos from Fig. 2.

The scattering response within 35 degrees of broadside incidence is consistent with scattering by a cylindrical shell.[10,13,21–24] The complicated overlapping echoes in this region primarily correspond to the response of helical quasi-compressional (s_0) and quasi-shear (T_0) surface waves which are reflected from the truncation of the cylinder most distant from the source. The response for tilt angles greater than 35 degrees, and hence greater than both the compressional and shear wave cutoff angles ($\gamma\approx16.5$ and 28.2 degrees, respectively) is the focus of this paper. It is primarily in this region that we apply simple time-frequency representations of the response at selected angles in order to identify and characterize the observed echoes.

III. TIME-FREQUENCY ANALYSIS

The scattering response in Fig. 2 is composed of overlapping echoes corresponding to different physical mechanisms. At each tilt angle a time-frequency representation was made of the echo response. Six particular angles were selected which highlight the scattering mechanisms under investigation; these are shown in Figs. 3(a)–(f). For the present analysis the simplest time-frequency representation, the short time Fourier transform (STFT) or spectrogram,[25] was found sufficient to separate the features of interest. Greater precision could be obtained for a parameter of interest (either time or frequency) by modifying the length or shape of the window function in the calculation of the STFT or an altogether different time-frequency representation could be employed.[25,17]

For each STFT in Fig. 3 the original time record (top) and the full un-windowed Fourier transform (right) are displayed. The STFT displays the energy density spectrum of the signal on a dB scale where 0 dB corresponds to the peak in the energy density spectrum at broadside incidence. Each STFT in Fig. 3 has the same normalization. A 40 μs Blackman window was employed in the calculation. The frequency response of the source and receiver system has been removed by dividing the result of the STFT at each time step

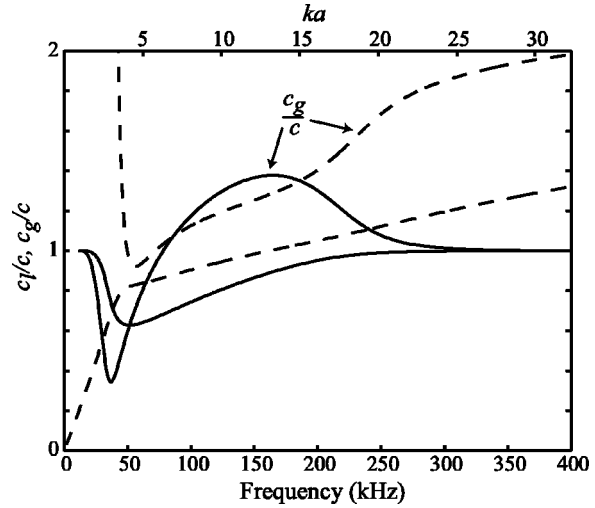


FIG. 4. Computed phase and group velocities for two axisymmetric modes of an infinite empty cylindrical shell submerged in water. The solid curves correspond to the subsonic antisymmetric solution a_{0-}, while the dashed lines correspond to the supersonic a_0. The subsonic branch exhibits extrema in the group velocity of 1.38^*c at $f=164.5$ kHz ($ka=13.28$) and 0.34^*c at $f=36.86$ kHz ($ka=2.98$), respectively.

by the magnitude of the spectrum of the incident acoustic pulse (see Fig. 5). The vertical white line marks the time of a direct echo from the nearest rear corner (corner **B** in Fig. 1 and time t_B in Fig. 2).

Figure 3(a) shows the response at broadside incidence. The initial broadband echo corresponds to the specular reflection. Following this in time by about 25 μs is another broadband echo resulting from the first circumnavigation of the a_0 leaky Lamb wave. Just after the a_0 echo near 130 kHz is the large first circumnavigation of the a_{0-} leaky Lamb wave commonly referred to as the coincidence frequency or midfrequency enhancement. Near this same frequency and repeating at intervals of about 60 μs are subsequent circumnavigations of the a_{0-}. For a detailed analysis of the scattering at broadside incidence for a thick-walled cylindrical shell see Refs. 26 and 27.

Figure 3(b) shows the echo response beyond the compressional and shear wave cutoff angles at 40.0 degrees. The primary scattering responses at this angle, neglecting the scattering from the nearest end and endcap, are due to the a_0 meridional ray and helical waves of the a_{0-}. At this angle the helical waves of the a_{0-} can be clearly seen as a sequence of six regularly spaced peaks between 100 and 200 kHz delayed with respect to the nearest rear corner timing. With knowledge of the group velocity of these waves it is possible to predict arrival times at the receiver using simple expressions (see Ref. 10 and Appendix B). The dashed magenta curves in Fig. 2 and Fig. 3(b) show these predictions for the timings of helical waves traversing the back side of the cylinder one through six times using an approximate value of the group velocity of 1.3^*c where $c=1.483$ mm/μs is the velocity of sound in water (see Fig. 4). Note that these helical wave responses are for waves which have reflected once from the far end of the cylinder. At this angle the meridional response of the a_0 is expected to be near 700 kHz, beyond the spectral range of the experiment.

The response of the a_0 meridional ray can be seen at 44 degrees tilt in Fig. 3(c) near the corner timing and centered at 470 kHz. It is composed of a single peak isolated from the other features. At this tilt angle the magnitude of the peak is roughly 11 dB below the level of the broadside incidence specular reflection. The ray responsible for this is described by path FHBI in Fig. 1. The meridional ray undergoes only one reflection from the end of the cylinder. No contributions from multiple traversals over the length of the cylinder are observed.

As can be seen from Figs. 3(b)–(f) the a_0 meridional response progresses toward lower frequencies as the tilt angle is increased. The timing, however, stays fixed with respect to the rear corner timing. The amplitude decreases slightly and then increases as the tilt angle reaches end-on incidence. The a_{0-} helical wave peaks remain in the same frequency band but decrease in number and amplitude as the tilt angle is increased. Also the delay between peaks decreases as expected from Fig. 2. At 71.2 degrees no a_{0-} helical waves can be observed. At 90 degrees, however, a periodic response emerges between 100 and 200 kHz, the period being roughly 230 μs. See below for a discussion of this feature.

IV. ANALYSIS AND DISCUSSION

The STFT plots in Fig. 3 show clearly the behavior of the end reflected a_0 meridional ray contribution to the backscattering. The backscattering response consists of a single pulse which is nearly coincident with the timing of a signal travelling direct through the water from the source to the nearest rear corner of the cylinder and back to the receiver. The a_0 response in this manner directly highlights one end of the finite cylinder.[1,2] Synthetic aperture images generated from the present data set clearly illustrate this.[28] There is no axial resonance associated with this feature. The a_0 leaky wave has a relatively high spatial damping rate for the present thick-walled shell such that axial resonances are not likely to be observed.[5] (The amplitude of an a_0 ray travelling the length of the cylinder along the meridian twice would be attenuated by a factor of nearly $e^{-24.0}$ without considering losses into the surrounding water upon reflection from the ends.) Because the damping rate is high it is evident that only a small region of the cylinder along the meridion near the end contributes significantly to the backscattering. The actual ray path should be considered to be FHBI rather than EGHBI, where only a length along the meridian of the cylinder on the order of $L_e=a/\mathrm{Im}[k_z a]$ contributes significantly to the backscattering. The attenuation length L_e is the length over which a ray is attenuated by a factor of e^{-1} through radiation, and $\mathrm{Im}[k_z a]$ is the imaginary part of the axial wave number radius product. Since $\mathrm{Im}[k_z a]$ is of order 1 over the range of frequencies considered[5] the attenuation length is of the order of the radius, which is small compared to the length ($L/a=12.0$) in the present case.

The a_{0-} helical waves are a special case of generalized surface guided helical waves on cylinders. The a_{0-} is subsonic with respect to water and is launched on the cylinder through tunnelling at grazing incidence to the cylinder, the curvature of the shell allowing for coupling. The echo curves

plotted in Fig. 2 describe the timing of these helical waves provided sufficient coupling exists to launch a wave. It is expected that the coupling decreases as the tilt angle approaches end-on incidence since the relative curvature of the shell likewise decreases.

For both the meridional a_0 and the helical a_{0-} the simple launching of waves by phase matching cannot be extended to end-on incidence. The fact that both the a_0 and a_{0-} responses are observed near and at end-on incidence demands a different coupling mechanism. The coupling mechanism is likely impulse loading at the end and flat endcap resulting in axisymmetric or nearly axisymmetric modes of the cylinder being generated. These waves propagate down the body of the cylinder with either no radiation loss, as is the case for the axisymmetric mode of the a_{0-}, or relatively high loss, as with the a_0. Radiation into the backscattering direction may then occur at the ends of the cylinder. The generation of cylinder modes by impulse end loading is well known.[29] In general a complicated set of signals is observed. These can be understood somewhat by referencing the relevant dispersion curves for the shell. Frequencies which correspond to high group velocities arrive first followed by those with lower velocities. Amongst these complicated signals often one portion arrives with a significantly greater amplitude. This is commonly referred to as the Airy phase.[29,30] The Airy phase signal corresponds to extrema in the group velocity dispersion.

To aid in interpreting the near end-on incidence response the dispersion curves for the two relevant axisymmetric modes of an infinite cylindrical shell corresponding to the current configuration have been plotted in Fig. 4. These have been calculated in the same fashion as in Ref. 5. The group velocity dispersion curve for the supersonic a_0 is simply increasing with frequency above its general cutoff frequency of 50 kHz and possesses no extrema that are not associated with cutoff. Given sufficient coupling one may expect higher frequency signals of the axisymmetric a_0 to arrive first followed by lower frequency components, although lower frequency components have somewhat greater damping. This may explain the feature in Fig. 3(f) near 200 kHz and 340 μs. The broadband features between 200 kHz and 500 kHz which are located between the specular reflection from the endcap (broadband feature near 50 μs) and the rear corner timing (the white line) may be the result of axisymmetric quasicompressional waves launched in the same fashion. These have considerably lower damping than the a_0 and possess much higher group velocities.

Impulsive end excitation of the subsonic a_{0-} offers a suitable explanation for the low frequency features in Figs. 3(e) and (f). The group velocity dispersion for the a_{0-} exhibits two extrema: a broad maximum at 164.5 kHz and a narrow minimum at 36.86 kHz. These may be expected to give rise to two Airy phase signals in the backscattering response, the higher frequency signal arriving first followed a considerable time later by the low frequency (and low speed) signal. Since the damping for this mode is identically zero, the backscattered levels may be large depending on the efficiency of the coupling at the ends and multiple traversals may enable axial resonances. Multiple reflections are observed for the signals between 120 kHz and 220 kHz at intervals of approximately 230 μs. This corresponds well to an axial resonance of the Airy phase at 165 kHz which has a group velocity maximum of $c_g = 1.38c$ (where $c = 1.483\,\text{mm}/\mu\text{s}$ is the velocity of sound in water) since $2L/c_g = 223.4\,\mu\text{s}$. The low frequency signal barely discernible below 50 kHz in Figs. 3(e) and (f) may be associated with the Airy phase signal due to the minimum in the group velocity dispersion at 37 kHz.

ACKNOWLEDGMENT

This research was supported by the Office of Naval Research.

APPENDIX A: BIPOLAR OPERATION OF A PVDF SHEET TRANSDUCER

Figure 5 shows a time record, and its spectrum (inset), of the acoustic pulse used in the experiment. This acoustic pulse is generated by applying a square electrical pulse of duration 1.6 μs to the PVDF sheet source (110 μm thick Kynar® polyvinylidene fluoride film, metallized on both sides with an Ag ink). Operating the sheet source in this manner is different from previous experiments[4] where an electrical step was applied to the source. In that mode of operation the resulting acoustic pulse is approximately unipolar. This gives increased low frequency performance. In the present mode of operation the acoustic pulse is bipolar, giving better high frequency performance. The behavior of this source can be understood through a simple lumped parameter circuit model which approximates the sheet as a capacitor with an associated series resistance.[4]

APPENDIX B: CORNER REFLECTION AND HELICAL WAVE ECHO RETURNS

This appendix section gives the expressions for the return times for geometrical propagation through the water from the sheet source to each of the four ''corners'' of the cylinder and back to the receiver. Also, a discussion is presented for calculating the echo timings for the a_{0-} helical waves.

For a fixed receiver a distance R away from the center of a cylinder the distances from each of the four corners of the

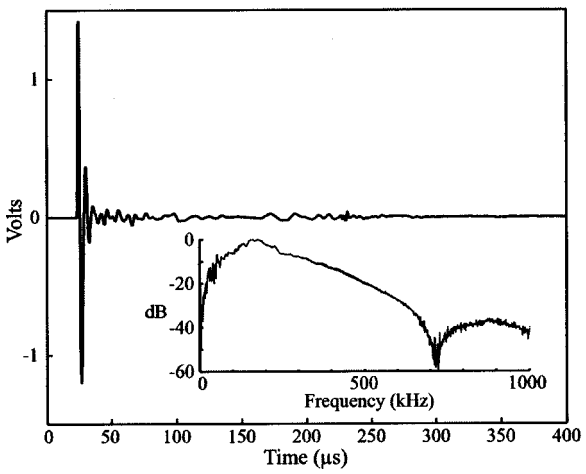


FIG. 5. Time record and spectrum (inset) of the incident acoustic pulse used in the experiments.

cylinder, **A**, **B**, **C** or **D**, to the receiver can be shown by simple geometry to be

$$R_{\underset{C}{A}}=[(r\sin\delta_-)^2+(R\mp r\cos\delta_-)^2]^{1/2}, \tag{B1a}$$

$$R_{\underset{D}{B}}=[(r\sin\delta_+)^2+(R\pm r\cos\delta_+)^2]^{1/2}, \tag{B1b}$$

where $r=[(L/2)^2+a^2]^{1/2}$ and $\delta_\pm=(90°-\gamma)\pm\tan^{-1}(2a/L)$. When the source is a flat sheet placed between the cylinder and the receiver (see Fig. 1) the travel times referenced to the timing of the specular reflection at broadside incidence ($\gamma=0$ degrees) are given by

$$t_{\underset{C}{A}}=[R_{\underset{C}{A}}\mp r\cos\delta_- -R+2a]/c, \tag{B2a}$$

$$t_{\underset{D}{B}}=[R_{\underset{D}{B}}\pm r\cos\delta_+ -R+2a]/c. \tag{B2b}$$

Echo return timings for the helical waves of the a_{0-} can be derived in a similar fashion to other surface guided helical waves with one essential change. The subsonic a_{0-} waves are launched at a point on the cylinder where the incident wave vector forms a tangent line with the curved surface of the cylinder. The expressions for helical wave echos given by Bao[10] may be used directly by setting the appropriate angle for generating surface waves to $\pi/2$. In Bao's notation the necessary changes and relations are $\gamma=\pi/2-\beta$, $\theta_s=\phi=\pi/2$, and $\alpha=\beta$. The resulting expressions for the echo timings are Bao's equations (4)–(7). The helical waves are distinguished by the number of times they traverse the back side of the cylinder. The echo timings plotted in Fig. 2 correspond to helical waves traversing one through six times. In computing these curves a constant value for the group velocity has been used. As can be seen from Fig. 4 the group velocity for the a_{0-} helical waves are clearly not constant over the frequency range of interest. In addition the group velocity for helical waves propagating on the shell at different helix angles generally differs, although in the present case this feature is only a small effect.[28] The anisotropy of helical wave parameters for the supersonic flexural helical wave on the present shell at similar frequencies has been shown by Blonigen[8,31] to be small but important, for example, in the formulation of a ray theory of helical flexural waves on cylindrical shells above the coincidence frequency. For the present approximation of the helical wave arrival times these effects are ignored and an average group velocity is assumed.

[1] G. Kaduchak, C. M. Wassmuth, and C. M. Loeffler, "Elastic wave contributions in high resolution acoustic images of fluid-filled, finite cylindrical shells in water," J. Acoust. Soc. Am. **100**, 64–71 (1996).

[2] S. S. Dodd, Ph.D. dissertation, University of Texas, Austin, Applied Research Laboratories, 1995.

[3] P. L. Marston, "Approximate meridional leaky ray amplitudes for tilted cylinders: End-backscattering enhancements and comparisons with exact theory for infinite solid cylinders," J. Acoust. Soc. Am. **102**, 358–369 (1997); **103**, 2236 (1998).

[4] S. F. Morse, P. L. Marston, and G. Kaduchak, "High-frequency backscattering enhancements by thick finite cylindrical shells in water at oblique incidence: Experiments, interpretation and calculations," J. Acoust. Soc. Am. **103**, 785–794 (1998).

[5] S. F. Morse and P. L. Marston, "Meridional ray contributions to scattering by tilted cylindrical shells above the coincidence frequency: ray theory and computations," J. Acoust. Soc. Am. **106**, 2595–2600 (1999).

[6] S. F. Morse and P. L. Marston, "Degradation of meridional ray backscattering enhancements for tilted cylinders by mode conversion: Wideband observations using a chirped PVDF sheet source," IEEE J. Ocean. Eng. **26**, 152–155 (2001).

[7] P. L. Marston, "Approximations for leaky wave amplitudes in acoustic imaging: applications to high frequency sonar," in *Acoustical Imaging*, edited by S. Lees and L. A. Ferrari (Plenum, New York, 1997), Vol. 23, pp. 369–374.

[8] F. J. Blonigen and P. L. Marston, "Leaky helical flexural wave scattering contributions from tilted cylindrical shells: ray theory and wave-vector anisotropy," J. Acoust. Soc. Am. **110**, 1764–1769 (2001).

[9] K. Gipson and P. L. Marston, "Backscattering enhancements due to reflection of meridional leaky Rayleigh waves at the blunt truncation of a tilted solid cylinder in water: Observations and theory," J. Acoust. Soc. Am. **106**, 1673–1680 (1999).

[10] X-L. Bao, "Echoes and helical surface waves on a finite elastic cylinder excited by sound pulses in water," J. Acoust. Soc. Am. **94**, 1461–1466 (1993).

[11] N. Touraine, L. Haumesser, D. Decultot, and G. Maze, "Analysis of the acoustic scattering at variable incidences from an extra thin cylindrical shell bounded by hemispherical endcaps," J. Acoust. Soc. Am. **108**, 2187–2196 (2000).

[12] L. Haumesser, A. Baillard, D. Decultot, and G. Maze, "Behavior of first guided wave on finite cylindrical shells of various lengths: Experimental investigation," J. Acoust. Soc. Am. **109**, 583–590 (2001).

[13] M. L. Rumerman, "Contribution of membrane wave reradiation to scattering from finite cylindrical steel shells in water," J. Acoust. Soc. Am. **93**, 55–65 (1993).

[14] N. Yen, L. R. Dragonette, and S. K. Numerich, "Time-frequency analysis of acoustic scattering from elastic objects," J. Acoust. Soc. Am. **87**, 2359–2370 (1990).

[15] L. R. Dragonette, D. M. Drumheller, C. F. Gaumond, D. H. Hughes, B. T. O'Connor, N. Yen, and T. J. Yoder, "The application of two-dimensional signal transformations to the analysis and synthesis of structural excitations observed in acoustical scattering," Proc. IEEE **84**, 1249–1262 (1996).

[16] W. H. Prosser and M. D. Seale, "Time-frequency analysis of the dispersion of Lamb modes," J. Acoust. Soc. Am. **105**, 2669–2676 (1999).

[17] M. Niethammer, L. J. Jacobs, J. Qu, and J. Jarzynski, "Time-frequency representation of Lamb waves using the reassigned spectrogram," J. Acoust. Soc. Am. **107**, L19–L24 (2000).

[18] P. Chevret, N. Gache, and V. Zimpfer, "Time-frequency filters for target classification," J. Acoust. Soc. Am. **106**, 1829–1837 (1999).

[19] G. C. Gaunaurd, D. Brill, and H. Huang, "Signal processing of the echo signatures returned by submerged shells insonified by dolphin "clicks:" Active classification," J. Acoust. Soc. Am. **103**, 1547–1557 (1998).

[20] D. S. Burnett, "A three-dimensional acoustic infinite element based on a prolate spheroidal multipole expansion," J. Acoust. Soc. Am. **96**, 2798–2816 (1994); "Prolate and oblate spheroidal acoustic infinite elements," Comput. Methods Appl. Mech. Eng. **158**, 117–141 (1998); "An ellipsoidal acoustic infinite element," *ibid.* **164**, 49–76 (1998).

[21] A. N. Norris and D. A. Rebinsky, "Acoustic coupling to membrane waves on elastic shells," J. Acoust. Soc. Am. **95**, 1809–1829 (1994).

[22] M. Tran-Van-Nhieu, "Scattering from a finite cylindrical shell," J. Acoust. Soc. Am. **91**, 670–679 (1991).

[23] F. Leon, F. Lecroq, D. Decultot, and G. Maze, "Scattering of an obliquely incident acoustic wave by an infinite hollow cylindrical shell," J. Acoust. Soc. Am. **91**, 1388–1397 (1992).

[24] V. V. Muzychenko and S. A. Rybak, "Low-frequency resonance scattering of sound by finite cylindrical shells (review)," Sov. Phys. Acoust. **34**, 325–333 (1988).

[25] L. Cohen, *Time-Frequency Analysis* (Prentice Hall, New Jersey, 1995).

[26] N. H. Sun and P. L. Marston, "Ray synthesis of leaky Lamb wave contributions to backscattering from thick cylindrical shells," J. Acoust. Soc. Am. **91**, 1398–1402 (1992).

[27] P. L. Marston and N. H. Sun, "Backscattering near the coincidence frequency of a thin cylindrical shell: Surface wave properties from elasticity theory and an approximate ray synthesis," J. Acoust. Soc. Am. **97**, 777–783 (1995).

[28] S. F. Morse, Ph.D. Dissertation, Washington State University, Department of Physics, 1998.

[29] K. F. Graff, *Wave Motion in Elastic Solids* (Dover, New York, 1991).

[30] W. M. Ewing, W. S. Jardetzky, and F. Press, *Elastic Waves in Layered Media* (McGraw-Hill, New York, 1957).

[31] F. J. Blonigen, Ph.D. dissertation, Washington State University, Department of Physics, 2001.

Broadband radiation modes: Estimation and active control

Arthur P. Berkhoff[a)]
TNO TPD/University of Twente, The Netherlands

(Received 16 July 2001; revised 10 December 2001; accepted 21 December 2001)

In this paper we give a formulation of the most efficiently radiating vibration patterns of a vibrating body, the radiation modes, in the time domain. The radiation modes can be used to arrive at efficient weighting schemes for an array of sensors in order to reduce the controller dimensionality. Because these particular radiation modes are optimum in a broadband sense, they are termed broadband radiation modes. Methods are given to obtain these modes from measured data. The broadband radiation modes are used for the design of an actuator array in a feedback control system to reduce the sound power radiated from a plate. Three methods for the design of the actuator are compared, taking into account the reduction of radiated sound power in the controlled frequency range, but also the possible increase of radiated sound power in the uncontrolled frequency range. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1451067]

PACS numbers: 43.40.Vn, 43.40.At, 43.40.Rj [EGW]

I. INTRODUCTION

Active control of sound radiated from structures using structural sensing involves controlling the vibration patterns of the structure that radiate sound efficiently. Several methods have been suggested to determine these vibration patterns. The methods can be based on the singular value decomposition of the Green's function[1,2] or an eigenvector analysis of a positive definite radiation operator.[3] The resulting vibration patterns are called radiation modes.[2,4] The associated singular values/eigenvalues can be interpreted as radiation efficiencies. Borgiotti[1] noted that the radiation modes are real valued and therefore, in principle, allow delay-free sensing of acoustic radiation. A sensor without delay is particularly important in feedback configurations. A complication for a real-time implementation is that both the radiation mode shapes as well as the radiation efficiencies depend on frequency. A solution was given by Borgiotti and Jones,[5] who demonstrated that minimizing the radiation modal error signals that are obtained by weighting the sensor signals with the radiation mode shapes at the highest controlled frequency is sufficient to reduce the sound also at lower frequencies. This is because the radiation modes at a certain frequency form a complete description of the radiating vibration patterns at lower frequencies.

At frequencies other than the frequency where the radiation modes have been determined, the normalization frequency,[6] the radiation modes do not diagonalize the radiation operator, except for special cases.[7] Therefore, in principle, if the radiation modes at a single normalization frequency are used, all self-radiation efficiencies and mutual radiation efficiencies have to be taken into account to be able to compute the exact radiated power at other frequencies. Simplifications were given,[8,9] demonstrating that good sound power reductions can be obtained by using only the self-radiation efficiencies. It was also shown that the choice of the radiation mode shapes is not critical, provided they form a complete set of basis functions for the radiating vibration patterns in the frequency range of interest.[9] In some cases, such as for free-field sound radiation, frequency-dependent weighting might not be strictly necessary.[9] In other cases, however, a frequency-dependent weighting of the radiation modal error signals was found to be necessary, especially in enclosed spaces.[10] There seems to be agreement that frequency-independent spatial filters can, in general, be used for broadband active noise control problems involving sound radiation from plates. Even if a frequency-dependent weighting of the radiation modal error signals is used, still a choice has to be made which normalization frequency to use.

In this paper we address the determination of the radiation mode shapes and associated efficiencies based on broadband sound radiation. The radiation modes obtained in this way are the optimum vibration patterns in an average sense evaluated over a predefined frequency band. This eliminates the need for the selection of a normalization frequency. The subsequent frequency-dependent weighting can be obtained by evaluating the self- and mutual radiation efficiencies of the resulting radiation modes. Instead of using a frequency domain formulation, in this paper we use a time domain formulation to determine the radiation modes. This leads to advantages for the identification of radiation modes from experimental data, in which robustness is a primary concern. The formulation can be used for the identification of multiple radiation modes from experimental data, using microphones in the farfield.

Preumont *et al.*[11] determined weighting coefficients for a piezoelectric sensor array for the estimation of the volume velocity, which is the strongest radiation mode at low frequencies. One of the extensions of the method in the present paper is that it also leads to weighting factors for higher-order radiation modes. Another extension is that the radiation modes can be determined from pressure measurements in the farfield instead of using a particular velocity distribution (viz., constant velocity) and assuming that this velocity distribution is responsible for the major part of the radiated acoustic power. Also in the experimental work of Gibbs *et al.*,[8,12] a theoretical model is used for the radiation into the

a)Electronic mail: berkhoff@tpd.tno.nl

acoustic environment. Simplifications of the actuator and sensor configuration were presented in Ref. 13.

Modal sensors and modal actuators based on spatially continuous transducers were discussed by Lee and Moon.[14] The present methods are based on discrete sensors and actuators, such as described by Clark,[15] Gawronski,[16] and Morgan,[17] in order to provide more flexibility by using programmable weighting coefficients. The emphasis in the present paper is on acoustic radiation and the connection with previous radiation mode theory.

The application of active structural acoustic control to flow-induced structural radiation of sound is described by Heatwole *et al.*,[18] where, in a simulation study, the sound pressure at a single point is reduced by using frequency-dependent radiation filters for each structural mode, similar to the approach of Baumann *et al.*[19] In the paper by Maury *et al.*,[20] matched, spatially continuous, radiation modal sensors and actuators are used in a simulation study in order to create a controller without spillover that consists of a gain times the inverse of the individual modal transfer functions, which are minimum phase.

As compared to the results in the paper of Morgan,[17] in the present paper we give an extension to causal systems suitable for broadband stochastic signals. As such, there is also a relationship with the paper of Baumann *et al.*[19] However, the present paper differs in a number of ways. First, in the present paper gives explicit results for the most efficiently radiating vibration patterns of a structure. The vibration patterns obtained in this way are determined only by the acoustics of the problem and not by structural properties. Hence, the structural part and the acoustic part of the problem are decoupled. Instead, the paper of Baumann *et al.* derives a set of frequency-dependent filters for combinations of structural vibration modes, which together provide a measure for acoustic radiation. Second, a proof is given of the completeness of the eigenvectors obtained from a broadband description of sound radiation and the relationship with the nesting principle[5] is demonstrated. The weighting schemes for the actuator array are optimized in a broadband sense for coupling to the acoustic field at low frequencies, while being constrained by a reduced coupling to the acoustic field at high frequencies leading to reduced spillover.

II. RADIATION MODES FOR SENSOR ARRAYS

The acoustic power radiated from a structure to the farfield can be derived from a measurement of the time-averaged squared sound pressure at a relatively large distance from the structure using a sufficient number of measurement positions.[21] Global sound pressure reduction inside an enclosure is related to the reduction of the potential energy in the enclosure, which can be estimated from the squared sound pressure at a number of positions evenly distributed over the interior of the enclosure. Thus, for some interesting problems involving sound radiation it seems appropriate to define a cost function,

$$\Pi = \mathsf{E}\{\mathbf{p}^T(t)\mathbf{p}(t)\}, \tag{1}$$

with

$$\mathbf{p}(t) = [p_1(t)p_2(t)\cdots p_N(t)]^T, \tag{2}$$

the $N\times 1$-dimensional vector of pressure signals at discrete time t and E the expected value operator, which, for a time-dependent matrix $\mathbf{F}(t)$ with elements that are assumed to be wide sense stationary and ergodic, is defined as

$$\mathsf{E}\{\mathbf{F}(t)\} = \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\mathbf{F}(t). \tag{3}$$

The pressure $p_n(t)$, $n=1,...,N$ is assumed to be related to the vibrations of the structure [which can be velocity or another quantity that provides sufficient information to predict $\mathbf{p}(t)$] through the relationship

$$p_n(t) = \sum_{m=1}^{M}\sum_{\tau=0}^{T-1} g_{mn}(\tau)q_m(t-\tau), \tag{4}$$

where $q_m(t)$ is the mth sensed quantity on the structure and where $g_{mn}(t)$ is the space–time Green's function for the problem. Further, T denotes the number of samples taken into account, i.e., the length of the impulse response of $g_{mn}(t)$. It is noted that in many definitions of the Green's functions for problems in acoustics an additional temporal differentiation has to be applied, such as for the Rayleigh integral.[21] However, in the present paper the additional differentiation and other temporal characteristics are incorporated in $g_{mn}(t)$.

Let us define the $M\times 1$-dimensional vectors,

$$\mathbf{g}_n(t) = [g_{1n}(t)g_{2n}(t)\cdots g_{Mn}(t)]^T, \quad n=1,...,N, \tag{5}$$

and the $N\times M$-dimensional matrix,

$$\mathbf{G}(t) = [\mathbf{g}_1(t)\mathbf{g}_2(t)\cdots\mathbf{g}_N(t)]^T. \tag{6}$$

Further, let us define an $M\times 1$-dimensional structural sensor vector as

$$\mathbf{q}(t) = [q_1(t)q_2(t)\cdots q_M(t)]^T. \tag{7}$$

We then have

$$\mathbf{p}(t) = \sum_{\tau=0}^{T-1}\mathbf{G}(\tau)\mathbf{q}(t-\tau). \tag{8}$$

Alternatively, Eq. (1) can be written as

$$\Pi = \text{tr.}\,\mathsf{E}\{\mathbf{p}(t)\mathbf{p}^T(t)\}, \tag{9}$$

where tr. denotes the trace operator. For notational convenience, it is assumed that any frequency dependence in the spectral properties of $\mathbf{q}(t)$ is incorporated in $\mathbf{G}(t)$ and that $\mathbf{q}(t)$ consists of independent white noise processes having equal variance σ. Then the following simplification can be used: $E\{\mathbf{q}(t)\mathbf{q}^T(t')\} = \sigma\delta_{tt'}\mathbf{I}$. Use of the properties that the trace of a square matrix equals the trace of the transpose of that matrix and that the trace of a matrix equals the sum of its eigenvalues leads to

$$\Pi = \text{tr.}\sum_{t=0}^{T-1}\mathbf{G}^T(t)\mathbf{G}(t) = \sum_i \lambda_i\left\{\sum_{t=0}^{T-1}\mathbf{G}^T(t)\mathbf{G}(t)\right\}, \tag{10}$$

where $\lambda_i\{\ \}$ denotes the ith eigenvalue of the matrix. Hence, the most efficient transfer of vibration to radiated sound power is governed by the largest eigenvalues and corre-

sponding eigenvectors of the matrix $\Sigma_{t=0}^{T-1}\mathbf{G}^T(t)\mathbf{G}(t)$.

For reasons of convenience, we turn to a frequency domain description where all variables now depend on angular frequency ω, are complex-valued and deterministic. The frequency domain equivalents of corresponding temporal variables are indicated by adding an overcaret sign. A discrete-time Fourier transform pair for matrices $\mathbf{P}(t)$ and $\hat{\mathbf{P}}(\omega)$ is defined as[22,23]

$$\hat{\mathbf{P}}(\omega)=\sum_{t=-\infty}^{\infty}\mathbf{P}(t)e^{-i\omega t/f_s}, \tag{11}$$

$$\mathbf{P}(t)=\frac{1}{2\pi f_s}\int_{-\pi f_s}^{\pi f_s}\hat{\mathbf{P}}(\omega)e^{i\omega t/f_s}\,\mathrm{d}\omega, \tag{12}$$

with ω the angular frequency in rad/s and f_s the sampling frequency in Hz.

At each angular frequency, the cost function can be expressed as

$$\hat{\Pi}(\omega)=\hat{\mathbf{p}}^H(\omega)\hat{\mathbf{p}}(\omega), \tag{13}$$

where a superscripted H denotes the Hermitian transpose. Using the frequency domain equivalent of Eq. (8), viz.,

$$\hat{\mathbf{p}}(\omega)=\hat{\mathbf{G}}(\omega)\hat{\mathbf{q}}(\omega), \tag{14}$$

the cost function becomes

$$\hat{\Pi}(\omega)=\hat{\mathbf{q}}^H(\omega)\hat{\mathbf{P}}_{\mathbf{GG}}(\omega)\hat{\mathbf{q}}(\omega), \tag{15}$$

where

$$\hat{\mathbf{P}}_{\mathbf{GG}}(\omega)=\hat{\mathbf{G}}^H(\omega)\hat{\mathbf{G}}(\omega). \tag{16}$$

The $M\times M$-dimensional matrix $\hat{\mathbf{P}}_{\mathbf{GG}}(\omega)$ contains information about the vibration patterns of $\hat{\mathbf{q}}(\omega)$ that are significant to $\hat{\mathbf{p}}(\omega)$. These vibration patterns can be obtained by an eigenvector analysis of $\hat{\mathbf{P}}_{\mathbf{GG}}(\omega)$, which is known from control theory (Ref. 24, p. 72). In the limit for an infinite number of pressure sensors at an infinite distance from the radiator, assuming that $\hat{\mathbf{q}}(\omega)$ represents a velocity vector, the matrix $\hat{\mathbf{P}}_{\mathbf{GG}}(\omega)$ approaches the matrix $\mathbf{R}(\omega)$ of Elliott and Johnson[4] up to a scaling factor.

The use of Eq. (11) shows that a transfer correlation matrix,

$$\mathbf{P}_{\mathbf{GG}}(\tau)=\sum_t\mathbf{G}^T(t)\mathbf{G}(t+\tau), \tag{17}$$

is the time domain equivalent of Eq. (16). It is noted that in these descriptions the order of the transpose operators and Hermitian operators are different from usual descriptions of spectral density matrices and corresponding correlation matrices, such as in Refs. 23, 25. From Eq. (12) it can be seen that the value of $\mathbf{P}_{\mathbf{GG}}(t)$ for $t=0$ is proportional to the integral over ω of the spectral density matrix (of transfer functions), and thus a measure of signal power (transfer) in a broadband sense. The term signal power is used here because the sum of squared pressures is not a power in a physical sense. As shown in Eq. (10), the most efficiently radiating, zero-delay vibration patterns in a broadband sense result from the decomposition of the zero lag transfer correlation matrix:

$$\mathbf{P}_{\mathbf{GG}}(0)=\mathbf{E}\mathbf{\Lambda}\mathbf{E}^T, \tag{18}$$

where the matrix $\mathbf{E}$ contains the broadband radiation mode shapes and the diagonal matrix $\mathbf{\Lambda}$ contains the broadband radiation efficiencies. Since $\mathbf{P}_{\mathbf{GG}}(0)$ is symmetric, positive definite, and real, the eigenvalues are also real. The matrices $\mathbf{E}$ and $\mathbf{\Lambda}$ are truncated to retain only the most significant, principal, components. In theory, the radiation modes obtained from the integral of $\hat{\mathbf{P}}_{\mathbf{GG}}(\omega)$ over ω are identical to that obtained from $\mathbf{P}_{\mathbf{GG}}(0)$. However, in practice, there is a difference if they are obtained from measured data, as will be made clear in the next section. Zero delay radiation modes for a certain frequency band can be obtained after bandpass filtering of the Green's functions that lead to $\mathbf{P}_{\mathbf{GG}}(t)$.

Error signals $\boldsymbol{\nu}(t)=[\nu_1(t)\nu_2(t)\cdots\nu_L(t)]^T$ for active control can be obtained by weighting the structural sensor vector $\mathbf{q}(t)$ with the radiation mode shapes in $\mathbf{E}$:

$$\boldsymbol{\nu}(t)=\mathbf{E}^T\mathbf{q}(t), \tag{19}$$

with possible subsequent frequency-dependent weighting by the matrix $\tilde{\mathbf{\Lambda}}(\omega)=\mathbf{E}^T\hat{\mathbf{P}}_{\mathbf{GG}}(\omega)\mathbf{E}$.[9] The corresponding frequency domain cost function becomes

$$\hat{\Pi}(\omega)=\hat{\boldsymbol{\nu}}^H(\omega)\tilde{\mathbf{\Lambda}}(\omega)\hat{\boldsymbol{\nu}}(\omega). \tag{20}$$

The resulting error signal $\mathbf{e}(t)=[e_1(t)e_2(t)\cdots e_L(t)]^T$, which is obtained by a frequency-dependent weighting of $\boldsymbol{\nu}(t)$, is expressed as

$$\mathbf{e}(t)=\sum_{\tau=0}^{T-1}\mathbf{K}(\tau)\boldsymbol{\nu}(t-\tau), \tag{21}$$

where $\mathbf{K}(t)$ is implicitly defined by

$$\tilde{\mathbf{\Lambda}}(\omega)=\sum_{\tau=-\infty}^{\infty}\sum_t\mathbf{K}^T(t)\mathbf{K}(t+\tau)e^{-i\omega\tau/f_s}. \tag{22}$$

If the full matrix $\tilde{\mathbf{\Lambda}}(\omega)$ is used, then the exact matrix $\hat{\mathbf{P}}(\omega)$ and the corresponding correlation matrix of Eq. (17) are reconstructed. Methods to find $\mathbf{K}(t)$ from rational matrices $\tilde{\mathbf{\Lambda}}(\omega)$ can be found in papers by, e.g., Youla;[26] methods with an application to acoustic radiation problems are also known.[27] In many cases it will be sufficient to use only the diagonal of $\tilde{\mathbf{\Lambda}}(\omega)$ or even no weighting at all. In the latter case, $\tilde{\mathbf{\Lambda}}(\omega)$ equals the identity matrix.

The eigenvectors in $\mathbf{E}$ are not exactly orthogonal for acoustic radiation at each frequency, but only approximately. They are, however, usable for broadband sound radiation control in the same sense as the radiation modes at the highest controlled frequency, as suggested for free space radiation by Borgiotti and Jones.[5] The essential observation to be made is that, because the radiation modes at a certain highest frequency are a complete set of basis functions for acoustic radiation at all lower frequencies, then also the eigenvectors obtained from an averaging procedure over frequency up to this highest frequency, that lead to Eq. (18), are a complete set of basis functions. A proof is given in the Appendix. Although the resulting eigenvectors are not radiating exactly independently, but only approximately, and, consequently, are not modes in a strict sense, they are termed broadband radiation modes here because of the close correspondence

with the frequency domain radiation modes. A property of the time domain technique is that the resulting radiation modes are real valued. For positive definite radiation operators, this also holds for frequency domain techniques but, in practice, using measured data from a finite number of field sensors at a finite distance, these radiation modes are not entirely real.[28]

III. ESTIMATION OF THE GREEN'S FUNCTION

The formulation of the previous section can be used for the identification of the radiation modes *in situ*. One method is based on the excitation of the structure with a number of different excitation patterns and the acquisition of the resulting structural sensor signals and pressure signals. A time–domain inverse filtering technique is used to extract the underlying Green's function. The time domain technique has the advantage that a single minimization procedure is used to obtain the complete impulse response. Instead, an averaging procedure in the frequency domain[28] based on inversions at each frequency can still be based on unreliable constituents.

Let k indicate the measurement; then we can define an error for the nth pressure signal as follows:

$$e_n^k(t) = p_n^k(t) - \sum_{\tau=0}^{T-1} \mathbf{g}_n^T(\tau)\mathbf{q}^k(t-\tau). \tag{23}$$

Using the average of the squared value of $e_n^k(t)$ over a sufficiently long period of time and summing the contributions of the different measurements leads to the cost function

$$J_n = \sum_k \mathsf{E}\{[e_n^k(t)]^2\}. \tag{24}$$

Let the correlation matrix $\mathbf{R}_{\mathbf{xy}}(\tau)$ be defined by

$$\mathbf{R}_{\mathbf{xy}}(\tau) = \mathsf{E}\{\mathbf{y}(t+\tau)\mathbf{x}^T(t)\}, \quad \tau = -T+1,...,T-1; \tag{25}$$

then the cost function can be written as

$$J_n = \sum_k \left[\mathbf{R}_{p_n^k p_n^k}(0) - \sum_{\tau=0}^{T-1} \mathbf{R}_{\mathbf{q}^k p_n^k}(\tau)\mathbf{g}_n(\tau) - \sum_{\tau=0}^{T-1} \mathbf{g}_n^T(\tau)\mathbf{R}^T_{\mathbf{q}^k p_n^k}(\tau) + \sum_{\tau=0}^{T-1}\sum_{\tau'=0}^{T-1} \mathbf{g}_n^T(\tau) \times \mathbf{R}_{\mathbf{q}^k\mathbf{q}^k}(\tau'-\tau)\mathbf{g}_n(\tau') \right]. \tag{26}$$

The derivative of the cost function with respect to the coefficients $\mathbf{g}_n(t)$ can be written as

$$\frac{\partial J_n}{\partial \mathbf{g}_n(\tau)} = \sum_k \left[\sum_{\tau'=0}^{T-1} \mathbf{g}_n^T(\tau')\mathbf{R}_{\mathbf{q}^k\mathbf{q}^k}(\tau-\tau') - \mathbf{R}_{\mathbf{q}^k p_n^k}(\tau) \right], \quad \tau = 0,...,T-1. \tag{27}$$

Requiring this derivative to be zero leads to the system of equations

$$\sum_k \sum_{\tau'=0}^{T-1} \mathbf{g}_n^T(\tau')\mathbf{R}_{\mathbf{q}^k\mathbf{q}^k}(\tau-\tau') = \sum_k \mathbf{R}_{\mathbf{q}^k p_n^k}(\tau), \quad \tau = 0,...,T-1, \tag{28}$$

from which $\mathbf{g}_n(t)$ can be solved. The Green's function matrix $\mathbf{G}(t)$ is then assembled from the different $\mathbf{g}_n(t)$ according to Eq. (6). Since the structure of the system of equations is block Toeplitz,[29] an efficient routine can be used for its solution.[30] A quadratic cost involving the coefficients of $\mathbf{g}_n$ was added in order to stabilize the solution. The cost function actually minimized was

$$J_n = \sum_k \mathsf{E}\{[e_n^k(t)]^2\} + \beta \sum_{\tau=0}^{T-1} \mathbf{g}_n^T(\tau)\mathbf{g}_n(\tau), \tag{29}$$

with β a weighting coefficient for the Green's function coefficients. If the Green's function contains strongly resonating behavior, such as in an enclosed space, then it may be advantageous to combine information regarding the poles of the system, such as with the subspace technique developed by De Moor, as described by Van der Veen *et al.*[31] In some cases it is possible and more practical to use a reciprocity method for measuring the Green's functions. Frequency domain reciprocity methods for the estimation of radiation modes are described in Ref. 28. In the reciprocity method the positions of the source and the receiver are interchanged. In that case there is no need to perform the inversion as described above. Combinations of inverse and reciprocity methods are also possible.[28]

IV. ACTUATOR MODES

The broadband radiation mode formulation can also be used for the design of signal processing schemes that drive an array of actuators. As compared to the radiation mode schemes for sensors, modal schemes for actuators have additional requirements. Like the sensor scheme, the actuator scheme is required to consist of a set of nearly independent driving patterns, the actuator modes, that have good sensitivity for sound radiation in the frequency range where sound power reductions have to be obtained. This frequency range is usually the low-frequency range. In addition, the actuator configuration should be designed in such a way that, in the frequency range where the near-field sensor configuration fails to predict the farfield sound field, usually in the high-frequency range, the actuator modes have a small sensitivity for sound radiation. In other words, the configuration should be designed in such a way that it has a minimum of spillover. In the following a method is given to satisfy these requirements using a given physical actuator configuration. As opposed to approaches in which the positions of the actuators are optimized,[32] the present formulation optimizes a signal processing scheme for a given actuator array. The advantage is that the system can be calibrated *in situ* with measured transfer functions using an identical actuator array, which can be used for different situations.

A. Cost function for actuator modes

An approach to reduce spillover effects is to minimize the modal contributions due to the secondary path, which are not taken into account by error sensors. As compared to the method of Morgan,[17] the present description gives an extension to broadband signals and an optimization for acoustic

radiation and not simply mechanical vibration. A further extension is that a method is presented to reduce spillover at high frequencies by defining two different sets of vibration patterns, one set of which is used for sound power reductions at low frequencies, while the other set is used for the reduction of spillover.

For the moment it will be assumed that the configuration is feedforward. Let $\mathbf{u}(t)=[u_1(t)u_2(t)\cdots u_K(t)]^T$ be the vector of actuator control signals, which is obtained by filtering the reference signal vector $\mathbf{x}(t)=[x_1(t)x_2(t)\cdots x_L(t)]^T$ with a matrix $\mathbf{W}(t)$ of finite impulse response filters:

$$\mathbf{u}(t)=\sum_{\tau=0}^{T-1}\mathbf{W}(\tau)\mathbf{x}(t-\tau). \tag{30}$$

Further, let $\mathbf{q}_s(t)$ be the vector of secondary signals at the sensors, and let $\mathbf{H}(t)$ be the matrix of transfer functions from actuator signal vector to sensor signal vector, as in

$$\mathbf{q}_s(t)=\sum_{\tau=0}^{T-1}\mathbf{H}(\tau)\mathbf{u}(t-\tau). \tag{31}$$

The radiation mode sensor signal vector for the observed radiation modes in $\mathbf{E}$ due to the secondary sources is

$$\boldsymbol{\nu}_s(t)=\mathbf{E}^T\mathbf{q}_s(t). \tag{32}$$

The corresponding signal vector after frequency-dependent weighting is given by

$$\mathbf{s}(t)=\sum_{\tau=0}^{T-1}\mathbf{K}(\tau)\boldsymbol{\nu}_s(t-\tau). \tag{33}$$

The total error signal vector is simply the sum of the primary signal vector and the secondary signal vector:

$$\mathbf{e}(t)=\mathbf{d}(t)+\mathbf{s}(t), \tag{34}$$

with $\mathbf{d}(t)$ the primary disturbance signal vector of the weighted radiation modal sensors. Thus, if the primary sensor signal vector is given by $\mathbf{q}_d(t)$, then $\mathbf{d}(t)$ can be obtained from the subsequent operations,

$$\boldsymbol{\nu}_d(t)=\mathbf{E}^T\mathbf{q}_d(t) \tag{35}$$

and

$$\mathbf{d}(t)=\sum_{\tau=0}^{T-1}\mathbf{K}(\tau)\boldsymbol{\nu}_d(t-\tau). \tag{36}$$

We also define error signals for unobserved modes. The radiation modal sensor signal vector due to the secondary sources for a selected set of unobserved radiation modes $\mathbf{E}'$, termed constraint modes, is

$$\boldsymbol{\nu}_{s'}(t)=\mathbf{E}'^T\mathbf{q}_s(t). \tag{37}$$

The corresponding signal vector after frequency-dependent weighting is given by

$$\mathbf{s}'(t)=\sum_{\tau=0}^{T-1}\mathbf{K}(\tau)\boldsymbol{\nu}_{s'}(t-\tau). \tag{38}$$

Finally, the error criterion is defined as

$$J=\mathrm{E}\{\mathbf{e}^T(t)\mathbf{e}(t)\}+\beta_1\,\mathrm{tr.}\left\{\sum_{\tau=0}^{T-1}\mathbf{W}^T(\tau)\mathbf{W}(\tau)\right\}+\beta_2\mathrm{E}\{\mathbf{u}^T(t)\mathbf{u}(t)\}+\beta_3\mathrm{E}\{\mathbf{s}'^T(t)\mathbf{s}'(t)\}. \tag{39}$$

The first part of the cost function minimizes the error signal on the observed modes, the second part minimizes the filter coefficients, the third part minimizes the control effort, and the fourth part minimizes the secondary path signals that are not observed by the error sensors. The minimization of this equation can be obtained by using efficient methods for block-Toeplitz matrices.[30] Usually, only one of the parameters β_1, β_2, or β_3 is used for the computation of $\mathbf{W}$.

B. Selection of constraint modes

As described in Sec. II, a method to obtain $\mathbf{E}$ is by low-pass filtering of the Green's functions if reductions at low frequencies have to be obtained. Usually, the constraint modes $\mathbf{E}'$ cannot be obtained from the same technique. This is because the modes in $\mathbf{E}'$ should be the modes that are efficient radiators at high frequencies only. In an approach in which the *in*significant eigenvectors are computed, based on low-pass filtering of the Green's functions, the low-pass filtering operation would eliminate information in the frequency range of interest. This is undesirable, especially for measured data. Also, simply high-pass filtering the Green's function and taking the significant eigenvectors is inadequate because the resulting radiation modes are not only efficient radiators at high frequencies, but also possibly efficient radiators at low frequencies. In the following a method is given to compute the modes $\mathbf{E}'$ that are the most efficient radiators in a broadband sense subject to the constraint that they are orthogonal to the modes $\mathbf{E}$ that have previously been found as the most efficient radiators at low frequencies. This implies that the modes $\mathbf{E}'$ are efficient radiators at high frequencies and inefficient radiators at low frequencies.

In this section it is shown how these two goals can be accomplished. In particular, the modes of $\mathbf{P}_{\mathbf{GG}}(0)$, obtained from possibly high-pass filtered Green's functions, are computed using the constraint that they are orthogonal to r vectors in an $M\times r$-dimensional matrix $\mathbf{E}$. If the desired modes are the columns of the M-row matrix $\mathbf{E}'$, then the constraint can be expressed as

$$\mathbf{E}^T\mathbf{E}'=\mathbf{0}. \tag{40}$$

Following Golub and van Loan[33] and assuming an $M\times M$-dimensional matrix $\mathbf{Q}$, an $M\times r$-dimensional matrix $\mathbf{S}$, and an $r\times r$ dimensional matrix $\mathbf{Z}$, the singular value decomposition of $\mathbf{E}$ is computed as

$$\mathbf{E}=\mathbf{QSZ}^T. \tag{41}$$

An $M\times M$-dimensional matrix $\mathbf{B}$ is formed from

$$\mathbf{B}=\mathbf{Q}^T\mathbf{P}_{\mathbf{GG}}(0)\mathbf{Q}, \tag{42}$$

where the Green's functions $\mathbf{G}$ leading to $\mathbf{P}_{\mathbf{GG}}(0)$ are possibly high-pass filtered. Assuming an $(M-r)\times(M-r)$-dimensional matrix $\mathbf{B}_{22}$, the matrix $\mathbf{B}$ is partitioned as follows:

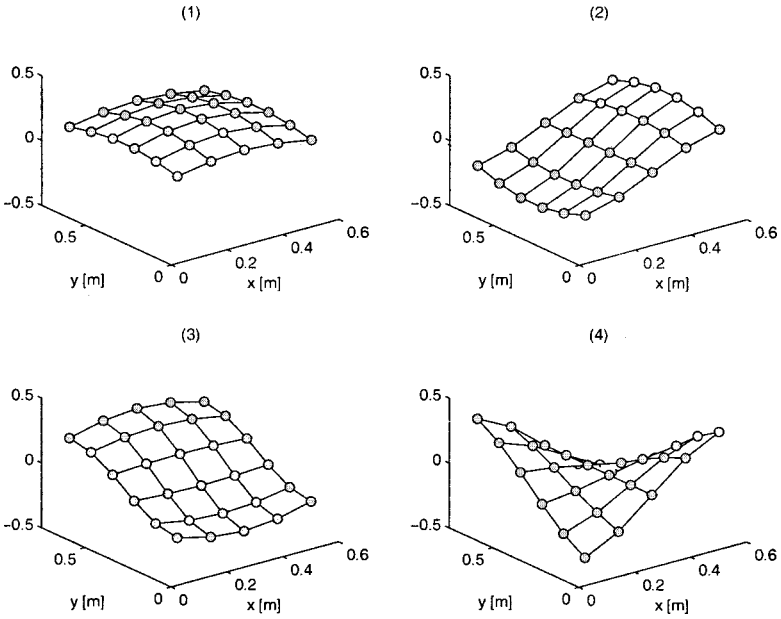


FIG. 1. The four radiation mode shapes with the largest eigenvalues obtained with simulations using a time domain technique using low-pass filtering at 450 Hz ($ka=3$).

$$\mathbf{B}=\begin{bmatrix}\mathbf{B}_{11} & \mathbf{B}_{12}\\ \mathbf{B}_{21} & \mathbf{B}_{22}\end{bmatrix}. \tag{43}$$

Eigenvectors $\mathbf{V}_{\mathbf{B}_{22}}$ are found from the decomposition

$$\mathbf{B}_{22}=\mathbf{V}_{\mathbf{B}_{22}}\mathbf{\Lambda}_{\mathbf{B}_{22}}\mathbf{V}_{\mathbf{B}_{22}}^{T}. \tag{44}$$

Upon construction of the $M\times(M-r)$-dimensional matrix,

$$\mathbf{Y}=\begin{bmatrix}\mathbf{0}\\ \mathbf{V}_{\mathbf{B}_{22}}\end{bmatrix}, \tag{45}$$

where the $\mathbf{0}$ is a zero matrix of dimensions $r\times(M-r)$, the $M-r$ constrained eigenvectors in the $M\times(M-r)$-dimensional matrix $\mathbf{E}'$ follow from

$$\mathbf{E}'=\mathbf{QY}. \tag{46}$$

The vectors in the columns of $\mathbf{E}'$ are the most significant modes in $\mathbf{P}_{\mathbf{GG}}(0)$ subject to the condition that they are orthogonal to the column vectors in $\mathbf{E}$. Simplifications of the computational procedure are possible if the vectors in $\mathbf{E}$ are orthonormal.

V. RESULTS

A. Radiation mode shapes for velocity sensing

As a simple example, radiation mode shapes were computed with Eq. (18) based on the monopole space–time domain Green's function for a baffled radiator [see, for instance, Eq. (5-2.1) in Ref. 21],

$$g_{mn}(\mathbf{t})=S_m\frac{\rho}{2\pi}\frac{\partial}{\partial\mathbf{t}}\frac{\delta(\mathbf{t}-r_{mn}/c)}{r_{mn}}, \tag{47}$$

where $\mathbf{t}$ denotes continuous time, ρ is the density of air, S_m is the area of an elemental radiator,[4] δ is a temporal Dirac impulse function, $c=343$ m/s is the speed of sound, and $r_{mn}=\|\mathbf{x}_m-\mathbf{x}_n\|$ is the distance between source coordinate $\mathbf{x}_m$ and receiver coordinate $\mathbf{x}_n$. Discrete-time representations $g_{mn}(t)$ were obtained by linear interpolation, using samples of Eq. (47) at a sampling frequency of 4096 Hz. Source points were assumed to be distributed over a rectangular area of 60 cm×75 cm, using 5×6 sources. The pressure was recorded at 12 positions on a hemisphere with a radius of 2 m according to the AS1217.6-1985 standard, as described in Ref. 34. The Green's function was low-pass filtered at a frequency $f=450$ Hz, which corresponds to $ka=3$, with $k=\omega/c$ the wave number, $\omega=2\pi f$ the angular frequency, and $a=0.3785$ m, being an effective radius of the total radiator. The low-pass filter and the differentiator were implemented with FIR filters of length 62. The resulting radiation modes are shown in Fig. 1 and the normalized eigenvalues in Fig. 2. In addition, radiation modes were computed with a frequency domain technique using $\mathbf{R}(\omega)$ at $ka=3$,[4] and with the same frequency domain technique using an averaging of $\mathbf{R}(\omega)$ over $0\leqslant ka\leqslant 3$. The time domain result and the averaged frequency domain result are nearly equal, as expected. Also, the eigenvalues are nearly equal. The shapes and eigenvalues of the frequency domain modes at $ka=3$ differ somewhat more, also as expected. Also, an inverse technique as described in Sec. III, was used to obtain radiation mode shapes. The numerical experiments were performed with 30 piezoelectric actuators attached to a plate. For each actuator, the resulting velocity at 30 positions on the plate was computed as well as the pressure at the 12 positions on a hemisphere. Again, the results were almost identical to that of Figs. 1 and 2.

The reciprocity method mentioned in Sec. III, was used to obtain the time domain Green's function $\mathbf{G}(t)$ and the resulting broadband radiation modes $\mathbf{E}$ from measured data. The measurements were performed in a semianechoic room. The measurement technique is the same as described in Ref. 28, except that a causal time–domain Green's function was estimated from broadband noise. As compared to the simulations, the positions of the sources and microphones were interchanged due to the reciprocity method. The sampling frequency was 4096 Hz and the number of samples for the estimation of the Green's functions was 256. The responses were truncated to 32 samples to remove contributions from

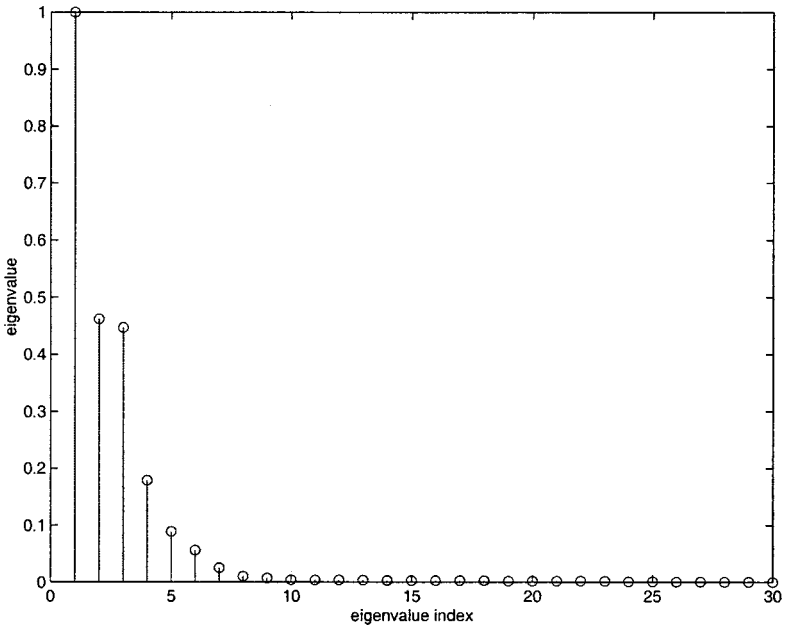


FIG. 2. Eigenvalues corresponding to Fig. 1.

reflections. Subsequently, the responses were low-pass filtered, where the low-pass filter was identical to that in the simulations. The measured radiation modes are shown in Figs. 3 and the normalized eigenvalues in Fig. 4. It can be seen that the shapes and eigenvalues of the first three measured radiation modes agree with the simulated shapes and eigenvalues in Figs. 1 and 2, respectively, except that the order of the second and third mode is interchanged. The shape of the fourth radiation mode differs somewhat more.

B. Design of a modal actuator with application to feedback ASAC

Numerical results are presented for the general cost function of Eq. (39). A simulation example is given of the design of an actuator array for use in active structural acoustic control. The configuration is identical to that in Ref. 35, which consists of nine piezoelectric patch actuators and nine collocated sensors.

For the error signal **e**, three radiation modes were used. The modes **E** were obtained by low-pass filtering the Green's function at 288 Hz, computing the eigenvectors of the transfer correlation matrix, and using the three strongest modes. The constraint modes $\mathbf{E}'$ were obtained by an eigenvector computation based on the unfiltered Green's function subject to the constraint that they had to be orthogonal to the three modes in **E**, as described in Sec. IV B. In the simulations, a simply supported aluminum sandwich plate of 6 mm thickness was used. The width and height were 60 and 75 cm, respectively. The plate density was taken to be $\rho_p = 870\,\mathrm{kg\,m^{-3}}$, the Young's modulus as $E = 3.6\times 10^{10}$ Pa, and the hysteretic damping as $\eta = 0.02$, except for the (1,1) plate

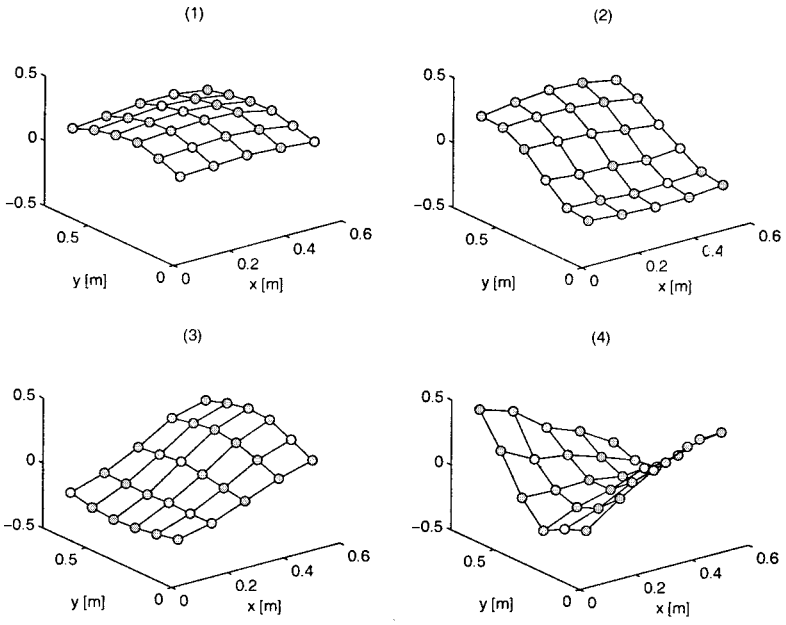


FIG. 3. The four radiation mode shapes with the largest eigenvalues obtained from measurements using a time domain technique with low-pass filtering at 450 Hz ($ka = 3$).

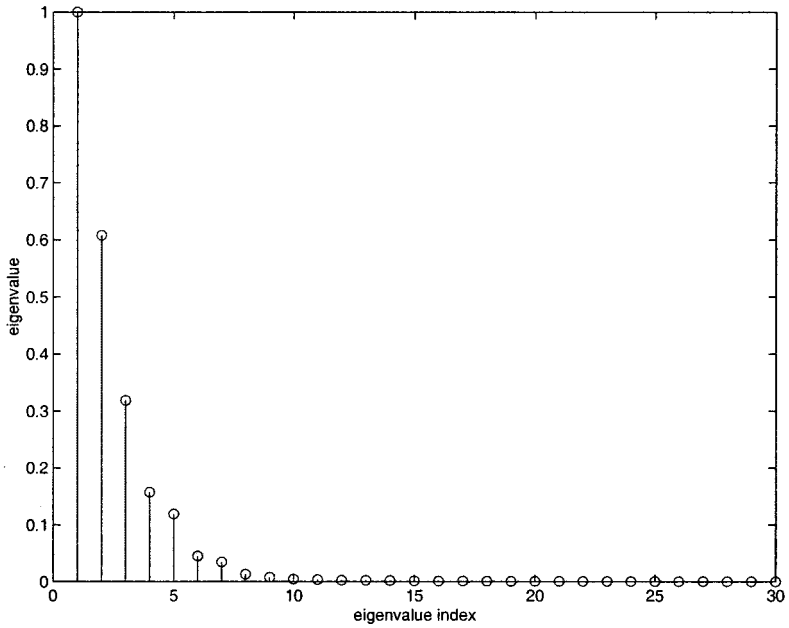


FIG. 4. Eigenvalues corresponding to Fig. 3.

mode for which a larger damping $\eta=0.1$ was used. The number of plate modes taken into account was 8×8. It is assumed that the actuators generate flexural waves only and not longitudinal waves. An analytical method for the computation of weighting coefficients for the piezoelectric sensor array was used, as described previously.[36] Together with the radiation modes $\mathbf{E}$ and $\mathbf{E}'$, these weighting coefficients could be used to derive radiation modal error signals and constraint signals, respectively, for a piezoelelectric sensor array. In Ref. 36, no frequency-dependent weighting was used. In order to improve the performance, for the present example a frequency-dependent weighting was used, which was based on the diagonal of Eq. (22) for both $\mathbf{E}$ and $\mathbf{E}'$.

The control configuration was based on Internal Model Control,[37] where the contribution of the secondary path on the detection sensor was subtracted in the controller (Fig. 5). In this way the controller coefficients $\mathbf{W}$ and the resulting performance of the system could be computed by using the techniques for feedforward control. The configuration was a feedback system where the signals $\mathbf{x}(t)$ were equal to the unfiltered, zero delay, radiation modal error signals $\boldsymbol{\nu}_d(t)$, and where the signals $\mathbf{d}(t)$ and $\mathbf{s}(t)$ were obtained by a frequency-dependent filtering of the radiation modal error signals, by using the diagonal of matrix $\mathbf{K}(t)$. The frequency dependency was implemented by a least-squares fit of the

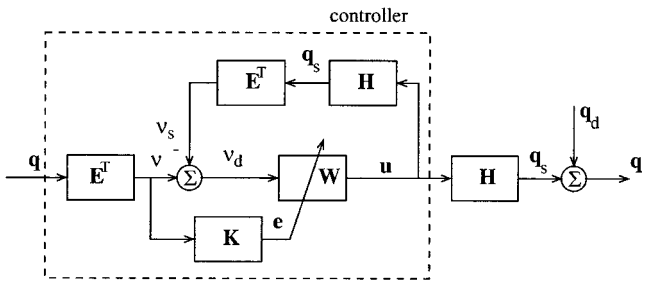


FIG. 5. Schematic diagram of an active feedback control system based on Internal Model Control.

frequency-dependent efficiency of the radiation modes, using the Matlab function firls. The order of the filters was chosen to be 10.

The incident field was an impulsive plane wave incident on the plate at directions $\theta=\pi/3$ and $\phi=\pi/3$ (Ref. 38) to the plate normal. The sampling frequency was 1 kHz and the length of each of the filters of $\mathbf{W}$ was 256, resulting in a total number of coefficients of $256\times3\times9=6912$. A controller delay of one sample was assumed. The radiated sound power was evaluated by computing the plate velocity before control and after control at 10×10 positions.

The influence on the performance of each of the individual parameters β_1, β_2, and β_3 was investigated. For each of these parameters, a value could be found for which the reduction of the broadband radiated sound power was maximum. The results that are shown are at these maximum values of reduction. The results are given in Fig. 6. The broadband reductions for coefficient weighting with β_1, effort weighting with β_2, and radiation modal constraint weighting with β_3 are 8.0, 6.7, and 9.9 dB, respectively. Especially at high frequencies the performance with modal constraint weighting is better than with the other two methods. For configurations having less symmetry than the present configuration the results substantially degrade for weighting with β_1 or β_2, while being relatively close to the above result for weighting with β_3. It can also be seen that the result for modal constraint weighting is the only result that does not lead to increases at any frequency. According to an approximate formula,[9] the maximum frequency for which significant reductions can be obtained using three radiation modes for a plate of 60 cm$\times$75 cm is approximately 300 Hz, which agrees with the results of Fig. 6.

Instead of using a simultaneous minimization of the radiation modal error signals, the cost function J was also minimized for each of the radiation modal error signals independently. Considering the optimal filters for a particular radiation modal error signal, the constraint modes were augmented with the remaining radiation modes, which are two

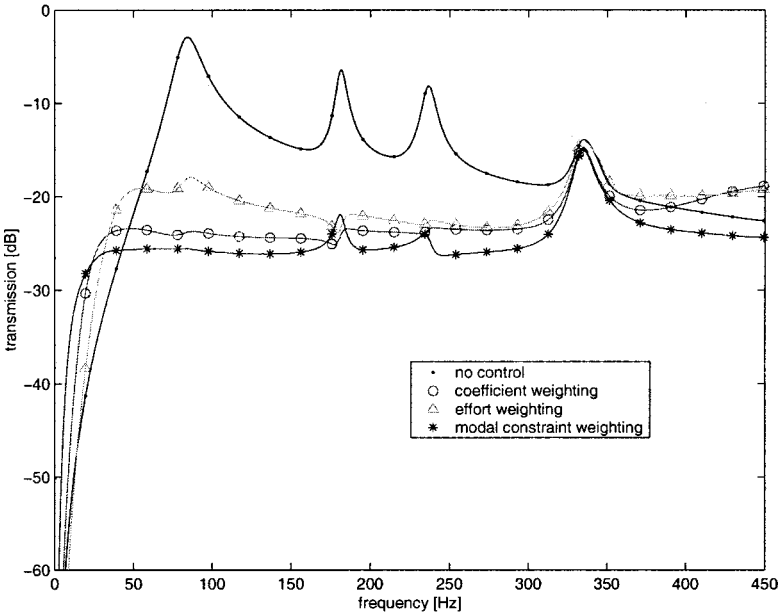


FIG. 6. Sound transmitted through a plate without control and with feedback control, comparing three different feedback control strategies.

modes in this case. The results were similar to the previous results with slightly smaller reductions, being 7.6, 6.7, and 8.5 dB, for weighting with β_1, β_2, and β_3, respectively. The advantage is that the systems obtained in this way are more robust than those obtained from the simultaneous minimization of the radiation modal error signals. The quantification of robustness was based on a method described by Elliott.[39] The maximum singular value of the open-loop gain at any frequency was 700 for the simultaneous optimization and approximately 20 for the independent minimization. It is noted that the latter configuration is a true modal actuator, and therefore the adaptive filter could be chosen to be diagonal. In other words, each radiation modal error signal could be controlled independently by the corresponding radiation modal actuator.

In the present examples the number of actuators is larger than the number of error signals. Also, for the case where the number of actuators was less than or equal to the number of error signals the performance of the modal weighting scheme was found to give the largest reduction of broadband radiated sound power. However, for that case the differences were smaller, especially between effort weighting and modal constraint weighting.

It is interesting to compare the present technique with the technique described by Elliott *et al.*[37] In the latter paper, a method is described to reduce the primary field at the microphones in the farfield by feedback of the structural sensor signals. A property of this method is that a control filter is required between each structural sensor and each actuator. For the present configuration this would lead to a nine times increase of the dimensionality of the adaptive controller, viz. from 3×3 to 9×9. The technique of Elliott *et al.* gives the possibility of limiting the control effort at high frequencies, but this would not lead to a more efficient implementation. A more efficient implementation of the sensor, valid for $ka<1$, is obtained by combining the structural sensor signals into a single radiation modal sensor signal, such as described by Sors[40] and Gibbs.[12,13] The techniques of the present paper give extensions for higher-order radiation modes (i.e., reductions also for $ka\geqslant1$), a method for the design of modal actuators, the reduction of controller dimensionality, a method to limit undesired sound power increases at high frequencies, and a suggested technique to obtain radiation modes *in situ*.

VI. CONCLUSIONS

In this paper methods have been presented to obtain the most efficiently radiating vibration patterns of a plate, termed radiation modes, by using a broadband formulation. It has been shown that a radiation mode formulation based on a description in the time domain can be used to arrive at efficient schemes for the reduction of sound radiated from plates. The resulting modes were called broadband radiation modes because they are the optimum modes in a broadband sense. This is in contrast with existing techniques that are based on radiation modes determined at a single frequency but that are often used to control broadband radiation. Frequency-dependent weighting schemes for the broadband radiation modes have been presented.

For a baffled plate, for which analytical descriptions of relevant Green's functions are available, it has been shown that the radiation modes for velocity sensing obtained from a low-pass filtered Green's function are equivalent to conventional radiation modes obtained from an averaging in the frequency domain. A method was given to determine the radiation modes *in situ*, in which case the relevant Green's function is usually unknown.

Finally, the broadband radiation modes were used to obtain driving schemes for a configuration consisting of arrays of piezoelectric actuators and piezoelectric sensors. Three methods for the derivation of the driving schemes were compared. It has been shown that an optimum filter based on constraints on a special set of vibration patterns, of which the outputs were weighted with frequency-dependent filters, gave better results than that obtained with coefficient weight-

ing and control effort weighting. These constraint patterns were computed in such a way that they were inefficient radiators at low frequencies and efficient radiators at high frequencies. In addition, the new weighting scheme was found to be less sensitive to changes in the configuration. In this way efficient modal sensors and modal actuators can be designed having a good tradeoff between large sound power reductions in the controlled frequency range and a minimum of increased sound power in the uncontrolled frequency range.

ACKNOWLEDGMENT

The author would like to thank Leonard van Lier for doing the measurements.

APPENDIX: COMPLETENESS OF AVERAGED RADIATION MATRICES

In this appendix a proof is given of the completeness of the eigenvectors of an averaged radiation matrix. In particular, it is shown that the subspace spanned by the significant eigenvectors of an averaged radiation matrix $\mathbf{R}$ equals the subspace spanned by the significant eigenvectors at the highest frequency over which the averaging is performed. Positive angular frequencies ω_k are assumed that are increasing with index k, i.e., $0<\omega_k<\omega_{k+1}$. The radiation matrix $\hat{\mathbf{R}}(\omega)$ is defined such that $\hat{\Pi}(\omega)=\hat{\mathbf{q}}^H(\omega)\hat{\mathbf{R}}(\omega)\hat{\mathbf{q}}(\omega)$ is the radiated power. Further, the radiation matrix $\hat{\mathbf{R}}(\omega)$ at $\omega=\omega_k$ is denoted by $\mathbf{R}_k$.

Assuming an eigenvector decomposition $\mathbf{R}_k=\mathbf{E}_k\mathbf{D}_k\mathbf{E}_k^T$, with $\mathbf{E}_k=\{E_1^k,E_2^k,\ldots,E_m^k\}$ and $\mathbf{D}_k=\mathrm{diag}\{d_1^k,d_2^k,\ldots,d_m^k\}$ a diagonal matrix with the eigenvalues in decreasing order, the range of $\mathbf{R}_k$ is defined as the span of its significant eigenvectors $\{E_1^k,E_2^k,\ldots,E_l^k\}$:

$$\mathrm{range}\,\mathbf{R}_k=\mathrm{span}\{E_1^k,E_2^k,\ldots,E_l^k\},\quad d_p^k>\epsilon,\ \forall p\in\{1,\ldots,l\}, \tag{A1}$$

where ϵ is a fixed positive threshold. Using $\overline{\mathbf{R}}=\Sigma_{k=1}^K\mathbf{R}_k$, the range of the averaged radiation matrix is given by[33]

$$\begin{aligned}\mathrm{range}\,\overline{\mathbf{R}}&=\mathrm{range}\sum_{k=1}^{K}\mathbf{R}_k\\&=\mathrm{range}\,\mathbf{R}_1+\mathrm{range}\,\mathbf{R}_2+\cdots+\mathrm{range}\,\mathbf{R}_K.\end{aligned} \tag{A2}$$

In view of the conjecture of Borgiotti and Jones:[5]

$$\mathrm{range}\,\mathbf{R}_k\subset\mathrm{range}\,\mathbf{R}_{k+1}(\mathrm{nesting\ \ property}), \tag{A3}$$

it is easily shown that Eq. (A2) leads to

$$\mathrm{range}\,\overline{\mathbf{R}}=\mathrm{range}\,\mathbf{R}_K, \tag{A4}$$

which is the desired result, because range $\mathbf{R}_K$ provides a complete set of basis functions for the reduction of sound radiation for all frequencies below ω_K. The above procedure also applies to an averaging procedure based on an integral over ω instead of a summation, which, in view of the integral of Eq. (12) and the identification $\hat{\mathbf{P}}(\omega)=\hat{\mathbf{R}}(\omega)$, demonstrates the completeness of the eigenvectors $\mathbf{E}$ obtained from Eq. (18). For many geometries the eigenvectors of $\mathbf{R}$ that are efficient radiators at low frequencies are efficient radiators also at high frequencies. Therefore these eigenvectors are weighted more heavily, i.e., have larger eigenvalues, than the eigenvectors of $\mathbf{R}$ that are efficient radiators mainly at high frequencies.

[1] G. V. Borgiotti, "The power radiated by a vibrating body in an acoustic fluid and its determination from boundary measurements," J. Acoust. Soc. Am. **88**, 1884–1893 (1990).
[2] D. M. Photiadis, "The relationship of singular value decomposition to wave-vector filtering in sound radiation problems," J. Acoust. Soc. Am. **88**, 1152–1159 (1990).
[3] A. Sarkissian, "Acoustic radiation from finite structures," J. Acoust. Soc. Am. **90**, 574–578 (1991).
[4] S. J. Elliott and M. E. Johnson, "Radiation modes and the active control of sound power," J. Acoust. Soc. Am. **94**, 2194–2204 (1993).
[5] G. V. Borgiotti and K. E. Jones, "Frequency independence property of radiation spatial filters," J. Acoust. Soc. Am. **96**, 3516–3524 (1994).
[6] K. A. Burgemeister, "Novel methods of transduction for active control of harmonic sound radiated from vibrating surfaces," Ph.D. thesis, University of Adelaide, 1996.
[7] B. S. Cazzolato and C. H. Hansen, "Structural radiation mode sensing for active control of sound radiation into enclosed spaces," J. Acoust. Soc. Am. **106**, 3732–3735 (1999).
[8] G. P. Gibbs, R. L. Clark, D. E. Cox, and J. S. Vipperman, "Radiation modal expansion: Application to active structural acoustic control," J. Acoust. Soc. Am. **107**, 332–339 (2000).
[9] A. P. Berkhoff, "Sensor scheme design for active structural acoustic control," J. Acoust. Soc. Am. **108**, 1037–1045 (2000).
[10] B. S. Cazzolato and C. H. Hansen, "Active control of sound transmission using structural error sensing," J. Acoust. Soc. Am. **104**, 2878–2889 (1998).
[11] A. Preumont, A. Francois, and S. Dubru, "Piezoelectric array sensing for real-time, broad-band sound radiation measurement," J. Vibr. Acoust. **121**, 446–452 (1999).
[12] G. P. Gibbs, K. W. Eure, and J. W. Lloyd, "Active control of turbulent boundary layer induced sound radiation from aircraft style panels," in *Proc. Active 99* (INCE, Fort Lauderdale, USA, 1999), pp. 837–848.
[13] G. P. Gibbs, R. H. Cabell, and J. Juang, "Controller complexity for active control of tbl induced sound radiation from panels," in *Proceedings of the 6th AIAA/CEAS Aeroacoustics Conference* (AIAA, Lahaina, Hawaii, 2000), pp. 1–11.
[14] C.-K. Lee and F. C. Moon, "Modal sensors/actuators," Trans. ASME **57**, 434–441 (1990).
[15] R. L. Clark, "Adaptive modal space control," J. Acoust. Soc. Am. **98**, 2639–2650 (1995).
[16] W. Gawronski, "Modal actuators and sensors," J. Sound Vib. **229**, 1013–1022 (2000).
[17] D. R. Morgan, "An adaptive modal-based active control system," J. Acoust. Soc. Am. **89**, 248–256 (1991).
[18] C. M. Heatwole, M. A. Franchek, and R. J. Bernhard, "Robust feedback control of flow-induced structural radiation of sound," J. Acoust. Soc. Am. **102**, 989–997 (1997).
[19] W. T. Baumann, F.-S. Ho, and H. H. Robertshaw, "Active structural acoustic control of broadband disturbances," J. Acoust. Soc. Am. **92**, 1998–2004 (1992).
[20] C. Maury, P. Gardonio, and S. J. Elliott, "Active control of the flow-induced noise transmitted through a panel," in Ref. 13.
[21] A. D. Pierce, *Acoustics—An Introduction to its Physical Principles and Applications*, 2nd ed. (Acoustical Society of America, Woodbury, NY, 1989).
[22] R. A. Roberts and C. T. Mullis, *Digital Signal Processing* (Addison-Wesley, Reading, MA, 1987).
[23] L. Ljung, *System Identification—Theory for the User* (Prentice–Hall, Englewood Cliffs, 1987).
[24] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control* (Wiley, Chichester, 1996).
[25] S. L. Marple, Jr., *Digital Spectral Analysis with Applications* (Prentice–Hall, Englewood Cliffs, 1987).
[26] D. C. Youla, "On the factorization of rational matrices," IRE Trans. Inf. Theory **IT-7**, 172–189 (1961).
[27] W. T. Baumann, W. R. Saunders, and H. H. Robertshaw, "Active suppression of acoustic radiation from impulsively excited structures," J. Acoust. Soc. Am. **90**, 3202–3208 (1991).

[28] A. P. Berkhoff, E. Sarajlic, B. S. Cazzolato, and C. H. Hansen, ''Inverse and reciprocity methods for experimental determination of radiation modes,'' in *Proceedings of the ICSV8*, edited by L. Cheng, K. M. Li, and R. M. C. So (Institute of Acoustics and Vibration, Hong Kong, 2001), pp. 1629–1636.
[29] J. Chun and T. Kailath, ''Generalized displacement structure for block-toeplitz, toeplitz-block, and toeplitz derived matrices,'' in *Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms* (Springer-Verlag, Berlin, 1991), pp. 215–236.
[30] E. A. Robinson, *Multichannel Time Series Analysis with Digital Computer Programs*, revised edition (Holden-Day, San Fransisco, 1978).
[31] A.-J. v.d. Veen, E. F. Deprettere, and A. L. Swindlehurst, ''Subspace based signal analysis using singular value decomposition,'' Proc. IEEE **81**, 1277–1308 (1993).
[32] R. L. Clark, ''Adaptive structures: compensators by design,'' in *Proceedings of the Active 99*, edited by S. Douglas (Institute of Noise Control Engineering of the USA, Fort Lauderdale, USA, 1999), pp. 63–72.
[33] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. (Johns Hopkins, Baltimore, 1989).
[34] D. A. Bies and C. H. Hansen, *Engineering Noise Control* (Unwin Hyman, London, 1988).
[35] A. P. Berkhoff, ''Active structural acoustic control using broadband radiation modes,'' in *Proc. Internoise 2001*, edited by R. Boone (Nederlands Akoestisch Genootschap, The Hague, The Netherlands, 2001), pp. 729–732.
[36] A. P. Berkhoff, ''Piezoelectric sensor configuration for active structural acoustic control,'' J. Sound Vib. **246**, 175–183 (2001).
[37] S. J. Elliott, T. J. Sutton, B. Rafaely, and M. E. Johnson, ''Design of feedback controllers using a feedforward approach,'' in *Proc. ACTIVE 95*, edited by S. D. Sommerfeldt and H. Hamada (INCE, Newport Beach, USA, 1995), pp. 863–874.
[38] C. R. Fuller, S. J. Elliott, and P. A. Nelson, *Active Control of Vibration* (Academic, London, 1996).
[39] S. J. Elliott, *Signal Processing for Active Control* (Academic, New York, 2001).
[40] T. C. Sors, ''Active structural acoustic control of sound transmission through a plate,'' Ph.D. thesis, University of Southampton, 2000.

Modeling of the human middle ear using the finite-element method

Takuji Koike and Hiroshi Wada
Department of Mechanical Engineering, Tohoku University, Aoba-yama 01, Sendai 980-8579, Japan

Toshimitsu Kobayashi
Department of Otolaryngology, Tohoku University School of Medicine, Sendai 980-0872, Japan

(Received 11 October 2000; revised 21 May 2001; accepted 19 December 2001)

In this study, a three-dimensional finite-element model (FEM) of the human middle ear was established, including features of the middle ear which were not considered in the previous model, i.e., the ligaments, tendons, I–S joint, loading of the cochlea, external auditory meatus (EAM), middle-ear cavities, etc. The unknown mechanical properties of these parts and the boundary conditions were determined so that the impedance obtained from the FEM analysis resembled the measurement values. The validity of this model was confirmed by comparing the motion of the tympanic membrane and ossicles obtained by this model with the measurement data, and the effects of the newly considered features on the numerically obtained results were examined. By taking the ligaments and tendons into account and assuming that the cochlea acts as a damper, with this model it was possible to realistically reproduce complex ossicular chain movement. It was found that the middle-ear cavities did not affect the vibration mode of the tympanic membrane. Although the EAM enhanced the sound pressure applied to the tympanic membrane compared with that at the entrance of the EAM, the pressure distribution on the surface of the tympanic membrane was not affected by the EAM. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1451073]

PACS numbers: 43.64.Bt, 43.64.Ha, 43.64.Yp [LHC]

I. INTRODUCTION

The finite-element method (FEM) is a powerful tool for analyzing middle-ear vibrations because the complicated shape of the middle ear can be modeled in detail. If a complete FEM model of the middle ear were constructed, spatial variations in displacement on the tympanic membrane, three-dimensional ossicular vibrations, and spatial pressure distributions in the middle-ear cavities and external auditory meatus could be clarified without direct measurements, which are difficult to perform. In addition, it would be possible to predict how the middle-ear function is affected by various kinds of middle-ear pathologies and to understand how individual differences in middle-ear structures affect that function.

A number of studies have investigated the vibration modes of the tympanic membrane and ossicles in humans (Williams and Lesser, 1990; Wada *et al.*, 1992; Beer *et al.*, 1999; Bornitz *et al.*, 1999; Eiber, 1999; Prendergast *et al.*, 1999; Ferris and Prendergast, 2000) and in animals (Funnell *et al.*, 1987; Ladak and Funnell, 1996) using the FEM. By these studies, it has been shown that the FEM is advantageous in the investigation of middle-ear motion. However, some features of the middle ear have not been considered in the models used in previous studies, e.g., complicatedly shaped middle-ear cavities and external auditory meatus. Furthermore, the validity of the system parameters and that of the boundary conditions have not been sufficiently confirmed.

In our previous study, a three-dimensional FEM model of the human middle ear was established (Wada *et al.*, 1992). In that model, the ossicular axis of rotation was assumed to be fixed and the loading of the cochlea on the stapes footplate was expressed by linear springs. However, Decraemer *et al.* (2000) subsequently reported that the instantaneous position of the rotation axis of the malleus varied at most frequencies within a cycle because the ossicles were supported by elastic ligaments and tendons. With regard to the loading of the cochlea, the cochlear impedance has been estimated to be damping dominant (Møller, 1965; Lynch *et al.*, 1982; Aritomo *et al.*, 1987; Zwislocki, 1965; Merchant *et al.*, 1996). Therefore, the numerically obtained vibration patterns of the ossicles in the previous study were different from those shown by the measurements, particularly at high frequencies.

In this study, the FEM model was modified to simulate real middle-ear vibrations. In order to reproduce the complex ossicular chain movement, the anterior malleal ligament, posterior incudal ligament, tensor tympani tendon, stapedial tendon, and stapedial annular ligament were modeled, and the loading of the cochlea on the stapes footplate was expressed by the damping. The external auditory meatus, tympanic cavity, aditus ad antrum, and tympanic antrum were also taken into consideration in this model in order to reproduce the real sound field. The unknown mechanical properties of these ligaments and tendons and the boundary conditions were determined by comparing the impedance of both the tympanic membrane and the stapes obtained by FEM analysis with those measured in fresh cadavers. To confirm the validity of this model, the motion of the tympanic membrane and ossicles obtained with this model was compared with that shown by the measurement data. Results indicated which of the features newly considered by this model were important for modeling the middle ear realistically.

 0001-4966/2002/111(3)/1306/12/$19.00 © 2002 Acoustical Society of America

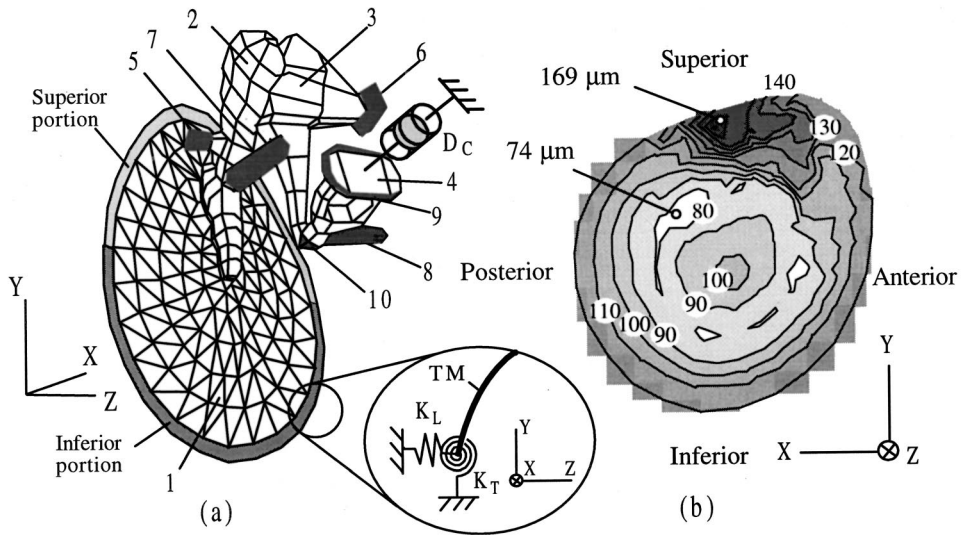


FIG. 1. FEM model of the human right middle ear. (a) Perspective view. 1 tympanic membrane; 2 malleus; 3 incus; 4 stapes; 5 anterior malleal ligament; 6 posterior incudal ligament; 7 tensor tympanic muscle; 8 stapedial muscle; 9 stapedial annular ligament; 10 I–S joint. Boundary conditions at the tympanic ring represented by linear springs, K_L, and torsional springs, K_T. The stiffness of these springs in the superior portion shown by light gray is assumed to be less than that in the inferior portion shown by dark gray. The cochlea is simply modeled as a damper, the coefficient of viscous damping of which is D_C. (b) Thickness distribution of the model of the tympanic membrane. Contours are drawn at intervals of 10 μm.

II. FINITE-ELEMENT MODEL

A. Improvements on the previous model

In the previous study, although the geometries of the tympanic membrane and ossicles were precisely reproduced based on the results of measurement (Kirikae, 1960; Kamifusa *et al.*, 1988), the ligaments and tendons which support the ossicles were ignored, and the axis of rotation of the ossicles was assumed to be a fixed straight line running from the anterior process of the malleus to the short process of the incus. The stiffness of the ossicular chain was simplified, namely, it was expressed by the stiffness of the cochlea which was represented by linear springs. Furthermore, the effect of the middle-ear cavities and that of the external auditory meatus were neglected in that model. The vibration modes of the tympanic membrane and ossicles at low frequencies resembled those shown by experimental results except for the relatively large displacement at the tympanic ring. However, the ossicular lever ratio was different from the measured one, probably due to the fixed rotational axis of the ossicles, the simplified stiffness of the ossicular chain, and the load of the cochlea. In this study, therefore, a finite-element program was developed, and the following features of the FEM model were newly considered.

(a) Modeling of the ligaments and tendons.
(b) Modeling of the cochlea as damping instead of stiffness.
(c) Modeling of the middle-ear cavities and the external auditory meatus.

Furthermore, the following features were improved so that the model more closely resembles the real middle ear.

(d) Thickness of the tympanic membrane.
(e) Boundary condition at the tympanic ring.
(f) Attachment between the tympanic membrane and malleus.
(g) Mobility of the incudo-stapedial (I–S) joint.

B. Geometry and mechanical properties

1. View of the middle-ear model and general parameters

Figure 1 shows a modified FEM model of the human right middle ear. As Funnell and Laszlo (1978) mentioned that it had not been necessary to introduce anisotropy to establish their middle-ear model, all parts of our model were also assumed to be isotropic. The Poisson's ratio was assumed to be 0.3. This value has been commonly used in structural analysis, and Funnell *et al.* also used it in their model.

The damping, except that of the cochlea, was expressed by the Rayleigh damping (Funnel *et al.*, 1987; Wada *et al.*, 1992). The system damping matrix C is given in terms of the mass and stiffness matrices as

$$C = \alpha M + \beta K, \tag{1}$$

where M and K are the system mass and stiffness matrices, respectively, and α and β are the damping parameters. The damping ratio ζ is given by

$$\zeta = (\alpha/\omega + \beta\omega)/2, \tag{2}$$

where ω is the angular frequency. Therefore, α and β control the damping at low and high frequencies, respectively.

The mechanical properties used in this model are summarized in Table I. Some values were based on previously reported measurement data, and the others were newly determined. Details of the model and the methods of determination of mechanical properties of each part are described separately in the following sections.

2. Tympanic membrane

Two hundred and thirty-two flat triangular plate elements were used to represent the tympanic membrane. The geometry of this model is the same as that of the previous one, but the distribution of the thickness of the tympanic membrane has been modified to more closely resemble that of the real tympanic membrane based on dimensions obtained from the study of Kamifusa *et al.* (1988). The Young's modulus and density of the tympanic membrane were based on the values obtained by Wada *et al.* (1990). In construction of the model of the tympanic membrane, both the pars tensa and pars flaccida were taken into consideration by assuming the Young's modulus of the pars flaccida to be 1/3 of that of the pars tensa (Lesser and Williams, 1988).

TABLE I. Mechanical properties applied to the FEM model.

Young's modulus (N/m^2)		
Tympanic membrane (pars tensa)[a]		3.34×10^7
Tympanic membrane (pars flaccida)		1.11×10^7
Anterior malleal lig.		2.1×10^7
Posterior incudal lig.		6.5×10^5
Tensor tympanic muscle		2.6×10^6
Stapedial annular lig.		4.9×10^5
Stapedial muscle		5.2×10^5
Incudostapedial joint		6.0×10^6
Ossicles[b]		1.2×10^{10}
Spring constant at the tympanic ring		
Linear spring (N/mm)		
Superior portion		3.0×10^3
Inferior portion		1.5×10^5
Torsional spring (N m/m)		
Superior portion		3.0×10^{-5}
Inferior portion		1.0×10^{-4}
Bulk modulus (N/m^2)		
Air		1.36×10^5
Density (kg/m^3)		
Tympanic membrane[a]		1.2×10^3
Ossicles[c]		$2.5-6.2\times10^3$
Ligaments and tendons		2.5×10^3
Air		1.18×10^3
Poisson's ratio		0.3
Damping parameters	α (s^{-1})	β (s)
Tympanic membrane	260	3.7×10^{-5}
Anterior malleal lig.	0.0	1.86×10^{-5}
Posterior incudal lig.	0.0	1.86×10^{-5}
Incudostapedial joint	0.0	5.0×10^{-4}
Other	0.0	3.72×10^{-4}
Cochlea	$Dc=8.91\times10^{-1}$ Ns/m	

[a]Values obtained by Wada *et al.* (1990).
[b]Value of human humerus obtained by Evans (1973).
[c]Values based on the measurement result by Kirikae (1960).

The boundary condition at the tympanic ring is represented by linear and torsional springs, the same as in the previous model. However, in the previous model, uniform springs were applied to the whole node at the periphery of the tympanic membrane. By contrast, in this model, due to the lack of the tympanic ring in the superior portion (Ars, 1989), the stiffness of the springs in the superior portion [shown by light gray in Fig. 1(a)] is assumed to be less than that in the inferior portion (shown by dark gray). These spring constants were determined using the same method as that described in the section ''Determination of unknown mechanical properties.''

3. Tympanic membrane–malleus attachment

At the umbo, the malleus handle is embraced by the fibrous layer of the tympanic membrane as shown in Fig. 2(a). In contrast, a fine filament of mucosa with a fibrous layer connects the tympanic membrane and malleus in the middle area of the malleus handle (Graham *et al.*, 1978). It is difficult to construct this fine structure by using the FEM because its representation requires many elements. Therefore, in the previous model, the malleus handle was rigidly attached to the tympanic membrane. In this model, assuming that the tympanic membrane–malleus attachment is intimate at the umbo and short process but tenuous in the middle of the malleus handle, the Young's modulus of the elements connecting the tympanic membrane with the tip of the malleus handle and the short process [shaded area in Fig. 2(b)] was considered to be equal to that of the tympanic membrane, and that of the elements connecting the tympanic membrane with the malleus between the umbo and the short

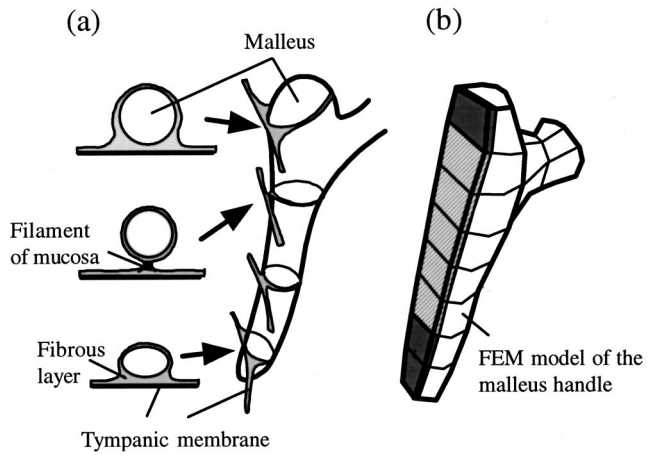


FIG. 2. Attachment of the malleus handle and tympanic membrane. (a) Cross section of the malleus handle. The gray area shows the fibrous layer of the tympanic membrane and mucosa. (b) FEM model of the malleus handle. The Young's modulus of the hatched elements is taken to be 1/1000 of that of the shaded elements because the state of the junction varies with the portion.

process [hatched area in Fig. 2(b)] was assumed to be 1/1000 of that of the tympanic membrane so that the vibration of the tympanic membrane was practically unrestricted by the malleus handle between the umbo and short process. This value was confirmed numerically to be the highest at which such an unrestricted vibration of the tympanic membrane could be realized.

4. Ossicles, ligaments, and tendons

Sixty-six hexahedral elements were applied to the ossicles. The Young's modulus and density of the ossicles were based on the values obtained by Kirikae (1960). In order to reproduce the complex movements of the ossicles, the anterior malleal ligament, posterior incudal ligament, tensor tympani tendon, stapedial tendon, and stapedial annular ligament were newly modeled using the hexahedral elements. The shapes of these ligaments and tendons were determined based on photographs of the middle-ear sections. The end points of the ligaments and tendons including the periphery of the stapedial annular ligament were completely fixed.

It has been reported that the malleus and incus vibrate as a solid body (Kirikae, 1960) and that the I–S joint is loose. In the previous model, this looseness of the I–S joint was not considered. In this model, therefore, the malleus and incus were rigidly connected, and a flexible I–S joint was realized by considering the Young's modulus of the I–S joint element to be smaller than that of the ossicles.

The mechanical properties of the ligaments, tendons and I–S joint were determined using the methods described in the section ''Determination of unknown mechanical properties.''

5. Damping of the cochlea

An experimental study of cochlea impedance in the cat (Møller, 1965) showed that the resistive component in the normal ear is derived mainly from the cochlea. Lynch *et al.* (1982) also reported that the input impedance of cats had a constant value of 200 GΩ (mks) and that its phase was zero at frequencies from 0.5 to 8.0 kHz. Aritomo *et al.* (1987) measured the cochlear impedance of the human temporal bone and certified that it was damping dominant at frequencies from 0.6 to 2.2 kHz. Moreover, Zwislocki (1965) reported that the human cochlear impedance was damping dominant and that its value was 35 GΩ. Merchant *et al.* (1996) also reported that the human cochlear impedance was damping dominant at frequencies between 0.1 and 5.0 kHz. They estimated that its value was approximately 50 GΩ in the frequency region between 0.1 and 2 kHz and increased with an increase in frequency up to 200 GΩ. In this study, the loading of the cochlea on the stapes footplate was expressed by the constant damping D_C, and its value was assumed to be 8.91×10^{-1} Ns/m so that the impedance of the cochlea becomes equivalent to 50 GΩ.

6. Middle-ear cavities and external auditory meatus

Figure 3 shows FEM models of the middle-ear cavities and external auditory meatus. Although the middle-ear cavities are complicated in structure, the tympanic cavity, aditus ad antrum, and tympanic antrum are simplified and expressed by combining rectangular solids. A large difference in volume of the middle-ear cavities exists between individual subjects, this volume varying from approximately 2000 to 22 000 mm^3 [Molvær *et al.* (1978)]. However, the volume of the tympanic cavity is within the range of about 500–1000 mm^3 (Voss *et al.*, 2000). In this model, the mastoid cells were ignored and the air volume of the tympanic cavity and that of the tympanic antrum were assumed to be 860 and 550 mm^3, respectively, based on the dimensions obtained by Kirikae (1960). The aditus ad antrum was modeled as a tube, and its length and cross-sectional area were taken to be 1.3 mm and 6.25 mm^2, respectively.

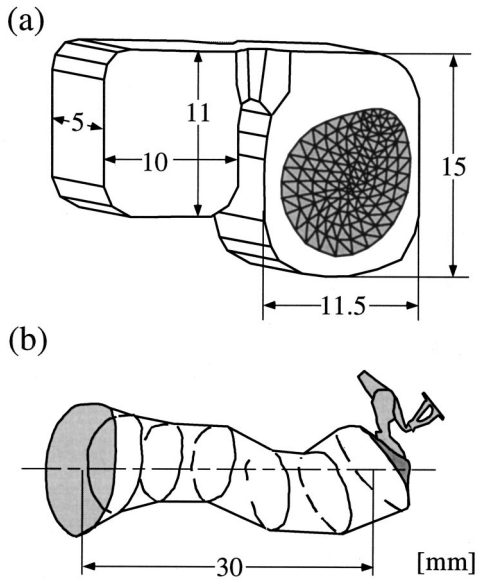


FIG. 3. FEM model of the middle-ear cavities and the external auditory meatus. (a) Tympanic cavity, aditus ad antrum, and tympanic antrum. (b) External auditory meatus.

The external auditory meatus (EAM) is expressed as a bent tube with rigid walls based on the dimensions obtained through CT scanning x-ray data (Egolf *et al.*, 1993). The length is considered to be 30 mm. The cross-sectional area in front of the tympanic membrane is 61 mm^2 and that at the midpoint of the EAM is 49 mm^2. The tympanic membrane forms an angle of 50 deg to the EAM.

These air parts, i.e., the tympanic cavity, aditus ad antrum, tympanic antrum, and EAM, were treated as elastic bodies without shear stiffness (Wilson *et al.*, 1983). The middle-ear cavities and the EAM were modeled using 956 and 696 hexahedral elements, respectively. As these elements are based on a pure displacement formulation, the displacement compatibility and equilibrium at the air–structure interface are automatically satisfied. At the boundaries between the air and the wall of the cavities or EAM, the displacement of air was restrained in the direction perpendicular to the surface of the walls and free in the direction parallel to them.

C. Determination of unknown mechanical properties

1. Tympanic ring, ligaments, tendons, and I–S joint

The spring constants at the tympanic ring and the mechanical properties of the ligaments, tendons, and I–S joint

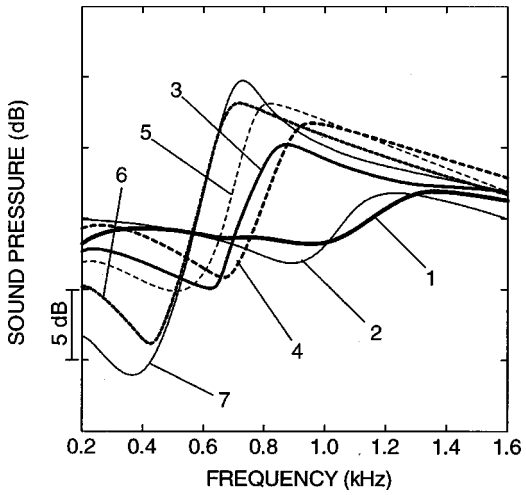


FIG. 4. Measurement results obtained from a manipulated temporal bone with a sweep frequency impedance meter (SFI). The vertical axis shows relative pressure change at the entrance of the external auditory meatus due to the impedance change of the middle ear, and the horizontal axis shows the frequency of the probe tone. 1 before manipulation; 2 after destruction of the cochlea; 3 I–S joint separation; 4 removal of incus; 5 cutting of the tensor tympani tendon; 6 cutting of the anterior malleal ligament; 7 malleus, incus, and stapes removed, tympanic membrane and tympanic ring intact.

are unknown. In this study, these mechanical properties, except those of the stapedial annular ligament, were determined by a comparison between the impedance changes at the tympanic membrane obtained from the FEM analysis and those of the manipulated temporal bones, which were obtained by sweep frequency impedance meter (SFI) (Wada *et al.*, 1990, 1998). Although a detailed description of the method of SFI measurement has already been given (Wada and Kobayashi, 1990), it will be briefly restated here. A sound of constant volume displacement between 0.1 and 2.0 kHz is presented to the EAM by an earphone installed in an SFI probe. The sound-pressure variation at the entrance of the EAM due to the impedance change of the tympanic membrane is detected by the probe microphone.

Manipulations of the temporal bone were performed in the following order. (1) The cochlea was destroyed. (2) The I–S joint was separated. (3) The incus was removed. (4) The tensor tympani tendon was cut. (5) The anterior malleal ligament was cut. (6) The malleus was removed from the tympanic membrane. After each manipulation, measurement was performed with the SFI.

Figure 4 shows the results obtained from a temporal bone with the SFI. The horizontal axis and the vertical axis indicate the probe frequency and the sound-pressure level in the EAM, respectively. Before manipulation, the sound-pressure curve (curve 1 in Fig. 4) varies considerably in the frequency region around 1.2 kHz. This frequency region is considered to be the resonance frequency region of the middle ear, and the pressure difference between the minimum of the curve (0.97 kHz) and its maximum (1.36 kHz) reflects the quantity of the volume displacement of the tympanic membrane at the resonant frequency, i.e., the mobility of the middle ear (Wada *et al.*, 1993, 1998). Therefore, differences among pressure curves 1–7 indicate how the resonance frequency and the mobility change due to each manipulation. As the resonance frequency is mainly related to the mass and stiffness components and the mobility is mainly related to the damping component, the difference between these curves reflects the difference in mass, stiffness, and damping components contributed by each manipulated part. Therefore, by constructing an FEM model of a manipulated middle ear at each stage and comparing the sound-pressure curves obtained from those models and those from the SFI measurement, the Young's modulus and damping parameters of the each part can be estimated.

As one example, the process of the determination of the Young's modulus of the tensor tympani tendon is shown here. The solid gray line in Fig. 5(c) shows the sound pressure curve obtained from a temporal bone, the incus of which had been removed, but the tensor tympani tendon of which had been left intact [see the illustration on the left in Fig. 5(a)], and the dashed gray line shows the sound-pressure curve after the tensor tympani tendon had been cut [see the illustration on the right in Fig. 5(a)]. These curves are the same as curves 4 and 5 in Fig. 4, respectively. Curve 1 shows the pressure generated at the entrance of the EAM of the FEM model excluding the incus and tensor tympani tendon [model on the right in Fig. 5(b)], when a constant displacement was applied to the node at the entrance of the EAM. Curves 2 and 3 show the pressure obtained from the FEM model excluding the incus but including the tensor tympani tendon [model on the left in Fig. 5(b)], when the Young's modulus of the tensor tympani tendon, E_{TT}, is 2.6×10^6 N/m^2 and 2.6×10^7 N/m^2, respectively. When E_{TT} is 2.6×10^6 N/m^2, the difference between the pressure curves obtained from the FEM models including and excluding the tensor tympani tendon is similar to that of the measurement. Therefore, the Young's modulus of the tensor tympani tendon was determined to be 2.6×10^6 N/m^2. In the same way, the spring constants at the tympanic ring and the Young's modulus and damping parameter of the ligaments, tendons, and I–S joint were determined as shown in Table I.

2. Stapedial annular ligament

In this model, the loading of the cochlea on the stapes footplate was expressed by damping instead of by stiffness which was assumed in the previous model, and the stiffness component loaded on the stapes was assumed to originate from the stapedial annular ligament. To determine the stiffness of the stapedial annular ligament, a stapes model consisting of only the stapes, stapedial annular ligament, and loading of the cochlea was established as shown in Fig. 6(a), and its acoustic impedance, Z_{SC}, was compared with the result of measurement. The volume displacement of the stapes footplate, Vs, was obtained using this stapes model when the constant pressure, P, was applied to the stapes footplate. Z_{SC} was calculated by

$$Z_{\mathrm{SC}}=P/2\pi fVs. \tag{3}$$

Figure 6(b) shows Z_{SC} obtained from the FEM analysis using a different Young's modulus of the stapedial annular ligament, E_{AL}, and that measured by Merchant *et al.* (1996). The value of Z_{SC} obtained from the measurement decreases with an increase in the frequency, having the minimum value

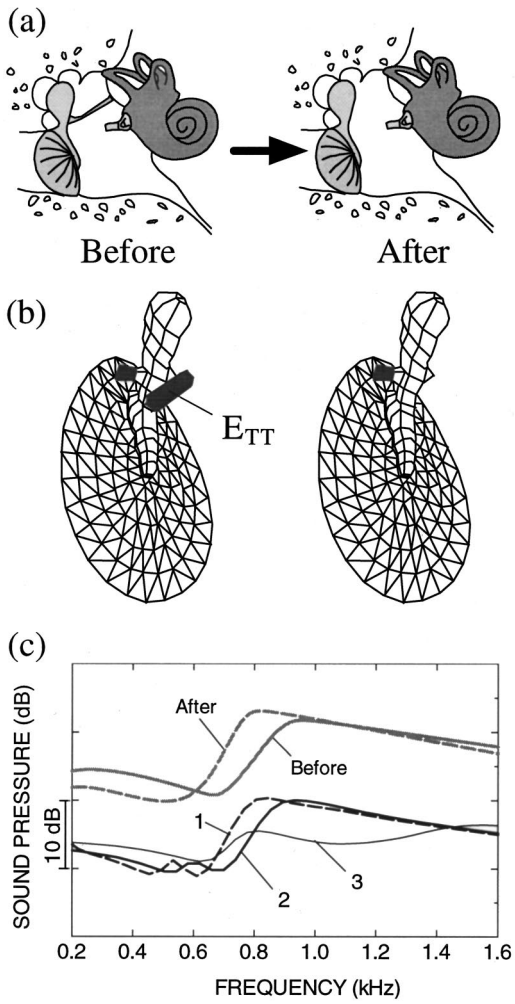


FIG. 5. Effect of the Young's modulus of the tensor tympani tendon, E_{TT}, on the sound-pressure curve. (a) An example of the manipulation. The tensor tympani tendon is cut at the portion which is attached to the malleus. (b) FEM models corresponding to each stage of the manipulation shown in (a). (c) Comparison of measured and calculated pressure curves. The gray lines show the pressure curves measured with the SFI before and after cutting of the tensor tympani tendon. 1 pressure curve obtained from the FEM model excluding the incus and tensor tympani tendon [model on the right in (b)]; 2 pressure curve obtained from the FEM model including the tensor tympani tendon [model on the left in (b)] when its Young's modulus, E_{TT}, is 2.6 $\times 10^6$ N/m^2; 3 pressure curve when $E_{TT}=2.6\times 10^7$ N/m^2.

around 3 kHz, and then increases up to a frequency around 7 kHz. The value of Z_{SC} and slope of the impedance curve obtained from the FEM analysis changed according to E_{AL}, especially at low frequencies. As the slope of the impedance curve is dominated by the stiffness of the stapedial annular ligament (Merchant *et al.*, 1996), the slope obtained from the FEM analysis has to be made to agree with that of the measurement by selecting an adequate E_{AL}. When $E_{AL}=4.9\times 10^5$ N/m^2, although Z_{SC} obtained from the FEM analysis is smaller than the measured value over the entire frequency region, its slope is similar to that of measurement at low frequencies. Therefore, E_{AL} is determined to be 4.9×10^5 N/m^2.

III. COMPARISON WITH EXPERIMENTAL OBSERVATIONS

To confirm the validity of the model of the middle ear modified in this study, the vibration mode of the tympanic membrane and ossicular motion obtained using this model were compared with the experimental data which have already been reported.

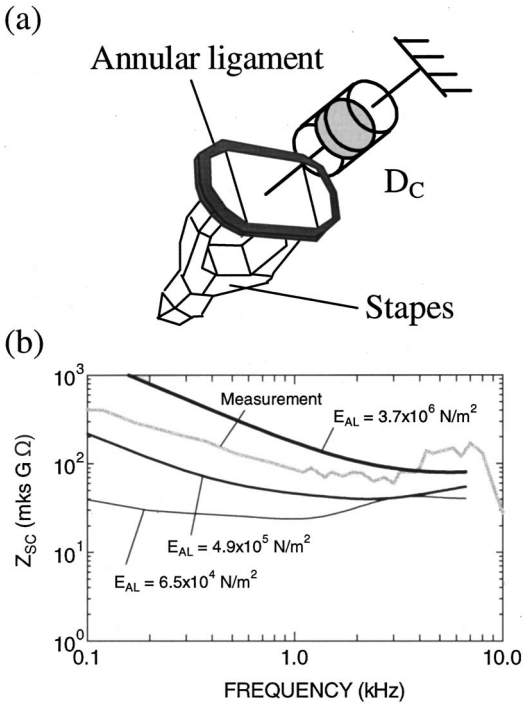


FIG. 6. Acoustic input impedance of the stapes and cochlea. (a) Stapes model. (b) Impedances of the stapes and cochlea obtained using the stapes model when three different Young's moduli of the stapedial annular ligament E_{AL} are used. Gray line shows the measurement result of Z_{SC} obtained by Merchant *et al.* (1996).

A. Vibration mode of the tympanic membrane

Figure 7 shows the numerically obtained vibration mode of the tympanic membrane and the one measured by Tonndorf and Khanna (1972) at low frequency. There are two maximal displacement amplitudes, one each in the anterior and posterior portions of the numerically obtained mode, and they are similar to those of the measured mode. Although the vibration mode was unclear, Tonndorf and Khanna reported that the vibration of the tympanic membrane started to become more complex above 3 kHz. Our results also showed a complicated mode above 3 kHz (the vibration modes obtained from the FEM analysis at high frequencies are shown in Fig. 12).

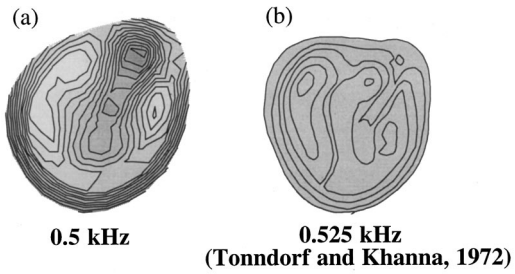


FIG. 7. Vibration modes of the tympanic membrane. (a) Numerical result. Frequency $f=0.5$ kHz. (b) Measurement result obtained by Tonndorf and Khanna (1972). $f=0.525$ kHz.

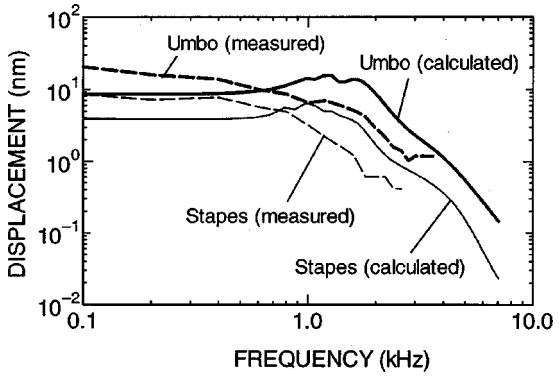


FIG. 8. Displacements of the umbo and stapes head. Sound pressure in front of the tympanic membrane is 80 dB SPL. The solid lines and dashed lines show the results obtained by FEM analysis and the measurement result obtained by Gyo *et al.* (1987), respectively. As the original measurement results are those when a sound pressure of 124 dB SPL was applied to the tympanic membrane, they are converted assuming the linearity of the middle-ear vibration.

B. Displacements of the umbo and stapes head

Figure 8 shows the numerically obtained displacements of the umbo and stapes head along the Z axis shown in Fig. 1 when a sound pressure of 80 dB SPL was applied to the tympanic membrane. Gyo *et al.* (1987) measured the displacements of the umbo and stapes head of the temporal bones using a video measurement system. Their results are also shown in Fig. 8 by the dashed lines. The numerically obtained displacement at the umbo was larger than that of the stapes head over the entire frequency region. Both displacements maintained almost constant values (8.7 and 3.8 nm, respectively) up to 0.7 kHz and had the maximum values at around 1 kHz. At frequencies over 1 kHz, this value decreased with an increase in the frequency. The displacements obtained using the present model were smaller than those obtained by the measurement at low frequencies. In addition, peaks around 1 kHz in the model were clearer than those in the measurement. These results mean that the total stiffness of the model was greater than that of the temporal bone used in the measurement and that the damping of the model was less than that in the temporal bone. However, Aritomo (1989) reported that the individual difference in displacement of both the umbo and stapes was larger at low frequencies (more than 2 times) and that some temporal bones showed a clear peak in amplitude at resonance frequency. Therefore, it was considered that the differences mentioned above were within the permissible range.

C. Vibration mode of the ossicles

Figure 9 depicts the vibration mode of the ossicles and loci of the tip of the malleus handle and the stapes head at 0.1, 2.0, and 4.0 kHz. At low frequencies around 0.1 kHz, the ossicles rotated around the fixed axis between the anterior malleal ligament and the posterior incudal ligament, and the tip of the malleus and stapes head had piston-like movements. With an increase in the frequency, the end of this axis of the malleus side shifted to the upper part of the malleus. In this state, neither end of the axis was fixed, and the axis moved within a cycle at this frequency as shown in Fig. 9(b). As a result, the tip of the malleus and stapes head had elliptical movements in addition to the piston-like movements. From 2.0 to 8.0 kHz, the axis of rotation shifted to the upper part of the ossicles, and the tip of the malleus and stapes head again manifested the piston-like movements.

Decraemer *et al.* (2000) reported that a fixed rotation axis was found only at low frequencies and that both the anterior process of the malleus and the posterior incudal process were not stationary at high frequencies in cats. Aritomo (1989) reported that the umbo and stapes head had piston-like movements at 0.1 kHz, which became elliptical with an increase in the frequency. Our results are in good agreement with these previous experimental results. However, at high frequencies, the axis of rotation is fixed again at the upper part of the ossicles in this FEM analysis, and this phenomenon was not detected by the measurements. Future experimental measurements could be designed to test the high-frequency axis of rotation in the human ear.

D. Lever ratio

Figure 10 shows the numerically obtained ossicular lever ratio using two different values of the cochlear damping, i.e., $D_C = 8.91 \times 10^{-1}$ and 0 N/m^2. In this study, the lever ratio $\mathrm{LR} = A_\mathrm{U}/A_\mathrm{S}$ was defined as the ratio of the displacement of the umbo, A_U, along the Z axis to that of the stapes head, A_S. The dashed lines in Fig. 10 show the lever ratio obtained by measuring the ossicular displacement before and after destruction of the cochlea (Gyo *et al.*, 1987).

When $D_\mathrm{C} = 8.91 \times 10^{-1}$ Ns/m, the numerically obtained lever ratio was nearly constant in the low-frequency region below 1.0 kHz. In the frequency region between 1.0 and 2.0 kHz, the lever ratio increased with an increase in the frequency and showed a maximum at 2.0 kHz. At frequencies higher than 2.0 kHz, the lever ratio decreased. By contrast, when $D_\mathrm{C} = 0$ Ns/m, the lever ratio was lower than that when $D_\mathrm{C} = 8.91 \times 10^{-1}$ Ns/m at high frequencies over 1.0 kHz, although both values were very similar below 1.0 kHz.

As depicted in Fig. 10, the ossicular lever ratio in the case of the intact cochlea had a maximum around 2.0 kHz. The movement of the rotation axis explains this phenomenon. As shown in Fig. 11, at low frequencies, the ossicles have rotation axis A, and the ratio of the distance between the rotational axis and the tip of the malleus handle L_AM to that between the same axis and the I–S joint L_{AI} is expressed by $R_A = L_{AM}/L_{AI}$. Around 2.0 kHz, the axis leans like axis B and the ratio is expressed by $R_B = L_{BM}/L_{BI}$. Figure 11 clearly shows that R_B is larger than R_A and leads to an explanation of why the ossicular lever ratio has a maximum around 2.0 kHz as shown in Fig. 10.

IV. EFFECT OF THE NEWLY CONSIDERED FEATURES ON THE RESULTS PREDICTED BY THE MODEL

In the present model, some features have been newly considered or improved compared with the previous model (Wada *et al.*, 1992) as explained in Sec. II. In this section, their effects on model predictions are summarized.

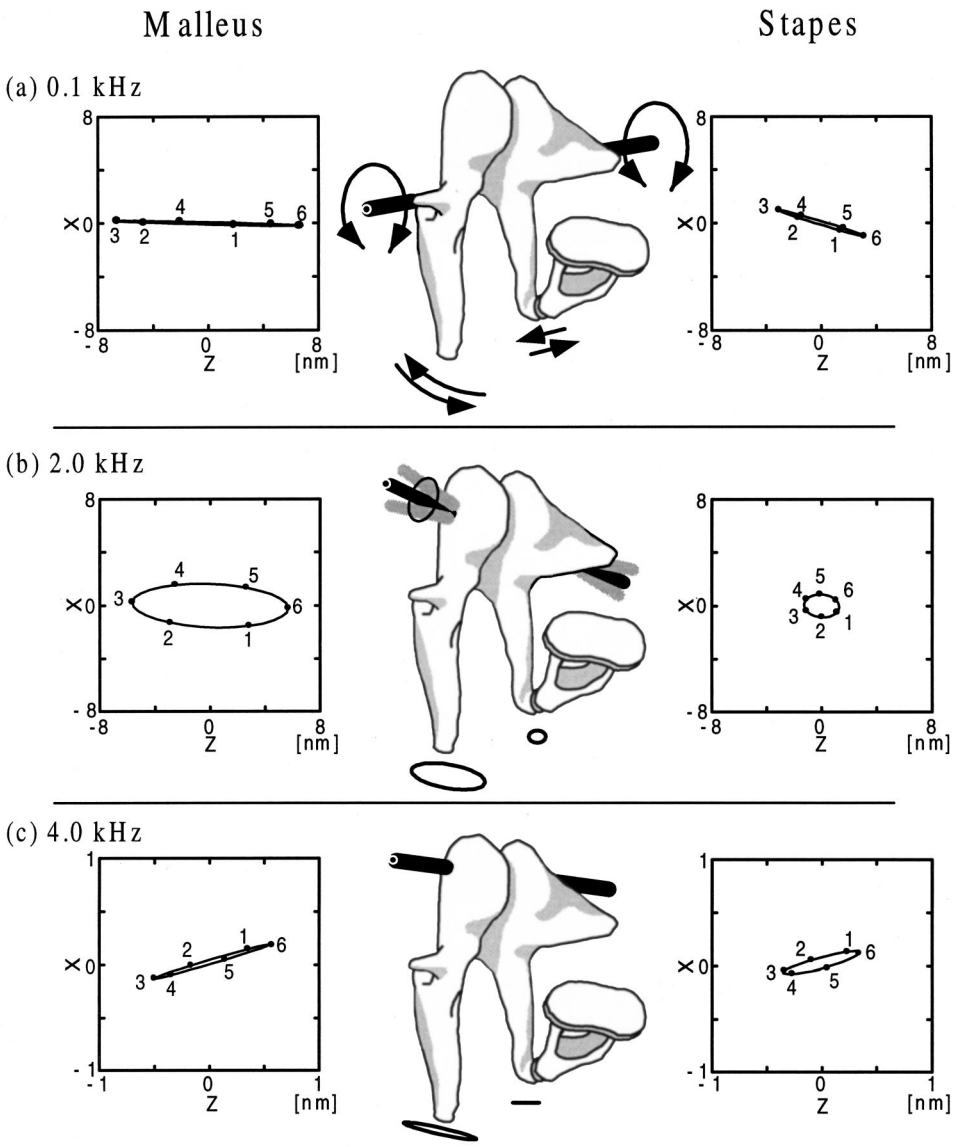


FIG. 9. Rotation axis of the ossicular chain. Loci of the tip of the malleus handle and the stapes head which are projected onto the X–Z plane, the axis directions of which are shown in Fig. 1, are depicted by ellipses. The origin of the coordinate axis of each graph is moved to the center of vibration of each part. Numbers on the loci represent the passage of time. (a) Frequency $f=0.1$ kHz. The ossicles rotate around the fixed axis between the anterior malleal ligament and the posterior incudal ligament. The tip of the malleus and stapes head have piston-like movements. (b) $f=2.0$ kHz. The end of the axis of the malleus side exists in the upper part of the malleus. Neither end of the axis is fixed, and the axis moves within a cycle. The tip of the malleus and the stapes head have elliptical movements. (c) $f=4.0$ kHz. A fixed axis exists at the upper part of the ossicles.

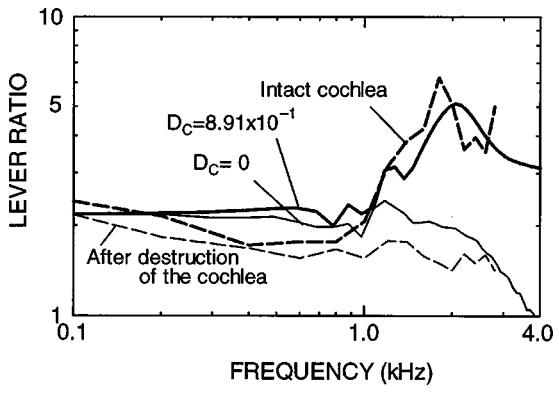


FIG. 10. Ossicular lever ratio versus frequency. The solid lines exhibit the numerical results obtained using two different values of the cochlear damping, D_C. The dashed lines demonstrate the measurement results obtained by Gyo *et al.* (1987) before and after destruction of the cochlea.

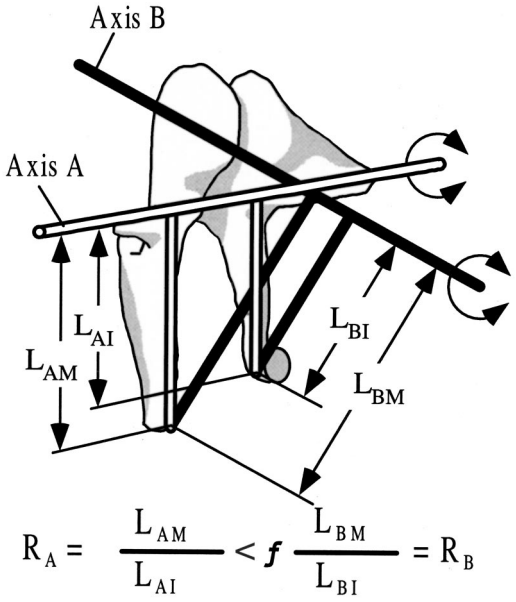


FIG. 11. Variation in the ossicular lever ratio caused by the change of the rotation axis.

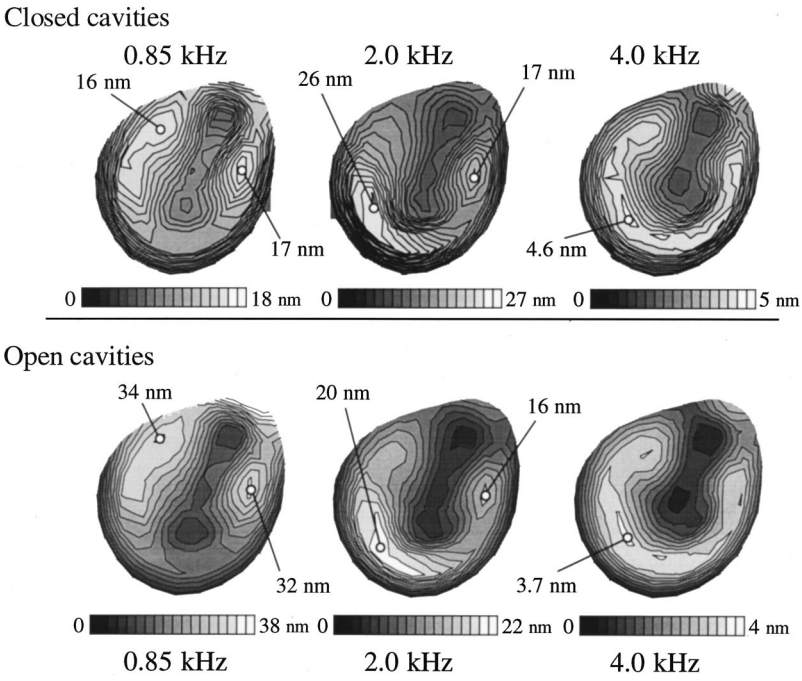


FIG. 12. Vibration modes of the tympanic membrane before and after opening of the middle-ear cavities. Sound pressure in front of the tympanic membrane is 80 dB SPL.

A. Modification of the tympanic membrane model

In the previous model, a relatively large displacement at the tympanic ring was shown, which was not in agreement with the measurement result of the vibration mode of the tympanic membrane by Tonndorf and Khanna (1972). This discrepancy was caused by the application of soft uniform springs to the periphery of the tympanic ring as a boundary condition. In the new model, as the spring constant in the inferior portion of the tympanic membrane was taken to be greater than that in the previous model, this unsuitable large displacement was reduced.

As shown in Fig. 7, not only the displacement at the tympanic ring, but also the overall vibration mode of the tympanic membrane became similar to the measurement by modifying the distribution of the thickness of the tympanic membrane to more closely resemble that of a real membrane. By contrast, although the displacement distribution around the malleus handle was slightly changed, the modification of the attachment between the malleus handle and the tympanic membrane did not greatly influence the vibration mode and displacement of the tympanic membrane.

B. Modeling of the ligaments, tendons, and I–S joint

In the previous model, as the ossicular axis of rotation was assumed to be fixed, the complex vibration modes of the ossicles shown in Fig. 9 were not reproduced. In the present model, by supporting the ossicles with the elastic ligaments and tendons and by assuming a loose I–S joint, the degree of restriction on ossicular motion was reduced, and although the validity of the vibration mode at high frequency has not been confirmed, it was possible to simulate realistic vibration of the ossicles. In particular, the stiffness of the I–S joint affected the middle-ear resonance frequency and ossicular lever ratio, and the stiffness of the anterior malleal ligament and the posterior incudal ligament affected the vibration mode of the ossicles. The complex movement of the ossicles was simplified when both stiffnesses were increased. Therefore, it is concluded that consideration of the elastic boundary condition of the ossicles is indispensable in the modeling of the middle ear.

C. Modeling of the cochlea as a damper instead of a spring

In our previous study (Wada *et al.*, 1992), the frequency characteristics of the numerically obtained ossicular lever ratio were different from those obtained by measurement of the temporal bone because the axis of rotation was assumed to be fixed and the loading of the cochlea on the stapes footplate was expressed by linear springs. In contrast, by modeling the ligament and tendons and assuming damping of the cochlea, the numerical and measurement results were in good agreement as shown in Fig. 10. Especially, the effect of the cochlear damping on the lever ratio at high frequencies was remarkable, and the lever ratio which increases with an increase in frequency could not be reproduced without assuming the cochlea to be a damper.

D. Effect of the middle-ear cavities on vibration of the tympanic membrane

In our previous model, the middle-ear cavities were not modeled. In this section, the effect of the middle-ear cavities on vibration of each portion of the tympanic membrane was examined. Figure 12 shows the vibration modes of the tympanic membrane viewed from the EAM before and after opening the middle-ear cavities. In this study, the vibration of the tympanic membrane after the cavities had been opened were obtained using a model which consists of the tympanic membrane, ossicles, ligaments, and tendons, but does not include the air elements. Before opening of the cavities, there were two maximal displacement amplitudes in the posterior and anterior portions at 0.85 kHz. The large-amplitude area in the posterior portion shifted to the inferior portion with an

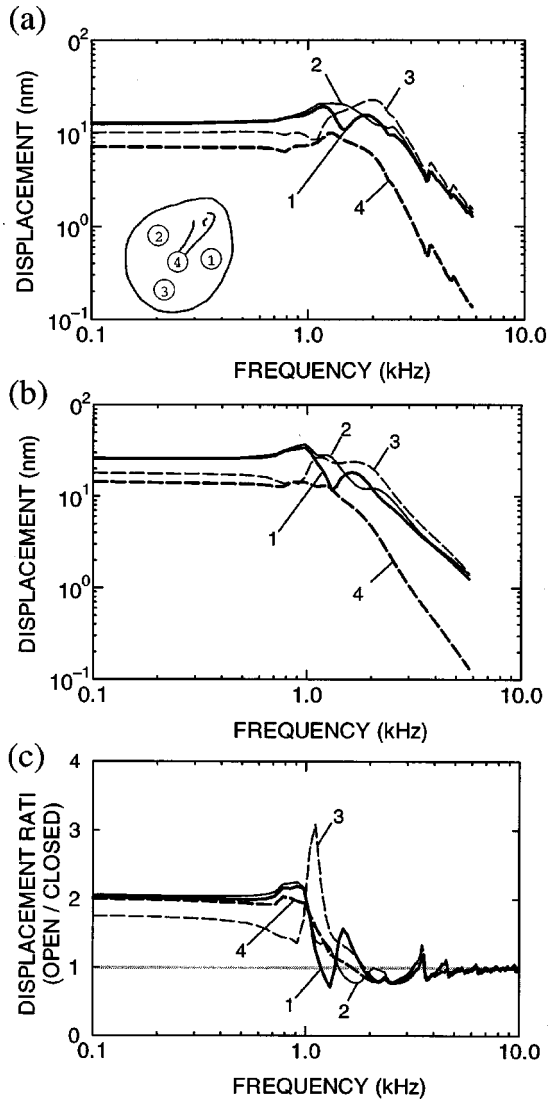


FIG. 13. Frequency response of the tympanic membrane vibration. Sound pressure in front of the tympanic membrane is 80 dB SPL. (a) Displacement amplitude when the middle-ear cavities are closed. 1 anterior portion; 2 posterior portion; 3 inferior portion; 4 umbo. (b) Displacement amplitude when the middle-ear cavities are open. (c) Displacement ratio of the displacements for the open and closed cavities.

increase in the stimulus frequency. At frequencies above 4 kHz, several small local maxima were seen at the pars tensa with complicated vibration modes. The vibration mode of the tympanic membrane after the cavities had been opened was similar to that before opening. However, a difference was found in the displacement of each portion of the tympanic membrane.

Figures 13(a) and (b) show the frequency responses of the displacement amplitudes at the anterior, posterior, and inferior portions of the tympanic membrane and umbo before and after the cavities had been opened, respectively, and Fig. 13(c) shows the ratio of the displacement of the open cavities to that of the closed cavities. When the cavities were closed, the posterior (curve 2) and inferior (curve 3) portions of the tympanic membrane showed a maximum displacement amplitude at around 1.2 and 2.0 kHz, respectively, and the anterior portion (curve 1) had two maxima at both frequencies. The displacement amplitude of the umbo (curve 4) had a

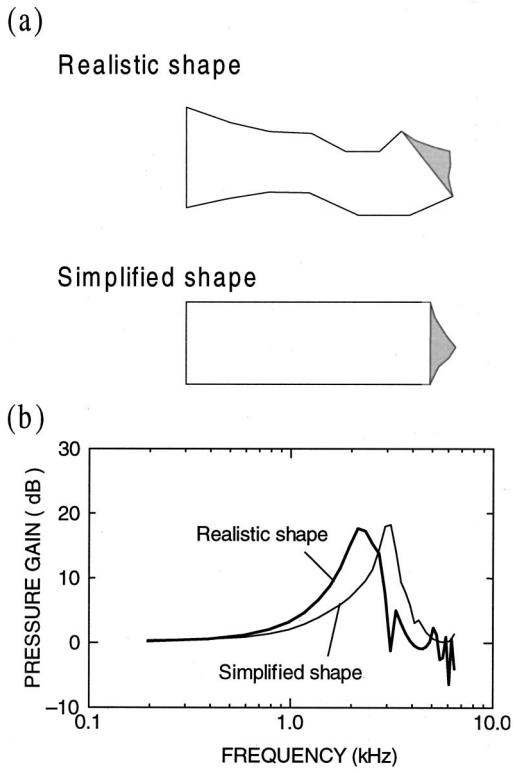


FIG. 14. Effect of the external auditory meatus on the pressure in front of the tympanic membrane. (a) Shape of the external auditory meatus. (b) Pressure gain caused by the external auditory meatus.

maximum at 1.2 kHz and was smaller than those of the other portions. When the cavities were open [Fig. 13(b)], although the general tendency in the frequency response of the displacement at each point was nearly the same as that when the cavities were closed, the peak of each response curve shifted to a lower frequency. As shown in Fig. 13(c), the displacement amplitudes of the tympanic membrane after the cavities had been opened were approximately two times larger than those before opening at frequencies below about 1.5 kHz. By contrast, the difference between them was small at high frequencies.

It has been reported that the effect of the middle-ear cavities on the vibration of the tympanic membrane is remarkable at low frequencies and that it behaves as a spring (Kirikae, 1960; Zwislocki, 1962; Voss *et al.*, 2000). In our results, the displacement of each part of the tympanic membrane was suppressed only at low frequencies when the middle-ear cavities were closed, because of the stiffness component contributed by the cavities to the tympanic membrane. However, the degree of the suppression at the inferior portion of the tympanic membrane was approximately 20 percent smaller than those at the other three portions [see Fig. 13(c)]. The reason why this difference occurs is unclear.

E. Pressure distribution in the external auditory meatus

The modeling of the complicatedly shaped EAM is also a new feature of this model compared with the previous one. In this section, therefore, its effect on the sound field in front of the tympanic membrane is examined. Figure 14 shows the

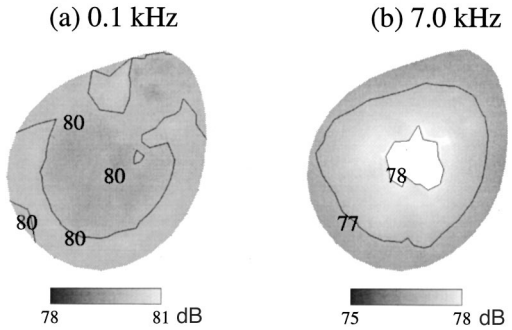


FIG. 15. Pressure distribution in front of the tympanic membrane. (a) $f = 0.1$ kHz. (b) $f = 7.0$ kHz.

numerically obtained ratio of the pressure at a distance of 2 mm from the umbo to that applied to the entrance of the EAM, which is complicated in shape as shown in Fig. 2. The ratio calculated using a simple-shaped EAM is also shown for comparison. The cross-sectional area of the simple-shaped EAM is constant and its length is the same as that of the complicatedly shaped EAM. The sound pressure at the tympanic membrane increased with an increase in frequency and had a peak around 2–3 kHz due to the resonance of the EAM.

The EAM is a kind of sound tube, and it has been reported that the resonance frequency of the EAM is different according to its length and that the pressure in front of the tympanic membrane becomes higher than that at the entrance of the EAM (Smith and Vernon, 1976). In addition, our results suggest that not only the length of the EAM but also the degree of complication in the shape of the EAM affects this pressure enhancement. As shown in Fig. 14, the resonance frequency of the EAM is affected by the change of the cross-sectional area, and the resonance of the complicatedly shaped EAM is lower than that of the straight one. The impedance of the air in the EAM mainly consists of a stiffness component, the same as in the case of the middle-ear cavity. However, at the narrow part of the EAM, it consists of not only the stiffness component but also a mass component. This mass component which exists in a realistically shaped EAM is thought to lower its resonance. Therefore, it is suggested that in case of an even more complicatedly shaped EAM, the relationship between the pressure at the entrance of the EAM and that in front of the tympanic membrane is more complicated. This matter should be taken into account when performing a hearing test or OAE measurement.

Figure 15 shows the pressure distribution in front of the tympanic membrane when a pressure of 80 dB is applied to the entrance of the EAM. The distribution of the pressure was almost uniform over the whole surface of the tympanic membrane at 0.1 kHz. Although the pressure at the center of the tympanic membrane was greater than that at the peripheral part at 7.0 kHz, the difference was within 2 dB. Therefore, the effect of the EAM on the pressure distribution in front of the tympanic membrane can be ignored in the frequency region below 7.0 kHz.

V. CONCLUSIONS

Our previously reported three-dimensional FEM model of a human middle ear was modified by including some new features, i.e., modeling of the ligaments, tendons, I–S joint, external auditory meatus, and middle-ear cavities, and by changing the distribution of the thickness and boundary condition of the tympanic membrane and the state of attachment of the malleus handle to the tympanic membrane. The unknown mechanical properties and the boundary condition of each part were determined by comparing the impedance of both the tympanic membrane and the stapes obtained by the FEM analysis with those measured in fresh cadavers. The effects of the newly considered features on the results predicted by the model were as follows.

(1) By modeling the ligaments, tendons, and I–S joint and by assuming the cochlea to be a damper, the new model presented in this report was able to reproduce complex ossicular chain movement and lever ratio. Therefore, the elastic condition at the ligaments, tendons, and I–S joint, and the damping of the cochlea are necessary in order to simulate the ossicular vibrations.
(2) The middle-ear cavities suppress the tympanic membrane vibration amplitude below 1.5 kHz. However, the overall vibration mode of the tympanic membrane was not remarkably changed before and after the middle-ear cavities had been opened.
(3) The frequency response of the ratio of the pressure in front of the tympanic membrane to that applied to the entrance of the external auditory meatus varies with its cross-sectional shape. However, the external auditory meatus does not affect the distribution of the pressure on the surface of the tympanic membrane.

The model herein presented is believed to have high validity and to be able to simulate the dynamic behavior of the middle ear realistically. This model, therefore, makes it possible to predict the middle-ear function in ears with various kinds of middle-ear pathologies.

ACKNOWLEDGMENTS

The authors acknowledge with thanks the various valuable comments of the reviewers. This work was supported by the Ministry of Education of Japan under Scientific Research Grant No. 09771327, as well as by grants from the Tateishi Science and Technology Foundation and Mikiya Science and Technology Foundation.

Aritomo, H., and Goode, R. L. (**1987**). ''Cochlear input impedance in fresh human temporal bones,'' 91st Meeting, American Academy of Otolaryngology, Chicago Research Forum.

Aritomo, H. (**1989**). ''Ossicular vibration in human temporal bones,'' J. Oto-Rhino-Laryngol. Soc. Japan **92**, 1359–1370.

Ars, B. (**1989**). ''Organogenesis of the middle ear structures,'' J. Laryngol. Otol. **103**, 16–21.

Beer, H. J., Bornitz, M., Hardtke, H. J., Schmidt, R., Hofmann, G., Vogel, U., Zahnert, T., and Hüttenbrink, K. B. (**1999**). ''Modelling of components of the human middle ear and simulation of their dynamic behavior,'' Audiol. Neuro-Otol. **4**, 156–162.

Bornitz, M., Zahnert, T., Hardtke, H. J., and Hüttenbrink, K. B. (**1999**). ''Identification of parameters for the middle ear model,'' Audiol. Neuro-Otol. **4**, 163–169.

Decraemer, W., and Khanna, S. (**2000**). ''New insights into vibration of the middle ear,'' in *The Function and Mechanics of Normal, Diseased and Reconstructed Middle Ears*, edited by J. J. Rosowski and S. N. Merchant (Kluwer, The Hague), p. 23.

Egolf, D. P., Nelson, D. K., Howell, H. C., and Larson, V. D. (**1993**). ''Quantifying ear-canal geometry with multiple computer-assisted tomographic scans,'' J. Acoust. Soc. Am. **93**, 2809–2819.
Eiber, A. (**1999**). ''Mechanical modeling and dynamical behavior of the human middle ear,'' Audiol. Neuro-Otol. **4**, 170–177.
Evans, F. G. (**1973**). *Mechanical Properties of Bone* (Thomas, Springfield, IL, 1973), p. 35.
Ferris, P., and Prendergast, P. J. (**2000**). ''Middle-ear dynamics before and after ossicular replacement,'' J. Biomech. **33**, 581–590.
Funnell, W. R. J., and Laszlo, C. A. (**1978**). ''Modeling of the cat eardrum as a thin shell using the finite-element method,'' J. Acoust. Soc. Am. **63**, 1461–1467.
Funnell, W. R. J., Decraemer, W. F., and Khanna, S. M. (**1987**). ''On the damped frequency response of a finite-element model of the cat eardrum,'' J. Acoust. Soc. Am. **81**, 1851–1859.
Graham, M. D., Reams, C., and Perkins, R. (**1978**). ''Human tympanic membrane–malleus attachment. Preliminary study,'' Ann. Otol. **87**, 426–431.
Gyo, K., Aritomo, H., and Goode, R. L. (**1987**). ''Measurement of the ossicular vibration ratio in human temporal bones by use of a video measuring system,'' Acta Otolaryngol. **103**, 87–95.
Kamifusa, K., Kodama, A., Oka, Y., and Ishii, T. (**1988**). ''Thickness of human normal ear drum,'' Ear Res. Jpn. **19**, 70–73.
Kirikae, I. (**1960**). *The Structure and Function of Middle Ear* (Tokyo University Press, Tokyo, 1960), p. 60.
Ladak, H. M., and Funnell, W. R. (**1996**). ''Finite-element modeling of the normal and surgically repaired cat middle ear,'' J. Acoust. Soc. Am. **100**, 933–944.
Lesser, T. H. J., and Williams, K. R. (**1988**). ''The tympanic membrane in cross section: A finite element analysis,'' J. Laryngol. Otol. **102**, 209–214.
Lynch III, T. J., Nedzelnitski, V., and Peake, W. T. (**1982**). ''Input impedance of the cochlea in cat,'' J. Acoust. Soc. Am. **72**, 108–130.
Møller, A. R. (**1965**). ''An experimental study of the middle ear and its transmission properties,'' Acta Otolaryngol. **60**, 129–149.
Merchant, S. N., Ravicz, M. E., and Rosowski, J. J. (**1996**). ''Acoustic input impedance of the stapes and cochlea in human temporal bones,'' Hear. Res. **97**, 30–45.
Molvær, O. I., Vallersnes, F. M., and Kringlebotn, M. (**1978**). ''The size of the middle ear and the mastoid air cell,'' Acta Otolaryngol **85**, 24–32.
Prendergast, P. J., Ferris, P., Rice, H. J., and Blayney, A. W. (**1999**). ''Vibro-acoustic modelling of the outer and middle ear using the finite-element method,'' Audiol. Neuro-Otol. **4**, 185–191.
Smith, C. A., and Vernon, J. (**1976**). *Handbook of Auditory and Vestibular Research Methods* (Thomas, Springfield, IL).
Tonndorf, J., and Khanna, S. M. (**1972**). ''Tympanic-membrane vibrations in human cadaver ears studied by time-averaged holography,'' J. Acoust. Soc. Am. **52**, 1221–1233.
Voss, S. E., Rosowski, J. J., Marchant, S. N., and Peake, W. T. (**2000**). ''Acoustic responses of the human middle ear,'' Hear. Res. **150**, 43–69.
Wada, H., and Kobayashi, T. (**1990**). ''Dynamical behavior of middle ear: Theoretical study corresponding to measurement results obtained by a newly developed measuring apparatus,'' J. Acoust. Soc. Am. **87**, 237–245.
Wada, H., Metoki, T., and Kobayashi, T. (**1992**). ''Analysis of dynamic behavior of human middle ear using a finite-element method,'' J. Acoust. Soc. Am. **92**, 3157–3168.
Wada, H., Ohyama, K., Kobayashi, T., Sunaga, N., and Koike, T. (**1993**). ''Relationship between evoked otoacoustic emissions and middle-ear dynamic characteristics,'' Audiology **32**, 282–292.
Wada, H., Koike, T., and Kobayashi, T. (**1998**). ''Clinical applicability of the sweep frequency measuring apparatus for diagnosis of middle ear diseases,'' Ear Hear. **19**, 240–249.
Williams, K. R., and Lesser, T. H. J. (**1990**). ''A finite element analysis of the natural frequencies of vibration of the human tympanic membrane. I,'' Br. J. Audiol. **24**, 319–327.
Wilson, E. L., and Khalvati, M. (**1983**). ''Finite elements for the dynamic analysis of fluid–solid system,'' Int. J. Numer. Methods Eng. **19**, 1657–1668.
Zwislocki, J. (**1962**). ''Analysis of the middle-ear function. I. Input impedance,'' J. Acoust. Soc. Am. **34**, 1514–1523.
Zwislocki, J. (**1965**). ''Analysis of some auditory characteristics,'' in *Handbook of Mathematical Psychology III*, edited by R. D. Luce, R. R. Bush, and E. Galanter (Wiley, New York), Chap. 15, p. 66.

Frequency specificity of chirp-evoked auditory brainstem responses

Oliver Wegner and Torsten Dau
Carl von Ossietzky Universität Oldenburg, Arbeitsgruppe Medizinische Physik, Graduiertenkolleg Neurosensorik, D-26111 Oldenburg, Germany

(Received 23 July 2001; accepted for publication 15 November 2001)

This study examines the usefulness of the upward chirp stimulus developed by Dau *et al.* [J. Acoust. Soc. Am. **107**, 1530–1540 (2000)] for retrieving frequency-specific information. The chirp was designed to produce simultaneous displacement maxima along the cochlear partition by compensating for frequency-dependent traveling-time differences. In the first experiment, auditory brainstem responses (ABR) elicited by the click and the broadband chirp were obtained in the presence of high-pass masking noise, with cutoff frequencies of 0.5, 1, 2, 4, and 8 kHz. Results revealed a larger wave-V amplitude for chirp than for click stimulation in all masking conditions. Wave-V amplitude for the chirp increased continuously with increasing high-pass cutoff frequency while it remains nearly constant for the click for cutoff frequencies greater than 1 kHz. The same two stimuli were tested in the presence of a notched-noise masker with one-octave wide spectral notches corresponding to the cutoff frequencies used in the first experiment. The recordings were compared with derived responses, calculated offline, from the high-pass masking conditions. No significant difference in response amplitude between click and chirp stimulation was found for the notched-noise responses as well as for the derived responses. In the second experiment, responses were obtained using narrow-band stimuli. A low-frequency chirp and a 250-Hz tone pulse with comparable duration and magnitude spectrum were used as stimuli. The narrow-band chirp elicited a larger response amplitude than the tone pulse at low and medium stimulation levels. Overall, the results of the present study further demonstrate the importance of considering peripheral processing for the formation of ABR. The chirp might be of particular interest for assessing low-frequency information. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1433805]

PACS numbers: 43.64.Qh, 43.64.Ri, 43.64.Bt [LHC]

I. INTRODUCTION

A number of direct and indirect approaches have been used for retrieving frequency-specific information from the auditory brainstem responses (ABR). These approaches include different stimulus paradigms as well as different signal processing techniques. Stimulation with filtered clicks or different tone pulses is normally used, and selective masking techniques are generally employed. A limiting factor for eliciting frequency-specific ABR in the frequency region below 2 kHz is related to cochlea mechanics and to the time-frequency uncertainty principle applied to the acoustic stimulus.

A straightforward approach to obtaining frequency-specific ABR has been the stimulation by brief tone pulses with a short rise and fall time (e.g., Kodera *et al.*, 1977; Suzuki *et al.*, 1977; Klein and Teas, 1978; Coats *et al.*, 1979; Purdy *et al.*, 1989; Conijn *et al.*, 1993; Beattie and Torre, 1997; Bunke *et al.*, 1998). As a compromise between frequency specificity and sufficient synchronization capability of the stimulus, Davis (1976) suggested the use of tone pulses with rise and fall times equal to two cycles of the stimulus frequency, and a plateau time equal to one cycle. It was found that high-frequency pulses (2 kHz and higher) elicit ABR which are similar to click-evoked responses (e.g., Terkildsen *et al.*, 1975; Gorga *et al.*, 1985; Laukli and Mair, 1986; Kileny, 1981; Conijn *et al.*, 1992b; van der Drift, 1987). Second, it has been shown that low-frequency pulses (below 2 kHz) of higher intensity, however, elicit ABR which include strong contributions originating from the more basal regions of the cochlea (e.g., Beattie and Kennedy, 1992; Gorga and Thornton, 1989). On the other hand, responses evoked by less intense low-frequency tone pulses are difficult to identify since the larger rise time required to obtain a sufficient narrow bandwidth of the acoustic stimulus is not effective in synchronizing neural discharges (Kramer and Teas, 1979; Laukli and Mair, 1986; Hoke *et al.*, 1991). Therefore, it was argued that ABR elicited by stimulation with brief tone pulses of frequencies below about 2 kHz are only poor predictors of low-frequency behavioral thresholds (Debruyene, 1982; Davis and Hirsh, 1976; Laukli, 1983a, b; Laukli *et al.*, 1988; Laukli and Mair, 1986; Scherg and Volk, 1983; Sohmer and Kinarti, 1984; Weber, 1987).

As a consequence, masking techniques have been suggested as an appropriate paradigm to obtain frequency-specific responses. The masker serves either to eliminate unwanted non-frequency-specific contributions to the ABR by selectively masking regions of the cochlea which are outside the region to be stimulated; e.g., by notched-noise masking or high-pass noise masking (Terkildsen *et al.*, 1975; Picton *et al.*, 1979; Stapells and Picton, 1981; Pratt and Bleich, 1982; Jacobson, 1983; Stapells *et al.*, 1990; Beattie and Kennedy, 1992; Beattie *et al.*, 1992; Conijn *et al.*, 1992a, b; Abdala and Folsom, 1995a, b; Oates and Stapells, 1997a). Alternatively, the neural activity in specified cochlea regions

 0001-4966/2002/111(3)/1318/12/$19.00 © 2002 Acoustical Society of America

can be selectively suppressed by computing off-line the difference waveform between the masked and unmasked responses, e.g., derived response technique (Don and Eggermont, 1978; Eggermont, 1976; Eggermont and Don, 1980; Kramer, 1992; Nousak and Stapells, 1992; Donaldson and Ruth, 1993; Don *et al.*, 1994, 1997; Oates and Stapells, 1997b) or pure-tone masking (Folsom, 1984, 1985; Pantev *et al.*, 1985; Klein, 1983; Klein and Mills, 1981a, b; Mackersie *et al.*, 1993; Wu and Stapells, 1994).

Using the high-pass noise masking derived ABR technique, Don *et al.* (1994) investigated the effect of the temporal variability in the neural conduction time and the effect of variability in the cochlear response times on wave-V amplitude of the compounded ABR to clicks. They adjusted for differences in neural conduction time (I-V delay) through compression or expansion of the derived response times and adjusted for differences in the cochlear response times through (individual) shifts of the derived ABR waveforms. Compensation for the I-V variability had little effect while compensation for cochlear response times greatly affected the amplitude of wave V of the compounded ABR. The study demonstrated the powerful influence of the temporal aspects of cochlear activation and response times on the component amplitude of the compounded ABR.

Recently, Dau *et al.* (2000) developed an upward chirp stimulus that theoretically produces simultaneous displacement maxima by canceling traveling-time differences along the cochlear partition. The equations determining the temporal course of the chirp were derived on the basis of a cochlea model (de Boer, 1980) and were calculated to be the inverse of the delay-line characteristic of the human cochlea partition. The fundamental relationship between stimulus frequency and place of maximum displacement was taken from Greenwood (1990). ABR evoked by the broadband chirp showed a larger wave-V amplitude than click-evoked responses. Dau *et al.* (2000) demonstrated that the ABR is not an electrophysiological event purely evoked by onset or offset of an acoustic stimulus, but that an appropriate temporal organization, determined by basilar-membrane traveling-wave properties, may significantly increase synchrony of neural discharges. The use of the upward broadband chirp enables the extension of activity to lower frequency regions whereas click synchrony is decreased in accordance with decreasing traveling velocity in the apical region of the cochlea.

The present article examines the usefulness of this chirp for estimating frequency-specific information. In the first experiment, ABR evoked by the broadband chirp in the presence of high-pass and notched-noise masking are compared with corresponding click-evoked responses for the same subjects. The second experiment investigates ABR obtained with narrow-band stimuli: Responses elicited by a low-frequency chirp are compared with tone-pulse evoked responses whereby the chirp and the tone pulse were designed to have similar duration and magnitude spectrum so that they mainly differ in their phase characteristic. The role of cochlear processing for brainstem responses and the possible application of the chirp for assessing low-frequency information are considered.

II. METHOD

A. Subjects

Nine normal-hearing subjects (one female and eight male) with audiometric thresholds of 15 dB HL or better and no history of hearing problems participated in the experiments. All subjects were between 24 and 36 years of age, and either volunteered or were paid for the experiment.

B. Apparatus

The experiments were carried out with a PC-based computer system which controlled stimulus presentation and recording of evoked potentials. A DSP-card (Ariel DSP32C) converted the digitally generated stimulus (25 kHz, 16 bit) to an analog waveform.

The masking noise in the two experiments was generated by feeding broadband white noise from a random noise generator (TDT WG2) to two cascaded filters (TDT PF1). The output of the second filter was attenuated (TDT PA4) and then added to the stimulus by a signal mixer (TDT SM3). The output of the signal mixer was connected to a digitally controlled audiometric amplifier, which presented the stimulus through an insert earphone (Etymotic Research ER-2) to the subject.

Electroencephalic activity was recorded from the scalp via silver/silver chloride electrodes, attached to the vertex (positive) and the ipsilateral mastoid (negative). The forehead served as the site for the ground electrode. Interelectrode impedance was maintained below 5 kΩ. Responses were amplified (80 dB) and filtered (95–1640 Hz, 6 dB/oct) with a commercially available ABR preamplifier (Hortmann Neurootometrie).[1] Extra amplification (Kemo VBF/40) was used to reach the optimum range for the A/D-converter. This amplification was in the range from 10 to 16 dB, resulting in a total amplification of 90–96 dB. The amplified signal was digitized by the DSP-card (25 kHz, 16 bit), which also performed artifact rejection and signal averaging. Responses were recorded for 40 ms following the stimulus onset.

C. Stimuli and procedure

All chirp stimuli used in the present study were generated on the basis of the equations described in Dau *et al.* (2000). In the first two experiments, a broadband chirp was used with a magnitude spectrum corresponding to that of the click. This chirp thus represents the ''flat-spectrum'' chirp as defined in Dau *et al.* (2000). Its nominal edge frequencies are 0.1 and 10.4 kHz, resulting in a duration of 10.48 ms. The chirp started and ended in zero phase and no windowing was applied. Figure 1 (upper panel) shows the digital waveform of the chirp (solid curve). The corresponding acoustic spectrum is given in the lower panel of the figure. Waveform and corresponding acoustic spectrum of the 80-μs click are indicated as dashed curves in Fig. 1. The spectra were obtained by coupling the ER-2 insert earphone to a Brüel and Kjær ear simulator (type 4157) with a $\frac{1}{2}$-in. condensor microphone

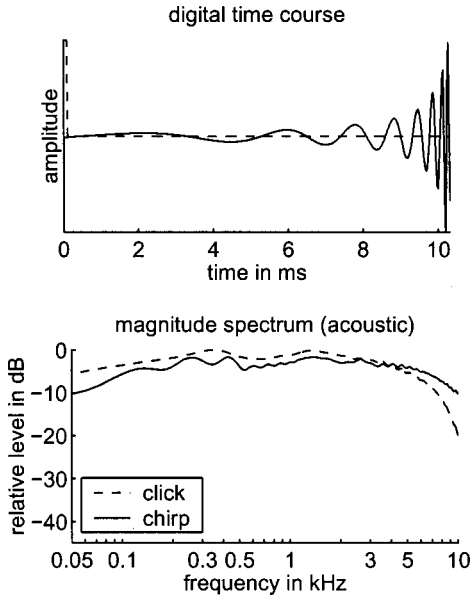


FIG. 1. Temporal course (upper panel) and acoustic spectra (lower panel) of the chirp (solid line) and the click (dashed line) used in experiment 1. The chirp was defined in Dau *et al.* (2000) as ''flat-spectrum chirp.'' Its acoustic spectrum is similar to that of the click stimulus.

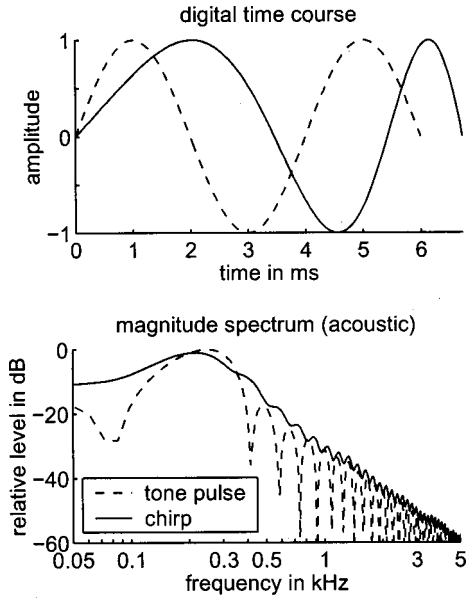


FIG. 2. Temporal course (upper panel) and acoustic spectra (lower panel) of the low-frequency stimuli used in experiment 2. The narrowband chirp (solid curve) and the 250-Hz tone pulse (dashed curve) were designed to exhibit three ''half-waves'' beginning and ending with zeros. The tone-pulse frequency was chosen in such a way that duration and magnitude spectrum were similar to those of the chirp.

(type 4134), a 2669 preamplifier, and a 2610 measuring amplifier. The spectra were derived from fast Fourier transforms (FFTs) of 100-trial time-domain averages of the stimulus over an analysis time of 64 ms using a sampling rate of 25 kHz (Stanford Research System SR780). The waveforms were not windowed prior to FFT. In the recordings, both stimuli were presented at a stimulation level of 40 dB HL, corresponding to a peak-equivalent sound pressure level of 87 dB for the click and 80 dB for the chirp stimulus. This 7-dB difference between click and chirp sound pressure level reflects temporal integration of signal energy involved in behavioral threshold measures that probably occurs at more central stages of auditory processing and is most likely not reflected in ABR. The level of only 40 dB HL was chosen for two reasons: (*i*) chirp and click still exhibit a clearly detectable wave V (Dau *et al.*, 2000) when presented without noise masking, and (*ii*) the overall level and loudness of the stimuli in combination with additional high-pass or notched-noise maskers remain within a comfortable range.

In the second part of the study, ABR were obtained using a low-frequency chirp with nominal edge frequencies of 100 and 480 Hz. Results were compared with ABR obtained with a 250-Hz tone pulse. Figure 2 shows the digital temporal course (upper panel) and the corresponding acoustic magnitude spectrum (lower panel) of the low-frequency chirp (solid curve) and the tone pulse (dashed curve). Both stimuli were designed to exhibit three ''half waves'' beginning and ending with zero. The tone-pulse frequency was chosen in such a way that duration and magnitude spectrum were similar to those of the low-frequency chirp. As can be seen in the lower panel of Fig. 2, the magnitude spectrum of the tone pulse is slightly narrower than that of the low-frequency chirp with the largest differences occurring at frequencies below about 100 Hz. The stimulation level was varied between 20 and 40 dB HL, in 5-dB steps. The peak-equivalent sound pressure level at hearing threshold (0 dB HL) was 40 dB for the tone pulse and 42 dB for the low-frequency chirp.

To determine the hearing level for the different stimuli, the absolute hearing thresholds were measured individually with an adaptive alternative forced choice (3AFC) procedure. The average over all subjects in the present study was considered as representing 0 dB HL.

The subject lay on a couch in an electrically shielded, soundproof room, and electrodes were attached. The subject was instructed to keep movement at a minimum, and to sleep if possible. The lights were turned out at the beginning of the session. Each session lasted between 1 and 2 h, depending on the subject's ability to remain still. The ear of stimulation was chosen randomly, i.e., for each subject one ear was chosen and then maintained. The acoustic signals were delivered at a repetition rate of 20 Hz for all stimulus conditions. A temporal jitter of ± 2 ms was introduced to minimize response superimposition from preceding stimuli. Thus the resulting interstimulus interval (ISI) was equally distributed between 48 and 52 ms. Each trial consisted of 3000 averages. For each stimulus condition, two independent trials were stored in separate buffers. These are illustrated as superimposed waveforms in the figures to show response replicability.

D. Experimental masking paradigms

Two different experimental masking paradigms were used in the first part of the study for retrieving frequency-specific activity: derived responses obtained with the procedure proposed by Don and Eggermont (1978) and responses obtained with the notched-noise masking method. Responses to the broadband chirp were compared with corresponding click-evoked responses. In the present study, white noise was

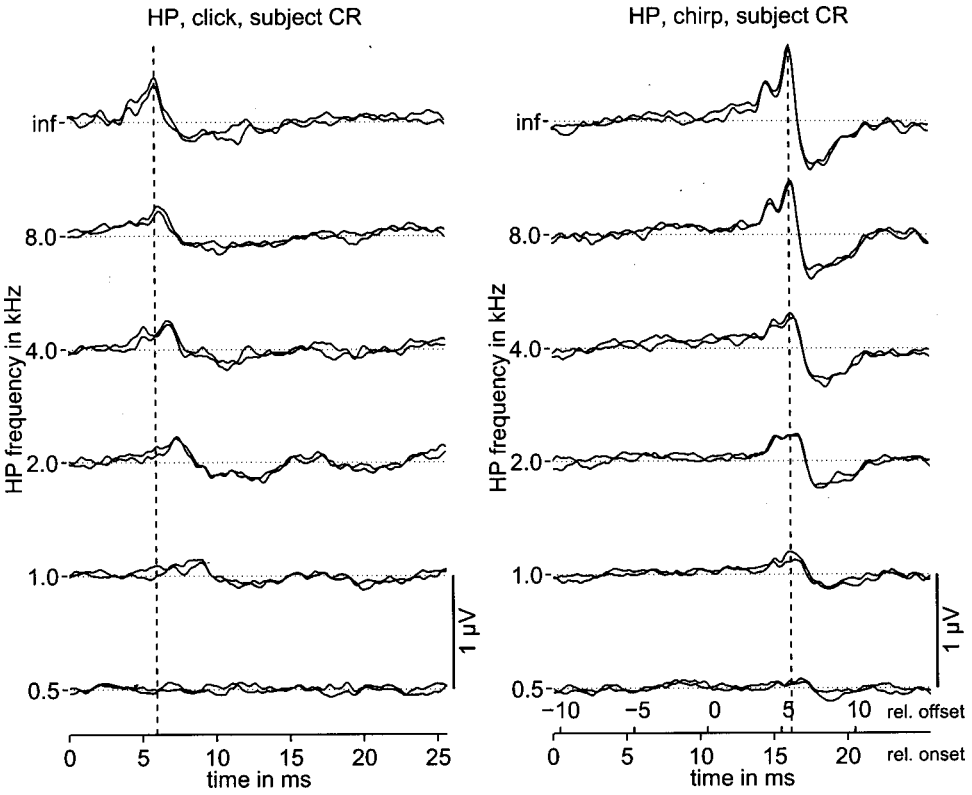


FIG. 3. Individual ABR evoked by the click (left panel) and the chirp (right panel) in the presence of high-pass noise masking. Parameter is the cutoff frequency of the noise. The unmasked responses (''inf'') are plotted on the top. To show response variability, the waveforms of the two independent buffers (3000 averages each) are shown on top of each other. The stimulus level was 40 dB HL. The stimulus presentation rate was 20/s. Subject: CR.

used as the masker which is different from the study of Don and Eggermont where pink noise was used.[2] In a first step, the level of unfiltered broadband noise which was sufficient to obliterate the brainstem response was determined and this ''masked'' activity was recorded (Purdy *et al.*, 1989; Conijn *et al.*, 1990, 1992a). This was done for a stimulus level of 40 dB HL. Without changing any attenuation levels, the brainstem responses were recorded with this noise filtered in the following way: In case of the derived-response method, the noise was high-pass filtered at values in the order 0.5, 1, 2, 4, and 8 kHz. Then, including the unmasked (no noise) and the completely masked responses, a total of seven recordings were obtained and stored. By successively subtracting response waveforms obtained in noise with high-pass cutoff frequencies separated by one octave, the narrow-band contributions to the ABR were obtained off-line, as suggested by Don and Eggermont (1978). In the case of the notched-noise method, the spectral notches of the noise were represented by the octave-wide regions at 0.5–1 kHz, 1–2 kHz, 2–4 kHz, and 4–8 kHz, respectively. These notches represent those spectral regions from which stimulus-evoked activity can effectively contribute to the recorded ABR. The spectrum levels of both high-pass noise and notched noise were nearly the same in the two experimental conditions (click: 32.6 dB, chirp: 32.5 dB).

In the second part of the study, responses evoked by the low-frequency chirp and the tone pulse were obtained. In order to ensure that neurons from more basal portions of the cochlea do not contribute to the evoked response, an additional set of recordings was obtained for the same stimuli with additional high-pass noise masking. The level of the unfiltered broadband noise which was sufficient to obliterate the brainstem response of the 40-dB HL signal was determined in a first step. Recordings were obtained using the noise high-pass filtered at 1 kHz without changing any attenuation levels. For the remaining signal levels it was assumed that the signal-to-noise ratio at ''masked threshold'' remains the same. Such a strategy has also been used by other investigators (e.g., Conijn *et al.*, 1990, 1992a).

E. Statistical analysis

Wave-V peak-to-peak amplitude was analyzed in all stimulus conditions. The amplitude was measured from the peak to the largest negativity following it. For each condition, wave-V amplitude was averaged across subjects. A Wilcoxon matched-pairs signed-rank test ($\alpha = 0.05$) was performed to verify whether the response amplitude differed significantly for the two comparison stimuli. Throughout the present article, responses are shown for sample subjects. Mean data for wave-V amplitude, averaged across the nine subjects, are summarized in additional figures.

III. RESULTS

A. Click- versus chirp-evoked responses using noise masking

The left panel of Fig. 3 shows, for subject CR, a series of brainstem responses to the click obtained in the presence of high-pass noise. As a reference, the unmasked response is shown as the top curve (indicated as ''inf'') in the figure. Wave-V is the only clear peak in the unmasked as well as in most of the noise masking conditions. However, wave-V amplitude is strongly reduced or absent for noise cutoff frequencies below 2 kHz. These observations are consistent with the results of Don and Eggermont (1978) apart from the fact that

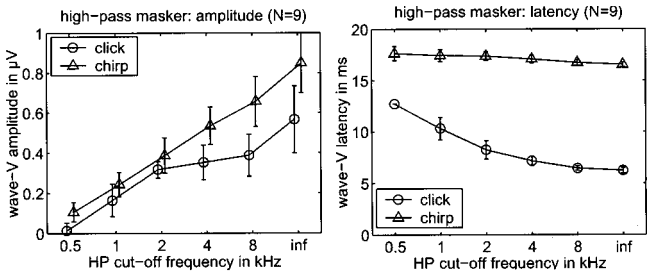


FIG. 4. Mean wave-V amplitude (left panel) and latency (right panel) for click- and chirp stimulation, as a function of the high-pass noise cutoff frequency (from Fig. 3). Error bars indicate the standard deviation.

their potential waveforms exhibited a larger wave-V amplitude and more distinct earlier waves I and III for the unmasked response caused by the higher stimulation level they used (60 dB SL instead of 40 dB in the present study). The dashed vertical line in Fig. 3 indicates wave-V latency for the unmasked response. As expected, there is a gradual increase in latency for the wave-V peaks as the high-pass cutoff frequency decreases, since the nonmasked contributions to the ABR are located in more and more apical portions of the cochlea (e.g., Laukli, 1983b; Burkard and Hecox, 1983; Don *et al.*, 1994).

The right panel of Fig. 3 shows the corresponding potential patterns evoked by the broadband chirp. Wave-V is clearly detectable in all conditions and larger than the click-evoked amplitude. The earlier waves (particularly wave III) are visible in the unmasked condition (top curve) and for the 8-kHz masking condition. Wave-V latency is roughly constant across all conditions. This latter finding was expected because of the specific phase characteristic of the chirp stimulus which was designed to compensate for travel-time differences on the BM (Dau *et al.*, 2000).

Figure 4 shows the mean results for click and chirp stimulation, averaged across the nine subjects. The left panel shows wave-V amplitude for the click (circles) and for the chirp (triangles), as a function of the cutoff frequency of the noise masker. For the click, wave-V amplitude increases with increasing cutoff frequency up to 2 kHz but tends to saturate for higher frequencies. In contrast, for the chirp wave-V amplitude increases continuously with increasing cutoff frequency. The chirp-evoked wave-V amplitude is significantly larger than the click-evoked one in all conditions ($N=9$, $\alpha=0.05$). The right panel of Fig. 4 shows the mean values for wave-V latency. It can be seen that for the click (circles) the latency decreases with increasing masker cutoff frequency while it remains roughly constant for the chirp (triangles).

Figure 5 shows the derived responses for subject CR, obtained by successively subtracting the responses from Fig. 3. The left panel shows results for the click and the right panel represents corresponding results for the chirp. Even though wave-V amplitudes of all responses from the high-pass masking conditions (from Fig. 3) were larger for the chirp than for the click, the amplitude of the largest component of the derived responses in Fig. 5 is similar for the two stimuli, or even larger for the click. The mean data for wave-V amplitude are shown in Fig. 7 (left panel). Statistical analysis revealed no significant difference between the response amplitudes obtained with click and chirp.

Finally, Fig. 6 shows click- and chirp-evoked responses obtained in the presence of a notched-noise (NN) masker. The response amplitudes are smaller than for the derived responses. This is most likely a consequence of the spread of excitation into the notch especially from the low-frequency noise band of the masker. The low-frequency part of the noise reduces the effective depth of the notch which is a much larger effect than the downward spread from the upper band (Picton *et al.*, 1979; Abdala and Folsom, 1995a; Beattie

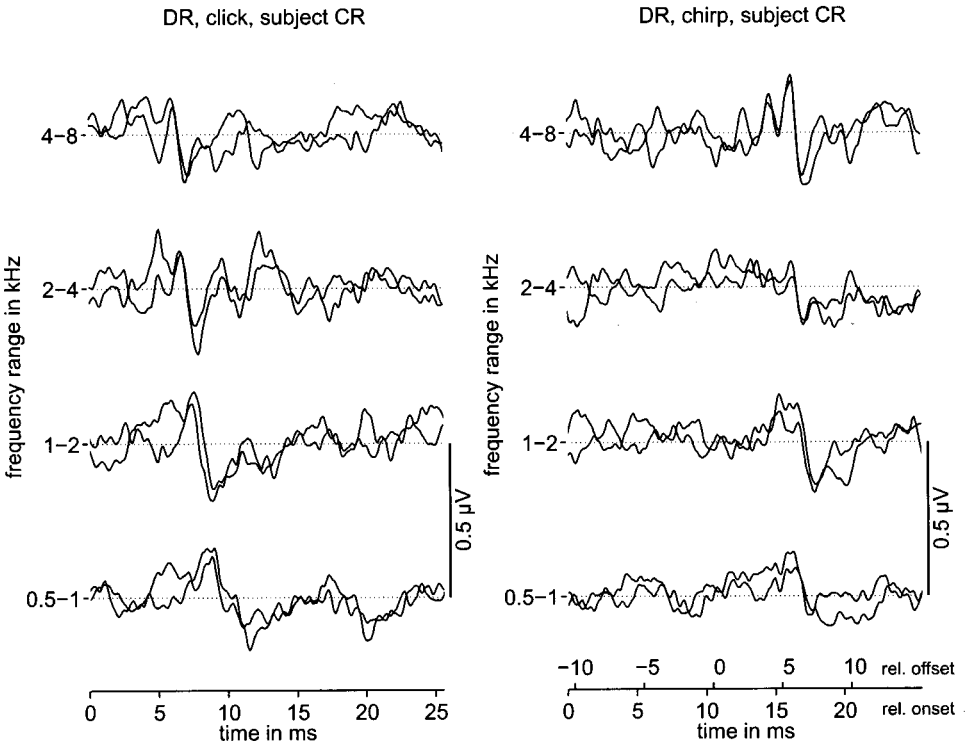


FIG. 5. Individual derived responses for click (left panel) and chirp stimulation (right panel). The responses were obtained by subtracting the high-pass masked ABR from the one shown immediately above it in Fig. 3.

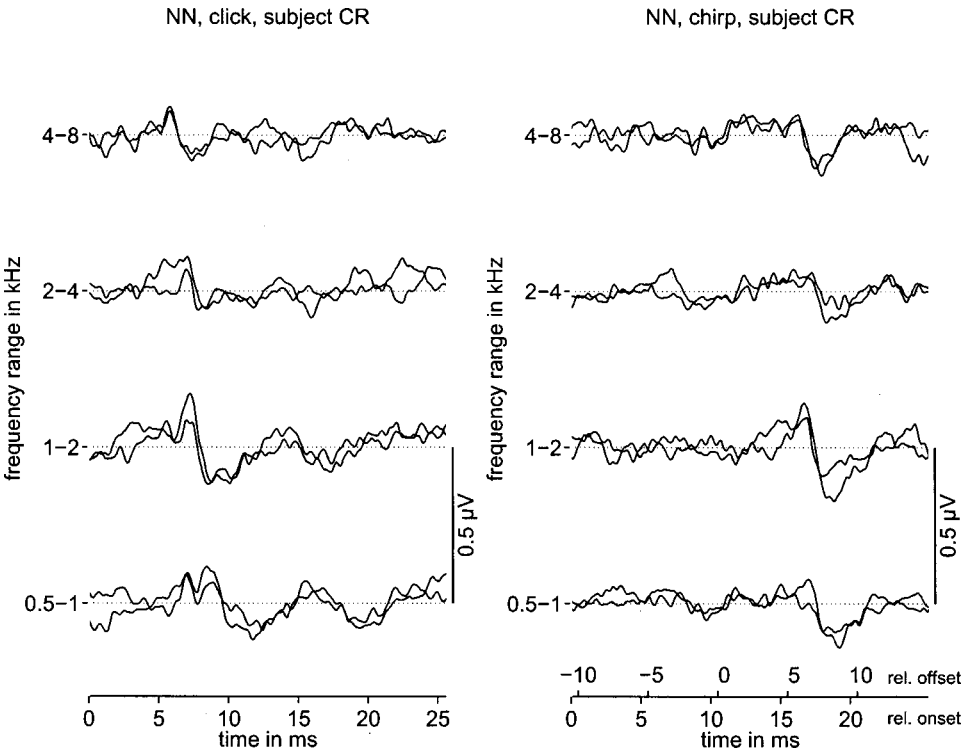


FIG. 6. Individual ABR to clicks (left panel) and chirps (right panel) in the presence of a notched-noise masker. The spectral notches of the noise were one octave wide in each condition.

et al., 1996). Wave-V amplitude is similar for the two stimuli in all NN conditions. The mean wave-V amplitude is plotted in the right panel of Fig. 7, as a function of the notch region of the noise masker. Wave-V amplitudes obtained with click and chirp stimulation do not differ significantly from one another.

In summary, larger response amplitudes were obtained for chirp than for click stimulation in the unmasked condition as well as in *all* high-pass masking conditions. However, no significant difference in response amplitude between click and chirp stimulation could be observed for the derived responses as well as for the notched-noise responses. This indicates that, even though neural synchrony is higher for the chirp than for the click, the contributions from one-octave wide frequency regions are not sufficient to obtain a significant difference in the far field.

B. Tone-pulse versus low-frequency chirp-evoked responses

A more direct approach to retrieve the frequency-specific ABR than the recording of click- or (broadband) chirp-evoked responses in combination with masking noise may be the use of brief tonal stimuli. In the present study, tone-pulse-evoked responses were compared with responses elicited by a low-frequency chirp (see Fig. 2).

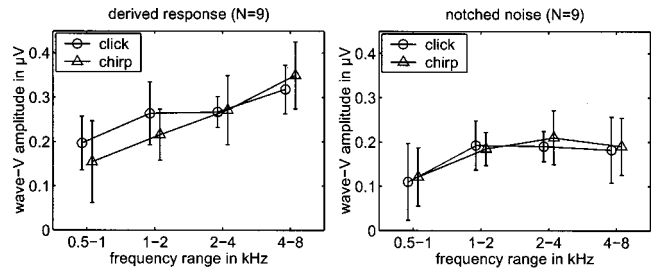


FIG. 7. Mean amplitudes for the click- and chirp-evoked derived responses (left panel), and for notched-noise masking responses (right panel) (from Figs. 5 and 6). Error bars indicate standard deviation.

Figure 8 shows individual results for the 250-Hz tone pulse (left panel) and the chirp (right panel). First, for all stimulation levels tested, the chirp elicits a larger wave-V amplitude than does the tone pulse. Second, wave-V latencies of the chirp-evoked responses are shifted by about 5 ms towards larger values relative to those obtained with the tone pulse. The responses for this particular subject have a larger amplitude than those for the other subjects but reflect the main trends observed in the mean data shown in Fig. 9. The left panel indicates wave-V amplitude as a function of the stimulation level. The chirp leads to a significantly larger wave-V amplitude than does the tone pulse, for the three lowest levels ($N=9$, $\alpha=0.05$). The right panel shows corresponding mean values for wave-V latency. For both stimuli, wave-V latency decreases with increasing level by about 2.6 ms, consistent with findings from other studies (e.g., Gorga *et al.*, 1988). More importantly, there is a nearly constant latency shift of 5.1 ms between tone-pulse and chirp stimulation. The reason for the latency difference is the same as for the observed difference in the responses to click and broadband chirp if plotted relative to the stimulus onset (Dau *et al.*, 2000): By successively stimulating lower and higher frequency components with a sweeping rate determined by cochlear travel-time properties, activity should be more dominated near the stimulus end for the chirp. In contrast, for the tone pulse, activity should be more dominated near the beginning of the stimulus. This will be discussed in more detail in Sec. IV.

In fact, because of the short duration of the stimuli and because no ramps were used, neurons tuned to medium

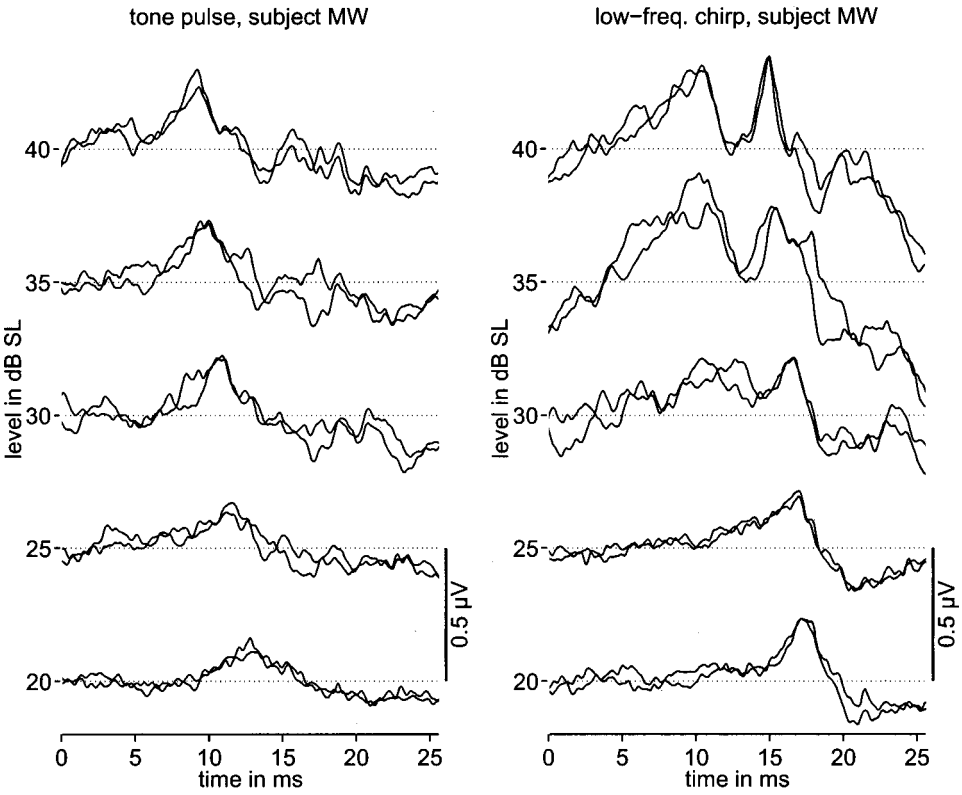


FIG. 8. Individual ABR to the 250-Hz tone pulse (left panel) and the low-frequency chirp (right panel). The stimulation level varied from 20 to 40 dB HL in 5-dB steps. Subject: MW.

and high frequencies may have also contributed to the response, particularly at the higher stimulation levels. For this reason, additional responses with a high-pass noise masker (at a cutoff frequency of 1 kHz) were recorded. Individual results are shown in Fig. 10, and mean data are represented in Fig. 11. The response amplitude is markedly reduced compared to the no-noise condition from the previous experiment, indicating that frequency components higher than about 1 kHz also contributed to the responses shown in Fig. 8. However, as for the unmasked conditions, statistical analysis revealed that for the three lowest stimulation levels, the chirp causes a significantly higher wave-V amplitude than does the tone pulse.

These results demonstrate that even a slight change in the phase characteristic of a stimulus can cause significant differences in the corresponding neural excitation. The low-frequency chirp leads to a higher synchronization of neural discharges which is also maintained at higher than peripheral stages of signal processing, and is reflected in a larger value of wave-V amplitude.

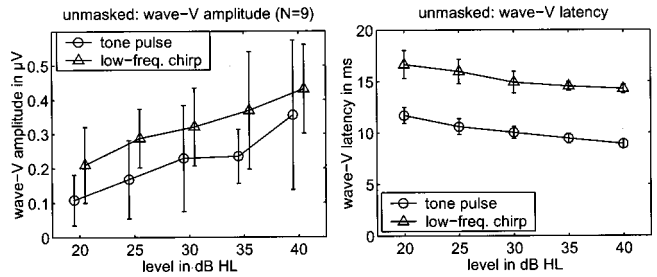


FIG. 9. Mean wave-V amplitude (left panel) and latency (right panel) for tone-pulse and low-frequency chirp stimulation. Error bars indicate standard deviation.

However, since the evoked potential amplitude is affected not only by the degree of synchronization of neural activity but also by the number of neural elements involved, one may argue that the differences in response amplitude are due to differences in the overall neural excitation of the two stimuli. As was shown in Fig. 2, tone pulse and low-frequency chirp differ somewhat in their magnitude spectrum, particularly at frequencies below about 100 Hz, which will lead to differences in their neural excitation patterns. This will be discussed in Sec. IV B.

IV. DISCUSSION

A. Broadband chirp versus click

In Dau *et al.* (2000) it was shown that, without any masking noise, the flat-spectrum chirp elicits a larger wave-V amplitude than the click for a large range of stimulation levels. In the present study, only one signal level (40 dB HL) was used. The results for the no-noise condition correspond to those in Dau *et al.* (2000). In order to investigate frequency-specific differences of ABR obtained with click and chirp stimulation, the stimuli of the present study were presented in combination with masking noise. First, since the chirp stimulus has been designed to minimize phase canceling effects of evoked activity, the chirp responses in all high-pass noise masking conditions are of larger amplitude than the click responses where some of the activity is still phase canceled. The finding that the chirp causes a significantly larger response even at the lowest cutoff frequency of 0.5 kHz makes the stimulus interesting for clinical use in assessing low-frequency information. Second, it was observed that for the chirp, wave-V amplitude increases with increasing noise cutoff frequency while for the click it tends to saturate

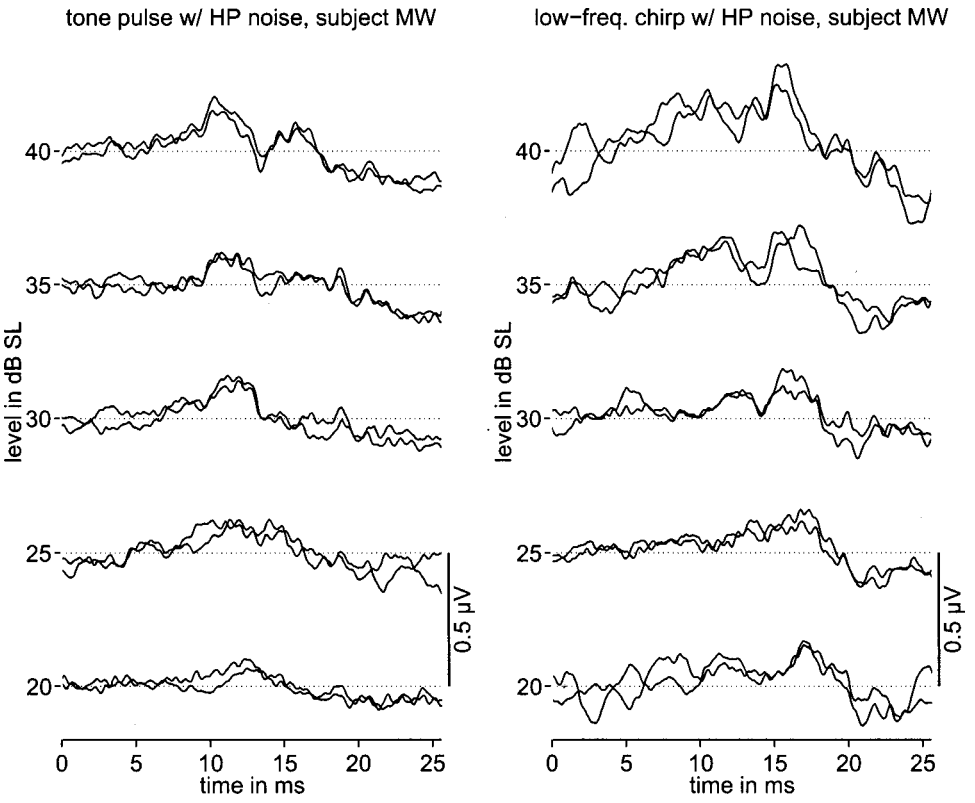


FIG. 10. Individual ABR to the 250-Hz tone-pulse and the low-frequency chirp. Stimulation as in Fig. 8 but with additional high-pass noise ($f_c = 1$ kHz) which masks contributions from high-frequency portions of the basilar membrane.

for frequency regions above about 1 kHz. This further confirms that, for the chirp, neural activity from the entire frequency range contributes to the evoked response while, for the click, neural activity is less synchronized across frequencies because of the travel-time differences on the basilar membrane.

To illustrate the effects of peripheral processing associated with cochlear response times, Fig. 12 shows simulated neural activity patterns for click stimulation (top panel) and for broadband chirp stimulation (bottom panel). After middle ear filtering, approximated by a second-order band-pass filter with cutoff frequencies of 0.3 and 7 kHz, the stimuli served as input to the recently developed auditory-nerve (AN) model by Heinz *et al.* (2001). The model is a modification of the physiologically based AN model by Zhang *et al.* (2001) which was developed for the cat and is itself an extention of the original Carney (1993) model. The model by Heinz *et al.* (2001) uses a human cochlear map according to Greenwood (1990), and the auditory filter bandwidth has been matched to humans based on psychophysical estimates of auditory filters (Glasberg and Moore, 1990). This model includes effects of level-dependent tuning, level-dependent phase, compression, suppression, and fast nonlinear dynamics on the response. The model is specifically designed to describe the time-varying discharge rate of the AN fibers for a given char-

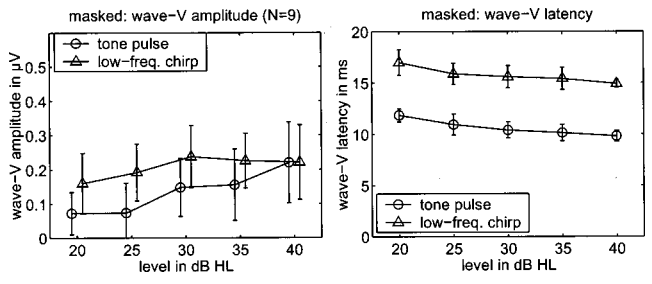


FIG. 11. Mean wave-V amplitudes (left panel) and latencies (right panel) for the conditions from Fig. 10. Error bars indicate standard deviation.

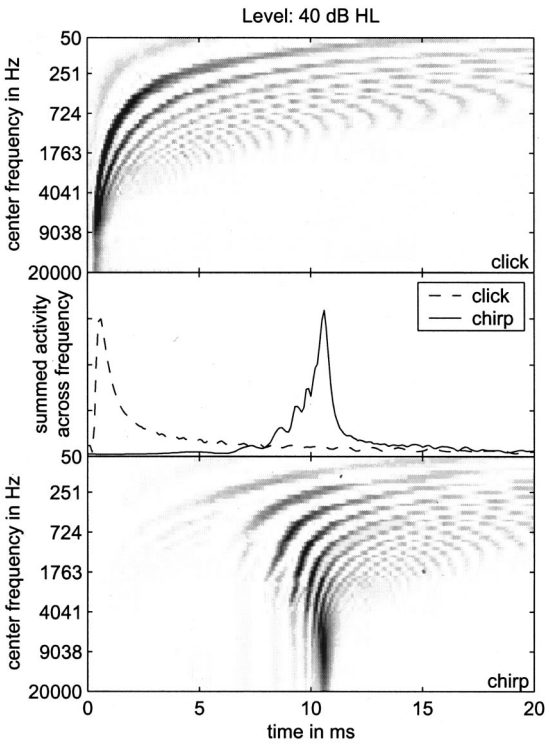


FIG. 12. Neural activity patterns for click (top panel) and chirp stimulations (bottom panel), obtained with the AN model of Heinz *et al.* (2001). The middle panel represents the summed neural activity across the frequency channels.

acteristic frequency (CF). A set of 60 model CFs was used in the present study. The CFs ranged from 0.1 to 10 kHz and were spaced according to a human cochlear map (Greenwood, 1990). Details of the processing stages can be found in Heinz *et al.* (2001).

A stimulus level of 40 dB HL was used for the following simulation since this level was also used in the experiments. It can be seen that for the click (top panel) the maxima of the neural activity are delayed at the lower frequencies relative to those at higher frequencies, while for the chirp (bottom panel) the maxima of neural excitation are more synchronized across frequency. Note that the simulation is based on a functional basilar-membrane model assuming Gammatone filters that treat the cochlea essentially as a bank of bandpass filters and ignore cochlear hydrodynamics while the chirp was developed based on spatial variations in cochlear geometry and mechanics that underlie the frequency-position map. Nevertheless, the resulting delay-line characteristic obtained with the different modeling concepts is similar. Within the framework of the Heinz *et al.* model, the latencies directly result from the filter bandwidths assumed. The middle panel of Fig. 12 shows the summed activity for the click (dashed curve) and for the chirp (solid curve), calculated by simply adding up the activities of all 60 frequency channels. The peak activity for the chirp is shifted (by about the duration of the chirp) relative to that of the click, and has a slightly larger amplitude. This agrees qualitatively with experimentally obtained compound action potentials (CAPs) for click and chirp stimulation in the guinea pig (Shore and Nuttall, 1985).

As was shown in the left panel of Fig. 7, if contributions from one-octave-wide frequency regions are investigated, there is no significant difference in efficiency anymore between chirp and click stimulation. Derived responses were found to be of essentially the same amplitude for the two stimuli. Similar results were obtained with the notched-noise masker (right panel of Fig. 7). Apparently, the better synchronization of neural activity obtained with the chirp is *not* sufficient to produce an advantage if only one-octave-wide frequency regions (or less) contribute to the response.

However, in all conditions tested in the present study, the broadband chirp was found to be as, or more effective than, the traditional click stimulus. In the conditions of high-pass masking alone the chirp always elicited a higher response than the click. The chirp may therefore be considered as the stimulus of choice for retrieving frequency-specific information as long as a broadband stimulus in combination with masking noise is considered.

B. Low-frequency chirp versus tone pulse

The results obtained with the narrow-band stimuli showed that a slight change in the phase characteristic can lead to large differences in the evoked ABR. Both the low-frequency chirp and the tone pulse consisted of only three half-waves and no ramps were used to introduce a smooth on- and offset. Thus, these stimuli exhibit a broadened magnitude spectrum compared to, e.g., tones of longer duration. It is important to note that the goal of the present study was not to develop an optimal *narrow-band* stimulus for retrieving low-frequency information from ABR. It is still controversial what narrow-band stimulus might represent the best choice. Instead, the present study attempts to emphasize the role of cochlear processing for the formation of ABR, particularly at low frequencies.

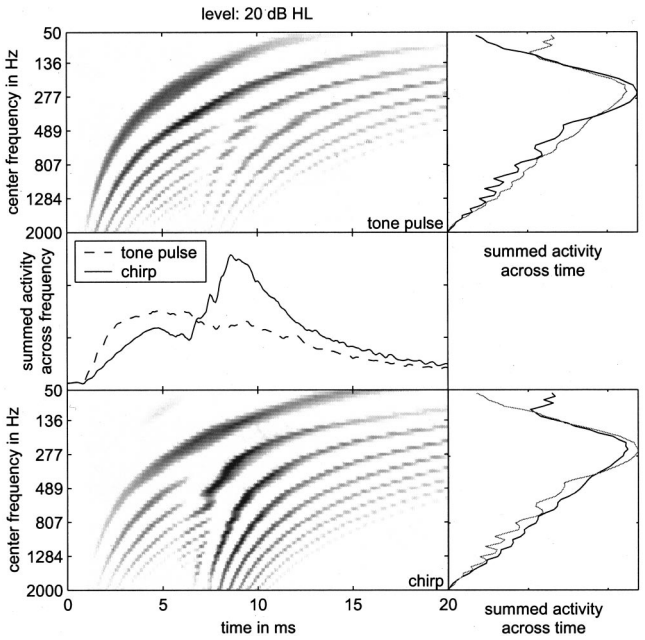


FIG. 13. Neural activity patterns for the 250-Hz tone pulse (top panel) and the low-frequency chirp (bottom panel), obtained with the same model as in Fig. 12. The middle panel represents the synchronized neural activity integrated across the frequency channels. The two panels on the right show corresponding neural activity within each frequency channel, integrated over time, in order to estimate the neural ''excitation pattern'' obtained with the tone pulse (top) and the chirp (bottom). For direct comparison, each of these panels also shows the pattern for the other stimulus, as indicated by the gray curve.

To illustrate the differences between the two narrow-band stimuli at the level of auditory-nerve processing, Fig. 13 shows the corresponding neural activity patterns obtained with the model described above. The top and bottom panels show the simulated output activity for tone-pulse and narrow-band chirp, respectively. A stimulus level of 20 dB HL was used in this case. The simulations nicely demonstrate the main characteristics observed in the data. First, for the chirp, neural activity is maximal near the end of the stimulus while it is shifted towards earlier times for the tone pulse. Second, the activity is temporally more concentrated and higher in amplitude for the chirp than for the tone pulse. Both observations are summarized in the corresponding neural activities summed across frequency as shown in the middle panel of Fig. 13. In order to also illustrate the amount of neural excitation as a function of frequency, the two panels on the right show the corresponding activity, integrated over time, in all excited frequency channels. For direct comparison, each of these panels also shows the pattern for the other stimulus, as indicated by the gray curve. The ''excitation patterns'' for the two stimuli are very similar. As for the magnitude spectra (from Fig. 2), the largest differences occur at frequencies below about 100 Hz where the chirp has more energy than the tone pulse. However, as can be seen in Fig. 13, this energy only contributes to the synchronized activity

at times *after* the maximum of synchronization. The simulations therefore suggest that the degree of synchronization but not the higher amount of overall neural activity is mainly responsible for the larger evoked response amplitude of the chirp.

However, for a deeper understanding of the relationship between peripheral processing and the formation of ABR patterns, additional assumptions about the neural processing at brainstem level as well as assumptions about the shape of the ''unitary'' responses of the different contributing neurons need to be considered (de Boer, 1975; Elberling, 1976; Goldstein and Kiang, 1958). Such modeling was beyond the scope of the present article but is currently being investigated (Dau, 2001).

C. Limitations of the chirp

There are some limitations of using the chirp that are considered in the following. For any given individual subject, the chirp designed from published functions regarding distance, frequency, and temporal maps in the cochlea is not necessarily optimal for that individual. That is, there probably is significant variation from subject to subject in the cochlear response time between frequency regions. Thus, the chirp may represent a compensation that is optimized for some mean delay of a group of individuals. Amplitude differences between individuals or between cochlear regions within a given individual may reflect how well the chirp represents the true cochlear response times across and within individuals and not solely the amount of activation. This issue may become more problematic when impaired cochleae are assessed in which case cochlear filter characteristics might vary as a function of the degree of damage. Thus, if the chirp response amplitude is being used to assess the amount of neural activity, those differences may also reflect the differences in the appropriateness of the chirp used.

However, in a series of accompanying experiments we found that a variation of the chirp parameters (such as the sweeping rate) that may result from the variability of physical parameters (such as the stiffness constant of the basilar membrane) had only little effect on the evoked response amplitude of the same subject. In addition, we also generated chirps based on estimates of the delay line characteristic from transiently evoked otoacoustic emissions (Shera and Guinan, 2000). In contrast, our original chirp was generated on the basis of parameters that were derived from post mortem experiments. Again, even though these two chirps differed considerably in duration and sweeping rate, the corresponding evoked potential amplitude hardly differed from one another for any subject tested so far. This suggests that the evoked potential measure does not seem sensitive enough to reflect *slight* changes in the physical properties of the cochlea. It seems sufficient to roughly match the (inverse) delay line characteristic of the cochlea in order to produce a much better response than the click (as long as the effective frequency range of the contributing neurons is not too narrow).

Of course, since chirps make an assumption about the response times along the cochlea, the actual delays down the cochlea in terms of peak activity cannot be estimated *directly* because it has already been established by the parameters of the chirp. In contrast, such information is directly available when using the click-evoked derived bands because one observes to a large extent the ''natural'' response of the system and differences due to cochlear damage can be studied. However, in the case of the chirp, such peak activity can be derived indirectly as the difference from the assumed characteristic in the chirp parameters. There is no information loss when using the chirp instead of the click.

D. Relation to other methods of frequency specificity

Since the sweep rate of the chirp is determined by basilar-membrane traveling-wave properties, the chosen ''nominal'' edge frequencies determine the duration of the chirp. A chirp consisting of only three halfwaves (starting and ending with zero crossings) will always be stretched over about two octaves, irrespective of the frequency region. It is possible that ABR obtained with more frequency selective stimuli allow for a better estimate of frequency specific information. For example, tone pulses as suggested by Davis (1976) with rise and fall times equal to two cycles of the stimulus frequency and a plateau time equal to one cycle may represent a good compromise between frequency specificity and sufficient synchronization capability of the stimulus. A direct comparison between such a stimulus and a narrow-band chirp was not attempted in the present study.

However, as mentioned in the Introduction, ABR elicited by stimulation with brief tone pulses of frequencies below 2 kHz has been shown to be only a poor predictor of low-frequency behavioral threshold, at least if presented without additional high-pass noise masking. Also, responses to longer-duration low-frequency tones, known as frequency following responses (FFR), do not represent frequency-specific information but most likely reflect synchronized activity from basally located regions (Janssen *et al.*, 1991; Dau, 2001) and can only be detected at relatively high stimulation levels above about 70 dB SPL (e.g., Hou and Lipscomb, 1979; Batra *et al.*, 1986). It has been suggested that instead of investigating transiently evoked responses and FFR, steady-state responses such as amplitude modulation following responses (AMFR) may serve as a better estimate of frequency-specific information even though the interpretation of the steady-state responses may be complicated because activity from brainstem and cortical generators are superimposed (e.g., Kuwada *et al.*, 1986; Griffiths and Chambers, 1991).

Whatever stimulus might represent the best choice for assessing frequency specific information from evoked potentials, particularly at low frequencies, the results from the present study suggest that the chirp stimulus produces a higher neural synchrony than other stimuli of similar magnitude spectrum. This was demonstrated for the broadband chirp in comparison with the click as well as for the band-limited chirp in comparison with the tone pulse.

V. SUMMARY AND CONCLUSIONS

(i) The broadband chirp elicited a larger wave-V amplitude than the click in the unmasked condition as well

as in all masking conditions where high-pass noise was presented in addition to the signal. The results demonstrate that the increased synchrony obtained with the chirp stretches over the entire frequency region. The chirp may be particularly interesting for clinical use in the low-frequency region below about 0.5–1 kHz.

(ii) The derived responses obtained with high-pass noise masking as well as the responses using notched-noise maskers indicate that the gain in synchrony within frequency regions of about one octave is not sufficient for the chirp to produce a significantly larger response amplitude than the click.

(iii) The low-frequency chirp elicited a larger wave-V amplitude at low and medium levels than a tone pulse with similar duration and magnitude spectrum. Wave-V latency differed by about 5 ms for these two stimuli. These observations as well as the differences between a click and a broadband chirp could be qualitatively explained in terms of the simulated neural activity patterns in the auditory periphery using a computational AN model.

(iv) Overall, the results further demonstrate the importance of cochlear processing for the formation of ABR. In order to obtain a deeper understanding of these effects, modeling work is needed, reflecting signal processing at a cochlear level and at subsequent brainstem stages, as well as assumptions about the contributions of single unit activity at these stages to the far-field response.

ACKNOWLEDGMENTS

We would like to thank our colleagues of the research group ''Medizinische Physik,'' the European Graduate School for ''Neurosensory Science and Systems,'' and the Special Research Program ''Neural basis of cognitive functions'' at the Universities of Oldenburg and Bremen for fruitful discussions on the content of this article. We thank Michael Heinz and Laurel Carney for all the discussions on their auditory nerve model. Manuel Don and an anonymous reviewer provided very helpful and fruitful criticism on an ealier version of this manuscript. This study was supported by the Deutsche Forschungsgemeinschaft (DFG).

[1]In the official data sheet of the preamplifier, a ''hard-wired'' high-pass cutoff frequency of 30 Hz is given. Unfortunately, we could not replicate this value and found a 3-dB cutoff of 95 Hz. The problem is that this setting will cut out a substantial portion of the wave-V amplitude which results in smaller responses overall, particularly for responses from lower frequency stimulus energy.

[2]It is not clear which noise type is more appropriate for the analysis of frequency-specific contributions. Pink noise with its 3-dB reduction of energy per octave excites the cochlea roughly uniformly on a logarithmic frequency scale while white noise may lead to some overmasking at high frequencies and undermasking at low frequencies. On the other hand, the two stimuli used in the present study, the click and the chirp, have a flat frequency spectrum and thus themselves do not excite the cochlea uniformly suggesting white noise as an appropriate masker. Also, white noise has been used in most of the notched-noise masking studies in the literature.

Abdala, C., and Folsom, R. C. (**1995a**). ''Frequency contribution to the click-evoked auditory brain-stem response in human adults and infants,'' J. Acoust. Soc. Am. **97**, 2394–2404.

Abdala, C., and Folsom, R. C. (**1995b**). ''The development of frequency resolution in humans as revealed by auditory brain-stem response recorded with notched-noise masking,'' J. Acoust. Soc. Am. **98**, 921–930.

Batra, R., Kuwada, S., and Maher, V. L. (**1986**). ''The frequency-following response to continuous tones in humans,'' Hear. Res. **21**, 167–177.

Beattie, R. C., and Kennedy, K. M. (**1992**). ''Auditory brainstem response to tone bursts in quiet, notch-noise, high-pass noise, and broadband noise,'' J. Am. Acad. Audiol **3**, 349–360.

Beattie, R. C., and Torre, P. (**1997**). ''Effects of rise-fall time and repetition rate on the auditory brainstem response to 0.5 and 1 kHz tone bursts using normal-hearing and hearing-impaired subjects,'' Scand. Audiol. **26**, 23–32.

Beattie, R. C., Franzone, D. L., and Thielen, K. M. (**1992**). ''Effects of notch noise bandwidth on the auditory brainstem response to clicks,'' J. Am. Acad. Audiol **3**, 269–274.

Beattie, R. C., Garica, E., and Johnson, A. (**1996**). ''Frequency-specific auditory brain-stem responses in adults with sensorineural hearing loss,'' Audiology **35**, 194–203.

Burkard, R., and Hecox, K. (**1983**). ''The effect of broadband noise on the human brainstem auditory evoked response. II. Frequency specificity,'' J. Acoust. Soc. Am. **74**, 1214–1223.

Bunke, D., von Specht, H., Mühler, R., Pethe, J., and Kevanishvili, Z. (**1998**). ''Der Einfluß der Reizanstiegszeit und der Hochpaßmaskierung auf die frühen auditorisch evozierten Potentiale,'' Laryngorhinootologie **77**, 185–190.

Carney, L. H. (**1993**). ''A model for the responses of low-frequency auditory-nerve fibers in cat,'' J. Acoust. Soc. Am. **93**, 401–417.

Coats, A. C., Martin, J. L., and Kidder, H. R. (**1979**). ''Normal short-latency electrophysiological filtered click responses recorded from vertex and external auditory meatus,'' J. Acoust. Soc. Am. **65**, 747–758.

Conijn, E. A. J. G., Brocaar, M. P., and van Zanten, G. A. (**1990**). ''Frequency specificity of the auditory brainstem response elicited by 1000-Hz filtered clicks,'' Audiology **29**, 181–195.

Conijn, E. A. J. G., Brocaar, M. P., and van Zanten, G. A. (**1992a**). ''Low-frequency specificity of the auditory brainstem response threshold elicited by clicks masked with 1590-Hz high-pass noise in subjects with sloping cochlear hearing losses,'' Audiology **31**, 272–283.

Conijn, E. A. J. G., Brocaar, M. P., and van Zanten, G. A. (**1993**). ''Frequency-specific aspects of the auditory brainstem response threshold elicited by 1000-Hz filtered clicks in subjects with sloping cochlear hearing losses,'' Audiology **32**, 1–11.

Conijn, E. A. J. G., Brocaar, M. P., van Zanten, G. A., and van der Drift, J. F. C. (**1992b**). ''Comparison between the frequency specificities of auditory brainstem response thresholds to clicks with and without high-pass masking noise,'' Audiology **31**, 284–292.

Dau, T. **(2001)**. ''The importance of basilar-membrane and auditory-nerve processing for the formation of auditory brainstem responses,'' 24th meeting of the Association for Research in Otolaryngology, p. 195.

Dau, T., Wegner, O., Mellert, V., and Kollmeier, B. (**2000**). ''Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion,'' J. Acoust. Soc. Am. **107**, 1530–1540.

Davis, H. (**1976**). ''Principles of electric response audiometry,'' Ann. Otol. Rhinol. Laryngol. Suppl. **28**, 1–96.

Davis, H., and Hirsch, S. K. (**1976**). ''The audiometric utility of brainstem responses to low-frequency sounds,'' Audiology **15**, 181–195.

de Boer, E. (**1975**). ''Synthetic whole-nerve action potentials for the cat,'' J. Acoust. Soc. Am. **58**, 1030–1045.

de Boer, E. (**1980**). ''Auditory physics. Physical principles in hearing theory I,'' Phys. Rep. **62**, 87–174.

Debruyene, F. (**1982**). ''Frequency specificity and on-effect in brainstem electric response audiometry,'' J. Otolaryngol. **11**, 267–270.

Don, M., and Eggermont, J. J. (**1978**). ''Analysis of the click-evoked brain stem potentials in man using high-pass noise masking,'' J. Acoust. Soc. Am. **63**, 1084–1092.

Don, M., Masuda, A., Nelson, R., and Brackmann, D. (**1997**). ''Successful detection of small acoustic tumors using the stacked derived-band auditory brain stem response amplitude,'' Am. J. Otol. **18**, 608–621.

Don, M., Ponton, C. W., Eggermont, J. J., and Masuda, A. (**1994**). ''Auditory brain-stem response (ABR) peak amplitude variability reflects individual differences in cochlear response times,'' J. Acoust. Soc. Am. **96**, 3476–3491.

Donaldson, G. S., and Ruth, R. A. (**1993**). ''Derived band auditory brainstem response estimates of traveling wave velocity in humans. I: Normal-hearing subjects,'' J. Acoust. Soc. Am. **93**, 940–951.
Elberling, C. (**1976**). ''Modeling action potentials,'' Rev. Laryngol. Otol. Rhinol. (Bord) **97**, 527–537.
Eggermont, J. J. (**1976**). ''Electrocochleography,'' in *Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neff (Springer-Verlag, New York), Vol. 5(3), pp. 625–705.
Eggermont, J. J., and Don, M. (**1980**). ''Analysis of the click-evoked brain stem potentials in man using high-pass noise masking,'' J. Acoust. Soc. Am. **68**, 1671–1675.
Folsom, R. C. (**1984**). ''Frequency specificity of human auditory brainstem potentials as revealed by pure-tone masking profiles,'' J. Acoust. Soc. Am. **75**, 919–924.
Folsom, R. C. (**1985**). ''Auditory brainstem responses from human infants: Pure-tone masking profiles for clicks and filtered clicks,'' J. Acoust. Soc. Am. **78**, 555–562.
Glasberg, B. R., and Moore, B. C. J.. (**1990**). ''Derivation of auditory filter shapes from notched noise data,'' Hear. Res. **47**, 103–138.
Goldstein, M. H., and Kiang, N. Y. S. (**1958**). ''Synchrony of neural activity in electric responses evoked by transient stimuli,'' J. Acoust. Soc. Am. **30**, 107–114.
Gorga, M., and Thornton, A. **(1989)**. ''The choice of stimuli for ABR measurements,'' Ear Hear **10**, 217–230.
Gorga, M. P., Kaminski, J. R., Beauchaine, K. A., and Jestead, W. (**1988**). ''Auditory brainstem responses to tone bursts in normal hearing subjects,'' J. Speech Hear. Res. **31**, 87–97.
Gorga, M. P., Worthington, D. W., Beauchaine, K. A., and Goldgar, D. E. (**1985**). ''Some comparisons between brain-stem response threshold, latencies and the pure tone audiogram,'' Ear Hear. **6**, 105–112.
Greenwood, D. D. (**1990**). ''A cochlear frequency position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.
Griffiths, S. K., and Chambers, R. D. (**1991**). ''The amplitude modulation-following response as an audiometric tool,'' Ear Hear. **12**, 235–241.
Heinz, M. G., Zhang, X., Bruce, I. C., and Carney, L. H. (**2001**). ''Auditory nerve model for predicting performance limits of normal and impaired listeners,'' Acoustics Research Letters Online **2**, 91–96.
Hoke, M., Pantev, C., Ansa, L., Lütkenhöner, B., and Herrmann, E. (**1991**). ''A time-saving BERA technique for frequency-specific assessment of the auditory threshold through tone-pulse series stimulation (TOPSTIM) with simultaneous gliding high-pass noise masking (GHINOMA),'' Acta Otolaryngol. Suppl. **482**, 45–56.
Hou, S. M., and Lipscomb, D. M. (**1979**). ''An investigation of the auditory frequency-following responses as compared to cochlear potentials,'' Arch. Oto-Rhino-Laryngol. **222**, 235–240.
Jacobson, J. T. (**1983**). ''Effects of rise time and noise masking on tone pip auditory brainstem responses,'' Sem. Hear. **4**, 363–372.
Janssen, T., Steinhoff, H.-J., and Böhnke, F. (**1991**). ''Zum Entstehungsmechanismus der Frequenzfolgepotentiale,'' Oto-Rhino-Laryngologia Nova **1**, 16–25.
Kilney, P. (**1981**). ''The frequency specificity of tone-pip evoked auditor brain stem responses,'' Ear Hear. **2**, 270–275.
Klein, A. J. (**1983**). ''Properties of the brain-stem response slow-wave component. II. Frequency specificity,'' Arch. Otolaryngol. **109**, 74–78.
Klein, A. J., and Teas, D. C. (**1978**). ''Acoustically dependent latency shifts of BSER (wave V) in man,'' J. Acoust. Soc. Am. **63**, 1887–1895.
Klein, A. J., and Mills, J. H. (**1981a**). ''Physiological (wave I and V) and psychological tuning curves in human subjects,'' J. Acoust. Soc. Am. **69**, 760–767.
Klein, A. J., and Mills, J. H. (**1981b**). ''Physiological and psychological measures from humans with temporary threshold shift,'' J. Acoust. Soc. Am. **70**, 1045–1053.
Kodera, K., Yamone, H., and Suzuki, J. (**1977**). ''The effect of onset, offset and rise-decay times of tone bursts on brainstem response,'' Scand. Audiol. **6**, 205–210.
Kramer, S. J. (**1992**). ''Frequency-specific auditory brainstem responses to bone-conducted stimuli,'' Audiology **31**, 61–71.
Kramer, S. J., and Teas, D. C. (**1979**). ''BSR (wave V) and N1 latencies in response to acoustic stimuli with different bandwidths,'' J. Acoust. Soc. Am. **66**, 446–455.
Kuwada, S., Batra, R., and Maher, V. L. (**1986**). ''Scalp potentials of normal and hearing-impaired subjects in response to sinusoidally amplitude-modulated tones,'' Hear. Res. **21**, 179–192.
Laukli, E. (**1983a**). ''Stimulus waveforms used in brainstem response audiometry,'' Scand. Audiol. **12**, 83–89.
Laukli, E. (**1983b**). ''High-pass and notch noise masking in suprathreshold brain-stem response audiometry,'' Scand. Audiol. **12**, 109–115.
Laukli, E., and Mair, I. W. S. (**1986**). ''Frequency specificity of the auditory brain-stem responses. A derived band study,'' Scand. Audiol. **15**, 141–146.
Laukli, E., Fjermedal, O., and Mair, I. W. S. (**1988**). ''Low-frequency and auditory brainstem response threshold,'' Scand. Audiol. **17**, 171–178.
Mackersie, C., Down, K. E., and Stapells, D. R. (**1993**). ''Pure-tone masking profiles for human auditory brainstem and middle-latency responses,'' Hear. Res. **65**, 61–68.
Nousak, J. M. K., and Stapells, D. R. (**1992**). ''Frequency specificity of the auditory brain stem response to bone-conducted tones in infants and adults,'' Ear Hear. **13**, 87–95.
Oates, P. A., and Stapells, D. R. (**1997a**). ''Frequency specificity of the human auditory brainstem and middle-latency responses to brief tones. I. High-pass noise masking,'' J. Acoust. Soc. Am. **102**, 3597–3608.
Oates, P. A., and Stapells, D. R. (**1997b**). ''Frequency specificity of the human auditory brainstem and middle-latency responses to brief tones. II. Derived response analysis,'' J. Acoust. Soc. Am. **102**, 3609–3619.
Pantev, C., Lagidze, S., Pantev, M., and Kevanishvili, Z. (**1985**). ''Frequency-specific contributions to the auditory brain stem response derived by means of pure-tone masking,'' Audiology **24**, 275–287.
Picton, T. W., Ouellette, J., Hamel, G., and Durieux-Smith, A. D. (**1979**). ''Brain-stem evoked potentials to tonepips in notched noise,'' J. Otolaryngol. **8**, 289–314.
Pratt, H., and Bleich, N. (**1982**). ''Auditory brainstem potentials evoked by clicks in notch-filtered masking noise,'' Electroencephalogr. Clin. Neurophysiol. **53**, 417–426.
Purdy, S. C., Houghton, J. M., Keith, W. J., and Greville, K. A. (**1989**). ''Frequency-specific auditory brainstem responses. Effective masking levels and relationship to behavioural thresholds in normal hearing adults,'' Audiology **28**, 82–91.
Scherg, M., and Volk, S. A. (**1983**). ''Frequency specificity of simultaneously recorded early and middle-latency auditory evoked potentials,'' Electroencephalogr. Clin. Neurophysiol. **56**, 443–453.
Shera, C. A., and Guinan, J. J. (**2000**). ''Frequency dependence of stimulus-frequency-emission phase: Implications for cochlea mechanics,'' in *Recent Developments in Auditory Mechanics*, edited by H. Wada, T. Takasake, K. Ikeda, K. Ohyama, and T. Koike (World Scientific, Singapore), pp. 381–387.
Shore, S. E., and Nuttall, A. L. (**1985**). ''High-synchrony cochlear compound action potentials evoked by rising frequency-swept tone bursts,'' J. Acoust. Soc. Am. **78**, 1283–1295.
Sohmer, H., and Kinarti, R. (**1984**). ''Survey of attempts to use auditory evoked potentials to obtain an audiogram,'' Br. J. Audiol. **18**, 237–244.
Stapells, D. R., and Picton, T. W. (**1981**). ''Technical aspects of brainstem evoked potential audiometry using tones,'' Ear Hear. **2**, 20–29.
Stapells, D. R., Picton, D. W., Perez-Abalo, M., Read, D., and Smith, A. (**1990**). ''Frequency specificity in evoked potential audiometry,'' in *The Auditory Brainstem Response*, edited by J. T. Jacobson (College-Hill, San Diego), pp. 147–177.
Suzuki, T., Hirai, Y., and Hiriuchi, K. (**1977**). ''Auditory brain stem responses to pure tone stimuli,'' Scand. Audiol. **6**, 51–56.
Terkildsen, K., Osterhammel, P., and Huis in't Velt, F. (**1975**). ''Farfield electrocochleography. Frequency selectivity of the response,'' Scand. Audiol. **4**, 167–172.
van der Drift, J. F. C., Brocaar, M. P., and van Zanten, G. A. (**1987**). ''The relation between the pure-tone audiogram and the click auditory brainstem response threshold in cochlear hearing loss,'' Audiology **26**, 1–10.
Weber, B. A. (**1987**). ''Assessing low-frequency hearing using auditory evoked potentials,'' Ear Hear. **8**, (suppl. 4), 49–54.
Wu, C.-Y., and Stapells, D. R. (**1994**). ''Pure-tone masking profiles for human auditory brainstem and middle latency responses to 500-Hz tones,'' Hear. Res. **78**, 169–174.
Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (**2001**). ''A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression,'' J. Acoust. Soc. Am. **109**, 648–670.

Auditory filter nonlinearity in mild/moderate hearing impairment

Richard J. Baker[a)]
Human Communication and Deafness, School of Education, University Of Manchester, Oxford Road, Manchester M13 9PL, United Kingdom

Stuart Rosen
Department of Phonetics and Linguistics, University College London, Wolfson House, 4 Stephenson Way, London NW1 2HE, United Kingdom

(Received 8 February 2000; revised 31 May 2001; accepted 10 December 2001)

Sensorineural hearing loss has frequently been shown to result in a loss of frequency selectivity. Less is known about its effects on the level dependence of selectivity that is so prominent a feature of normal hearing. The aim of the present study is to characterize such changes in nonlinearity as manifested in the auditory filter shapes of listeners with mild/moderate hearing impairment. Notched-noise masked thresholds at 2 kHz were measured over a range of stimulus levels in hearing-impaired listeners with losses of 20–50 dB. Growth-of-masking functions for different notch widths are more parallel for hearing-impaired than for normal-hearing listeners, indicating a more linear filter. Level-dependent filter shapes estimated from the data show relatively little change in shape across level. The loss of nonlinearity is also evident in the input/output functions derived from the fitted filter shapes. Reductions in nonlinearity are clearly evident even in a listener with only 20-dB hearing loss. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1448516]

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

I. INTRODUCTION

A fundamental property of the peripheral auditory system is that it operates as a kind of frequency analyzer. It can therefore be thought of in terms of a bank of overlapping bandpass filters. Because of their central and obligatory role in determining the nature of any further auditory processing, much effort has gone into characterizing the properties of these *auditory filters*. In hearing-impaired (HI) listeners with sensorineural hearing loss, the overwhelming consensus is that frequency selectivity is reduced, and that such a reduction has important implications for understanding the communication difficulties experienced by such listeners (e.g., de Boer and Bouwmeester, 1974; Leshowitz and Lindstrom, 1977; Pick *et al.*, 1977; Wightman *et al.*, 1977; Florentine *et al.*, 1980; Tyler *et al.*, 1982; Tyler *et al.*, 1984; Glasberg and Moore, 1986; Faulkner *et al.*, 1990; Laroche *et al.*, 1992; Sommers and Humes, 1993a, 1993b; see Moore, 1995, 1998 for reviews).

Several studies have also assessed the relationship between frequency selectivity and absolute threshold (Tyler *et al.*, 1982; Pick and Evans, 1983; Glasberg and Moore, 1986; Lutman *et al.*, 1991; Bergman *et al.*, 1992; Laroche *et al.*, 1992). Generally, selectivity decreases with increasing hearing loss, at least for losses above 30–40 dB HL. Below this level the selectivity remains approximately constant, with little correlation of filter bandwidth (quantified by the equivalent rectangular bandwidth, or ERB) with threshold (e.g., Fig. 9 of Glasberg and Moore, 1986, and Figs. 3 and 4 of Laroche *et al.*, 1992). Indeed, Pick and Evans (1983) noted a marked dissociation between frequency selectivity and threshold at 4 kHz—some listeners had very little hearing loss, but a significant increase in filter bandwidth. Similarly, West and Evans (1990) investigated selectivity and sensitivity in young adults with and without a history of exposure to amplified music. They found that in the subjects more regularly exposed to loud music the filter bandwidths were 10%–15% wider than in the least exposed, while there was little difference in sensitivity between the two groups. Thus, while moderate hearing losses typically lead to a concomitant reduction in selectivity, this may not necessarily be the case for mild hearing losses. Indeed, loss of selectivity may occur without significant loss of sensitivity.

In the physiological domain it has also been shown that cochlear damage affects both sensitivity and selectivity. The outer hair cells (OHCs) appear to be the prime mediators of the cochlear amplifier (e.g., Davis, 1983; Dallos, 1992), and damage to them leads to cochleae that are more linear, and show reduced sensitivity and frequency selectivity, than is found in an intact cochlea (e.g., Dallos and Harris, 1978; see Ruggero, 1991 and Patuzzi, 1996, for reviews).

Thus, current evidence suggests that OHC damage is reflected in, and is probably causally related to, loss of sensitivity and loss of frequency selectivity. As well as the loss of sensitivity and selectivity due to OHC damage, there is also the question of how changes in filtering across level are affected by OHC damage. It is only recently, however, that attempts have been made to clarify how filtering changes with level in normal-hearing (NH) listeners. While significant effort was initially put into showing whether or not filter shapes change at all with stimulus level in NH listeners (Patterson, 1971; Weber, 1977; Pick, 1980; Lutfi and Patterson, 1984), relatively little attention has been paid to the accurate characterization of this phenomenon.

a)Electronic mail: richard.baker@man.ac.uk

 0001-4966/2002/111(3)/1330/10/$19.00 © 2002 Acoustical Society of America

TABLE I. Absolute threshold of five hearing-impaired listeners measured using standard pure-tone audiometry. Values in bold face italics represent the thresholds for the frequencies at which the listeners' frequency selectivity was measured using notched-noise masking.

		Threshold (dB HL) (kHz)										
Subject	Age	0.125	0.25	0.5	0.75	1	1.5	2	3	4	6	8
BL	64	40	45	50	55	55	55	***50***	55	60	60	55
CH	24	10	5	5	⋯	5	⋯	***20***	⋯	15	⋯	5
CS	59	15	15	***10***	15	30	35	***35***	40	45	65	80
ET	24	⋯	15	5	⋯	10	20	***30***	⋯	35	⋯	40
MF	50	20	20	10	10	***15***	20	***50***	60	60	70	60

Moore and Glasberg (1987; also Glasberg and Moore, 1990) attempted to tie together the existing notched-noise masking data from various studies to provide a coherent description of how auditory filters change shape as a function of frequency and stimulus level. Moore and Glasberg (1987) argued, on the basis of an excitation pattern model derived from their filter shapes, that the filter shape is determined primarily by the signal level at the input of the filter, and that filters whose shape depends on their own output do not produce realistic excitation patterns. We have shown previously, however, that this is not true. Using a technique of fitting filter shapes to the notched-noise masking data as an explicit function of stimulus level (PolyFit procedure—Rosen and Baker, 1994; Rosen *et al.*, 1998), we have been able to describe how the filter shapes vary with stimulus level much more accurately. Using this technique, we have also shown that when the variations in the low-frequency skirt of the filter are taken into account, an output-controlled nonlinearity can produce realistic excitation patterns (Baker *et al.*, 1998). Recently, Glasberg and Moore (2000) have also found an output-controlled nonlinearity to more adequately account for notched-noise masking results than an input-controlled one.

Such an issue cannot be treated as a minor detail, as it has ramifications both in the design of experiments, and the interpretation of their results. Rosen *et al.* (1998) also argued that using a fixed-level probe tone is more appropriate than the fixed-level notched-noise masker typically used to measure auditory filter shapes at a single level in NH listeners. Related to this issue, Leek and Summers (1993) also used the level-dependent fitting procedure of Rosen *et al.* (1992) to help clarify changes in filtering in the presence of background noise. They showed that the use of probe-level-dependent filter shapes better accounted for their data (measured at a single fixed masker level, with different levels of background noise) than a level-independent filter shape.

For listeners with OHC damage, the cochlea will behave more linearly, with filter shapes changing less with increasing stimulus level (Stelmachowicz *et al.*, 1987; Murnane and Turner, 1991). The data of Moore *et al.* (1985; their Table V) also suggest a reduction in the change of filter shape with level in impaired ears, although they did not discuss this aspect of their data. In such cases the confound between measurement technique and nonlinearity will be less important, as the cochlea is likely to behave in a more linear manner. Thus, notched-noise masking experiments in which the masker level is fixed will result in filter shapes similar to those in which the probe tone level is fixed.

The aim of the present study is to use the notched-noise masking technique coupled with the PolyFit analysis to quantify the changes in frequency selectivity in mild/moderate hearing loss over a range of stimulus levels, and to relate these to measurements of selectivity in NH listeners.

II. METHODS

A. Listeners

Table I presents summary data concerning the 5 HI listeners who participated in the study. The degree of hearing impairment at 2 kHz ranged from 20 to 50 dB HL. All testing was done monaurally, and in a single ear per listener.

CH was a 24-year-old male with a reported history of exposure to loud sounds as a musician in a rock band. ET was a 24-year-old female with a mild, bilateral high-frequency sensorineural hearing loss confirmed in childhood. The remaining three listeners (BL, CS, and MF) were recruited from an adult audiology clinic. None of the five listeners demonstrated any middle-ear involvement in their impairment as assessed by bone-conduction thresholds and tympanometry.

B. Threshold estimation

Masked thresholds were determined for sinusoidal probe tones in the presence of notched-noise maskers with variable notch widths. The notches were placed both symmetrically and asymmetrically about the probe frequency, and either the probe level or the noise level was varied to determine thresholds. For listeners CH and ET, a two-interval, two-alternative forced-choice (2I-2AFC) paradigm with feedback was used to estimate the 79% point on the psychometric function (Levitt, 1971). From a starting level at which the probe was clearly audible, the varying sound, either probe or masker, was initially changed in 5-dB steps, with step size decreasing by 1 dB after each turnaround. Once the step size reached 2 dB, it remained constant for a further eight turnarounds, the mean of which was taken as the threshold. For the remaining three listeners a 2I-2AFC maximum-likelihood procedure was used to estimate the 90% point on the psychometric function (see Baker and Rosen, 2001, for a description of the procedure used). In both cases, listeners responded on a button box, with illuminated buttons indicating presentation intervals and providing feedback.

TABLE II. Stimulus levels used in notched-noise masking experiments at 2 kHz (and also at 0.5 kHz for listener CS and 1 kHz for listener MF—frequencies at which the listeners' absolute threshold was less than 20 dB HL).

Listener	Absolute threshold (dB HL)	Frequency (kHz)	Fixed masker spectrum levels (dB SPL)	Fixed probe levels (dB SPL)
BL	50	2	40,50, 60	60,70, 80
CH	20	2	30,40,50,60	40,50,60,70
CS	35	2	40,50, 60	50,60,70,80
ET	30	2	40,50, 60	50,60, 70
MF	50	2	40,50, 60	60,70, 80
CS	10	0.5	30, 50	30,50, 70
MF	15	1	30, 50	30,50, 70

For each particular combination of notch width and fixed probe or fixed masker level, at least two thresholds per listener were obtained. Where two measurements of the same condition in the same listener differed by more than 4 dB, one or two further measurements were taken and the mean of all measurements used.

C. Stimulus configurations

The outside edges of the masker noise were fixed at $\pm 0.8 \times f_p$ [400 and 3600 Hz for a probe frequency (f_p) of 2 kHz]. A maximum of 16 different notch conditions was used, 6 symmetric and 10 asymmetric. The frequencies (f) of the edges of the notch (the inner edges of the two noise bands) are specified in normalized frequency units (g) relative to the probe frequency as given by $g=(|f-f_p|)/f_p$. In the symmetric conditions, both notch edges were placed at normalized values of 0.0, 0.1, 0.2, 0.3, 0.4, and 0.5. In the asymmetric condition, one of the notch edges was set at a normalized value of 0.0, 0.1, 0.2, 0.3, and 0.4, while the other was set to 0.2 normalized units further away (0.2, 0.3, 0.4, 0.5, and 0.6). When the masker level was fixed, a subset of noise spectrum levels (N_0) was chosen, ranging from 20–60 dB SPL in 10-dB steps, and when the probe level was fixed, a subset of probe levels (P_s) was chosen, ranging from 30–80 dB SPL in 10-dB steps. The specific values used for each listener depended on the nature of the hearing loss, and are detailed in Table II. In addition to measurements at 2 kHz, CS and MF were also tested at frequencies where their sensitivity was in the normal range (0.5 kHz for CS and 1 kHz for MF).

D. Stimulus generation

All the stimuli were computer generated at a sampling frequency of 20 kHz. The time waveform of the probe consisted of a steady-state portion of 360 ms plus 20-ms raised-cosine onset and offset. The probe was temporally centered within the masker, which consisted of a 460-ms steady-state portion with 20-ms cosine-squared onset and offset. To generate the masker, the desired frequency spectrum was defined by setting all the spectral components (spaced at intervals of 0.61 Hz) within the appropriate frequency limits to have equal amplitudes while those outside were set to zero. Nonzero components had their phase randomized uniformly in the range of $0-2\pi$ radians. An inverse FFT was then applied to generate the time waveform. At the start of each threshold determination, a 3.2768-s buffer of noise was generated for use during that test. On each trial, a 500-ms portion of the buffer was chosen randomly for each of the two masker intervals within each trial.

The probe and masker were played out through separate channels of a stereo 16-bit D/A converter (12-bit D/A for listener CH) and attenuated independently under computer control before being electrically mixed. The signal was then presented monaurally via Etymotic ER2 insert earphones to listeners CH and ET and via Beyer DT48 circumaural headphones to the remaining listeners. For listeners using the Beyer DT48 headphones, the noise spectrum was shaped at the synthesis stage to give a flat spectrum of appropriate spectrum level as measured in a B&K 4157 ear simulator.

E. Analyses

All analyses were performed on an individual basis using the mean threshold for each notch/level combination. A variety of models was fitted to each data set, using the Poly-Fit technique described in detail by Rosen *et al.* (1998). All of the models were variants of the asymmetric roex(p,w,t) model (Patterson *et al.*, 1982). These included simplified models in which, for example, the upper half of the filter was described with a roex(p) shape whereas the lower half was a complete roex(p,w,t) shape. It is also necessary to estimate k, the signal-to-noise ratio necessary for detection at the output of the filter. All of these parameters can be arbitrary polynomial functions of the level of the masker or the probe, but we have never investigated models with more than a quadratic dependence on level (that is, three coefficients per parameter to be estimated). Finally, we also estimate an absolute threshold by allowing a further parameter to be added to the predicted masked threshold in power terms (see Rosen *et al.*, 1998 for details). This parameter was added to take into account the fact that for low-level wide notch conditions the probe-tone level is governed by the listener's absolute threshold. In the present study this estimated threshold parameter was always within 5 dB of the listener's pure-tone threshold at that frequency.

III. RESULTS

A. Growth-of-masking functions

1. Data and predictions

The notched-noise masking experiments carried out across a range of stimulus levels allow the data to be plotted

as growth-of-masking (GOM) functions for each notch configuration. In producing these GOM functions we follow the convention of Lutfi and Patterson (1984) in plotting signal-to-noise ratio at threshold versus masker spectrum level. This has the advantage over plotting signal level versus masker level in that a simple linear filter with constant detector efficiency will result in a set of horizontal parallel lines, one for each of the different notch conditions. For a filter that broadens with level and a constant detector efficiency, the GOM functions for the no-notch condition will be approximately horizontal, while the GOM functions for the wider notch conditions will show a positive slope (for data above absolute threshold).

Figure 1 shows such GOM functions for each of the five HI listeners, and also for the average of the three NH listeners described in Sec. II A of Rosen *et al.* (1998). The curves through the data points show the fits to the data obtained using the PolyFit procedure described later in Sec. B. For the NH listeners there is a clear convergence of GOM functions as stimulus level is increased—indicating a broadening of the underlying filter. For listener CH, with the mildest hearing loss (20 dB HL), there appears to be less convergence than for the NH listeners. This is even clearer for listeners ET, CS, and BL, where the GOM functions (and the predictions obtained from the PolyFit procedure) are approximately parallel. For listener MF there appears to be a slight divergence of the GOM functions, indicating a slight narrowing of the filter with increasing stimulus level. These changes in the slopes of the GOM functions relative to the NH listeners are quantified in the following section.

2. Estimates of amplitude compression

If it is assumed that maskers near the tone frequency undergo the same compression as the tone, whereas maskers well below the tone frequency are subject to little compression, it is possible to use the GOM functions to estimate the degree of amplitude compression arising from the underlying cochlear mechanisms. Oxenham and Plack (1997) have used such a technique in forward masking. They measured GOM functions for a 6-kHz probe tone with a 6-kHz masker and with a 3-kHz masker. The idea is that with the 6-kHz masker, both the tone and the masker will be subject to the same ''CF'' compression. However, when the masker is at 3 kHz it will not be subject to compression at the 6-kHz place while the probe will. Thus, a 10-dB increase in the 3-kHz masker will result in a greater than 10-dB increase in probe level at masked threshold. Using this technique, Oxenham and Plack (1997) estimated a basilar membrane input/output (I/O) function slope of about 0.25, similar to the slope measured directly in physiological experiments.

Making similar assumptions about the nature of the underlying filtering mechanism (i.e., linearity at frequencies well below CF and also that the high-frequency side of the filter is approximately level independent), similar estimates can be made from the data given in Fig. 1. Taking the thresholds for the conditions where the lower masker band is at least 1 octave below the probe tone frequency—notch conditions (0.6,0.4), (0.5,0.3), and (0.5,0.5)—Fig. 1 shows that the GOM functions are approximately parallel when points near absolute threshold are excluded. Using these three notch conditions, estimates of the slopes of the GOM function for the low-frequency skirt of the filter were derived from a simple linear regression. Here, it was assumed that the GOM functions from the three conditions had identical slopes, but were permitted to have different intercepts. Only data points 6 dB or more above absolute threshold were used. Examples of the fits are shown in Fig. 2 for two of the HI listeners and also from the NH listeners of Rosen *et al.* (1998) (plotted as signal level versus masker spectrum level).

The solid horizontal line shows the absolute threshold, while the solid diagonal lines show the linear fits to the data for the (0.5,0.3) and (0.6,0.4) notch conditions. The dashed diagonal lines in the lower two plots indicate the slope of GOM functions for the NH listeners.

Since the slope of the GOM functions for the no-notch condition is approximately one (zero when signal to masker ratio is plotted versus masker spectrum level), the reciprocal of the slope of the GOM functions obtained as above gives an estimate of the slope of the I/O function of the underlying cochlear mechanism (although this may be somewhat confounded by the presence of suppression in simultaneous masking). Table III shows the estimated slopes of the GOM functions using the linear fits described above and also the derived I/O function slopes. In addition to the data for the normal and impaired listeners at 2 kHz, estimates are given at 0.5 and 1 kHz from listeners CS and MF, respectively, i.e., frequencies where absolute thresholds for these listeners were within the normal range. For comparison, Table III also includes estimates at these two frequencies for the two listeners of Baker *et al.* (1998).

The estimated I/O function slope for the NH listeners at 2 kHz is 0.578, indicating compression approaching 2:1. However, compression was reduced for listener CH (compression ratio approximately 1.25:1), whose absolute threshold at 2 kHz was 20 dB HL. For listeners ET, CS, and BL the derived I/O functions approach linearity (1.03:1, 1.1:1, and 1.1:1, respectively). For listener MF, the GOM functions show a degree of divergence; this is reflected in an I/O function slope of greater than unity indicating an expansive nonlinearity. The derived slopes for the listeners CS and MF at 0.5 and 1 kHz were similar to those obtained previously from NH listeners.

B. Level-dependent filter shapes

Using the PolyFit procedure described by Rosen and colleagues (Rosen and Baker, 1994; Rosen *et al.*, 1998) the notched-noise masking data were fitted with level-dependent filter shapes based on the rounded exponential (roex) family described by Patterson *et al.* (1982). Any of the fitted parameters could be allowed to be constant or a linear or quadratic function of stimulus level (either masker or probe level). Results are presented for probe-dependent models because they have been found to give a better fit to notched-noise data than masker-dependent models (see also Rosen *et al.*, 1998). Each data set was fitted with a set of filter shapes of varying complexity. Typically a complex model was first fitted to the data. One of the parameters was then removed from the model and the data refitted. If the error in the fit

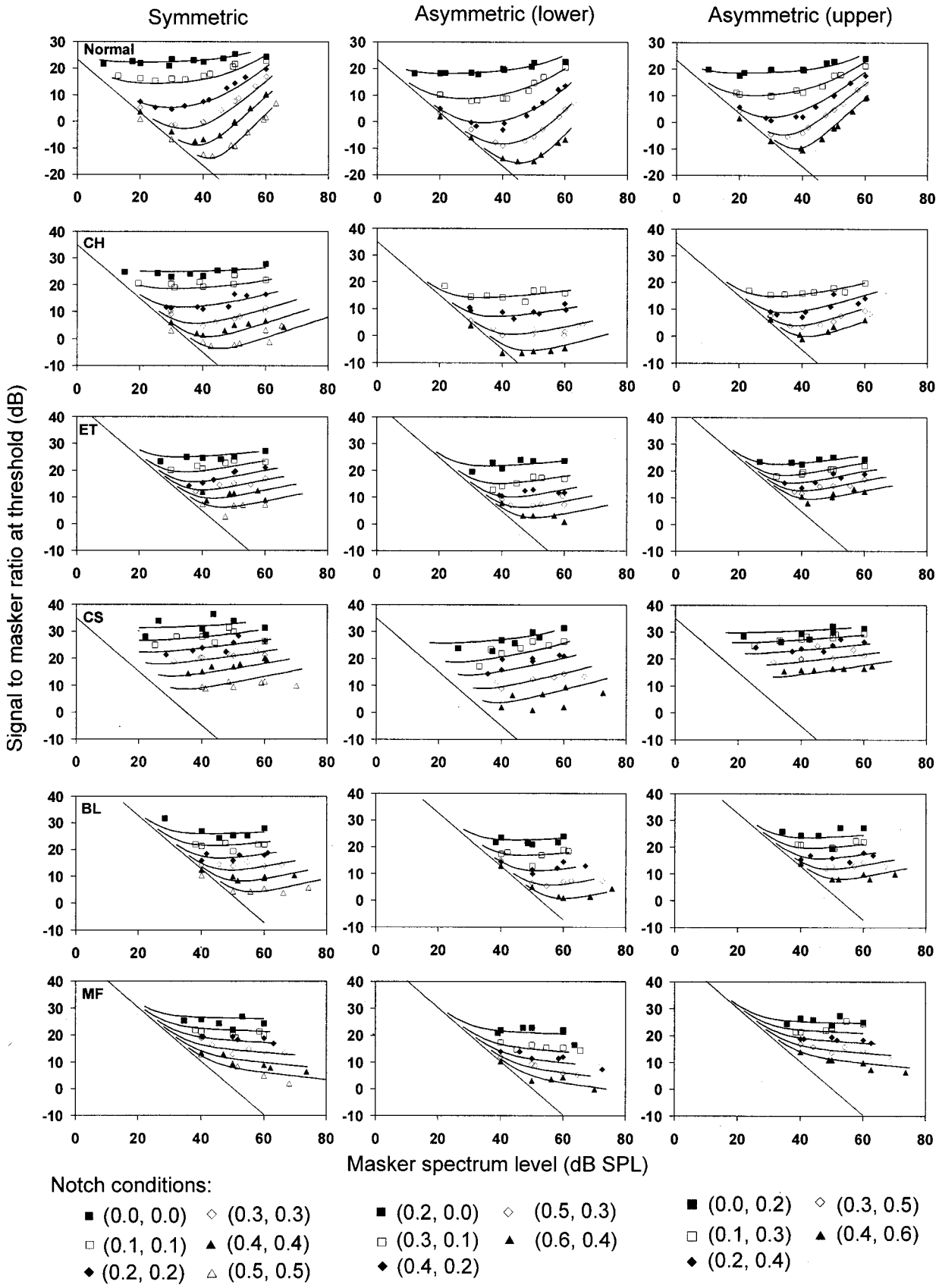


FIG. 1. Growth-of-masking functions for normal-hearing listeners (average of three listeners from Rosen *et al.*, 1998) and hearing-impaired listeners. The notch widths are indicated at the bottom of each column. The solid curves show the predictions derived from the filter shapes fitted to the data as described in Sec. III B. Note that for any particular listener a single model describes all the data. Thus, for one row of plots the solid lines are not independent of each other, but are all determined by the same fit to the data. The diagonal solid line represents the listener's absolute threshold.

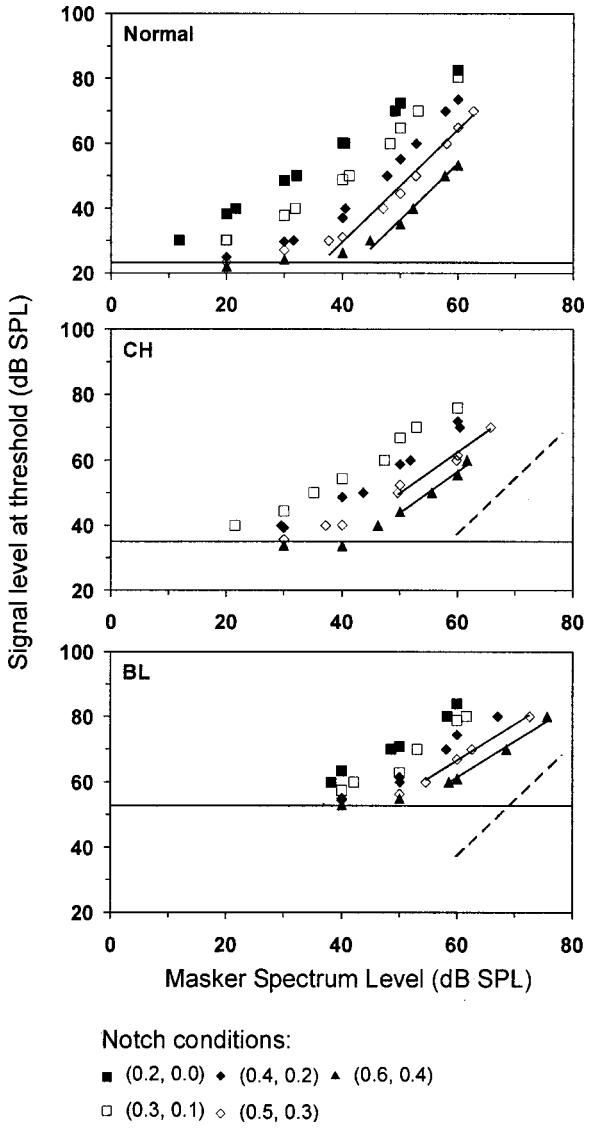


FIG. 2. Examples of growth-of-masking functions with data taken from the middle column of Fig. 1 and replotted as signal level against masker level. The solid diagonal lines show two of the three linear fits to the wide notch data [the third being for the (0.5,0.5) notch condition not shown here]. The dashed diagonal lines in the lower two plots indicate the slope of the normal-hearing listeners' functions. The horizontal solid line represents the listeners' absolute threshold.

worsened then that parameter was deemed necessary to the model. Thus, it was possible to see which parameters were important to accurately describe the data and which were relatively superfluous. The models that fitted each listener's data best vary in structure since the degree of nonlinearity evident in the filtering varies. The listeners with a greater degree of hearing loss show filtering which can be described by models with fewer free parameters than those where a greater degree of nonlinearity is present. The filter shapes obtained using the "best" model for each listener are shown in Fig. 3.

The filter shapes obtained by Rosen *et al.* (1998) for averaged data from 3 NH listeners are presented at the top of Fig. 3. For these data the high-frequency side of the filter could be described by a one-parameter, roex(p), shape in which the slope parameter (p_u) was allowed to be a quadratic function of probe-tone level. The lower frequency side of the filter was better described by a three-parameter, roex(p,w,t), shape in which the slope of the filter in the passband (p_l) was constant while the slope of the skirt (t_l) was a linear function of probe level. Also, the point of intersection between the passband and tail of the filter (w_l) was required to be a linear function of probe level, with the tip-to-tail ratio decreasing as probe level increased. Following the nomenclature of Rosen *et al.* (1998), this model is referred to as $p1312x2x$, where p specifies that the parameters are dependent on the probe-tone level, and the digits refer to the number of fitted coefficients in the polynomial that describes how that parameter varies with tone level. The parameters are specified in the order p_l, p_u, k, w_l, w_u, t_l, t_u (x indicates that a parameter is not required in the fit). In order to estimate the changes in gain at CF as stimulus level is increased, we used the approach of Rosen *et al.* (1998). This technique relies on the assumption that the gain of the filter is constant well below the filter's CF, as evidenced in basiler membrane measurements. We thus normalize all filter shapes to have the same gain (an arbitrary 0 dB) at a frequency that is 0.4 times the filter CF.

The rest of the panels in Fig. 3 show the filter shapes for the five HI listeners, along with their abso-lute thresholds. The number of parameters required to adequately describe the filter shapes varied between 4 and 7 for the HI listeners

TABLE III. Growth-of-masking slopes for the low-frequency skirt of the filter derived using linear fits as described in the text. The reciprocal of the GOM function slope gives an estimate of the slope of the underlying input–output function. The final column gives the slope of the I/O function estimated from the fitted filter shapes as described in Sec. III C.

Subject	Frequency (kHz)	Threshold, dB HL	GOM slope	I/O slope (GOM)	I/O slope (filter)
Normal—Rosen *et al.* (1998)	2	8	1.731	0.578	0.47
CH	2	20	1.257	0.796	0.72
ET	2	30	1.028	0.973	0.81
CS	2	35	1.105	0.905	0.77
BL	2	50	1.108	0.903	0.80
MF	2	50	0.652	1.534	1.18
Normal—Baker *et al.* (1998)	0.5	9	1.284	0.779	0.55
CS	0.5	10	1.371	0.729	0.54
Normal—Baker *et al.* (1998)	1	5	1.776	0.563	0.52
MF	1	15	1.709	0.585	0.34

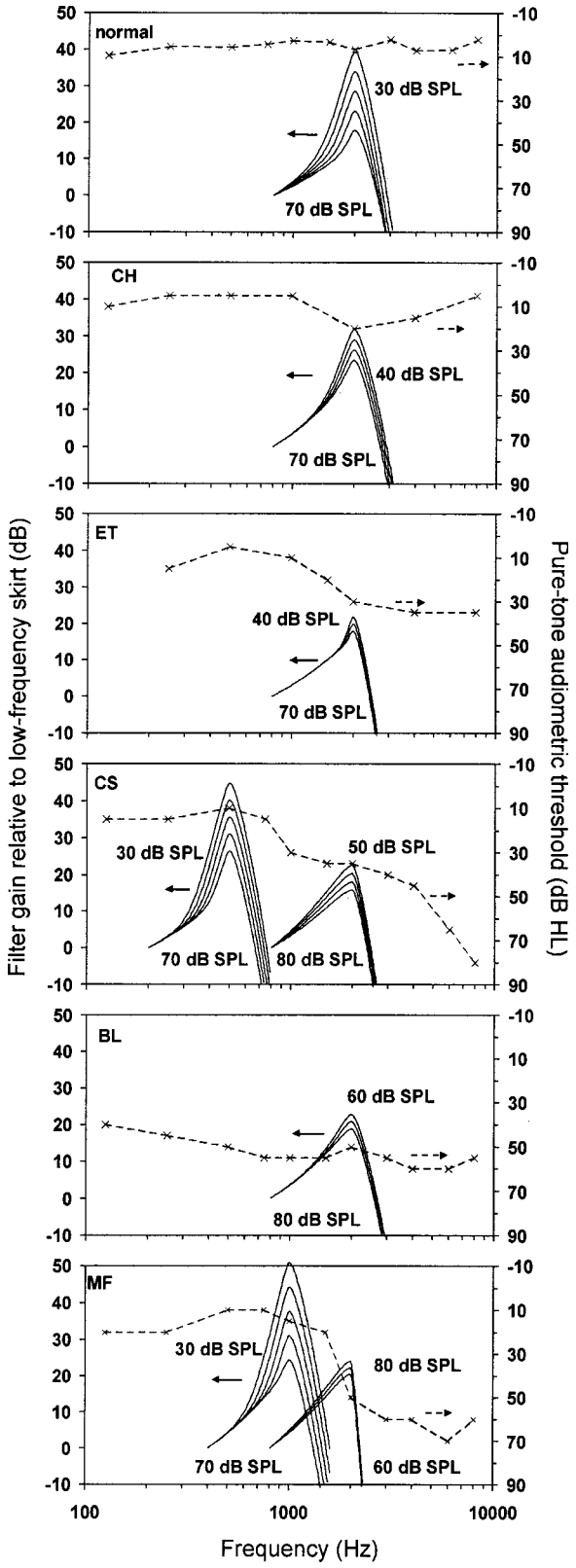


FIG. 3. Filter shapes fitted to the data using the PolyFit procedure. The filter shapes are all normalized to 0 dB at the lowest frequency point (0.4×CF) and are plotted at probe-level intervals of 10 dB within the ranges indicated. The dashed lines, associated with the right-hand ordinate, indicate the listeners' audiograms.

compared with 9 for the NH listeners, and the parameter structures of those plotted in Fig. 3 are given in the sixth column of Table IV. These "best-fitting" models were selected according to the strategy described above such that the simplest shape is chosen that does not result in a large increase in the rms error of the fit. Clearly all of the HI listeners show a reduced change in gain at CF, reflecting a more linear filter compared with those of NH listeners. It is also apparent that the two listeners who were tested at frequencies of normal sensitivity (CF and MF) display an unimpaired degree of nonlinearity at these frequencies. All the listeners, apart from MF, show filters that broaden and decrease in gain with increasing stimulus level. The 2-kHz measurement of MF shows slight changes in the reverse direction, with the filter at 60 dB SPL being slightly broader than that at 80 dB SPL. It must be emphasized, however, that the reduced degree of nonlinearity in the HI listeners means that the filter shapes can, in some cases, be fitted almost as well with filter shapes that don't change at all with stimulus level.

Table IV gives the goodness of fit (rms error) for three different filter shape models to the data of each listener. The model with the greatest number of parameters (p2212222) obviously fits the data best. Removing parameters from this model to find the best fit gives a marginally worse fit to the data (columns 6 and 7 of Table IV). Removing all the level dependency from the fitted model (p1111111, column 8) further increases the rms error. As the fitted shapes are describing the same underlying sets of data, it is to be expected that the more linear the underlying filter, the less the difference would be in the goodness of fit of the linear and nonlinear models. For the NH listeners the linear model shows a 515% increase in rms error relative to the most complex model. For the HI listeners this difference is much reduced, ranging from 75% in the least impaired to 12% in the most impaired, again providing evidence for a more linear filter.

C. Estimated compression

We have already used GOM functions to estimate the slope of the underlying basilar-membrane I/O function. Our technique of normalizing filter shapes at a frequency significantly below CF (as illustrated in Fig. 3) also allows an estimate of this I/O function by measuring the change in gain at CF across the range of stimulus levels used. Figure 4 plots the estimated I/O function slope against the absolute threshold for the five HI listeners, and also for the nine NH listeners measured at 2 kHz by Rosen *et al.* (1998). For the NH listeners the estimated slope is about 0.47, giving a compression ratio just over 2:1. This estimated slope shows little variation over a 25-dB range of absolute thresholds (−5 to 20 dB HL). For the HI listeners (with absolute thresholds ranging from 20 to 50 dB HL) the range is from 0.72 to 1.18 (see the last column of Table III).

While the actual values of the I/O slopes derived from the filter shapes differ from those estimated from the GOM-functions, the general pattern is the same. Each of the HI

TABLE IV. Details of goodness of fit of various filter shapes to the notched-noise masking data at 2 kHz. Complexity of fitted filter shapes is denoted, for example, in the form p2212222 or p1111111, where p indicates that the parameters change as a function of probe-tone level (m—masker spectrum level). The digit denotes that a parameter is constant (1), a linear function (2—2 fitted coefficients) or a quadratic function (3—3 fitted coefficients). "x" indicates that a parameter is not required. The order of the parameters is pl, pu, k, wl, wu, tl, tu, where "l" indicates the lower, and "u" the upper side of the filter. The percentages in brackets show the percentage increase in rms error relative to the p2212222 model.

Subject	Threshold (dB HL)	Number of thresholds	p2212222	m2212222	Best model	rms error for best model	p1111111 (linear)
NH	5	158	0.81	1.47 (81%)	p1312x2x	0.86 (6%)	4.98(515%)
CH	20	110	1.19	1.47 (24%)	p1112x1x	1.20 (1%)	2.08 (75%)
ET	30	94	0.99	1.13 (14%)	p1112x1x	1.07 (8%)	1.69 (71%)
CS	35	111	1.68	2.14 (27%)	p211x2x1	1.76 (5%)	2.45 (46%)
BL	50	96	1.16	1.24 (7%)	p1112x1x	1.25 (8%)	1.48 (28%)
MF	50	93	1.45	1.31(−10%)	p211xxxx	1.60(10%)	1.63 (12%)

listeners shows values less than those of NH listeners and MF shows a slightly expansive nonlinearity. The discrepancies that do exist between these two measures may be due to the fact that the values estimated from the GOM functions use only a limited subset of the data, whereas those estimated from the fitted filter shapes use the entire data set. A final point from Fig. 4 is that results from listener CH suggest a markedly less compressive cochlea than in one of the NH listeners (listener MB from Rosen *et al.*, 1998) despite the fact that their absolute thresholds are the same. Clearly the labels NH and HI for these two listeners are misleading, since they have the same absolute threshold. However, the impaired filtering of CH suggests that such a classification based on absolute threshold alone does not tell the whole story.

D. Equivalent rectangular bandwidth

From the fitted filter shapes it is possible to provide a simple quantitative measure of selectivity in terms of filter bandwidth. The measure that has typically been used is the ERB. For a probe-level-dependent filter shape the ERB can be calculated over a range of individual probe levels. Figure 5 shows how the ERB, calculated from the best filter models shown in Fig. 3 (and given in Table IV), changes as a function of probe-tone level. For the NH listeners the ERB remains approximately constant at low levels where the low-frequency skirt of the filter has little influence on the amount of energy the filter would pass. It is not until the probe level is increased above 70 dB SPL that the low-frequency skirt makes a significant contribution to the amount of energy passed by such a filter. For the HI listeners there are three major points worth noting. First, at low probe levels the ERB is always broader than the NH listeners. Second, the ERBs measured in the impaired ears change by smaller amounts than in the NH ears—the filter is more linear. Third, at high levels the ERBs measured in the impaired ears are more similar to those of the normal ears.

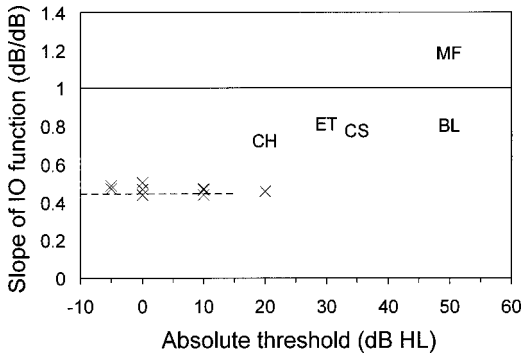


FIG. 4. Input/output function slopes derived from the filter shapes shown in Fig. 3 for the HI listeners. The "×"s indicate measurements from the nine normal-hearing listeners described by Rosen *et al.* (1998) and the dashed line represents the "mean of three" normal-hearing listeners described in the same study. A slope of value 1.0 indicates a linear filter, and a value of 0.5 indicates a compression of 2:1.

IV. DISCUSSION

The emphasis of this study was to characterize the way in which auditory filters in HI listeners change in shape as stimulus level increases. While previous studies have demonstrated that cochlear hearing loss results in broadened auditory filters when using the notched-noise masking technique, such studies have nearly all used a fixed masker-level paradigm. In characterizing filter shapes for ears where OHC loss is severe the choice of fixed masker-level or fixed probe-level paradigm will have little consequence, as the cochlea in question will be approximately linear. However, when comparing these "linear" filters with those of NH listeners, the choice of paradigm in the NH listeners is crucial. Rosen *et al.* (1998) have shown that it is more appropriate to use a fixed probe level than a fixed masker level and argued that

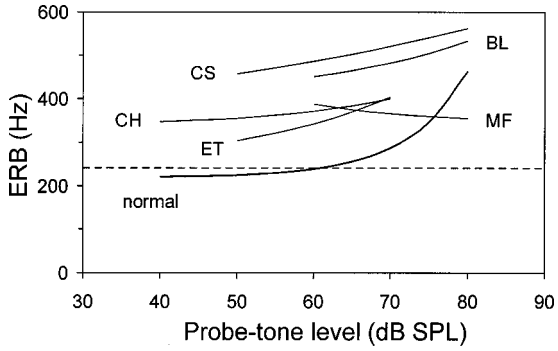


FIG. 5. Equivalent rectangular bandwidths derived from the fitted filter shapes given in Fig. 3. The dashed line represents the value of 241 Hz from the equation given by Glasberg and Moore (1990) for a frequency of 2 kHz.

the filter shape appears to be controlled by its own output, rather than by its input, via some sort of feedback mechanism on the basilar membrane. The same argument was followed by Baker *et al.* (1998) to produce a realistic output-controlled excitation pattern model. More recently, Glasberg and Moore (2000) have suggested that the filter shape "*depends on the power at its output divided by the gain*" and that "*this dependence may arise because filtering on the BM involves a feedback mechanism*." The likely mediators of such a feedback mechanism are the OHCs, which provide a "fast mechanical positive feedback" to the basilar membrane (Patuzzi, 1996). Here, we have used the same procedure as Rosen *et al.* (1998) to describe filter shapes from listeners with mild to moderate sensorineural hearing loss in a way that allows direct comparison with the NH data.

We have been able to estimate the degree of compression from normal and impaired ears, first using an approach somewhat similar to that of Oxenham and Plack (1997), and second indirectly from the change in gain at the CF of the fitted filter shapes. As expected, both these methods show a reduced degree of compression in HI listeners compared to NH listeners. For the HI listeners the GOM functions are more parallel across the different notch widths than in the NH listeners, indicating a smaller change in filter shape across level.

This reduced nonlinearity is evident in the filter shapes fitted to the HI data (Fig. 3). For the impaired listeners, the "tip-to-tail" gain is about 20 dB (slightly more for CH, the least-impaired listener), which is similar to the higher level filters from the NH listeners. However, the change in tip-to-tail gain increases with decreasing stimulus level much more slowly in HI listeners than in NH listeners.

One thing that is somewhat unusual is the degree of loss of nonlinearity present in listener CH. His threshold at 2 kHz is 20 dB HL, yet the degree of compression is only about 1.26:1. Listener MB from Rosen *et al.* (1998) had the same absolute threshold, but showed a compression ratio of just over 2:1 (right-most "×" in Fig. 4). It is tempting to speculate that these differences reflect mainly an OHC loss in listener CH and a loss not involving OHCs in MB, although other factors such as differences in middle-ear efficiency cannot be ruled out.

Subject MF shows evidence of a slightly expansive, rather than compressive, nonlinearity. The higher level filter shapes are slightly sharper than the lower level shapes. This in itself may be of little significance as a linear filter shape can fit the same data almost as well. However, there are two other points worth noting from MF's filter shapes in Fig. 3. First, the high-frequency side of the filter at 2 kHz appears to be abnormally steep. Second, the degree of compression in the filter measured at 1 kHz in this listener seems abnormally large (26.4-dB change in gain for a 40-dB change in probe level giving a compression ratio of 2.9:1). The reason for these differences is not clear.

Quantitative comparison of the selectivity of these listeners using the ERB clearly shows that at low stimulus levels the HI ears show less selectivity than NH ears. However, at higher levels the NH filters broaden and become more similar to those measured in impaired ears. The impaired ears themselves show a much smaller change in ERB with increasing stimulus level. One problem with using the ERB as a measure of selectivity is that it is dominated by the filter tip and is little affected by the skirt. Rosen *et al.* (1998) showed that the low-frequency skirt of the filter is an important part of the filter shape, as is also the case in psychoacoustic tuning curves (PTCs). At low stimulus levels the low-frequency skirt has little effect on the ERB. It is not until the stimulus reaches higher levels that the ERB increases significantly (as the tip-to-tail ratio becomes small). A similar effect is observed with PTCs. Nelson *et al.* (1990) showed that for forward-masked PTCs the $Q_{10\,dB}$ remains relatively constant at low levels, and only begins to decrease at high stimulus levels where the tip-to-tail ratio decreases. For the HI listeners in the present study, the effect of the impairment is to reduce the sharpness (and gain) of the tip at lower stimulus levels—resulting in a reduced tip–to–tail ratio, and hence a larger ERB than for NH listeners at the same level. At higher stimulus levels, the ERBs of the NH and HI are more similar.

The compression ratios derived from the fitted filter shapes and from the GOM functions in NH listeners are approximately 2:1. Such values are typical for estimates using simultaneous masking (e.g., Stelmachowicz *et al.*, 1987). In forward masking compression ratios are more typically of the order of 4 or 5:1 (Oxenham and Plack, 1997). This latter figure is more comparable to the compression ratios measured directly from the basilar membrane. Previous comparisons of selectivity measured using both forward and simultaneous masking at a single level suggest that a greater degree of nonlinearity may be revealed by forward masking than by simultaneous masking, and that using forward masking to compare NH and HI listeners will result in a larger difference in the derived compression ratios (e.g., Moore and Glasberg, 1986). Using the PolyFit procedure to characterize changes in selectivity across level using forward masking in NH listeners should reveal a greater degree of nonlinearity than is observed in simultaneous masking, and should lead to measurements of compression that are more directly comparable to those of Oxenham and Plack (1997) and to the physiological measurements.

V. CONCLUSIONS

In summary, we have used notched-noise masking and the PolyFit procedure to characterize auditory filter shapes as a function of stimulus level in mild/moderate hearing impairment. Comparison of the fitted filter shapes to previous results from NH listeners shows that the filter shapes in HI listeners are broader and change less with increasing stimulus level. Estimates of basilar-membrane input–output function slopes from masking functions and from the fitted filter shapes are in broad agreement with each other, suggesting compression ratios of approximately 2:1 in NH listeners and a reduced degree of compression in HI listeners.

ACKNOWLEDGMENTS

We are grateful to Dr. B. Cadge and the staff at the Adult Audiology Clinic, Royal National Throat Nose and Ear Hos-

pital, London for their help in recruiting listeners for this study. Most of all, we thank the listeners themselves for their participation in what are often tedious tasks. This work was supported by the Wellcome Trust (Grant No. 046823/Z/96). We would also like to thank Sid Bacon, Brian Moore, and an anonymous reviewer for their comments and suggestions.

Baker, R. J., Rosen, S., and Darling, A. M. (**1998**). ''An efficient characterization of human auditory filtering across level and frequency that is also physiologically reasonable,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. R. Palmer, A. Rees, A. Q. Summerfield, and R. Meddis (Whurr, London), pp. 81–87.

Baker, R. J., and Rosen, S. (**2001**). ''Evaluation of maximum-likelihood threshold estimation with tone-in noise masking,'' Br. J. Audiol. **35**, 43–52.

Bergman, M., Najenson, T., Korn, C., Nirit, H., Erenthal, P., and Sacharatov, E. (**1992**). ''Frequency selectivity as a potential measure of noise damage susceptibility,'' Br. J. Audiol. **26**, 15–22.

de Boer, E., and Bouwmeester, J. (**1974**). ''Critical bands and sensorineural hearing loss,'' Audiology **13**, 236–259.

Dallos, P. (**1992**). ''The active cochlea,'' J. Neurosci. **12**, 4575–4585.

Dallos, P., and Harris, D. M. (**1978**). ''Properties of auditory nerve responses in the absence of outer hair cells,'' J. Neurophysiol. **14**, 365–383.

Davis, H. (**1983**). ''An active process in cochlear mechanics,'' Hear. Res. **9**, 79–90.

Faulkner, A., Rosen, S., and Moore, B. C. J. (**1990**). ''Residual frequency selectivity in the profoundly hearing-impaired listener,'' Br. J. Audiol. **24**, 381–392.

Florentine, M., Buus, S., Scharf, B., and Zwicker, E. (**1980**). ''Frequency selectivity in normally-hearing and hearing impaired observers,'' J. Speech Hear. Res. **23**, 646–669.

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.

Glasberg, B. R., and Moore, B. C. J. (**2000**). ''Frequency selectivity as a function of level and frequency measured with uniformly exciting noise,'' J. Acoust. Soc. Am. **108**, 2318–2328.

Laroche, C., Hetu, R., Tran Quoc, H., Josserand, B., and Glasberg, B. (**1992**). ''Frequency selectivity in workers with noise-induced hearing loss,'' Hear. Res. **64**, 61–72.

Leek, M., 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.

Leshowitz, B., and Lindstrom, R. (**1977**). ''Measurement of nonlinearities in listeners with sensorineural hearing loss,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 283–292.

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

Lutfi, R. A., and Patterson, R. (**1984**). ''On the growth of masking asymmetry with stimulus intensity,'' J. Acoust. Soc. Am. **76**, 739–745.

Lutman, M. E., Gatehouse, S., and Worthington, A. G. (**1991**). ''Frequency resolution as a function of hearing threshold and age,'' J. Acoust. Soc. Am. **89**, 320–328.

Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (Oxford University Press, Oxford).

Moore, B. C. J. (**1998**). *Cochlear Hearing Loss* (Whurr, London).

Moore, B. C. J., and Glasberg, B. R. (**1986**). ''Comparisons of frequency selectivity in simultaneous and forward masking for subjects with unilateral cochlear impairments,'' J. Acoust. Soc. Am. **80**, 93–107.

Moore, B. C. J., and Glasberg, B. R. (**1987**). ''Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns,'' Hear. Res. **28**, 209–225.

Moore, B. C. J., Laurence, R. F., and Wright, D. (**1985**). ''Improvements in speech intelligibility in quiet and in noise produced by two-channel compression hearing aids,'' Br. J. Audiol. **19**, 175–187.

Murnane, O., and Turner, C. W. (**1991**). ''Growth of masking in sensorineural hearing loss,'' Audiology **30**, 275–285.

Nelson, D. A., Chargo, S. J., Kopun, J. G., and Freyman, R. L. (**1990**). ''Effects of stimulus level on forward-masked psychophysical tuning curves in quiet and in noise,'' J. Acoust. Soc. Am. **88**, 2143–2151.

Oxenham, A. J., and Plack, C. J. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.

Patterson, R. D. (**1971**). ''Effect of amplitude on auditory filter shape,'' J. Acoust. Soc. Am. **49**, 81a.

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

Patuzzi, R. (**1996**). ''Cochlear micromechanics and macromechanics,'' in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer, New York), pp. 186–257.

Pick, G. F. (**1980**). ''Level dependence of psychophysical frequency resolution and auditory filter shape,'' J. Acoust. Soc. Am. **68**, 1085–1095.

Pick, G. F., and Evans, E. F. (**1983**). ''Dissociation between frequency resolution and hearing threshold,'' in *Hearing—Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartman (Springer, Berlin), pp. 393–398.

Pick, G. F., Evans, E. F., and Wilson, J. P. (**1977**). ''Frequency resolution in patients with hearing loss of cochlear origin,'' in *Psychophysics and Physiology of Hearing*, edited by E. F. Evans and J. P. Wilson (Academic, London), pp. 273–281.

Rosen, S., Baker, R. J., and Kramer, S. (**1992**). ''Characterizing changes in auditory filter bandwidth as a function of level,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, K. Horner, and L. Demany (Pergamon, Oxford), pp. 171–177.

Rosen, S., and Baker, R. J. (**1994**). ''Characterizing auditory filter nonlinearity,'' Hear. Res. **73**, 231–243.

Rosen, S., Baker, R. J., and Darling, A. M. (**1998**). ''Auditory filter nonlinearity at 2 kHz in normal listeners,'' J. Acoust. Soc. Am. **103**, 2539–2550.

Ruggero, M. A. (**1991**). ''Physiology and coding of sound in the auditory nerve,'' in *The Mammalian Auditory Pathway: Neurophysiology*, edited by A. N. Popper and R. R. Fay (Springer, New York).

Sommers, M. S., and Humes, L. E. (**1993a**). ''Auditory-filter shapes in normal-hearing, noise-masked normal and elderly listeners,'' J. Acoust. Soc. Am. **93**, 2903–2913.

Sommers, M. S., and Humes, L. E. (**1993b**). Erratum: ''Auditory-filter shapes in normal-hearing, noise-masked normal and elderly listeners'' [J. Acoust. Soc. Am. **93**, 2903–2914], **94**, 2449–2450.

Stelmachowicz, P. G., Lewis, D. E., Larson, L. L., and Jesteadt, W. (**1987**). ''Growth of masking as a measure of response growth in hearing-impaired listeners,'' J. Acoust. Soc. Am. **81**, 1881–1887.

Tyler, R. S., Hall, J. W., Glasberg, B. R., Moore, B. C. J., and Patterson, R. D. (**1984**). ''Auditory filter asymmetry in the hearing impaired,'' J. Acoust. Soc. Am. **76**, 1363–1368.

Tyler, R. S., Wood, E. J., and Fernandes, M. (**1982**). ''Frequency resolution and hearing loss,'' Br. J. Audiol. **16**, 45–63.

Weber, D. L. (**1977**). ''Growth of masking and the auditory filter,'' J. Acoust. Soc. Am. **62**, 424–429.

West, P. D. B., and Evans, E. F. (**1990**). ''Early detection of hearing damage in young listeners resulting from exposure to amplified music,'' Br. J. Audiol. **24**, 89–103.

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.

Auditory stream segregation on the basis of amplitude-modulation rate[a)]

Nicolas Grimault[b)] and Sid P. Bacon
Psychoacoustic Laboratory, Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85287-1908

Christophe Micheyl
UMR CNRS 5020, Neurosciences & Systèmes Sensoriels, Université Claude Bernard Lyon 1, 50, avenue Tony Garnier, 69366 Lyon cedex 07, France

(Received 30 August 2001; revised 26 December 2001; accepted 28 December 2001)

In this study, auditory stream segregation based on differences in the rate of envelope fluctuations—in the absence of spectral and temporal fine structure cues—was tested. The temporal sequences to segregate were composed of fully amplitude-modulated (AM) bursts of broadband noises A and B. All sequences were built by the reiteration of a ABA triplet where A modulation rate was fixed at 100 Hz and B modulation rate was variable. The first experiment was devoted to measuring the threshold difference in AM rate leading subjects to perceive the sequence as two streams as opposed to just one. The results of this first experiment revealed that subjects generally perceived the sequences as a single perceptual stream when the difference in AM rate between the A and B noises was smaller than 0.75 oct, and as two streams when the difference was larger than about 1.00 oct. These streaming thresholds were found to be substantially larger than, and not related to, the subjects' modulation-rate discrimination thresholds. The results of a second experiment demonstrated that AM-rate-based streaming was adversely affected by decreases in AM depth, but that segregation remained possible as long as the AM of either the A or B noises was above the subject's AM-detection threshold. The results of a third experiment indicated that AM-rate-based streaming effects were still observed when the modulations applied to the A and B noises were set individually, either at a constant level in dB above AM-detection threshold, or at levels at which they were of the same perceived strength. This finding suggests that AM-rate-based streaming is not necessarily mediated by perceived differences in AM depth. Altogether, the results of this study indicate that sequential sounds can be segregated on the sole basis of differences in the rate of their temporal fluctuations in the absence of other temporal or spectral cues. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1452740]

PACS numbers: 43.66.Ba, 43.66.Jh, 43.66.Mk [NFV]

I. INTRODUCTION

Auditory stream segregation, or streaming, refers to a phenomenon whereby a temporal sequence of sounds is organized into two or more perceptual auditory entities or ''streams,'' which can be individually followed by listeners and are generally considered as arising from different acoustic sources (Miller and Heise, 1950; Bregman, 1978). The voice of a single person among a crowd, or the sound of a violin among those of other instruments in the orchestra, provide everyday examples of auditory streams. In laboratory situations, streaming is generally studied using the stimulus paradigm devised by Van Noorden (1975). In this paradigm, listeners are presented with sequences of iterated ABA triplets where A and B represent different sounds. In the early work of Van Noorden, A and B were pure tones differing in frequency. Van Noorden observed that when the frequency difference between A and B was small and the tempo was low, listeners generally perceived the sequence of ABA triplets as a single melodic stream. In contrast, when the frequency difference was large and the tempo fast enough, the sequence was perceived as consisting of two streams, one formed by the A tones, the other by the B tones. Van Noorden defined the temporal coherence boundary (TCB) as the point above which at least two streams are always heard, and the fission boundary (FB) as the point below which only one stream is always heard.

Following these early demonstrations of auditory streaming, several studies have shown that streaming could be elicited using stimuli other than pure tones and/or parameters other than frequency. In particular, streaming has been obtained with harmonic complex tones differing in their fundamental frequency ($F0$) (Singh, 1987; Bregman *et al.*, 1990) or frequency region (Bregman and Levitan, 1983), as well as with tones or bands of noise differing in their temporal or spectral characteristics (Bregman and Levitan, 1983; Bregman, 1990; Bregman *et al.*, 1999, 2001; Cusack and Roberts, 2000).

In view of the vast diversity of stimulus parameters ap-

[a)]Portions of this research were presented at the 141st meeting of the Acoustical Society of America [N. Grimault, S. P. Bacon, and C. Micheyl, ''Auditory streaming without spectral cues,'' J. Acoust. Soc. Am. **109**, 2288(A) (2001)].

[b)]Author to whom correspondence should be addressed, UMR CNRS 5020, Université Claude Bernard Lyon 1 50, avenue Tony Garnier, 69366 Lyon cedex 07, France. Electronic mail: nicolas.grimault@olfac.univ-lyon1.fr

 © 2002 Acoustical Society of America

parently able to promote streaming, it soon became important to determine the essential underlying factors and mechanisms of streaming. Hartmann and Johnson (1991) were the first to propose that streaming depends, primarily, on peripheral auditory filtering. Schematically, sounds that successively excite the same sets of auditory channels are perceived as a single stream; sounds that excite different peripheral auditory channels are assigned to different perceptual streams. This so-called ''channeling theory'' of streaming can account for most of the experimental data obtained in the studies cited above. And even if the results cannot be accounted for by a strict version of the channeling theory, they presumably can be accounted for by a ''milder'' version of it, according to which sounds that produce substantially different excitation patterns in the peripheral auditory system form different streams, whereas sounds that produce similar peripheral excitation patterns fall in the same stream. The latter view has recently received support from experimental data obtained in both psychoacoustical (Rose and Moore, 1997, 2000) and neurophysiological (Fishman *et al.*, 2001) investigations. Furthermore, the channeling theory has inspired computer models of auditory streaming, which can successfully account for a large number of experimental results (Beauvois and Meddis, 1996; McCabe and Denham, 1997).

There are, however, various results in the literature that cannot be accounted for by the channeling theory. In particular, several studies conducted over the past four years have convincingly demonstrated that sequences of harmonic complex tones exciting the same auditory filters can be perceptually segregated based on differences in $F0$ (Vliegen and Oxenham, 1999; Vliegen *et al.*, 1999; Grimault *et al.*, 2000, 2001). The results of these studies are striking as they included conditions in which the stimuli, although having a different $F0$, elicited very similar spectral patterns of excitation in the peripheral auditory system. This was achieved by using complex tones whose harmonics were filtered in the same frequency region and were not resolved by the auditory periphery. Furthermore, a continuous background noise was added in order to mask distortion products generated in the cochlea, which might have introduced $F0$-related spectral cues. Therefore, the results of these studies not only demonstrate that sounds which successively excite the same peripheral auditory filters can be perceptually segregated, they also indicate that streaming can occur in the absence of spectrum-related changes in the activity of the peripheral auditory system over time.

The next logical step consists of trying to determine in more detail which cues other than spectral cues could mediate the perceptual segregation of sounds. The only known alternative to spectral encoding is temporal encoding, and thus it seems likely that when streaming is not mediated by spectral cues, it is mediated by temporal cues. However, it remains to be determined exactly which types of temporal cues can mediate streaming. Two main types of temporal information that can be encoded in the auditory system are fine structure and envelope. Both types of temporal information could potentially mediate streaming. In the present study, we focused on investigating the role of the latter. The first experiment was devoted to testing whether changes in the rate of temporal envelope fluctuations could give rise to streaming. The purpose of the second experiment was to determine the influence of modulation depth on the modulation-rate-based streaming effects observed in the first experiment. Finally, the third experiment was performed in order to test whether the streaming effects obtained in the first two experiments were mediated by perceived differences in modulation strength.

II. EXPERIMENT 1: SEGREGATION OF FULLY MODULATED BROADBAND NOISES

A. Rationale

Previous studies (Vliegen and Oxenham, 1999; Vliegen *et al.*, 1999; Grimault *et al.*, 2000) have shown that streaming can be mediated solely by temporal cues. The characteristics of the unresolved harmonic complexes used in these earlier studies were such that both the temporal fine structure and the temporal envelope could be encoded in the auditory system. In experiment 1, we tested whether perceptual streaming can occur solely on the basis of differences in the temporal envelope of successive sounds. To prevent listeners from using spectral cues, we used broadband noise carriers. Envelope cues were produced by modulating the amplitude of the stimuli at different rates. If streaming can be observed with such stimuli, it will demonstrate that streaming can be induced solely on the basis of temporal envelope fluctuations.

B. Subjects

Five subjects (three females and two males) aged between 22 and 30 years participated. All but one had absolute thresholds in the test (right) ear at or below 15 dB HL (ANSI, 1996) at octave frequencies between 125 and 8000 Hz. The remaining subject (NG, first author) had a slight hearing loss at 8000 Hz (20 dB HL). Three subjects (CG, SH, and JB) were paid for their participation, whereas subject DE received university credit. Two subjects (NG and CG) had previous experience in psychoacoustic tasks, including streaming tasks. Indeed, they participated in Grimault *et al.* (2000). All the other subjects had no prior experience in streaming tasks, and thus they took part in a preliminary 1.5-h familiarization session. During this session, they performed a classical streaming task involving a constant-stimulus procedure and iterated sequences of both pure tones and of the same amplitude-modulated broadband noises that were used in the first experiment. The actual experiment itself was completed in two or three test sessions of approximately 1.5 h each.

C. Stimuli and apparatus

The stimuli used in this experiment consisted of temporal sequences of sinusoidally amplitude-modulated bursts of broadband noise. The stimulus sequences were formed by a repeating ABA-pattern, where A and B corresponded to bursts having different modulation rates, and - represented a 20-ms silent interval. The modulation depth was maintained at 100% throughout the experiment. Each burst was 100 ms

in duration, including rise and fall cosine ramps of 10 ms each. The stimuli were generated digitally and recorded on a CD-R prior to testing. The bursts used in this experiment were presented at a spectrum level of 50 dB SPL.

The stimuli were played through a 16-bit digital-to-analog converter (TDT DD1) at a sampling rate of 44.1 kHz and passed through an anti-aliasing filter (TDT FT6-2; −60 dB at 1.15 times the corner frequency) with a corner frequency set to 8 kHz. A continuous background noise was added in order to mask the perception of distortion products which could have been elicited by the amplitude-modulated noise stimuli (Wiegrebe and Patterson, 1999), even though the perception of such distortion products was unlikely. This masker was produced by successively feeding the output of a white noise generator (TDT WG1) to an anti-aliasing filter (TDT FT6-2) with a corner frequency of 8 kHz and to a low-pass filter (TDT PF1; −25 dB at 1.15 times the corner frequency) with a corner frequency of 900 Hz (i.e., 100 Hz above the highest modulation rate used in the experiment). The spectrum level of the masker (within the pass band) was 50 dB SPL.

The signals and the background noise were independently led to two separate programmable attenuators (TDT PA4). The outputs of the attenuators were summed (TDT SM3) and led to the right earpiece of Sony MDR 7506 headphones via a preamplifier (TDT HBC) and a resistor box aimed to prevent electrical cross talk.

D. Procedures

Two different experimental procedures were used for the measurement of perceptual stream segregation in this experiment. The first procedure was a traditional constant-stimulus procedure. In this procedure, 13 different iterating ABA-stimulus sequences, corresponding to 13 different modulation rates of the B noises, were used. The modulation rate of the B noise varied between 100 and 800 Hz across sequences. Its actual value could be 0, 0.3, 0.5, 0.7, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2, 2.5, or 3 oct above the modulation rate of the A noise. The overall sequence duration was fixed at 6.4 s. The modulation rate of the A noise was constant at 100 Hz throughout the experiment. Following the presentation of each sequence, the subject was required to indicate on a computer keyboard if one or two streams had been heard. The 13 conditions were presented 40 times, in four blocks, in random order.

The second experimental procedure was devised to obtain reliable estimates of the TCB and FB in a short period of time. This procedure will be referred to as the ''timing procedure.'' It involved the use of repeating ABA-sequences wherein the modulation rate of A was constant at 100 Hz, and the modulation rate of B varied either from 100 to 800 Hz (blocks of ''increasing'' sequence) or from 800 to 100 Hz (blocks of ''decreasing'' sequence). Ten different variation rates were used, leading to ten different sequence durations, ranging from 10.88 to 33.6 s. The different variation rates were produced by varying, simultaneously, the number of consecutive ABA-triplets in which the amplitude modulation (AM) rate of the B noises remained constant (from 1 to 5), and the variation step (from 0.03 to 0.3 oct). Each of the ten sequences was presented three times, in one block, in a pseudo-randomized order. Subjects were required to press a button on a response box (TDT) as many times as they wanted during the presentation of the sound sequence to indicate whether they heard one stream or two. Each time the button corresponding to the perception of one or two streams was pressed, the current modulation rate of the B noise was recorded. The subjects produced on average around one response per second. Based on these responses, two distributions were defined. The distribution of one-stream (two-streams) responses was the number of one-stream (two-streams) responses across the ten variation-rate conditions and three repetitions distributed in 0.1-oct-wide bins. The FB was then estimated from the distributions of responses obtained with the ''decreasing'' sequences, and the TCB was estimated from the distributions of responses obtained with the ''increasing'' sequences. Each estimate of these boundaries using this timing procedure was based on approximately 660 responses, whereas the threshold estimate was based on approximately 520 responses using the constant-stimulus procedure.

AM-rate discrimination thresholds (AMRDTs) were also measured. Threshold estimates were obtained using a three-interval, two-alternative, forced-choice procedure with a two-down one-up adaptive rule tracking the 70.7% correct point on the psychometric function (Levitt, 1971). On each trial, following a reference interval containing an AM noise burst with a rate of 100 Hz, the subject was presented two successive AM noise bursts, one with an AM rate of 100 Hz and the other with an AM rate of $100+\Delta F$ Hz, in random order. The subject picked the interval which was thought to contain the higher modulation rate. Differences in modulation rate were increased or decreased by a factor of 2 until the fourth turnpoint and by $\sqrt{2}$ thereafter. The procedure stopped after 16 turnpoints. The threshold was estimated as the geometric mean of the last 12 turnpoints, expressed as a percentage of the nominal modulation rate. Trial-by-trial visual feedback was provided. The noise bursts had the same spectral and temporal characteristics as those used in the streaming task.

E. Results and discussion

The results from the constant-stimulus procedure were similar across subjects, and thus Fig. 1 shows how the proportion of two-stream responses averaged across subjects varied as a function of the AM rate difference between the A and B noises. These average data were fitted using a two-parameter cumulative Gaussian function (Dai, 1995). The resulting function was used to estimate the streaming threshold, defined as the *x*-value yielding 0.5 proportion of two-streams responses. A one way repeated-measures ANOVA, applied to the individual data, revealed that the subjects' responses depended in a highly significant way on the AM-rate difference [$F(12,48)=60.58$, $p<0.001$]. These results are in very good agreement with the idea that it is possible to segregate sequential sounds on the basis of differences in the rate of the temporal fluctuations that they contain, in the absence of spectral or other temporal cues. Interestingly, the low proportion of two-streams responses with a one-octave

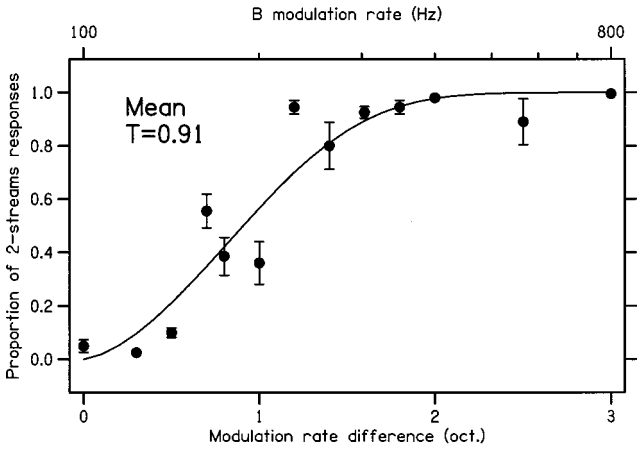


FIG. 1. Proportion of two-stream responses as a function of the modulation rate difference between A and B noises in the constant-stimulus procedure (see text for details). The modulation rate of A was 100 Hz, and the modulation rate of B varied from 100 to 800 Hz. The vertical bars represent the standard error across subjects. A two-parameter cumulative Gaussian function was fitted to the data (solid line) and used to extrapolate the streaming threshold (*T*).

modulation-rate difference could be related to the octave error phenomenon reported in a pitch judgment experiment with interrupted noises (Pollack, 1968). That would indicate that pitch *per se* rather than modulation rate *per se* was the basis for segregation.

AMRDTs were measured to determine whether the streaming thresholds induced by AM-rate differences were related to the subjects' ability to perceive differences in AM rate. Table I gives the streaming thresholds and the AMRDTs. Streaming thresholds varied across subjects by only 0.16 oct. Note that the mean (0.89 oct) of the individual thresholds differs slightly from the threshold (0.91) derived from the average results shown in Fig. 1. The measured AMRDTs[1] were significantly smaller than the streaming thresholds $[t(4)=16,\ p<0.001]$ and not correlated with them (Pearson's $r=-0.72$, $p=0.17$). The lack of a significant relationship between the two types of thresholds suggests that streaming induced by differences in AM rate is not related to fine AM-rate discrimination performance. This is consistent with the fact that the frequency difference necessary for iterated ABA-sequences of pure tones to be heard as two streams is much larger than the frequency difference limens at the considered frequencies.

Figure 2 shows the individual results obtained using the timing procedure. Each panel represents the difference between the distributions of two-streams and one-stream responses for a given subject. The first and second columns respectively correspond to the increasing and decreasing sequence conditions. The difference distributions were fitted using a cumulative Gaussian function with three parameters: asymptotic value, slope, and threshold. The resulting functions were used to estimate the streaming threshold, defined as the point at which the fitting curve crossed the *x*-axis.

A paired *t* test $[t(4)=4.51,\ p=0.01]$ applied to the data in Fig. 2 demonstrates that the individual thresholds measured using the increasing sequences were always higher than those measured using the decreasing sequences. This observation may be explained by the fact that subjects "held on" to the initial percept (i.e., that of one stream for increasing sequences and that of two streams for decreasing sequences) for as long as possible. For all subjects but one (DE), the streaming thresholds estimated previously using the constant-stimulus procedure (Table I) were between the two thresholds measured using the timing procedure (Fig. 2). The mean thresholds for increasing and decreasing sequences in the timing procedure were 1.03 (S.E.=0.06) and 0.75 oct (S.E.=0.03), respectively; the mean threshold for the constant-stimulus procedure was 0.89 oct (S.E.=0.03). Van Noorden (1975) has described the effect of tone repetition time on both the FB and the TCB. A repetition time of 100 ms led to a distance of about 2.27 semitones between FB and TCB. With modulated noises, the distance between the mean thresholds for increasing and decreasing sequences was found to be equal to 2.26 semitones. These observations are consistent with the notion that the streaming thresholds measured using the decreasing and increasing sequences reflected the FB and TCB, respectively, and that the streaming thresholds measured with the constant-stimulus procedure reflected the use of a neutral internal criterion.

The results from both the constant-stimulus procedure and the timing procedure are in very good agreement with the idea that envelope cues are a useful feature for streaming. It is of interest to compare the AM-rate-based streaming results from the current experiment with F0-based streaming results of Grimault *et al.* (2000). The data from their three unresolved conditions[2] have been reanalyzed using the same fitting method that was used in the current experiment. The average streaming threshold estimated this way from their data was 0.94 oct (S.E.=0.07). This is very close to that obtained in the present study using the same constant-stimulus procedure. This close correspondence between the streaming thresholds measured with unresolved harmonics differing in F0 and broadband noises differing in AM rate might be interpreted as reflecting the fact that streaming in both cases relies upon envelope cues. With unresolved harmonics, the envelope cue could come from the fact that the output of the auditory filter that processes two or more har-

TABLE I. Individual streaming thresholds in octaves obtained with the constant-stimulus procedure (first row) and individual amplitude modulation rate discrimination thresholds (AMRDT). The standard error of the AMRDT measurements was 0.02 or smaller for each subject. The two last columns show the mean and the standard error (S.E.) across subjects.

	NG	CG	DE	SH	JB	Mean	S.E.
Constant stimulus procedure (oct)	0.98	0.82	0.91	0.92	0.84	0.89	0.03
AMRDT (oct)	0.07	0.14	0.11	0.16	0.2	0.14	0.02

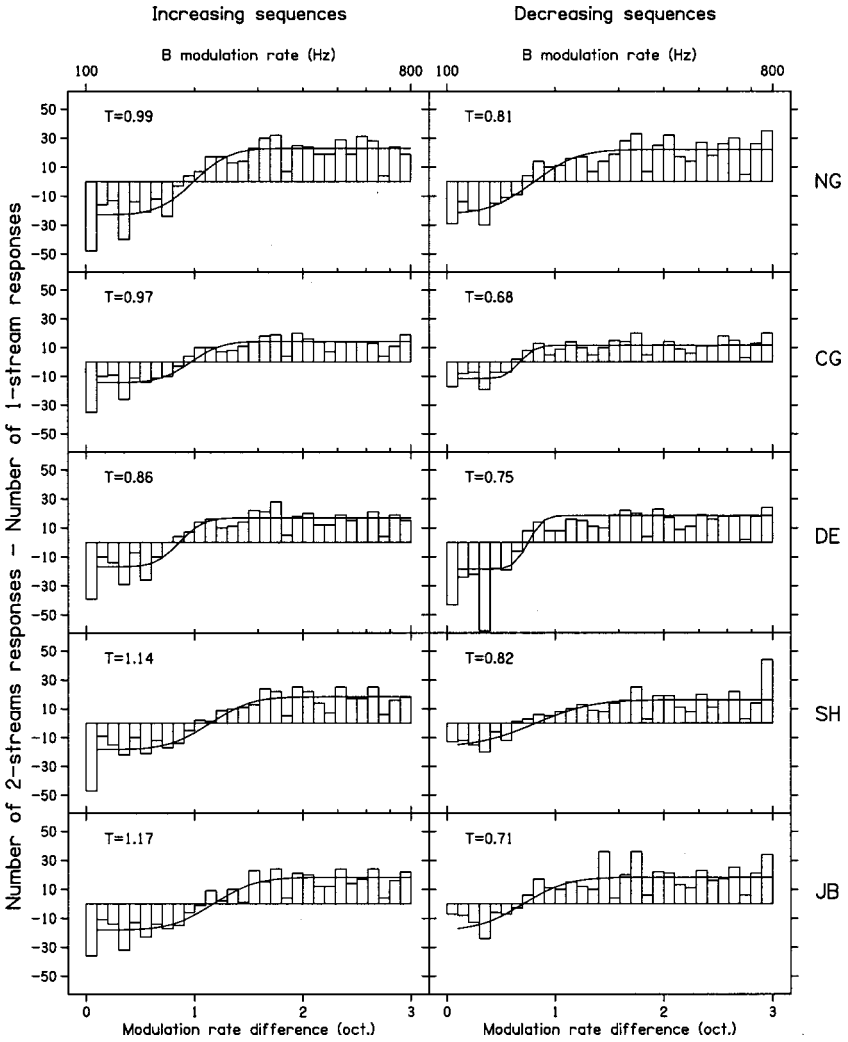


FIG. 2. Distribution in 0.1-oct-wide bins of the two-stream responses minus the one-stream responses for each subject (rows) and for both increasing (first column) and decreasing (second column) sequences (see text for details). In each panel, a three-parameter cumulative Gaussian function was fitted to the data (solid line) and used to extrapolate the streaming threshold (*T*).

monics may fluctuate in amplitude at a rate equal to *F*0. We cannot exclude the possibility, however, that the streaming with unresolved harmonics involved the use of fine-structure cues. The fact that the pitch of sinusoidally amplitude-modulated noise (Burns and Viemeister, 1976, 1981) is generally weaker than the pitch of unresolved harmonic complexes—despite the fact that the two stimuli have similar streaming thresholds—raises the possibility that modulation rate rather than pitch *per se* is fundamentally important to the streaming of either AM noises or unresolved harmonic complexes.

III. EXPERIMENT 2: EFFECT OF MODULATION DEPTH

A. Rationale

The results obtained in experiment 1 indicate that differences in the rate of temporal envelope fluctuations can be used for the purpose of auditory stream segregation in the absence of other temporal or spectral cues. The second experiment was aimed at investigating the effects of modulation depth on these AM-rate-based streaming effects. If such effects are present only at large modulation depths, like the one used in experiment 1, they should be of very limited use in daily life. In contrast, if streaming can be achieved with small modulation depths, this would argue for their potential usefulness in the natural auditory environment.

B. Subjects

The same five subjects participated in this experiment. All measurements were completed in two to three test sessions of approximately 1.5 h each.

C. Stimuli, apparatus and procedure

The stimuli, apparatus, and timing procedure were similar to those in experiment 1, except that the modulation depth (*m*) of the A and B noise bursts varied from 0.8 (80%) to 0.2 (20%) in steps of 0.2 (20%). An arbitrary value in dB, defined as $20 \log (m)$ was associated with each linear value of *m*. The modulation depth varied across blocks in a pseudo-randomized order.

Modulation detection thresholds (MDT) were also measured, using noise bursts with identical spectral and temporal characteristics as those used for the measurement of streaming thresholds. The two-interval, two-alternative, forced-choice procedure was the same as in Bacon and Gleitman (1992).

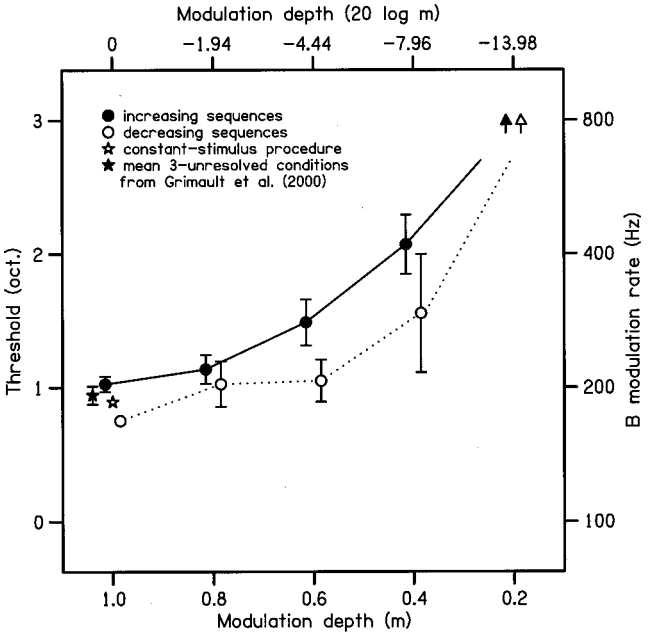


FIG. 3. Streaming thresholds in octaves obtained from either the timing procedure with increasing sequences (filled circles) or decreasing sequences (unfilled circles) or the constant-stimulus procedure (unfilled star) as a function of the modulation depth of A and B noises. The two arrows indicate that thresholds could not be measured, even at the largest difference in rate. In all these conditions, the vertical bars represent the standard error across subjects. Previous results from Grimault *et al.* (2000) have also been plotted (filled star). This data point corresponds to the mean streaming threshold from their three unresolved complex tone conditions. The vertical bars associated with this particular data point represent the standard error across experimental conditions.

D. Results and discussion

Figure 3 shows the mean streaming thresholds as a function of modulation depth for the increasing (filled circles) and decreasing (unfilled circles) sequences. These thresholds were estimated using the same fitting procedure as in experiment 1. The data corresponding to $m=1$ are replotted from experiment 1. The threshold from experiment 1 using the constant-stimulus procedure is shown as an unfilled star. The mean streaming thresholds obtained for unresolved harmonics in Grimault *et al.* (2000), discussed in the context of experiment 1, are represented by filled stars. The filled and unfilled arrows (at 3 oct or 800 Hz) indicate cases where the fitting procedure failed to provide a reliable threshold estimate, because the difference between the two response distributions was almost flat for 7 out of 10 conditions (5 subjects$\times$2 sequence directions). In such cases, it is unclear whether streaming would have occurred for larger modulation rate differences.

A two-way repeated-measures ANOVA (modulation depth$\times$sequence direction) applied to the data obtained with the timing procedure ($m=1.0-0.2$) revealed a highly significant effect of modulation depth [$F(3,12)=11.69$, $p<0.001$]: the smaller the depth, the smaller the proportion of two-stream responses. Although the thresholds from the increasing sequences were higher than those from the decreasing sequences in 14 out of 20 conditions, the main effect of direction of modulation rate variation was not significant [$F(1,4)=2.60$, $p=0.18$].

TABLE II. Group mean modulation detection threshold (MDTs) in dB for each modulation rate. Also shown is the standard error (S.E.) and the number (N) of subjects contributing to each threshold.

Modulation rate (Hz)	100	200	400	800
MDT (dB)	−13.81	−10.69	−8.49	−5.66
S.E. (dB)	1.43	2.67	2.11	0.88
N	5	5	4	2

The modulation detection thresholds averaged across subjects are shown in Table II. The third line of this table indicates the number of subjects contributing to the modulation detection threshold. At rates above 200 Hz, some subjects could not reliably detect modulation.[3]

The comparison between the modulation detection thresholds reported in Table II and the streaming data plotted in Fig. 3 reveals, not surprisingly, that when the modulations of the A and B noises were not detectable, thresholds could not be measured. More interestingly, it also reveals that the subjects were able to segregate the A and B noises as long as the modulation applied to either was detectable. For example, when the modulation applied to the A noises was detectable but that applied to the B noises was at detection threshold or even below (namely, in the condition where $m=0.4$), the subjects were still able to segregate the A and B noises. In this situation, streaming may have been mediated by a timbre difference between the A and B noises related to the fact that the former were warbling whereas the latter where not. The robustness of streaming with envelope cues suggests the potential usefulness of these cues in the natural auditory environment.

IV. EXPERIMENT 3: EFFECT OF MODULATION STRENGTH

A. Rationale

In the two previous experiments, streaming could either depend directly on differences in modulation rate between the A and B noises, or it could be mediated by differences in their perceived strength of modulation, which covaried with modulation rate differences (Fastl, 1982; Daniel and Weber, 1997). In this last experiment, two different approaches were used in an attempt to partial out the potential influence of modulation strength. The first approach involved the generation of stimulus sequences specifically tailored for each subject with the modulation depth of the A and B noises set at a constant number of dB above the subject's modulation-detection threshold at the considered rate. However, a fixed dB sensation level (SL) may not correspond to the same modulation strength at each modulation rate, and thus the second approach involved setting the modulation depths of the A and B noises for each subject in such a way that their modulation was judged to be of equal strength.

B. Subjects

Respectively four (NG, CG, DE, and SH) and three (NG, CG and DE) of the subjects of experiment 1 partici-

TABLE III. Individual streaming thresholds from experiment 3 obtained with the timing procedure with both increasing (column 8) and decreasing (column 9) sequences. The sequences have been built using the first or the second method (column 2) described in the text in order to equalize the modulation strength of A and B within each sequence. The individual SLs (method 1) are shown in column 3; the modulation strength (method 2) is shown in column 4. The modulation depth values for the two methods are indicated in columns 5–7.

Subject	Method	dB SL	Modulation strength	Modulation depth (20 log *m*)			Streaming threshold (oct)	
				100 Hz	200 Hz	400 Hz	Increasing	Decreasing
NG	1	10	···	−6.5	−11.2	−2.7	1.51	1.27
	2	···	2.2	−6.7	−3.0	−0.3	1.01	0.96
CG	1	6.9	···	−7.0	−1.2	0.0	1.55	0.91
	2	···	2.5	−9.5	−7.0	−2.0	1.49	0.92
DE	1	7.4	···	−9.9	0.0	−3.7	0.68	0.48
	2	···	2.0	−9.6	−6.8	0.0	1.17	0.89
SH	1	3.3	···	−6.4	−6.7	0	0.75	0.99
	2	···	···	···	···	···	···	···

pated in the first and second approaches. All measurements were completed in one to two test sessions of approximately 1.5 h each.

C. Stimuli, apparatus, and procedure

The stimuli, apparatus, and procedure were the same or similar to those used in experiment 1. The timing procedure described earlier was used with increasing and decreasing sequences where the modulation rate of the B noise varied either from 100 to 400 Hz or from 400 to 100 Hz. Different stimulus sequences were generated for the different conditions and the different subjects, and recorded on individual CD-Rs. The modulation depth, expressed in dB, was no longer constant within each sequence. In the first condition, it was set at a constant value in dB above the individual modulation detection threshold, measured in experiment 2, at the considered AM rate (100, 200, and 400 Hz). Modulation detection thresholds corresponding to other modulation rates of the B noises were estimated using linear interpolation. The individual SLs and the corresponding modulation depths are shown in Table III. These SLs were individually chosen to be close to the absolute value of the larger (worst) detection threshold in dB in the range 100–400 Hz. In the second condition, the modulation depths of the A and B noises were set so as to produce equal and constant modulation strength sensations in all modulation-rate conditions. To achieve this, subjects were first required to rate the strength of modulations applied to the same kind of noise bursts used in previous experiments. To provide the subjects with a comparison reference, the bursts of modulated noise were always preceded by a burst of unmodulated noise with an assigned arbitrary strength of 1. This was done at modulation rates of 100, 200, 400, and 800 Hz, as well as at modulation depths ranging from −22 dB (8%) to 0 dB (100%). All conditions were presented three times, in random order, in a single block. Example results obtained using this scaling procedure are shown in Fig. 4 for subject NG. On the basis of these results, a given modulation strength was then chosen for each subject as approximately the highest modulation strength value common to the curves for 100, 200, and 400 Hz (the dotted line in Fig. 4). In cases where the curve for a given rate was nonmonotonic and crossed this modulation strength value more than once, the lowest modulation depth where it crossed was chosen. Here 800 Hz was excluded because the modulation was often not detectable, and the modulation strength was generally too low. The modulation strength values used for the different subjects are reported in Table III, along with the corresponding modulation depths. The modulation depths corresponding to intermediate modulation rates were estimated by means of linear interpolation.

D. Results and discussion

The individual streaming thresholds are shown in Table III. Despite the fact that A and B noises evoked identical or similar modulation strength sensations, all subjects consistently reported hearing two streams in some of the conditions tested. This suggests that streaming can occur on the basis of

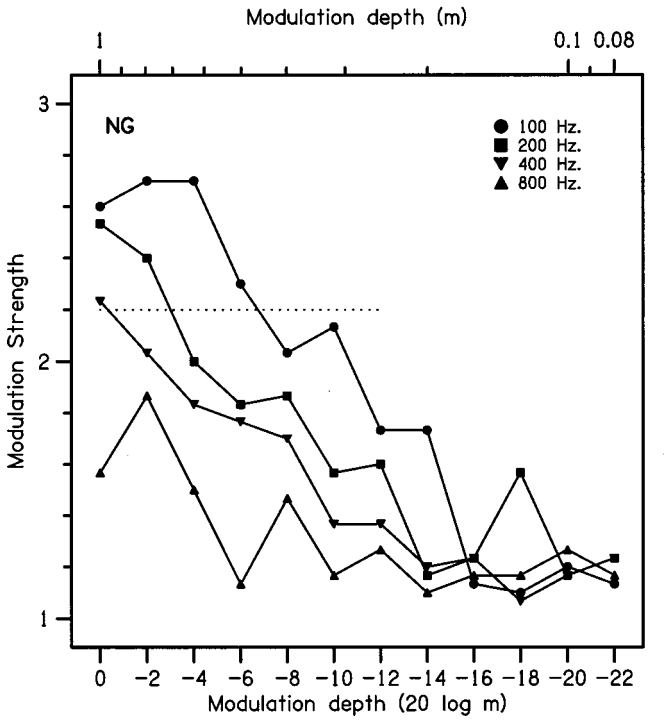


FIG. 4. Modulation strength curves for subject NG with modulation rates from 100 to 800 Hz (see text for details). The horizontal dotted line corresponds to the modulation strength used in experiment 3.

modulation rate in the absence of salient differences in modulation strength, although it is possible that slight differences in perceived modulation strength between A and B noises still existed. Altogether, the results of this and the previous two experiments strongly suggest that streaming can occur on the basis of differences in the rate of temporal envelope fluctuations.

V. GENERAL DISCUSSION

The results of this study indicate that it is possible to segregate sequential sounds on the sole basis of differences in the rate of the temporal fluctuations that they contain, in the absence of spectral or other temporal cues. This effect is still present when there are no physical or perceived differences in AM depth, suggesting that it is based on differences in the rate of envelope fluctuations *per se*. However, this does not imply that differences in the depth or strength of envelope fluctuations may not be used for the purpose of stream segregation; this possibility still needs to be tested in future studies.

The streaming effect was adversely affected by decreases in the depth of the amplitude fluctuations. The minimum difference in AM rate required for the subjects to hear the sequences as two streams increased with decreases in modulation depth. However, the effect was very robust in the sense that sequential segregation remained possible as long as subjects could detect the modulation applied to either of the two stimuli making up the sequences. This suggests that rate-based streaming may be possible in the natural environment, where the depth of envelope fluctuations in signals, like speech, may be attenuated by background noise or reflections.

The results of this study have implications for theories and models of auditory stream segregation. From a general point of view, these results cannot be accounted for by the channeling theory of streaming, according to which stream segregation can only occur when sounds excite different peripheral auditory filters. Indeed, in the present study, stream segregation was found to be possible with sounds that successively excited the same sets of peripheral auditory channels. Furthermore, since the long-term spectra of the stimuli used in this study were identical, the results generally argue against theories or models of auditory streaming based exclusively on long-term representations of the power spectra of the stimuli.

The present results suggest that future models of auditory stream segregation might involve channels in the modulation domain (Dau *et al.*, 1997a, b; Ewert and Dau, 2000), in addition to channels in the spectral domain. Within the framework of such a refined model, stream segregation might be predicted to occur when successive sounds excite different spectral or modulation channels.

ACKNOWLEDGMENTS

Special thanks to Douglas Edge for help running subjects, to Bob Carlyon for the original idea of the timing procedure for the measurement of streaming thresholds, and to Steve McAdams and Christian Lorenzi for helpful remarks on the results of this study. Finally, we thank two anonymous reviewers who helped us improve the article's quality. This work was supported by a post-doctoral fellowship from the Fondation pour la Recherche Médicale (FRM) awarded to the first author and a grant from NIDCD (DC01376) awarded to the second author.

[1]Using broadband noise carriers, Formby (1985) obtained AMRDTs of about 0.07–0.08 oct at 100 Hz. The slightly higher thresholds in the current experiment (between 0.07 and 0.2 oct) might be explained either by shorter stimulus duration (100 ms here versus 500 ms in Formby's study) or by less training (no training in rate discrimination here versus 1.5 h of training in Formby's study). This latter explanation is consistent with the fact that the lowest threshold (0.07 oct) obtained here was for the only subject (NG) who had previous experience in modulation rate discrimination.

[2]The three unresolved complex tones used by Grimault *et al.* (2000) were generated using a $F0$ of 88 Hz and filtering regions MID (1375–1875 Hz) and HIGH (3900–5400 Hz) or a $F0$ of 250 Hz and the same HIGH filtering region.

[3]Following data description of MDT with broadband noise carriers in the literature (e.g., Viemeister, 1979; Bacon and Gleitman, 1992), we could expect thresholds between about −25 and −15 dB at 100 Hz and between about −10 and −5 dB at 800 Hz. The thresholds measured in the current study are then higher (worse) than those former estimations. This could be presumably due to the shorter duration, although it is worth noting that, for these higher rates, Viemeister (1979, Fig. 6) did not see an effect of duration when the gated carrier was reduced from 500 to 250 ms. This, of course, does not mean that a further reduction (below 250 ms) would not have an effect.

American National Standards Institute (**1996**). ANSI S3.6-1996, "Specifications for audiometers" (ANSI, New York).

Bacon, S. P., and Gleitman, R. M. (**1992**). "Modulation detection in subjects with relatively flat hearing losses," J. Speech Hear. Res. **35**, 642–653.

Beauvois, M. W., and Meddis, R. (**1996**). "Computer simulation of auditory stream segregation in alternating-tone sequences," J. Acoust. Soc. Am. **99**, 2270–2280.

Bregman, A. S. (**1978**). "Auditory streaming is cumulative," J. Exp. Psychol. Hum. Percept. Perform. **4**, 380–387.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).

Bregman, A. S., and Levittan, R. (**1983**). "Stream segregation based on fundamental frequency and spectral peak. I: Effects of shaping by filters," unpublished manuscript, Psychology Department, McGill University.

Bregman, A. S., Ahad, P. A., and Van Loon, C. (**2001**). "Stream segregation of narrow-band noise bursts," Percept. Psychophys. **63**, 790–797.

Bregman, A. S., Colantonio, C., and Ahad, P. A. (**1999**). "Is a common grouping mechanism involved in the phenomena of illusory continuity and stream segregation?" Percept. Psychophys. **61**, 195–205.

Bregman, A. S., Liao, C., and Levitan, R. (**1990**). "Auditory grouping based on fundamental frequency and formant peak frequency," Can. J. Psychol. **44**, 400–413.

Burns, E. M., and Viemeister, N. F. (**1976**). "Nonspectral pitch," J. Acoust. Soc. Am. **60**, 863–869.

Burns, E. M., and Viemeister, N. F. (**1981**). "Played-again SAM: Further observations on the pitch of amplitude modulated noise," J. Acoust. Soc. Am. **70**, 1655–1660.

Cusack, R., and Roberts, B. (**2000**). "Effects of differences in timbre on sequential grouping," Percept. Psychophys. **62**, 1112–1120.

Dai, H. (**1995**). "On measuring psychometric functions: A comparison of the constant-stimulus and adaptive up-down methods," J. Acoust. Soc. Am. **98**, 3135–3139.

Daniel, P., and Weber, R. (**1997**). "Psychoacoustical roughness: Implementation of an optimized model," Acustica **83**, 113–123.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997a**). "Modeling auditory processing of amplitude modulation. I. Modulation detection and masking with narrowband carriers," J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997b**). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration in modulation detection," J. Acoust. Soc. Am. **102**, 2906–2919.

Ewert, S. D., and Dau, T. (**2000**). "Characterizing frequency selectivity for envelope fluctuations," J. Acoust. Soc. Am. **108**, 1181–1196.

Fastl, H. (**1982**). ''Fluctuation strength and temporal masking patterns of amplitude-modulated broadband noise,'' Hear. Res. **8**, 59–69.
Fishman, Y. I., Reser, D. H., Arezzo, J. C., and Steinschneider, M. (**2001**). ''Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey,'' Hear. Res. **151**, 167–187.
Formby, C. (**1985**). ''Differential sensitivity to tonal frequency and to the rate of amplitude modulation of broadband noise by normally hearing listeners,'' J. Acoust. Soc. Am. **78**, 70–77.
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.
Grimault, N., Micheyl, C., Carlyon, R. P., Arthaud, P., and Collet, L. (**2001**). ''Perceptual auditory stream segregation of sequences of complex sounds in subjects with normal and impaired hearing,'' Br. J. Audiol. **35**, 173–182.
Hartmann, W. M., and Johnson, D. (**1991**). ''Stream segregation and peripheral channeling,'' Music Percept. **9**, 155–184.
Levitt, H. (**1971**). ''Transformed up-down methods in psychoacoustic,'' J. Acoust. Soc. Am. **49**, 467–477.
McCabe, S. L., and Denham, M. J. (**1997**). ''A model of auditory streaming,'' J. Acoust. Soc. Am. **101**, 1611–1621.
Miller, G. A., and Heise, G. A. (**1950**). ''The trill threshold,'' J. Acoust. Soc. Am. **22**, 637–638.
Pollack, I. (**1968**). ''Periodicity pitch for interrupted white noise—fact or artifact?'' J. Acoust. Soc. Am. **45**, 237–238.
Rose, M. M., and Moore, B. C. J. (**1997**). ''Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **102**, 1768–1778.
Rose, M. M., and Moore, B. C. J. (**2000**). ''Effect of frequency and level on auditory stream segregation,'' J. Acoust. Soc. Am. **108**, 1209–1214.
Singh, P. G. (**1987**). ''Perceptual organization of complex-tones sequences: a tradeoff between pitch and timbre?'' J. Acoust. Soc. Am. **82**, 886–899.
Van Noorden, L. P. A. S. (**1975**). ''Temporal coherence in the perception of tone sequences,'' unpublished doctoral dissertation, Eindhovern University of Technology.
Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based on modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.
Vliegen, J., and Oxenham, A. J. (**1999**). ''Sequential stream segregation in the absence of spectral cues,'' J. Acoust. Soc. Am. **105**, 339–346.
Vliegen, J., Moore, B. C. J., and Oxenham, A. J. (**1999**). ''The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task,'' J. Acoust. Soc. Am. **106**, 938–945.
Wiegriebe, L., and Patterson, R. D. (**1999**). ''Quantifying the distortion products generated by amplitude-modulated noise,'' J. Acoust. Soc. Am. **106**, 2709–2718.

Spectral loudness summation as a function of duration[a)]

Jesko L. Verhey[b)] and Birger Kollmeier
Medizinsche Physik and Graduiertenkolleg Psychoakustik, Universität Oldenburg, D-26111 Oldenburg, Germany

(Received 12 April 2001; revised 27 September 2001; accepted 19 December 2001)

Loudness was measured as a function of signal bandwidth for 10-, 100-, and 1000-ms-long signals. The test and reference signals were bandpass-filtered noise spectrally centered at 2 kHz. The bandwidth of the test signal was varied from 200 to 6400 Hz. The reference signal had a bandwidth of 3200 Hz. The reference levels were 45, 55, and 65 dB SPL. The level to produce equal loudness was measured with an adaptive, two-interval, two-alternative forced-choice procedure. A loudness matching procedure was used, where the tracks for all signal pairs to be compared were interleaved. Mean results for nine normal-hearing subjects showed that the magnitude of spectral loudness summation depends on signal duration. For all reference levels, a 6- to 8-dB larger level difference between equally loud signals with the smallest ($\Delta f = 200$ Hz) and largest ($\Delta f = 6400$ Hz) bandwidth is found for 10-ms-long signals than for the 1000-ms-long signals. The duration effect slightly decreases with increasing reference loudness. As a consequence, loudness models should include a duration-dependent compression stage. Alternatively, if a fixed loudness ratio between signals of different duration is assumed, this loudness ratio should depend on the signal spectrum.
© *2002 Acoustical Society of America.* [DOI: 10.1121/1.1451065]

PACS numbers: 43.66.Cb, 43.66.Mk, 43.66.Ba [SPB]

I. INTRODUCTION

Loudness is defined as the psychophysical correlate of sound intensity which also depends on other physical parameters, such as, e.g., spectral content of the signal. Many studies have shown that the loudness of a sound with constant intensity which is restricted to a certain frequency band depends on the width of the band. While the bandwidth shows virtually no effect on loudness until a critical bandwidth is reached, the loudness significantly increases with bandwidth if this value is exceeded (e.g., Fletcher and Munson, 1933; Zwicker and Feldtkeller, 1955; Zwicker *et al.*, 1957; Scharf, 1959, 1961; Cacace and Margolis, 1985; Schneider, 1988; Hübner and Ellermeier, 1993). This effect is called (spectral) ''loudness summation'' due to the mechanism believed to be responsible for this effect (e.g., Fletcher and Steinberg, 1924; Zwicker and Fastl, 1999): for sounds which excite more than one critical band the (partial) loudness is determined in each critical band separately and then summed across frequency. Since the relation between loudness and intensity is compressive, this sum yields larger loudness values than the loudness of a sound with the same intensity, but with a subcritical bandwidth. According to this concept, the magnitude of spectral loudness summation[1] depends on two parameters: (*i*) the frequency resolution of the auditory system (i.e., the width of the critical bands) and (*ii*) the size of the compression. A higher compression and a higher frequency resolution yield a larger effect of signal bandwidth on loudness.

Loudness summation is not only found across frequency but also over time. Studies on the effect of signal duration and repetition rate of short pulses on loudness showed that loudness is integrated up to a ''critical duration'' of about 100 ms (Port, 1963a, b; Reichardt and Niese, 1965; Zwicker, 1966; Reichardt and Niese, 1970; Poulsen, 1981; Kumagai *et al.*, 1982a, b, 1984; Sone *et al.*, 1986; Namba, 1987; Takeshima *et al.*, 1988; Ogura *et al.*, 1991, 1993). The effect is usually referred to as temporal integration of loudness or *temporal* loudness summation (in contrast to *spectral* loudness summation as described above). Munson (1947) already modeled this effect by assuming that loudness is integrated over time by a leaky integrator.[2]

Most studies have investigated only one aspect of loudness summation, i.e., either purely temporal or purely spectral. For example, in most studies on spectral loudness summation long durations of about 1 s were used. Only a few studies have considered temporal aspects of spectral loudness summation. Port (1963a) found that the loudness of a 5-ms noise band spectrally centered at 2 kHz increases with increasing bandwidth in the same way as the loudness of a 1200-ms noise band does, assuming the absolute loudness of the reference is the same for the two durations. Zwicker (1965) compared the loudness of a 5-kHz tone with that of a white noise for durations of 2 and 500 ms, with level as the parameter. He found in agreement with Port (1963a) that the level difference necessary to produce equal loudness for the short signals is the same as for the long signals, given an identical reference loudness.

Scharf (1970) concluded from the vanishing effect of duration on the magnitude of spectral loudness summation found by Port (1963a) and Zwicker (1965) that the critical bandwidth does not change with signal duration. Note, however, that this conclusion also implies that the compression is the same for the different durations since both parameters determine the magnitude of spectral loudness summation (see above).

[a)]This article is a revised version of parts of the dissertation of the first author (Verhey, 1999).

[b)]Current address: Center for the Neural Basis of Hearing, The Physiological Laboratory, University of Cambridge, Downing Street, Cambridge CB2 3EG, United Kingdom. Electronic mail: jlv23@cam.ac.uk

However, in two recent studies on temporal integration of loudness, it was claimed that the loudness function must have a different slope for short (5 ms) and long (200 ms) tones at the same loudness (Florentine *et al.*, 1996; Buus *et al.*, 1997). The authors based their argument on the finding that the magnitude of temporal summation of loudness for a 1-kHz tone is level dependent. The level difference between the short and the long tone at the point of equal loudness increased with increasing level as the level of the short tone varied from 10 to about 70 dB SPL. This increase in level difference implies that the loudness function of the short tone must have a shallower slope (i.e., a different compression) than the loudness function of the long tone at the same absolute loudness. Following the explanation for spectral loudness summation as outlined above, this finding suggests that (in this level range) the magnitude of spectral loudness summation should be larger for short than for long sounds, if the same reference loudness is used. This is qualitatively in line with result from Boone (1973). His data show greater magnitude of spectral loudness summation for 18- than for 1000-ms-long signals at a reference loudness level of 72 phon (see also Sec. IV).

In fact, although it is not mentioned explicitly in the literature, models that assume a nearly instantaneously acting spectral integration of loudness prior to the temporal integration (e.g., Ogura *et al.*, 1993; Stone *et al.*, 1997) should predict different loudness summation for short and long signals if a level-dependent compression is assumed. At the same loudness, the short signals have a considerably higher intensity than the long signals and as a consequence the compression for the short signals may be different from that for the long signals. This should be especially salient at low to medium levels, where the slope of the loudness function varies considerably with level (e.g., Zwislocki, 1965; Florentine *et al.*, 1996; Zwicker and Fastl, 1999) and generally is lower at higher levels. Thus, loudness models with a realistic level-dependent compression stage should predict a greater magnitude of spectral loudness summation for short signals than for long signals (given the same reference loudness) in this level range. Note that they would not predict a different magnitude of spectral loudness summation at the same reference level, since the compression would be the same for short and long signals at the same level.

The present study investigates if the magnitude of spectral loudness summation depends on the duration of the signal (a) at the same reference level and (b) at the same reference loudness. The effect of bandwidth on loudness is measured for three durations. The level difference between a test signal and the equally long reference signal at the point of equal loudness was measured for all durations at three reference levels. The difference between the reference levels of 10 dB was chosen in order to permit a comparison of the magnitude of spectral loudness summation across durations not only at the same reference level but also at approximately the same reference loudness.

II. METHODS

A. Procedure

Stimuli with different bandwidths were matched in loudness to a reference signal with fixed bandwidth and level using an adaptive two-interval, two-alternative forced-choice procedure. In each trial the listeners heard two sounds, the reference and the test signal, which were separated by a 500-ms silent interval. Test and reference signals were presented in random order and with equal *a priori* probability. The listeners indicated which signal was louder by pressing the corresponding key on a keyboard. A simple one-up one-down procedure was used, which converges at the 50% point of the psychometric function (Levitt, 1971). If the listener indicated that the test signal was the louder one, its level was reduced in the next trial, otherwise it was increased. At the beginning, the step size was 8 dB. It was divided by 2 after each reversal in the adaptive level tracking procedure. At a step size of 2 dB it was held constant for the next four reversals. The level difference between test and reference signal needed to obtain the same loudness for one track was determined by calculating the median of the levels at these last four reversals. Three matches were obtained for each listener and pair of stimuli.

To reduce biases that occur when stimuli from only one stimulus pair are matched in loudness in a series of trials, several interleaved adaptive tracks were used. Hence, concurrent loudness matches were obtained for all stimulus pairs tested (Florentine *et al.*, 1996). On each trial, the track was chosen randomly from all possible tracks, i.e., from all tracks that had not yet been terminated. To ensure that the interleaved tracks converge at roughly the same time the random choice of tracks is further restricted by requiring that each track be selected once in random order before any track could be reselected. If one track is terminated, the rule is applied to the choice of trials from the remaining unterminated tracks.

Three different starting levels of the test signal at the beginning of each track were tested for each stimulus pair: 20 dB above, 20 dB below, and at the reference level (0 dB difference). In a series of trials, the tracks were uniformly distributed across the three starting levels. All starting levels had the same frequency of occurrence in a series of trials.

Although the procedure is comparable to the procedure proposed by Florentine *et al.* (1996) and Buus *et al.* (1997, 1998), there are some details which differ considerably. In the present experiment, the reference signal was the same within each series of trials and its level was fixed. It was argued by Florentine *et al.* (1996) that a fixed reference could bias the measurement. However, in Verhey (1999) it was shown that the the fixed reference does not influence the results for the procedure used in the present study. Another difference is the choice of the starting level. Whereas in the present article the tracks started at 20, 0, or −20 dB relative to the reference level, Florentine *et al.* used a starting level of 10 dB above the expected value. However, at least for the procedure used in the present study, the starting level has a negligible effect (smaller than 3 dB) on the results of spectral

loudness summation measurements with bandpass-noise signals (Verhey, 1999).

B. Evaluation of the data

The average of the data for the three different starting levels is considered the point of subjective equality. To examine the statistical significance of the effects of stimulus variables and differences among subjects, an ANOVA for repeated measures was performed (SPSS Release 9.0.1). The dependent variable was the level difference between equally loud test signal and reference signal. *Post hoc* Scheffé tests for contrast (SPSS Release 9.0.1) were performed to explore sources of significant effects and interactions.

C. Stimuli and apparatus

The stimuli were generated digitally with a sampling rate of 32 kHz. Stimulus generation and presentation were controlled by a silicon graphics workstation (INDY), which also sampled the listeners' responses and controlled the procedure. The software package SI was used for signal generation and control of the experiments, which was developed at the University of Göttingen. The stimuli were D/A converted (16 bits), and then preamplified and low-pass filtered at 16 000 Hz with a computer-controlled audiometric amplifier. The subjects sat in a double-walled, sound-attenuating booth. The stimuli were presented diotically via a Sennheiser HD 25 headphone without free-field equalization.

Test and reference signals were bandpass noise signals with a flat spectrum geometrically centered at 2 kHz. The reference bandwidth was 3200 Hz. Three reference levels were used: 45, 55, and 65 dB SPL. The bandwidths of the test signals were 200, 400, 800, 1600, and 6400 Hz. For a reference level of 45 dB SPL, a test-signal bandwidth of 3200 Hz also was used. Signals were gated with 2.5-ms raised-cosine ramps. Test and reference signals had the same duration. Three durations were used: 10, 100, and 1000 ms.

For a reference level of 45 dB SPL, Gaussian bandpass noise was used. To generate the signals, a broadband noise was digitally filtered by setting the magnitude of the Fourier coefficients to zero outside the desired passband. For a reference level of 55 and 65 dB SPL, noise with smaller envelope fluctuations (''low-noise noise'') was used instead of a Gaussian noise. This was necessary in order to avoid overload distortions at the highest signal levels employed. To reduce the envelope fluctuations, the bandpass noise was generated by (i) generating a Gaussian bandpass noise, (ii) restricting the bandwidth by setting the magnitude of the Fourier coefficients to zero outside the desired passband, (iii) dividing the time waveform by the envelope of the time signal on a sample by sample basis, and (iv) restricting the bandwidth of the resulting signal as described in (ii). An iteration of this procedure was described by Kohlrausch *et al.* (1997) to generate low-noise noise. The standard deviation σ relative to the mean m of the resulting noise is approximately half of the ratio σ/m for Gaussian noise.

D. Subjects

Nine subjects (five males, four females) participated in the experiment. They ranged in age from 24 to 29 years. Seven subjects were members of the ''Graduiertenkolleg Psychoakustik'' or the ''Arbeitsgruppe Medizinische Physik'' at the University of Oldenburg; one of them was the author JV. The other two subjects (AR, MT) were paid volunteers. Five subjects (CR, JA, JV, KT, OW) had previous experience with equal-loudness judgments. All subjects had normal audiograms (i.e., absolute threshold in quiet $\leqslant$15 dB HL) and no previous history of any hearing problems.

III. RESULTS

A. Individual data

Figures 1–3 show individual data and standard deviations for nine subjects for three reference levels: 45 dB SPL (Fig. 1), 55 dB SPL (Fig. 2), and 65 dB SPL (Fig. 3), respectively. As a function of the bandwidth, the figures show the level difference ΔL between the test signal and the reference signal needed to produce the sensation of equal loudness. A positive ΔL means that the test signal must be higher in level than the reference. The figures show ΔL for signal durations of 1000 ms (squares), 100 ms (circles), and 10 ms (triangles).

For each duration, the data points are connected with lines to indicate that they all produce the same loudness. Note that the loudness is different for each duration, since the reference duration is matched to the test-signal duration and the data are shown for the same reference level and not for the same reference loudness. A comparison at the same reference loudness will be performed in the Sec. IV A.

For a test-signal bandwidth of 3200 Hz, the level difference between test and reference signal was only measured for a reference level of 45 dB SPL. As expected, the data of almost all subjects show no level difference when test signal and reference signal have the same bandwidth (3200 Hz). Hence, for the reference level of 55 and 65 dB SPL, a level difference of 0 dB is assumed for this bandwidth.

In general, all subjects show a positive ΔL for test-signal bandwidths smaller than the reference bandwidth (3200 Hz) and negative ΔL for a test-signal bandwidth of 6400 Hz. Thus, the level for the sound at the point of equal loudness decreases as the bandwidth increases. This reflects the effect of bandwidth of the signal on loudness known as spectral loudness summation (see Introduction).

Large individual differences are observed in the amount of spectral loudness summation. For example, with a 1000-ms signals at a reference level of 55 dB SPL (Fig. 2, squares), the difference between the 200- and the 6400-Hz-wide signals range from 5 dB (subject AR) to 23 dB (subject JV). Despite this variability, the effect of duration is clear: for most of the subjects the magnitude of spectral loudness summation is greater for 10-ms signals than for 1000-ms signals. As expected, especially large differences are obtained for the narrowest bandwidth (i.e., the largest difference between reference and test-signal bandwidth), whereas for bandwidths close to the reference bandwidth the effect of duration is less clear. In general, the data for 100-ms signals is in between those for 10- and 1000-ms signals. Six subjects show this duration effect at all reference levels. One subject (CR) shows it at 55 and 65 dB SPL but not at 45 dB SPL. Only two subjects (OW and JV) show no effect of duration on the magnitude of spectral loudness summation.

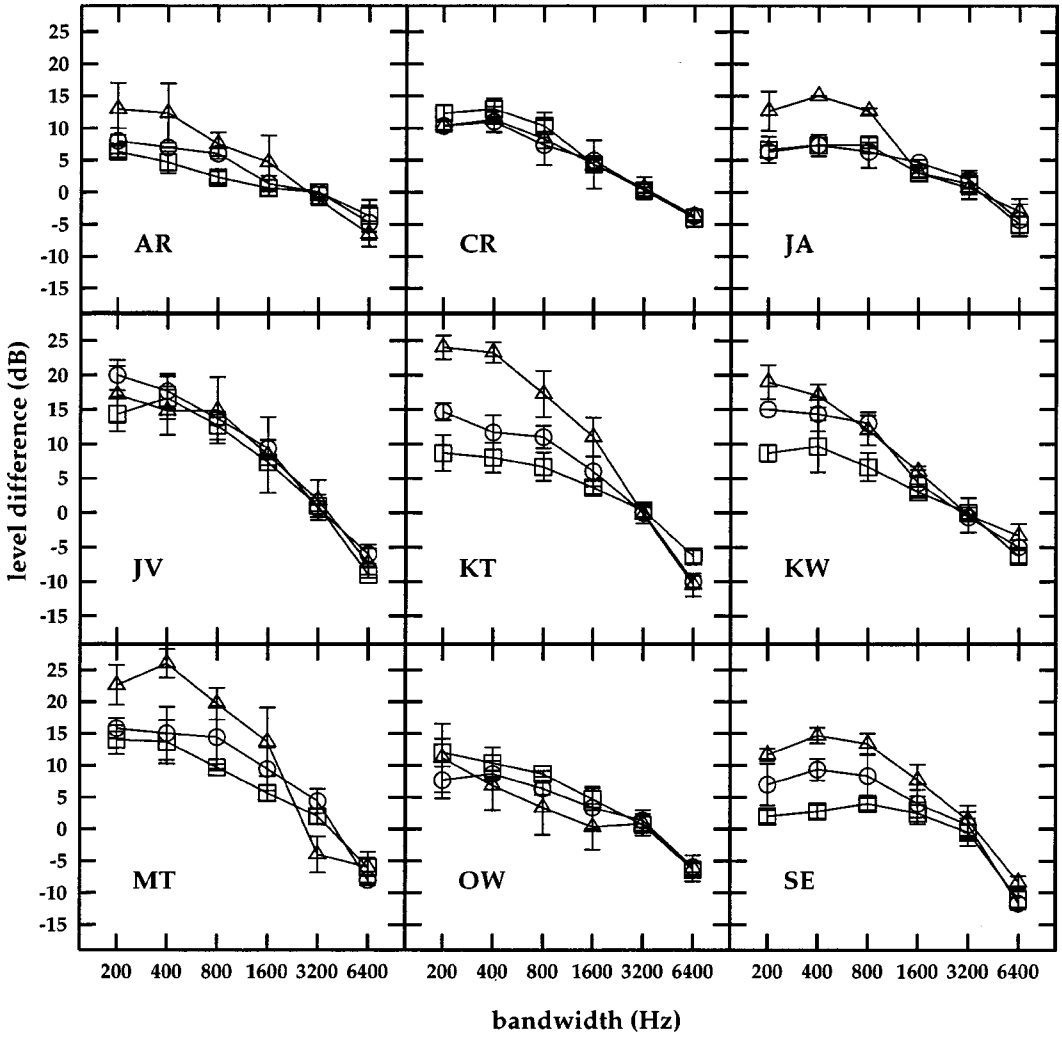


FIG. 1. Loudness summation for bandpass noise signals from nine listeners. The reference level was 45 dB SPL. The reference bandwidth was 3200 Hz. Signal duration was either 1000 ms (squares), 100 ms (circles), or 10 ms (triangles). The level differences between the level of test and reference signal needed to obtain equal loudness are plotted as a function of the bandwidth of the test signal. The vertical bars show plus and minus one standard deviation of the mean.

Although individual differences in the amount of loudness summation are large, subjects are reasonably consistent in their loudness judgments. In general, subjects who show a large amount of loudness summation for one reference level show a large amount of loudness summation for all reference levels used in the present study. The average intraindividual standard deviation is 1.9 dB (range is from 0 to 6.5 dB). Note that no systematic increase in standard deviation is observed for increasing difference between reference bandwidth and test-signal bandwidth. Apparently, the subjects had no difficulties in matching the loudness of the signals although (at least for the narrow test-signal bandwidths) the signals differ considerably in their quality.

B. Group data

Figure 4 shows average data and interindividual standard deviations from the nine subjects for the three different reference levels. Generally the error bars are large when a narrow-band test signal is compared to the broadband reference signal. This reflects the differences among the subjects in the magnitude of loudness summation. Although large individual differences are apparent in Figs. 1–3, the average data for signals of 1000 ms (Fig. 4, squares) are consistent with the results presented in the literature for long signals. For example, Zwicker *et al.* (1957), Port (1963a), and Cacace and Margolis (1985) measured maximum ΔL's from 12 to 20 dB at medium levels. In the present study, the average ΔL between equally loud signals with the smallest bandwidth and the largest bandwidth is 15 dB (65 dB reference level) to 16 dB (45 and 55 dB reference level).

A larger amount of spectral loudness summation is obtained for short signals than for long signals. For all reference levels, a 6 to 8 dB larger level difference between the reference and the 200-Hz-wide test signal is found for 10-ms signals than for 1000-ms signals.

These effects are supported by the ANOVA shown in Table I. All stimulus variables have highly significant effects on the level difference between the test and reference signal. As expected, the bandwidth has a highly significant effect on the level difference. A Scheffé *post hoc* test shows no significant effect between 200 and 400 Hz ($p>0.01$). For all other bandwidth pairs the level difference was significantly

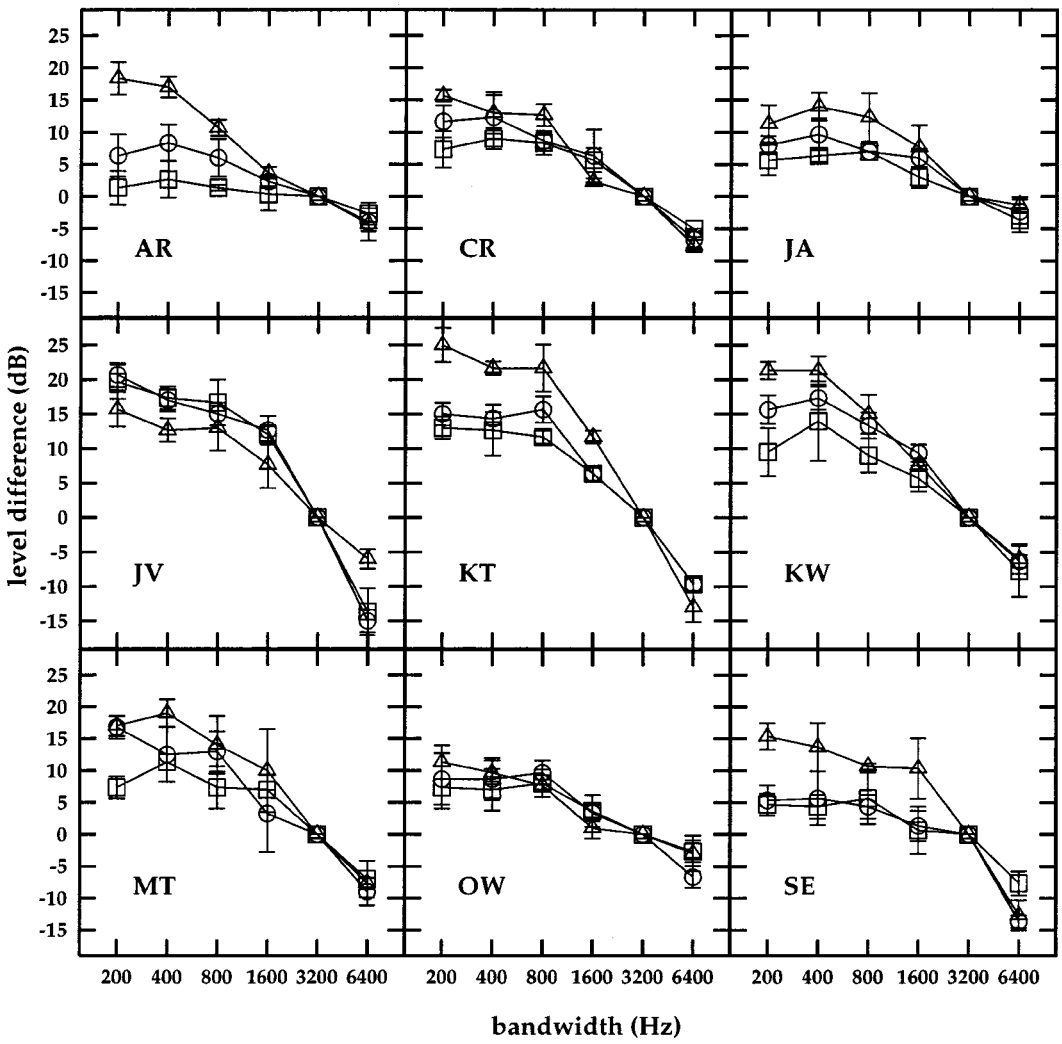


FIG. 2. Same as Fig. 1 but for a reference level of 55 dB SPL.

larger for the signal with the smaller bandwidth ($p < 0.0001$). The effect of duration shows that, overall, the amount of spectral loudness summation depends on the duration of the signal ($p < 0.0001$). The Scheffé *post hoc* tests show that the amount of loudness summation is significantly greater for 10-ms signals than for 100- and 1000-ms signals, and it is also greater for 100-ms signals than for 1000-ms signals ($p < 0.0001$). As indicated by the interaction of bandwidth and duration, bandwidth and subject, and bandwidth, subject, and duration, the amount of spectral loudness summation depends on the duration of the signal and it depends on the subject. The significant interaction between bandwidth and level indicates that the spectral loudness summation depends on the reference level. A Scheffé *post hoc* test shows a significantly smaller amount of loudness summation for a reference level of 65 dB SPL than for 45 and 55 dB SPL, but no significant difference between the results for the reference levels 45 and 55 dB SPL ($p > 0.01$). The interactions of duration and reference level and of duration, bandwidth and reference level are not significant. This indicates that within the investigated level range, the effect of duration on spectral loudness summation does not depend on the reference level.

IV. DISCUSSION

A. Loudness summation as a function of duration at the same reference loudness

In contrast to the present study, most of the data described in the literature (Port, 1963a; Zwicker, 1965) did not show an effect of duration on loudness summation (see Introduction). There is only one study in which a duration-dependent loudness-summation effect was found (Boone, 1973). As the reference signal, Boone (1973) used a 1-kHz tone at a level of 72 dB SPL. The test signals were bandpass noise at various bandwidths with a flat spectrum geometrically centered at 1 kHz. Boone found an increasing loudness summation effect with decreasing signal duration. To produce equal loudness for the reference tone and a test signal with a bandwidth of 1800 Hz, he measured a level difference of 16 dB for a duration of 18 ms, whereas he measured only 9 dB for a duration of 1000 ms. This finding that the level difference for short-duration signals is nearly twice the level difference for long-duration signals is in line with the experimental results of the present study.[3] Interestingly, Boone discussed only the relationship between bandwidth and duration. He focused on the unexpected large critical bandwidth

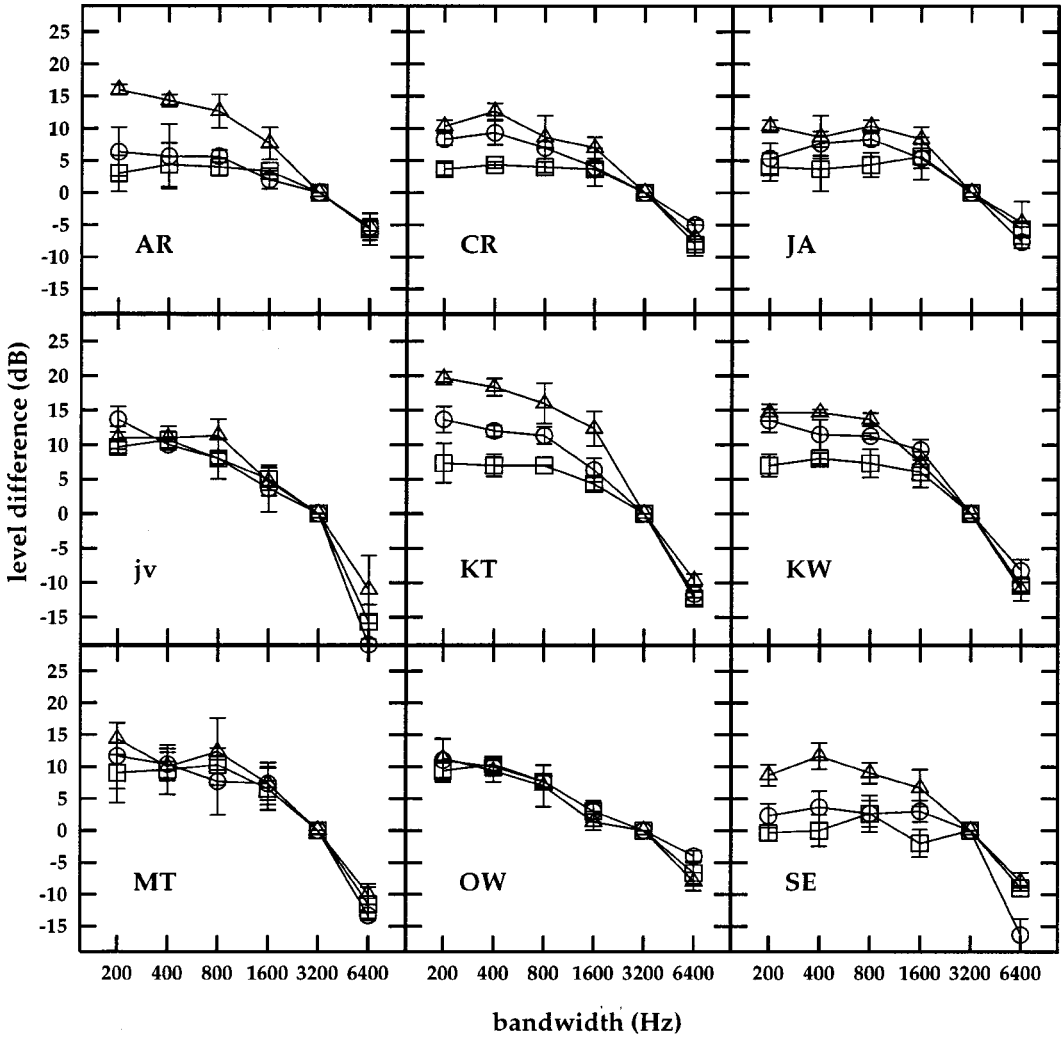


FIG. 3. Same as Fig. 1 but for a reference level of 65 dB SPL.

of 500 Hz derived from his data at a center frequency of 1 kHz, which he thought was due to an unacceptably large range of results from the 16 test subjects. This might explain why he did not discuss the large discrepancy in the amount of loudness summation between his results and those of other authors.

Boone (1973) and Port (1963a) measured loudness summation for different durations at nearly the same reference loudness, whereas in the present study the reference level (in dB SPL) was kept constant for the different durations. Therefore, a direct comparison might be problematic. However, the level difference between an equally loud 10-ms signal

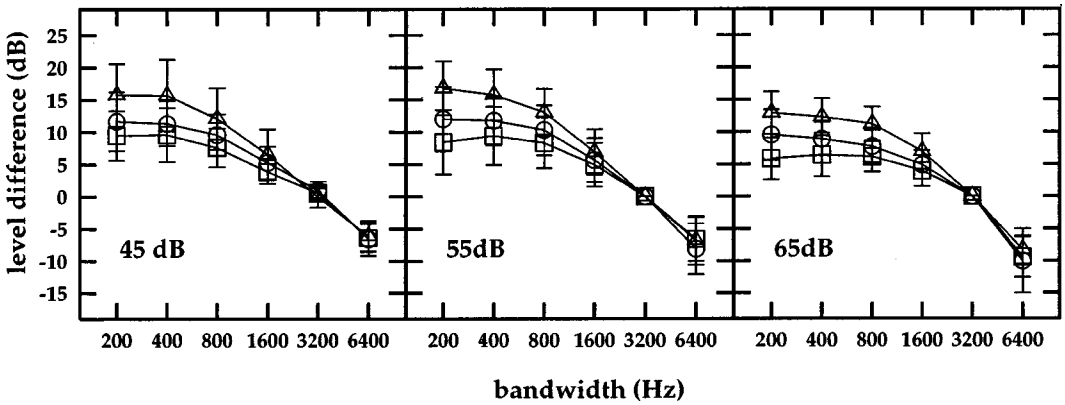


FIG. 4. Averaged loudness summation for bandpass-noise signals for a reference level of 45 dB SPL (left panel), 55 dB SPL (middle panel), and 65 dB SPL (right panel). Signal duration was either 1000 ms (squares), 100 ms (circles), or 10 ms (triangles). The level differences between the level of test and reference signal needed to obtain equal loudness are plotted in the same manner as in Fig. 1. The vertical bars show plus and minus one interindividual standard deviation from the mean.

TABLE I. Three-way analysis of variance for repeated measures of loudness matching. The dependent variable is the level difference between test and reference sound at equal loudness. The fixed factors are bandwidth of the test signal (bw; five steps: 200, 400, 800, 1600, and 6400 Hz), signal duration (dura; three steps: 10, 100, and 1000 ms), level of the reference signal (Lref; three steps: 45, 55, and 65 dB). Subjects (subj; nine steps: AR, CR, JA, JV, KT, KW, MT, OW, SE) is a random factor.

Source	Sum of squares	d.f.	Mean square	F	Prob$>F$
Const	43420.6	1	43420.6	84.81	$\leqslant$0.0001
bw	61732.0	4	15433.0	107.37	$\leqslant$0.0001
dura	3660.3	2	1830.2	13.99	$\leqslant$0.0001
Lref	1179.7	2	589.9	8.00	0.004
subj	4095.7	8	512.0	1.68	0.129
bw * dura	1253.3	8	156.7	7.58	$\leqslant$0.0001
bw * Lref	355.5	8	44.4	4.01	0.001
dura * Lref	15.6	4	3.9	0.19	0.940
bw * dura * Lref	83.0	16	5.2	0.65	0.837
bw * subj	4599.6	32	143.7	6.05	$\leqslant$0.0001
dura * subj	2093.4	16	130.8	3.99	$\leqslant$0.0001
bw * dura * subj	1322.1	64	20.7	2.59	$\leqslant$0.0001
Lref * subj	1179.3	16	73.7	3.18	0.002
bw * Lref * subj	708.9	64	11.1	1.39	0.059
dura * Lref * subj	644.0	32	20.1	2.52	$\leqslant$0.0001
bw * dura * Lref * subj	1020.6	128	8.0	1.10	0.226

and a 1000-ms signal amounts to about 10 dB (e.g., Poulsen, 1981). Thus, since the different reference levels used in the present study are separated by 10 dB, it is possible to compare the amount of loudness summation for short and long signals not only at the same reference level but also at approximately the same reference loudness. Since similar amounts of spectral loudness summation were found at a reference level of 45 and 55 dB, the duration effect in spectral loudness summation is the same if the data of 10-ms signals at a reference level 55 dB SPL are compared with the data of 1000-ms signals at the same loudness (3.3 sone), i.e., at a reference level of 45 dB SPL. The level difference between the narrowest (200 Hz) signal and the reference signal is again about 7 dB larger for the 10-ms signals than for the equally loud 1000-ms signals. Interestingly, it amounts to only 4 dB for 10-ms signals at a reference level of 65 dB SPL compared to the data for the equally loud 1000-ms signals at a reference level of 55 dB (5.7 sone). This indicates that (in the investigated level range) the duration effect in spectral loudness summation slightly decreases with increasing reference loudness whereas it is approximately independent of the reference level.

B. Further evidence for a duration-dependent loudness summation at the same reference loudness

The combination of experimental results on temporal integration of loudness for different bandwidths with results on loudness summation for stationary signals indirectly gives information about temporal aspects of loudness summation. One can construct equal-loudness level surfaces where one axis represents duration, one represents bandwidth, and the other axis represents the level. On such an equal-loudness level surface, level differences for equally loud signals can be estimated that have not directly been measured (e.g., short signals with different bandwidths).

In this way, it is possible to extract information about temporal aspects of loudness summation from the data presented by Florentine *et al.* (1996). They measured the level difference between equally loud 5- and 200-ms stimuli (1-kHz tone or white noise) for a wide range of levels. If one combines their results with results on loudness summation for stationary signals presented by Zwicker *et al.* (1957), it is possible to derive loudness summation for short signals at different levels. For a 5-ms white-noise signal with a level of 50 dB, the level of an equally loud 5-ms sinusoidal signal should be 18 dB higher (see the Appendix for a computation of the level difference). The loudness-summation effect is 6 dB larger than for long signals (12 dB). The same loudness-summation effect is obtained for a level of the short noise signal of 40 dB. The increase in loudness summation effect for short signals derived from Florentine *et al.*'s data at moderate levels agrees with the results of the present study.

The difference in loudness summation for short and long signals derived from Florentine *et al.*'s data decreases with increasing level. For a level of 70 dB for the short noise signal, the maximum spectral loudness summation for short signals is only 2 dB larger than for long signals. Thus, in qualitative agreement with the present data (see Sec. IV A), the duration effect decreases with increasing loudness. Finally, for $L_{\text{short}}(\text{noise}) = 90\,\text{dB}$, the loudness summation for short signals is even 5 dB smaller than the loudness summation for long signals.

Florentine *et al.* (1996) and Buus *et al.* (1997) proposed the equal-loudness-ratio model to account for their data on temporal integration of loudness. Figure 5 in Buus *et al.* (1997) shows the loudness function for 5- and 200-ms 1-kHz tones and broadband noises derived from Florentine *et al.*'s (1996) data. In this figure, the horizontal distance between the loudness function for the equally long tone and noise yields a prediction of the amount of spectral loudness summation. The same magnitude of spectral loudness summation for 5- and 200-ms signals is only predicted for 0.8 sone.[4] For smaller loudness values, spectral loudness summation for 200-ms signals is smaller than for the 5-ms signals. The dif-

ference in spectral loudness summation between 5- and 200-ms signals increases as the loudness decreases up to a maximum at about 0.1 sone. The reversed duration effect is predicted for loudness values higher than 0.8 sone, i.e., spectral loudness summation is always greater for the 200-ms signals than for the 5-ms signals with a maximal difference around 2 sone. Thus, also on the basis of this model, the amount of spectral loudness summation depends on the signal duration at the same loudness in qualitative agreement with the data in the present study. Note, however, that the present study indicates a greater magnitude of spectral loudness summation for short than for long signals at higher loudness values than expected from Fig. 5 in Buus *et al.* (1997) (see previous section and footnote 4). This discrepancy may be due to the different durations used in the two studies. If the maximal spectral loudness summation is compared between 10- and 100-ms signals at about the same loudness, a 5-dB difference is predicted for the lower loudness value and less than a 1-dB difference for the higher loudness value (see previous section for details of the comparison).

Garnier *et al.* (1999) measured temporal integration and spectral loudness summation in normal-hearing and hearing-impaired listeners with a categorical scaling method. They measured loudness functions for long (300-ms) broadband noise and a 1.6-kHz tone. In addition they measured loudness for a 16.5-ms broadband noise. Unfortunately, a direct calculation of spectral loudness summation with short signals is impossible since they did not measure the loudness function of a 16.5-ms 1.6-kHz tone. However, if the data of Garnier *et al.* (1999) are combined with the temporal integration data for tones from Buus *et al.* (1997) and Florentine *et al.* (1996), a 3 to 10 dB greater spectral loudness summation effect for a 16.5-ms than for a 300-ms signal is estimated for normal-hearing subjects in the range from very soft to comfortable loudness.

C. Loudness summation as a function of duration at the same reference level

In the present study, loudness summation was originally measured for short and long signals at the same reference level. Since Zwicker (1965) measured loudness summation for a wide range of levels, loudness summation can also be evaluated at the same level using his data. As in the present study, he obtained a different loudness summation for short and long signals at the same level in a level range of the broadband noise from 30 to 50 dB SPL. However, in contrast to the present data, loudness summation at the same reference level was smaller for short than for long signals. For example, to obtain equal loudness for a 5-kHz tone pulse at 55 dB SPL and a white-noise pulse at various durations of both signals, he obtained a 5 dB smaller level difference for 2-ms signals compared to 500-ms signals. It is unclear why such a large difference between results from the present study and the results from Zwicker occur. The differences in the experimental procedures might explain the different results. It was shown in Verhey (1999) that the use of a simple adaptive procedure might produce a difference in loudness summation for short and long signals of 1 to 4 dB, depending on the starting level of the test signal. A similar effect may reduce or even reverse the duration effect in the procedure he used.

The equal-loudness ratio model proposed by Florentine *et al.* (1996) assumes that the loudness ratio between short and long signals with the same level is independent of the level. In addition, Buus *et al.* (1997) assumed that it is also independent of the spectral content of the signals. In other words, the vertical distance between the loudness curves for signals with the same spectra but different durations is independent of the stimulus level and independent of the spectrum when plotted as log(sone) against the stimulus level. This is shown in Fig. 5 of Buus *et al.* (1997) for 5- and 200-ms broadband-noise stimuli and 1-kHz tones. In addition, this figure shows that the equal-loudness ratio model predicts for a given level of one of the sounds (e.g., the broadband noise) a *duration-independent* horizontal distance between the loudness curves for signals with the same duration but different spectra (broadband noise and tone). Hence, the model predicts the same magnitude of spectral loudness summation for different durations when the data are compared at the same reference level. The data in the present study contradict this prediction.

D. Temporal effects of spectral loudness summation in the context of loudness models for impulsive sounds

Although some results indicated that the temporal integration of loudness depends on physical parameters of the signal such as center frequency (Boone, 1973), bandwidth (Port, 1963a, b), and level (Poulsen, 1981), most loudness models assume a single overall temporal weighting (e.g., Port, 1963a, b; Ogura *et al.*, 1993; Stone *et al.*, 1997), probably for reasons of simplicity. Within these models, spectral effects are described by spectral analysis (critical-band filtering, compression, summation across critical bands) before temporal integration. It is assumed that the spectral analysis, which is responsible for loudness summation, works nearly instantaneously (e.g., Ogura *et al.*, 1993), whereas the overall temporal weighting has longer time constants of about 100 ms. Such a model would predict (as already mentioned in the Introduction) that the magnitude of spectral loudness summation is independent of signal duration, if it is measured at the same reference level. The results of the present study (measured at the same reference level) contradict this prediction.

A more complex model for calculating loudness of temporally variable sounds was proposed by Zwicker and Fastl (1999). In contrast to the models described above, they assumed a duration-dependent low-pass filtering within each critical band before summation across critical bands followed by an overall time weighting with a low-pass filter (τ about 100 ms). They based their assumption of a noninstantaneous spectral-analysis stage on results of experiments with strongly frequency-modulated sounds (Zwicker, 1974, 1977) and on experiments where the influence of temporal and spectral structures of sequences of tone bursts on loudness was investigated (Zwicker, 1969). The low-pass filter within each critical band had a short rise time (few millisec-

onds) and a long decay time. The decay is determined by experiments on forward masking. However, similar to the more simple models described previously, this model cannot account for the the duration-dependent spectral loudness summation at the same reference level, since only the decay time varies with duration, which is not important for the experiments in the present study.

One possibility to describe the results of the present study is to include a duration-dependent compression in the spectral analysis stage of the loudness models, i.e., higher compression at the beginning of a signal. Higher compression at the beginning of the signal would result in a larger amount of loudness summation for short signals as found in the experimental data.

E. Proposed modification of the equal-loudness ratio model

The equal-loudness ratio model predicts a duration-dependent spectral loudness summation at the same reference loudness (see Sec. IV A). This is a consequence of the different slope of the loudness function for broadband noise and tones at the same loudness. However, it predicts an amount of loudness-summation effect which is independent of duration at the same reference level, in contrast to the data in the present study (see Sec. IV C). The reason for this prediction is the assumption of an equal-loudness ratio which is independent of the spectral content. In the previous section, a different compression for short and long signals was proposed, i.e., the loudness ratio decreases as the intensity increases, which is in contrast to the basic assumption of the equal-loudness-ratio model. However, it is possible to modify the model to account for the present data by assuming that different level-independent loudness ratios exist for signals with different spectra. In order to predict a higher amount of loudness summation for short signals, the ratio should decrease as the bandwidth increases. Irrespective of the spectral content of the signal, Buus *et al.* (1997) assumed a loudness ratio of about 3 between a 5-ms signal and a 200-ms signal with the same level. If, for example, instead of 3 a loudness ratio of 2 is used for broadband noise (but keeping a loudness ratio of 3 for tones), an approximately 5 dB higher level difference would be predicted for short signals than for long signals at the same noise level in the range from 40 to 60 dB SPL.[5] Thus, such a modified model is in principle able to predict a duration-dependent magnitude of loudness summation at the same loudness and at the same reference level, in agreement with the data.

V. SUMMARY AND CONCLUSION

The data in the present study show that the amount of spectral loudness summation depends on signal duration. In the investigated level range, the duration effect was found to be independent of the reference level. It slightly decreases with loudness. At the same loudness, a duration-dependent amount of loudness summation is also extracted from recently published loudness data for short and long tones and noises. Present loudness models of time-varying sounds can account for the duration effect at the same loudness, if they account for the level dependence of spectral loudness summation for stationary signals. However, they cannot predict a duration effect at the same reference level. The reason for this failure is that their spectral-analysis stage acts nearly instantaneously and is duration independent. A model which contains a duration-dependent compression should in principle account for the present data. Alternatively, within the framework of a modified equal-loudness-ratio model, the loudness ratio at the same level between short and long signals with the same spectrum should depend on the signal spectrum. Further studies should be performed to test how the duration effect in loudness summation depends on the level for very high and very low levels.

ACKNOWLEDGMENTS

We would like to thank all colleagues of the ''Graduiertenkolleg Psychoakusik'' and the ''Arbeitsgruppe Medizinsche Physik'' at the University of Oldenburg for their fruitful discussions concerning this study. We also thank Carsten Reckhardt and Oliver Wegner for their many suggestions regarding the contents of this article. The authors gratefully acknowledge Torsten Dau for his critical reading of earlier versions of this article. The work was supported by the DFG (Graduiertenkolleg Psychoakustik, KO 942/12).

APPENDIX: ESTIMATION OF THE DURATION EFFECT FROM EQUAL-LOUDNESS-LEVEL SURFACES

The combination of experimental results on temporal integration of loudness for different bandwidths, as, e.g., shown in Florentine *et al.* (1996), with results on loudness summation for stationary signals, as shown in, e.g., Zwicker *et al.* (1957), provides information about spectral loudness summation for short signals.

First, for a specified level of the short signal the level difference between equally loud short (5 ms) and long (200 ms) white-noise signals is derived from Fig. 3 in Florentine *et al.* (1996). For example, at a level of 50 dB for the short broadband noise signal, the level difference between equally loud short and long broadband noise signals is 12 dB, i.e., the level of the equally loud long noise signal is 38 dB. At this level, for stationary signals Zwicker *et al.* (1957) measured a level difference of about 12 dB between a broadband noise signal and an equally loud noise signal with a subcritical bandwidth centered at 1420 Hz. Therefore, to produce equal loudness with the long noise signal, the level of a long 1420-Hz tone has to be 50 dB. According to the isophones for tones (e.g., Zwicker and Fastl, 1999), a long 1000-Hz tone [as used in Florentine *et al.* (1996)] is equally loud to a 1420-Hz tone at about the same level. Because the level differences are shown as a function of the level of the short signal, the level of an equally loud short tone cannot be derived directly from Fig. 3 in Florentine *et al.* (1996). Instead, it is necessary to find a level of the short tone, where the level difference yields 50 dB for the level of the long tone. This is the case for a level of 68 dB. Since the reference signal (long 1000-Hz tone) is equally loud to the long broadband noise, which itself is equally loud to the short broadband noise, it is now possible to calculate the level difference

between the short 50-dB broadband noise and the short 1000-Hz tone at the point of equal loudness. It amounts to 18 dB.

[1]Usually the effect of bandwidth on loudness is quantified by measuring the level difference between equally loud sounds with different bandwidths. Throughout this article the term ''amount of spectral loudness summation'' or ''magnitude of spectral loudness summation'' refers to this level difference as the size of the effect. The effect itself is called ''spectral loudness summation.''

[2]There is considerable confusion in the literature as to the exact numerical value of the time constant. Munson (1947) assumed a value of 200 ms. A similar value (150 ms) is used by Zwicker (1966). However, some studies indicate time constants which are much smaller. For example, Port (1963a) proposed a time constant of 70 ms and Niese (1959) fitted his data with a time constant of only 25 ms. Some other studies indicate an even larger time constant than 200 ms (e.g., Takeshima *et al.*, 1988).

[3]To compare the present data with Boone's data, we have to transform the absolute bandwidths in Fig. 2 to bandwidths relative to the center frequency CF. On this scale we obtain a level difference of about 17 dB (duration=10 ms) and 9.5 dB (duration=1000 ms) between a signal with a subcritical bandwidth ($\Delta f/CF<0.2$) and a signal with a relative bandwidth $\Delta f/CF=1.8$ (i.e., an absolute bandwidth of 3600 Hz at a center frequency of 2 kHz).

[4]In the figures from Buus *et al.* (1997), a loudness value of 0.8 sone corresponds to a level of the 200-ms 1-kHz tone of 57 dB. This value is a factor of 4.35 lower than the expected sone value for this level of 3.5 sone (e.g., Zwicker and Fastl, 1999). However, the derived loudness function can be vertically shifted by a single multiplicative factor (4.35) to more reasonable loudness values without changing the shape of the functions to account for the data of Florentine *et al.* (1996).

[5]The ratio of the assumed equal-loudness ratio and the measured level difference between short and long signals at the point of equal loudness determines the slope of the derived loudness curves. Thus, for an exact calculation of the predicted difference in the magnitude of spectral loudness summation between short and long signals, the loudness curves for the broadband noise have to be recalculated for such a modified model. However, the qualitative conclusion concerning the prediction of this modified version of the equal loudness ratio model for the duration effect in spectral loudness summation would not be altered by this recalculation.

Boone, M. (**1973**). ''Loudness measurements on pure tone and broad band impulsive sounds,'' Acustica **29**, 198–204.

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 Müsch, H. (**1998**). ''Loudness function for tones at low levels derived from loudness summation,'' in *Psychophysical and Physiological Advances in Hearing*, edited by A. e. a. Palmer (Whurr, London), pp. 449–457.

Cacace, A. T., and Margolis, R. H. (**1985**). ''On the loudness of complex stimuli and its relationship to cochlear excitation,'' J. Acoust. Soc. Am. **78**, 1568–1573.

Fletcher, H., and Munson, W. (**1933**). ''Loudness, its definition, measurement and calculation,'' J. Acoust. Soc. Am. **5**, 82–108.

Fletcher, H., and Steinberg, J. (**1924**). ''Loudness of a complex sound,'' Phys. Rev. **24**, 306.

Florentine, M., Buus, S., and Poulsen, T. (**1996**). ''Temporal integration of loudness as a function of level,'' J. Acoust. Soc. Am. **99**, 1633–1644.

Garnier, S., Micheyl, C., Berger-Vachon, C., and Collet, L. (**1999**). ''Temporal loudness integration and spectral loudness summation in normal-hearing and hearing-impaired listeners,'' Acta Otolaryngol. (Stockh) **119**, 154–157.

Hübner, R., and Ellermeier, W. (**1993**). ''Additivity of loudness across critical bands: A critical test,'' Percept. Psychophys. **54**, 185–189.

Kohlrausch, A., Fassel, R., van der Heyden, M., Kortekaas, R., van de Par, S., Oxenham, A., and Püschel, D. (**1997**). ''Detection of tones in low-noise noise: further evidence for the role of envelope fluctuations,'' Acust. Acta Acust. **83**, 659–669.

Kumagai, M., Ebata, M., and Sone, T. (**1982a**). ''Comparison of loudness of impact sounds with and without steady duration (A study on the loudness of impact sound. II),'' J. Acoust. Soc. Jpn. (E) **3**, 33–40.

Kumagai, M., Ebata, M., and Sone, T. (**1982b**). ''Comparison of loudness of impact sounds with and without steady duration (A study on the loudness of impact sound. III),'' J. Acoust. Soc. Jpn. (E) **3**, 111–118.

Kumagai, M., Suzuki, Y., and Sone, T. (**1984**). ''Comparison of loudness of impact sounds with and without steady duration (A study on the loudness of impact sound. V),'' J. Acoust. Soc. Jpn. (E) **5**, 31–36.

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

Munson, W. (**1947**). ''The growth of auditory sensation,'' J. Acoust. Soc. Am. **19**, 584–591.

Namba, S. (**1987**). ''On the psychological measurement of loudness, noisiness and annoyance: A review,'' J. Sound Vib. **116**, 491–507.

Niese, H. (**1959**). ''Die Trägheit der Lautstärke in Abhängigkeit vom Schallpegel,'' Hochfrequenztech. U. Elektroakust. **68**, 143–152.

Ogura, Y., Suzuki, Y., and Sone, T. (**1991**). ''A temporal integration model for loudness perception of repeated impulsive sounds,'' J. Acoust. Soc. Jpn. (E) **12**, 1–11.

Ogura, Y., Suzuki, Y., and Sone, T. (**1993**). ''A new method for loudness evaluation of noises with impulsive components,'' Noise Control Eng. **40**, 231–240.

Port, E. (**1963a**). ''Über die Lautstärke einzelner kurzer Schallimpulse,'' Acustica **13**, 212–223.

Port, E. (**1963b**). ''Zur Lautstärkeempfindung und Lautstärkemessung von pulsierenden Geräuschen,'' Acustica **13**, 224–233.

Poulsen, T. (**1981**). ''Loudness of tones in a free field,'' J. Acoust. Soc. Am. **69**, 1786–1790.

Reichardt, W., and Niese, H. (**1965**). ''Die Addition der Schallerregungen in den einzelnen Frequenzgruppen bei impulsiven Geräuschen,'' Acustica **16**, 295–304.

Reichardt, W., and Niese, H. (**1970**). ''Choice of sound duration and silent intervals for test and comparison signals in the subjective measurement of loudness level,'' J. Acoust. Soc. Am. **47**, 1083–1090.

Scharf, B. (**1959**). ''Loudness of complex sounds as a function of the number of components,'' J. Acoust. Soc. Am. **31**, 783–785.

Scharf, B. (**1961**). ''Loudness summation and spectrum shape,'' J. Acoust. Soc. Am. **34**, 228–233.

Scharf, B. (**1970**). ''Critical bands,'' in *Foundations of Modern Auditory Theory*, edited by J. V. Tobias (Academic, New York), Vol. I, pp. 159–202.

Schneider, B. (**1988**). ''The additivity of loudness across critical bands: A conjoint measurement procedure,'' Percept. Psychophys. **43**, 211–222.

Sone, T., Suzuki, Y., Kumagai, M., and Takahashi, T. (**1986**). ''Loudness of a single burst of impact sound: Results of round robin test in Japan (I),'' J. Acoust. Soc. Jpn. (E) **7**, 173–182.

Stone, M., Moore, B., and Glasberg, B. (**1997**). ''A real-time DSP-Based loudness Meter,'' in *7th Oldenburg Symposium on Psychoacoustics*, edited by A. Schick and M. Klatte (BIS, Oldenburg), pp. 587–602.

Takeshima, H., Suzuki, Y., Kono, S., and Sone, T. (**1988**). ''Growth of the loudness of a tone burst with duration up to 10 seconds,'' J. Acoust. Soc. Jpn. (E) **9**, 295–300.

Verhey, J. L. (**1999**). *Psychoacoustics of Spectro-temporal Effects in Masking and Loudness Perception* (BIS-Verlag, Oldenburg), ISBN 3-8142-0662-2, http://docserver.bis.uni-oldenburg.de/publikationen/bisverlag/verpsy99/verpsy99.html.

Zwicker, E. (**1965**). ''Temporal effects in simultaneous masking and loudness,'' J. Acoust. Soc. Am. **38**, 132–141.

Zwicker, E. (**1966**). ''Ein Beitrag zur Lautstärkemessung impulshaltiger Schalle,'' Acustica **17**, 11–22.

Zwicker, E. (**1969**). ''Der Einfluß der zeitlichen Struktur von Tönen auf die Addition von Teillautheiten,'' Acustica **21**, 16–25.

Zwicker, E. (**1974**). ''Loudness and excitation patterns of strongly frequency modulated tone,'' in *Senstation and Measurement, Paper in honor of S. S. Stevens* (Reidel, Dordrecht), pp. 587–602.

Zwicker, E. (**1977**). ''Procedure of calculating loudness of temporally variable sounds,'' J. Acoust. Soc. Am. **62**, 675–682.

Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics*, 2nd ed., Springer Series in Information Sciences (Springer, Berlin).

Zwicker, E., and Feldtkeller, R. (**1955**). ''Über die Lautstärke von gleichförmiger Geräuschen,'' Acustica **5**, 303–316.

Zwicker, E., Flottorp, G., and Stevens, S. (**1957**). ''Critical bandwidth and loudness summation,'' J. Acoust. Soc. Am. **29**, 548–557.

Zwislocki, J. J. (**1965**). ''Analysis of some auditory characteristics,'' in *Handbook of Mathematical Psychology*, edited by R. D. Luce, R. R. Bush, and E. Galanter (Wiley, New York), pp. 1–97.

Informational masking with small set sizes

Virginia M. Richards and Zhongzhou Tang
Department of Psychology, 3815 Walnut Street, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Gerald D. Kidd, Jr.
Department of Communication Disorders and Hearing Research Center, 635 Commonwealth Avenue, Boston University, Boston, Massachusetts 02215

(Received 6 July 2001; revised 12 November 2001; accepted 27 November 2001)

Informational masking refers to interference in the detectability of a sound, or discrimination of some property of a sound, beyond that which can be attributed to interactions at the auditory periphery. In the current experiments the signal to be detected was a tone added to a 6-tone masker, and informational masking was introduced by randomly choosing the frequencies of the tones that comprise the masker. The primary question was whether small numbers of maskers could replace randomly drawn maskers without sacrificing the underlying detection schemes adopted by observers. Similar to the method used by Wright and Saberi [J. Acoust. Soc. Am. **105**, 1765–1775 (1999)], detection thresholds were measured for different masker set sizes, where set size refers to the number of 6-tone maskers from which any one masker was drawn. Set sizes of 3, 6, 12, and 24 were tested as well as conditions in which the maskers were chosen at random. In addition, observers' memory for maskers was coarsely evaluated. Large differences in thresholds were found across observers and across different masker sets. Even for set sizes of 24, the memory test suggests some recognition of maskers for some observers. *Post hoc* analysis of the data included an evaluation of the relative contribution of different frequencies using a single linear model. As a base for comparison, a linear model fitted to each condition was also evaluated. Although the data were fitted better using many rather than one linear model, the reduction in quality of fit was modest. This result suggests substantial consistency in decision strategies regardless of masker set size. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1445790]

PACS numbers: 43.66.Dc [MRL]

I. INTRODUCTION

For the detection of a tone added to a masker, a distinction frequently invoked is between energetic and informational masking (cf. Watson, 1987; Lutfi, 1989; Kidd *et al.*, 1994; Neff, 1995). The term energetic masking indicates masking associated with presumed peripheral interactions—instances in which the tone to be detected is not sufficiently well represented at the level of the auditory nerve to solve the task. The term informational masking refers to masking that occurs even if the tone signal is robustly represented at the periphery, but nonetheless observers fail to reliably detect the presence of the tone. The energetic/informational masking distinction is sometimes described in terms of peripheral/central masking (see Kidd *et al.*, 2001, for a recent description of these distinctions).

In recent years the study of informational masking has relied on a variety of experimental procedures, but the detectability of a tone added to a multi-tone masker has perhaps received the greatest attention. In these experiments, the frequencies of the tones that comprise the masker were randomly assigned either on each trial (the masker was the same across presentation intervals) or on each stimulus presentation. The popularity of this procedure owes, in part, to the results of an experiment reported by Neff and Green (1987). In their experiment observers detected a tone added to a multi-tone masker, where the number of tones in the masker was parametrically varied. Regardless of the number of tones comprising the maskers, the maskers were held at equal power. The frequencies of the tones that comprised the masker were randomly drawn for each presentation of the masker. While the results were somewhat different depending on the signal frequency, several trends were apparent. There was considerable masking even for the 2-tone maskers, and the amount of masking increased as the number of tones in the masker increased to 10 or so. Further increases in the number of tones in the masker had relatively little impact on thresholds until the number of tones reached 20 or 50 tones depending on signal frequency. Further increases in the number of masker components led to lower thresholds, with an apparent asymptote equal to the amount of masking associated with a noise masker of equal power. Thus, the function relating the number of tones in the masker and amount of masking was roughly an inverted ''u.'' Even when the frequency region around the signal tone was ''protected'' such that the maskers were not allowed to fall into the protected region (thereby reducing energetic masking), detection thresholds remained high when the number of maskers was 10 or so (Neff and Callaghan, 1986; see also Oh and Lutfi, 1999).

Subsequent work using this procedure indicates substantial individual differences. For example, when the frequencies of the tones comprising the masker were drawn randomly, Neff and Dethlefs (1995) found that observers' thresholds varied by 40 dB or more when the maskers were comprised of 20 or fewer tones. For observers with ''high thresholds,'' efforts to lower thresholds by making the signal

more salient (presumably aiding perceptual segregation) or by reducing the uncertainty in the task were sometimes met with substantial changes in threshold. Effects as large as 40 dB are reported by Neff and Dethlefs (1995), who also note that the magnitude of the effects depends on a variety of stimulus factors. For observers with ''low thresholds'' such manipulations had relatively less impact. Presumably the more sensitive observers were limited by energetic rather than informational masking, and thus efforts to reduce the amount of informational masking were not successful. With regard to the reasons for informational masking, Neff *et al.* (1993) note that observers with high thresholds '' . . . appear to be 'holistic' listeners, treating complexes as units, and needed much more signal energy to be able to separate the signal frequency from the rest of the complex . . . '' (Neff *et al.*, 1993, p. 3124).

In the current experiment the intent was to determine whether informational masking can be successfully studied using small numbers of maskers rather than maskers drawn at random from a very large pool of maskers. Here, the term ''successfully'' is intended to imply that the results obtained in experiments using a restricted number of maskers generalize to the case in which a large number of maskers are tested. The advantage in using small numbers of maskers is that models of detection may be studied in detail because the stimuli and observers' responses can be manipulated and compared to models of interest (cf. Green, 1964). Given the individual differences in informational masking studies, an alternative approach is to study a large number of observers, and use summary statistics as a means of revealing consistent trends in the data set (Neff and Dethlefs, 1995).

Wright and Saberi (1999) examined informational masking in an experiment in which the pool of maskers was small. In their experiment thresholds were measured as a function of the number of maskers in the pool of available maskers (set size). Masker set sizes ranged from 2 to 10 maskers. Across all conditions, the total number of maskers was 10, and the maskers used when the set size was 2 were also used when the set size was 4, etc. Even though their study used few maskers, Wright and Saberi's results are in accord with other informational masking studied. For example, they reported sizeable individual differences in thresholds, as occurs when the number of potential maskers is large. For all masker set sizes, threshold estimates were higher when the two intervals of the 2IFC had different versus the same maskers. Additionally, increases in set size led to higher thresholds.

In the current experiment, thresholds for the detection of a tone added to a 6-tone masker were measured using masker sets with different numbers of maskers. The number of different 6-tone maskers in a set, which will be referred to as *set size*, ranged from 3 to 24, and in an additional condition maskers were randomly generated prior to each presentation (Ran condition). Our primary interest is to compare thresholds for small and large set sizes in an effort to evaluate whether strategies vary substantially with large changes in set size. Additionally, as described below, the data were fitted using linear models. These factors led to an experimental design that varies from Wright and Saberi (1999) in several ways, including the following two. First, the maskers were randomly generated for the different sets rather than being drawn from a limited set of maskers. Second, on each interval the maskers were chosen with replacement. As a result, on some trials the masker might be the same in the two intervals of the 2IFC procedure. For a masker set size of 3, this occurred on approximately $\frac{1}{3}$ of the trials; when the masker set size was 24, the two intervals had the same masker on only $\frac{1}{24}$ of the trials. Subsequent analysis of the data includes a comparison of thresholds based on trials in which the maskers were the same versus different in the two intervals. In addition, recognition memory for the maskers was evaluated.

After the presentation of the results, fits of linear models to the data are considered. Although only briefly considered, the results form the backbone of efforts to determine whether observers' strategies vary significantly depending on whether the masker set size is small or large. To the degree that linear models fitted to the data depend on masker set size, it would suggest differences in detection strategies. If the models do not differ substantially, at most the results would suggest a similarity of processing across set size.

II. EFFECTS OF SET SIZE

Observers detected a 1000-Hz tone added to a 6-tone complex. Masker set sizes of 3, 6, 12, and 24 were tested, where a set size of 3 implies that the thresholds were tested when the maskers were drawn, with replacement, from a set of just three different 6-tone maskers. In addition, in the Ran condition the maskers were randomly generated prior to each presentation. Maskers were generated by summing 6 tones of equal amplitude and random phase, with the frequencies of the 6 tones being randomly chosen from frequencies ranging from 100 to 920 and 1080 to 6000 Hz uniformly distributed on log frequency. The frequencies between 920 and 1080 were ''protected'' to reduce the magnitude of energetic masking of the 1000-Hz tone. There were three masker sets of size 3, and two masker sets of size 6, 12, or 24. The different masker sets are denoted as 3A, 3B, 3C, 6A, 6B, 12A, 12B, 24A, 24B, and Ran. The maskers used in the different sets are unique. Thus, set 3A has no relation to set 6A, etc.; the use of the same letter is arbitrary.

III. METHODS

A. Primary experiment

In a two-interval, forced-choice trial, the signal was as likely to be in one as the other temporal interval. Observers indicated their choice of signal interval using a keyboard press, and visual feedback indicated the correctness of observers' responses. The signal level was altered using a 2-down, 1-up staircase procedure (Levitt, 1971). The initial signal level was approximately 10 dB above the ultimate threshold. For Obs 4, whose thresholds are the highest, the initial signal level was frequently nearer the ultimate threshold. Following three reversals the initial 4-dB step size was reduced to 2 dB. The track was terminated after 50 trials. Averaging the last even number of reversals excluding at least the first three yields a threshold estimate for 71% cor-

rect. For the current data set, each threshold estimate is based on at least six reversals. The ultimate track-based threshold reported is based on the average of 20 threshold estimates.

Additional analysis includes an estimation of the signal levels required for 75% correct detections using fits to psychometric functions drawn from the adaptive tracks (Dai, 1995). Three psychometric functions were fitted to the data; one associated with all trials, one associated with trials in which the maskers presented in the two intervals were different, and one for trials in which the maskers presented in the two intervals is the same. For the random condition, where the likelihood of having the same masker in two intervals is vanishingly small, only one psychometric function was fitted to the data. An adaptive tracking procedure was used rather than the method of constant stimuli because it successfully presents a reasonable range of signal levels even when thresholds vary substantially across observers and conditions (see also Dai, 1995).

There is little basis for preferring one parametrization of a psychometric function over another. We fitted the function $d' = (x/\alpha)^{\beta}$ where α and β are free parameters and x is the signal level in dB (Dai, 1995). The d' value was estimated from the percent correct scores using a numerical approximation to the p-to-z transform (Abramowitz and Stegun, 1964). The value minimized in the fitting procedure was a weighted sum of the squared errors, where the weights were associated with binomial variance (Dai, 1995, describes the procedure for the function used in the current study). Sixty-eight percent confidence limits for the threshold estimates were generated using resampling methods. All resampling simulations are based on 1000 randomly drawn psychometric functions that were fitted, and thresholds estimated (cf. Efron and Tibshirani, 1993). Thresholds estimated using fits to the psychometric functions are derived from the last 20 tracks.

The stimuli were digitally generated and presented through two channels of a 16-bit DAC using a sampling rate of 20 000 samples/s, low-pass filtered at 7 kHz using matched filters (KEMO VBF 8), and presented diotically by way of Sennheiser HD410SL headphones. The interstimulus interval was approximately 390 ms when the set size was 24 or less. In the Ran condition, the interstimulus interval was nearer 440 ms. The maskers were presented at 57.8 dB SPL, 50 dB SPL per tone.

All observers practiced for at least 3 h prior to data collection. During this time observers practiced in a variety of the conditions tested. Data collection was blocked. For Obs 2 and 4, data collection began with the set size of 3, then 6, etc., and finally the Ran condition. For Obs 1 and 3, the order of data collection was the opposite. For each set size, two observers ran the different exemplars in one order and the other two observers ran in the opposite order (e.g., 6A then 6B vs 6B then 6A). The exception to this protocol is that for Obs 1 and 3 set 24B was tested after all conditions were completed.

Before data collection in a new condition, observers practiced for at least 2 h. Then 30 50-trial tracks were run, with additional data being collected if practice effects were apparent in the final 20 threshold estimates.

The observers ranged in age from 20 to 27 years and were paid for participation. None had prior experience in psychoacoustic tasks, although Obs 3 is the second author. Tests were conducted with the observer seated in a double-walled sound attenuated booth. Observers had thresholds in quiet of 15 dB HL or better for frequencies ranging from 250 to 8000 Hz save thresholds of 20 dB HL at 6000 Hz in the left ear for Obs 1 and 25 dB HL at 8000 Hz in the left ear for Obs 4.

B. Supplemental experiment: Memory test

In most cases, observers participated in a memory test after finishing a condition. In the memory test a total of 30 trials were presented, and of those 6 maskers were chosen from the just-completed condition and 24 maskers were foils. The same foils were used for all memory tests. When the masker set had just 3 maskers, each masker was repeated twice, and when the masker set had more than 6 maskers, only 6 were presented (the same 6 for all observers). Observers were asked to indicate whether or not (yes/no) each of the presented sounds was in the previous experiment. Observers were not informed of the number of targets, nor was feedback presented. Hits are defined as the correct recognition of maskers from the just-completed experimental condition. False alarms (FA) are instances in which a foil is incorrectly identified as having been in the previous condition. The hit and false alarm rates were converted to d' using the equation $d' = z_{\text{hit}} - z_{\text{FA}}$. When the hit rates were 1 but the FA rates were greater than 0, or when the FA rate was 0 and the hit rates less than 1, the d' value was arbitrarily set to 2. When the hit rate was 1 and the FA rate 0, the d' value was arbitrarily set to 4. Due to an error in the program, data were not properly collected in condition 24B. Moreover, due to experimenter oversight, Obs 1 did not complete the memory test for conditions 12B and 24A.

IV. RESULTS

Figure 1 shows the thresholds as a function of the set size. These thresholds are estimated using the adaptive tracks directly; i.e., signal levels required for 71% correct signal detections. Each panel shows the results for a different observer. Error bars indicate the standard error of the mean across 20 replicate estimates. In most instances the error bars are smaller than the symbols.

The most striking result is the variation in thresholds across observers. For Obs 1 and 2 the pattern of thresholds is somewhat similar across the different set sizes as well as across the different exemplars within a single set size. This is apparent in the ordering of the symbols across the multiple sets and exemplars within a set. For Obs 3 thresholds are roughly the same regardless of set size and exemplar. For Obs 4 thresholds are very high, although differences in sensitivity are apparent across different set exemplars for set sizes of 3 and 6. Due to individual differences, interferences regarding the relative potency of different masker tokens are not easily made. Factors such as the amount of masking associated with a particular masker and the likelihood that any

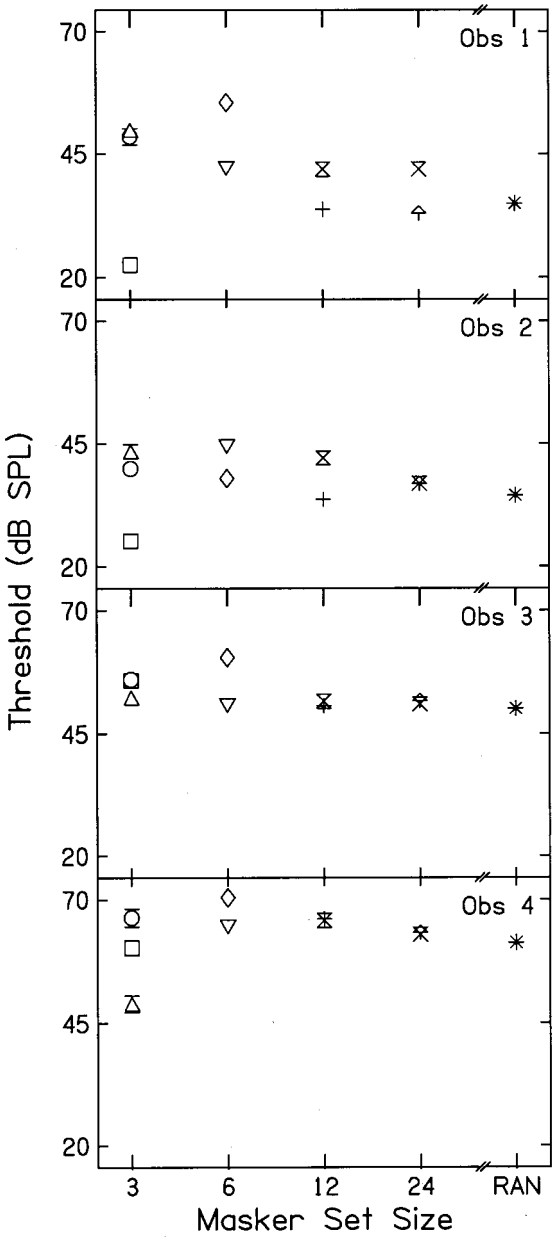


FIG. 1. Thresholds as a function of set size and set exemplar are plotted separately for the four observers. Error bars indicate the standard errors of the mean across 20 threshold replicates. Different symbols are used to indicate different set exemplars. The masker sets are 3A (□), 3B (○), 3C (△), 6A (▽), 6B (◇), 12A (⧖), 12B (+), 24A (⧗), 24B (↑), and Ran (*).

one masker is heard as having an added signal appear to depend on the strategy an observer adopts and how that strategy interacts with the individual maskers.

Averaging across exemplars for each set size, thresholds are generally the same across set sizes. For small set sizes there are substantial differences in thresholds depending on exemplar, whereas for larger set sizes the thresholds tend to be similar across exemplars. Presumably this reflects statistical sampling; for large set sizes the masker set provides an approximation of the population associated with random masker samples. For a set size of 24, thresholds are generally the same regardless of masker set (A or B), except for Obs 1. It seems likely that for Obs 1 the lower thresholds measured for set 24B reflects long-term practice effects (Neff and Dethlefs, 1995)—she returned to set 24B after thresholds were estimated for set sizes of 12, 6, and 3. A repeated-measures ANOVA with set as a fixed factor (24A, 24B, Ran) failed to indicate a significant effect of set [$F(2,6)=3.1$, $p>0.1$]. For Obs 3 and 4, when the set size is 12, the two different masker sets generate similar thresholds, a result not apparent for the more sensitive Obs 1 and 2. Overall, these results suggest that masker set sizes of 12–24 provide a reasonable approximation to the population of possible maskers associated with the Ran condition. It does not guarantee that the strategy an observer uses is the same in the Ran, 24, and 12 conditions, but that result seems likely. The opposite is not reasonably concluded; the fact that different masker sets of size 3 lead to different thresholds indicates that the individual maskers that make up the set have an impact on the thresholds, but does not indicate that the strategy is substantially different from that applied when the maskers are randomly drawn.

TABLE I. *D*-prime values for the recognition task are shown for the different observers and set size/exemplar. Averages and standard errors of the mean are indicated in the final two rows.

	3A	3B	3C	6A	6B	12A	12B	24A
Obs 1	0.98	1.52	2.00	0.43	1.40	NA	0.97	NA
Obs 2	2.00	2.35	1.93	1.52	1.40	0.67	0.24	0.97
Obs 3	0.00	0.00	0.67	0.00	−0.42	−0.22	0.81	0.75
Obs 4	4.00	1.93	0.75	1.15	2.00	−0.76	0.98	2.12
AVG	1.74	1.45	1.34	0.77	1.09	−0.10	0.75	1.28
SEM	(0.86)	(0.51)	(0.36)	(0.34)	(0.52)	(0.42)	(0.17)	(0.39)

Psychometric functions were fitted to the trials in which the same masker was presented in the two intervals and trials in which different maskers were presented in the two intervals. Recall that the number of trials in which the maskers were the same in the two intervals falls inversely with set size. Thus, the estimate of thresholds associated with "same" trials (e.g., the same masker in the two intervals) becomes unstable as the set size increases. Monte Carlo simulations indicate that stable threshold estimates are available in many of the conditions tested. In conditions 12B, 24A, and 24B, confidence intervals were unusually large for at least one of the observers when the maskers were the same in the two intervals. For that reason those conditions were not considered further. In condition 6A and when the maskers were the same in the two intervals, the 68% confidence interval was very large for Obs 1. Nonetheless, Obs 1's threshold estimate in condition 6A is included in the following analysis. A repeated-measures ANOVA with masker set size (3A, 3B, 3C, 6A, 6B, 12A) and trial type (same versus different masker) as fixed variables did not reveal significant main effects ($p\approx0.07$ and $p\approx0.07$, respectively). The interaction did not approach significance ($p\approx0.3$). Thus, there is no evidence to indicate shifts in sensitivity associated with similarity of maskers across intervals.

The results of the memory task are shown in Table I. *D*-prime values are shown for each observer and set. The final two rows show the mean and standard error of the mean across observers. For Obs 3 the d' values are both negative and positive, indicating poor recognition of the masker re-

gardless of set size. For the other three observers, the d' values are almost always positive. Moreover, for those observers increases in set size are not met with consistent decreases in d' values. There is little doubt that Obs 1, 2, and 4 remembered the maskers well when the set size was 3, and substantial masker recognition is maintained even when the set size is 24. None of the foil maskers was consistently chosen as having been heard in recent conditions.

Two aspects of the memory data deserve note. First, doing well on the memory task does not predict detection sensitivity. Obs 4, whose thresholds are the highest, performed as well in the memory task as any of the observers. Second, Obs 1, 2, and 4 demonstrate substantial knowledge of the individual maskers. As a result, one might imagine that these observers compare each sound against a long-term memory of each of the maskers. Such a decision process would generate thresholds that depend little upon whether the maskers are the same versus different in the two intervals. That prediction holds in the current experiment but is inconsistent with the data of Wright and Saberi (1999) who used only 10 maskers, but report substantial differences in masking depending on the degree of masker uncertainty across intervals. Our results are also inconsistent with the results of Neff and Dethlefs (1995, see also Neff and Callaghan, 1986), who report differences nearer 11 dB when the maskers are randomly drawn from a large pool of potential maskers. Undoubtedly the difference between the current results and those of Neff and colleagues is that in the current experiment the trials in which the maskers were the same versus different across intervals were intermixed, whereas they used blocked procedures.

It is probably the case that whether or not observers can remember the individual tokens, task constraints alter the way in which observers approach the detection task. When the maskers are always the same in the two observation intervals (even if maskers vary across trials) thresholds are lower than when the maskers are randomly chosen on each interval (cf. Neff and Dethlefs, 1995; Wright and Saberi, 1999). When it happens by chance that the maskers are same in the two intervals, as occurred in the current experiment, observers appear not to take advantage of that similarity even when it occurs on a substantial number of trials (e.g., set size of 3). The increase in uncertainty associated with variations in maskers across intervals might alter the strategy observers apply to the task (Wright and Saberi, 1999). For example, when the maskers are the same in the two intervals, observers may rely relatively more on short-term memory to detect the added tone. In contrast, when the maskers are likely to be different in the two presentation intervals, observers may depend more on comparison with long-term standards in order to detect the added tone. Whether or not an explanation of this sort ultimately holds, the link between stimulus uncertainty and strategy/memory processing requires consideration.

V. FITS TO LINEAR MODELS

During the course of experiment I parameters associated with each trial were saved to disk. The information included the maskers presented in each interval, the signal level, the interval with added signal, and the observer's response. Here, those data are briefly considered. The primary question to be addressed is whether observers' strategies vary substantially depending on the masker set size.

Neff and colleagues have suggested that an important difference between observers, in terms of their thresholds in informational masking paradigms, is that some are relatively more holistic, presumably integrating information across frequency even though it is disadvantageous to do so (cf. Neff *et al.*, 1993). Such holistic listeners' thresholds might not vary compared to their counterparts in tasks other than informational masking tasks. One way to imagine this is to suggest that observers differ in the degree that they are ideal, that is, differ in the degree to which they rely on information just at the signal frequency and ignore information at other, inappropriate, frequencies. This notion has substantial predictive power. First, it is obvious that if an observer is paying relatively little attention to the signal frequency and relatively more attention to the nonsignal frequencies, his/her thresholds will be higher than an observer with the opposite emphasis. Second, an observer who pays more attention to nonsignal frequencies is expected to generate higher thresholds when the maskers are different versus the same in the two intervals. If an observer is incorporating energy at frequencies distant from the signal, changes in the masker across the intervals of the trial leads to increased variability of the decision variable. When the maskers are the same across intervals, however, effects of energy at remote frequencies might cancel out across the two intervals of a trial. This prediction holds for a linear model in which the signal is treated as a separate variable from variables associated with the maskers.

For the Ran condition the linear model under consideration is essentially a ''frequency'' model. This model considers not an arbitrary representation of the masker, but its representation on a frequency axis. Recall that the components that made up the maskers could take on values ranging from 100 to 6000 Hz, save a frequency region near the 1000-Hz signal frequency. This range was divided into 28 ''bins,'' each bin being equal in width on a logarithmic axis. The middle bin was geometrically centered at the signal frequency, 1000 Hz, and its width was the width of the protected region, 160 Hz. The power in each bin was transformed to a dB scale so as to approximate Weber-Law behavior in each frequency bin. The resulting decision variable, DV, is assumed to be the weighted sum of the levels in the frequency bins. In addition, it is assumed that the signal is weighted, and when the signal occurs in the first interval the weight is negative. Thus, DV is given by

$$DV = \left(\sum_{i \neq s} \alpha_i X_i \right)_2 - \left(\sum_{i \neq s} \alpha_i X_i \right)_1 \pm \alpha_s X_s + b,$$

where α_i is the linear weight associated with frequency bin i and b is the bias. Subscripts 1 and 2 indicate the first and second intervals, and α_s is the signal weight. The bias, b, provides the match between the expected value of DV and the mean of the observer's responses (1.5 here; the responses are coded as 1 or 2, depending on whether the observer indicated the signal was presented in interval 1 or interval 2).

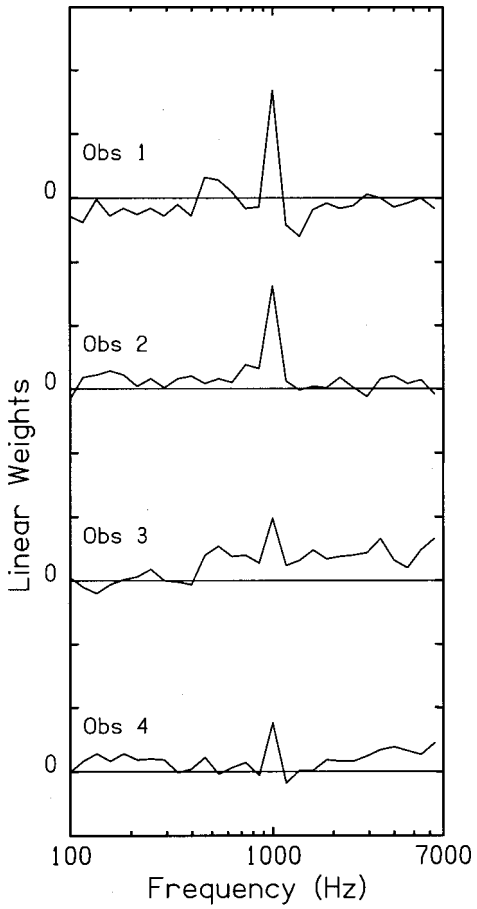


FIG. 2. Fitted linear weights are plotted as a function of frequency for the Ran condition. The solid horizontal lines indicate a weight of zero for each observer. Weights are estimated to within a scalar as indicated by the absence of ordinate label.

This model is similar to others that have been used to evaluate detection of a tone added to noise and various profile analysis experiments (e.g., Berg and Green, 1990). Indeed, in the Ran condition the analysis is much like those used in the past, except that we used a general linear model that minimized the squared deviations between *DV*'s and the observers' responses.[1]

Two concerns regarding the interpretation of the weights derived using the linear model in the Ran condition deserve note. First, because the distribution of energy in the signal bin across trials is different from the distribution in the other bins, its weight is most susceptible to deviations. Second, it might be the case that the observer's decision rule is not stationary. For example, very high or very low signal intensities might lead to differences in the weight of the signal component relative to the other components.

In the conditions other than the Ran condition, 28 frequency bins is more than the number of degrees of freedom available to be accounted for. As a result, the number of parameters in the linear model changes with the number of maskers in the set. For example, when the masker set size is 6, the 28-bit frequency axis is transformed to allow a six-dimensional orthogonal description of the stimuli (plus a bias value; see footnote 1). Regardless of set size, the signal level is one of the independent variables.

Linear weights (α's) and biases are estimated by using a linear model in which the squared deviations between *DV* and the observer's response on each trial is estimated, and the sum of the squared deviations is minimized. For the 20 sets of 50-trial tracks, the first 4 trials were removed, yielding a total of 920 trials for each observer in each condition.

Figure 2 shows the linear weights (α's) as a function of frequency fitted to the data in the Ran condition. Averaging across observers, this fitted linear model accounted for 35% of the variance in the observers' responses. Comparing Figs. 1 and 2, it is apparent that observers with higher signal weights have lower thresholds. For Obs 3 the pattern of weights potentially indicates the integration of substantial energy at frequencies away from the signal frequency. For the other observer with poorer sensitivity, Obs 4, this is not as obvious.

For all set sizes, the level of the signal component was one of the variables in the linear model. While the signal weights vary from observer to observer (cf. Fig. 2), for any one observer the estimated signal weights are fairly constant across set sizes. This suggests at least some consistency in strategy across conditions. Even though the variation in signal weights is small, for Obs 1 and 2 correlations between the weight at the signal frequency and thresholds across conditions are high ($\rho = -0.95$ and -0.98, respectively). For the less sensitive listeners, Obs 3 and 4, the correlations between

TABLE II. Proportion of variance accounted for when the data are fitted using the linear model fitted separately to the individual conditions (Separate) compared to when the linear model fitted to the Ran data is used to predict the data obtained in the other conditions (Ran).

		3A	3B	3C	6A	6B	12A	12B	24A	24B
Obs 1	Separate	0.46	0.30	0.34	0.31	0.28	0.39	0.49	0.36	0.41
	Ran	0.44	0.30	0.17	0.24	0.25	0.33	0.33	0.30	0.25
Obs 2	Separate	0.40	0.31	0.41	0.32	0.29	0.33	0.35	0.32	0.38
	Ran	0.17	0.18	0.23	0.22	0.19	0.24	0.28	0.22	0.20
Obs 3	Separate	0.24	0.40	0.27	0.40	0.35	0.43	0.50	0.37	0.46
	Ran	0.23	0.36	0.09	0.32	0.15	0.35	0.37	0.24	0.41
Obs 4	Separate	0.25	0.31	0.36	0.17	0.21	0.20	0.26	0.23	0.32
	Ran	0.23	0.10	0.32	0.14	0.16	0.12	0.21	0.15	0.20
AVG	Separate	0.34	0.33	0.35	0.30	0.28	0.34	0.40	0.32	0.39
SEM		(0.05)	(0.02)	(0.03)	(0.05)	(0.03)	(0.05)	(0.06)	(0.03)	(0.03)
AVG	Ran	0.27	0.24	0.20	0.23	0.19	0.26	0.30	0.23	0.26
SEM		(0.06)	(0.06)	(0.05)	(0.04)	(0.02)	(0.05)	(0.04)	(0.03)	(0.04)

weights at the signal frequency and thresholds are nonnegative (ρ=0.04 and 0.15, respectively). This result is consistent with the suggestion that the more sensitive observers are nearer ''signal energy detectors'' than their less sensitive counterparts and those less sensitive observers rely relatively more on the individual maskers.

Table II shows the properties of variance accounted for when set sizes ranging from 3 to 24 are fitted separately (''Separate'' entries). Across conditions and observers, separately fitted linear models account for 34% of the variance in observers' responses. The values range from 17% (Obs 4, condition 6A) to 50% (Obs 3, condition 12B). By comparison, assuming an average of 75% correct responses, one would expect a linear model to account for at most 65% of the variance in the observers' responses.[2]

The rows labeled as ''Ran'' in Table II indicates the proportion of variance accounted for when the linear model fitted to the Ran data (weights shown in Fig. 2) is used to predict responses in the other conditions.[3] On average, the cost of going from a model fitted to the individual data sets (Separate) to a single model fitted to the Ran data set (Ran) is an average reduction in the proportion of variance accounted for in the individual conditions of 0.1. Most noteworthy is that the linear model fitted to the Ran data provides reasonably good fits even when the set size is 3. The stability of the Ran model across set size suggests stability in observers' strategies—if observers' strategies depend substantially on set size, it is unlikely (but not impossible) that a single linear model could account for the data in the different conditions.

In another exploration, the data in the Ran and set size 24 data sets were fitted using a linear model with 13 rather than 28 frequency bins so that the combination weights estimated in these two conditions could be directly compared. An ANOVA failed to reveal significant differences in the estimated linear weights. That result, plus the fact that thresholds in the Ran condition and when the set sizes is 24 are not reliably different, suggest that observers' decision strategies are similar whether the maskers are drawn from a very large pool versus a pool size of 24.

VI. SUMMARY AND CONCLUSIONS

Consistent with past results, the current experiment indicates large individual differences in informational masking, even when the number of maskers tested is small and the maskers are reliably recognized. The results also suggest that when maskers are drawn from a set size of 24, the particular masker set tested is relatively unimportant. In terms of statistical sampling, masker set sizes of 12 to 24 appear to provide a reasonable substitute for randomly drawn maskers. Such a conclusion is tempered by the fact that only two masker sets were tested. Finally, thresholds estimated by separating the observers' responses into two pools, trials in which the maskers were either different or the same across the two presentation intervals, did not reveal substantial differences in amount of masking. Restricting the comparison to the thresholds estimated when the set size is 3, where the possibility that the maskers are the same in the two intervals is $\frac{1}{3}$, thresholds were on average only 2 dB lower when the maskers were the same versus different in the two intervals.

Analyses of the trial-by-trial data provide little compelling evidence to support the hypothesis that observers change their decision strategies as the set sizes are altered. This appears to hold even for set sizes as small as 3. Noting that for some of the analyses there was a tendency for the results measured using a set size of 3 to vary slightly from the other conditions (e.g., the memory d' values were somewhat higher), it seems reasonable to suggest that there is little cost in examining informational masking using set sizes as small as 6.

ACKNOWLEDGMENTS

This work was supported by Grant Nos. R01 DC 02012 and R01 DC04545 from the National Institute of Health. Christine Mason, two anonymous reviewers, and associate editor Dr. Marjorie Leek provided valuable comments on an early version of this manuscript.

[1]First consider the linear model fitted to the Ran data. For each trial, the decision variable is as defined in the text. The linear model predicts a decision variable,

$$\hat{DV} = \begin{bmatrix} \hat{DV}_1 \\ \hat{DV}_2 \\ \vdots \\ \hat{DV}_n \end{bmatrix} = \begin{bmatrix} Y_1 & 1 \\ Y_2 & 1 \\ \vdots & \vdots \\ Y_n & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ b \end{bmatrix}.$$

Here, n is the number of trials, and each Y_k is a vector of differences in level across the two intervals at each frequency bin. Because there are 28 bins, each Y_k has 28 entries. Overall, the matrix with the Y's has dimension $n \times 29$. α is a column vector with length equal to the number of bins, and b, the bias, is a real number. The vector containing the α's has dimension 29×1. Using a simpler notation, and referring to these two matrices as Y and α, one has $\hat{DV} = Y\alpha$. The observer's response for each trial may be described as a column vector with n entries,

$$T = \begin{bmatrix} T_1 \\ T_2 \\ \vdots \\ T_n \end{bmatrix}.$$

Recall that the T's take on values of either 1 or 2, indicating first or second interval.

To estimate the values of the α's and of b, the sum of the squared deviations, $\Sigma(\hat{DV} - T)^2$, is minimized. This leads to weights that are essentially the same as when the value to be minimized is $\Sigma(\hat{T} - T)^2$. Using the simplified matrix notation introduced above, the least square estimator for minimizing $\Sigma(\hat{DV} - T)^2$ is given by $Y'Y\alpha = Y'T$ (cf. Neter *et al.*, 1990). The entries of the matrix α can then be estimated: $\alpha = [Y'Y]^{-1}Y'T$.

The linear fit might be visualized as follows. Imagine that for each stimulus interval a variable X is formed, and the decision variable, DV, is the difference $X_2 - X_1$ where 1 and 2 indicate the first and second intervals, respectively. Imagine that when the signal is in the first interval the DV is usually negative and when the signal is in the second interval the DV is usually positive. Likewise, the observer's responses will tend to be 1 when the signal is in the first interval and tend to be 2 when the signal is in the second interval. Note that for this example the mean of the DV is zero but the mean of the observer's responses is 1.5—the bias term allows the mean of the predicted DV to shift. Consider just the predicted DV's associated with trials in which the signal is in the first interval. By altering the weights, the mean and the spread of the predicted DV values can be altered. As a result, the distribution of the predicted DV becomes most like (in a least-squares sense) the distribution of the observer's responses. When the signal is in the second interval, the change in the DV is symmetric.

It should be noted that combination weights could be fitted for each of

the 28 frequency bins only if the stimulus set has sufficient degrees of freedom. As an example consider the case when there are only 3 maskers in a set. Three maskers will present at most 18 unique masker frequencies (six tones for each of the three maskers) plus the signal frequency. Additionally, for any one masker the presence of any one tone predicts the frequencies of the other 5 tones. For a masker set of size k there are only k orthogonal dimensions available to the linear model.

Mathematically, the degrees of freedom present in the stimuli for small masker set sizes is reflected in the rank of matrix $Y'Y$, which has a rank of the number of maskers plus 1 (the number of maskers plus the signal component plus the bias minus 1). In practice, the linear model fitted to the data (except in the Ran condition) proceeded as follows. The matrix Y was derived across all trials and for 28 frequency bins. Then, the matrix $Y'Y$ was transformed using a QR decomposition (cf. Strang, 1980) in which $Y'Y$ was rewritten as $Y_Q Y_R$ where Y_Q is an unitary (invertible) matrix and Y_R is an upper triangular matrix. Having rewritten $Y'Y$ as $Y_Q Y_R$, the least squares estimator becomes $Y_Q Y_R \alpha = Y'R$, or $Y_R \alpha = Y_Q^{-1} Y'R$. Whenever a diagonal entry of Y_R was zero, the corresponding element of α was also assigned a value of zero.

[2]The linear model generates a continuous rather than dichotomous variable. Depending on the percent of correct responses in a 2IFC task, the percent of variance that can be accounted for in the data set varies. If one assumes that the variables are statistically independent, the proportion of variance that can be accounted for is provided by the squared correlation coefficient,

$$\rho^2 = \frac{[\sqrt{2/\pi} e^{-(1/2)z^2} + z(2P-1)]^2}{1+z^2},$$

where ρ is the correlation coefficient, P is the percent correct in a two-interval forced choice task, and z is the z-score associated with P. A sketch of the proof derived by Zhongzhou Tang follows.

In a 2IFC task, assume that the signal is as likely to be in each of the two intervals. For each interval a variable X is formed, and the normally distributed decision variable, DV, is the difference $X_2 - X_1$ where 1 and 2 indicate the first and second intervals, respectively. Note that the distribution of DV is symmetric about the $DV=0$ point depending on whether the signal occurs in the first versus second interval.

Without loss of generality, let T be the response

$$T = \begin{cases} -1 \cdots DV < 0 \\ 1 \cdots DV > 0 \end{cases}.$$

Assuming T is unbiased, the expected value of T, like the expected value of DV, is 0.

The squared correlation coefficient between the decision variable and the response is given by

$$\rho^2 = \frac{[\Sigma_i (DV_i - \overline{DV})(T_i - \overline{T})]^2}{\Sigma_i (DV_i - \overline{DV})^2 \Sigma_i (T_i - \overline{T})^2} = \frac{(\Sigma_i DV_i T_i)^2}{\Sigma_i DV_i^2 \Sigma_i 1},$$

where i varies from 1 to n, the total number of trials. DV and T share the same sign, so

$$\rho^2 = \frac{(\Sigma_i |DV_i|)^2}{\Sigma_i DV_i^2 \Sigma_i 1} = \frac{(\Sigma_i |DV_i|)^2/n^2}{\Sigma_i DV_i^2/n}.$$

First consider the denominator. Note that $\Sigma_i DV_i^2/n$ is the expected value of the squared decision variable (as $n \to \infty$). Due to the symmetry associated with the signal present in interval 1 versus interval 2 (i.e., DV_2^2 is the same), the numerator becomes $\sigma^2 + \mu^2$, where $\pm\mu$ is the expected value of the decision variable when the signal is in the first (minus) or second (plus) interval and σ^2 is the variance associated with the distributions when the signal is in the first or second interval.

Next, consider the numerator. Note that $(\Sigma_i |DV_i|)^2/n^2 = (\Sigma_i |DV_i|/n)^2$ is the squared expected value of the absolute value of the decision variable (as $n \to \infty$). Note that the distribution of $|DV|$ is the same whether the signal is in the first or second interval. Additionally, the DVs ($DV1$ and $DV2$, depending on whether the signal in first or second interval) are separately normally distributed. This leads to $(\Sigma_i |DV_i|)^2/n^2 = [E(|DV|)]^2 = [\sqrt{2/\pi}\sigma e^{-(1/2)(\mu/\sigma)^2} + \mu(2P-1)]^2$, where μ is the mean of the normal distribution when the signal is added to the second interval.

Taking the ratio of the numerator and the denominator,

$$\rho^2 = \frac{(\Sigma_i |DV_i|)^2/n^2}{\Sigma_i DV_i^2/n} = \frac{[(2/\pi)\sigma e^{-(1/2)(\mu/\sigma)^2} + \mu(2P-1)]^2}{\mu^2 + \sigma^2}.$$

Dividing the numerator and denominator by σ^2, for a standard normal variable z,

$$\rho^2 = \frac{[(2/\pi) e^{-(1/2)z^2} + z(2P-1)]^2}{1+z^2}.$$

[3]The predicted response associated with any one DV depends on the sign of the DV relative to the bias (i.e., above or below). Thus, scaling the DV does not alter the predicted response. Nonetheless, scaling the linear weights does impact the proportion of variance accounted for in the responses by the linear model. The values shown in Table II are the maximum proportion of variance accounted for in the different conditions when the weights in the Ran condition are optimally scaled.

Abramowitz, M., and Stegun, I. A. (**1964**). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* (Dover, New York).

Berg, B. G., and Green, D. M. (**1990**). "Spectral weights in profile listening," J. Acoust. Soc. Am. **88**, 758–766.

Dai, H. (**1995**). "On measuring psychometric functions: A comparison of the constant-stimulus and adaptive up-down methods," J. Acoust. Soc. Am. **98**, 3135–3139.

Efron, B., and Tibshirani, R. J. (**1993**). *An Introduction to the Bootstrap*, Monographs on Statistics and Applied Probability 57 (Chapman and Hall, New York).

Green, D. M. (**1964**). "Consistency of auditory detection judgments," Psychol. Rev. **71**, 392–407.

Kidd, G., Mason, C. R., Deliwala, P. S., Wood, W. S., and Colburn, H. S. (**1994**). "Reducing informational masking by sound segregation," J. Acoust. Soc. Am. **95**, 3475–3480.

Kidd, G., Jr., Mason, C. R., and Arbogast, T. L. (**2001**). "Similarity, uncertainty and masking in the identification of nonspeech auditory patterns," J. Acoust. Soc. Am. (in press).

Lutfi, R. A. (**1989**). "Informational processing of complex sound. I. Intensity discrimination," J. Acoust. Soc. Am. **86**, 934–944.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 167–177.

Neff, D. L. (**1995**). "Signal properties that reduce masking by simultaneous, random-frequency maskers," J. Acoust. Soc. Am. **98**, 1909–1920.

Neff, D. L., and Callaghan, B. P. (**1986**). "Effective properties for multicomponent simultaneous maskers under conditions of uncertainty," J. Acoust. Soc. Am. **83**, 1833–1838.

Neff, D. L., and Dethlefs, D. L. (**1995**). "Individual differences in simultaneous masking with random-frequency, multicomponent maskers," J. Acoust. Soc. Am. **98**, 125–134.

Neff, D. L., and Green, D. M. (**1987**). "Masking produced by spectral uncertainty with multicomponent maskers," Percept. Psychophys. **41**, 408–415.

Neff, D. L., Dethlefs, T. M., and Jesteadt, W. (**1993**). "Informational masking for multicomponent maskers with spectral gaps," J. Acoust. Soc. Am. **94**, 3112–3126.

Neter, J., Wasserman, W., and Kutner, M. H. (**1990**). *Applied Linear Statistical Models*, 3rd ed. (IRWIN, Homewood, IL).

Oh, E. L., and Lutfi, R. A. (**1999**). "Nonmonotonicity of informational masking," J. Acoust. Soc. Am. **106**, 3521–3528.

Strang, G. (**1980**). *Linear Algebra and its Applications* (Academic, New York).

Richards, V. M., and Zhu, S. (**1994**). "Relative estimates of combination weights, decision criteria and internal noise based on correlation coefficients," J. Acoust. Soc. Am. **95**, 423–434.

Watson, C. S. (**1987**). "Uncertainty, informational masking, and the capacity of immediate auditory memory," in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ).

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.

Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns

Gerald Kidd, Jr., Christine R. Mason, and Tanya L. Arbogast
Department of Communication Disorders and Hearing Research Center, Boston University, 635 Commonwealth Avenue, Boston, Massachusetts 02215

(Received 8 June 2001; revised 2 October 2001; accepted 7 December 2001)

This study examined whether increasing the similarity between informational maskers and signals would increase the amount of masking obtained in a nonspeech pattern identification task. The signals were contiguous sequences of pure-tone bursts arranged in six narrow-band spectro-temporal patterns. The informational maskers were sequences of multitone bursts played synchronously with the signal tones. The listener's task was to identify the patterns in a 1-interval 6-alternative forced-choice procedure. Three types of multitone maskers were generated according to different randomization rules. For the least signal-like informational masker, the components in each multitone burst were chosen at random within the frequency range of 200–6500 Hz, excluding a ''protected region'' around the signal frequencies. For the intermediate masker, the frequency components in the first burst were chosen quasirandomly, but the components in successive bursts were constrained to fall in narrow frequency bands around the frequencies of the components in the initial burst. Within the narrow bands the frequencies were randomized. This masker was considered to be more similar to the signal patterns because it consisted of a set of narrow-band sequences any one of which might be mistaken for a signal pattern. The most signal-like masker was similar to the intermediate masker in that it consisted of a set of synchronously played narrow-band sequences, but the variation in frequency within each sequence was sinusoidal, completing roughly one period in a sequence. This masker consisted of discernible patterns but not patterns that were part of the set of signals. In addition, masking produced by Gaussian noise bursts—thought to produce primarily peripherally based ''energetic masking''—was measured and compared to the informational masking results. For the three informational maskers, more masking was produced by the maskers comprised of narrow-band sequences than for the masker in which the frequencies were not constrained to narrow bands. Also, the slopes of the performance-level functions for the three informational maskers were much shallower than for the Gaussian noise masker or for no masker. The findings provided qualified support for the hypothesis that increasing the similarity between signals and maskers, or parts of the maskers, causes greater informational masking. However, it is also possible that the greater masking was a consequence of increasing the number of perceptual ''streams'' that had to be evaluated by the listener. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1448342]

PACS numbers: 43.66.Dc, 43.66.Mk, 43.66.Ba [MRL]

I. INTRODUCTION

Virtually all real-world listening environments are comprised of multiple sources of sound, either sounds that arise from independent sources or sounds that follow multiple paths (e.g., echoes) from a single source, or, perhaps most typically, both at once. For a listener in a multisource sound field, the task is often to selectively attend to a desired source while ignoring unwanted sources. Accomplishing this task involves detecting and segregating the various sound sources and evaluating whether they deserve attention. The sound field may be both complex *and* dynamic as sound sources come and go and change over time, often requiring that attention be distributed or redirected. Attending to one particular source and ignoring unwanted sources may be a challenging task if the desired sound is soft relative to the unwanted sounds, if the unwanted sounds are difficult to ignore, or if the listener is uncertain about which sound to attend to. The factors that influence signal selection in multisource listening environments have been of great interest to auditory scientists dating at least from Cherry's (1952) framing of the ''cocktail party problem'' (see also Pollack and Pickett, 1958) continuing through contemporary times as evidenced by the recent works of Bregman (1990), Yost (1991, 1997), Yost *et al.* (1996), Darwin and Carlyon (1995), Hawley *et al.* (1999), Freyman *et al.* (1999), Brungart (2001), and Brungart and Simpson (2001), among others.

In multisource listening, sounds interfere with one another. *Masking* is the term used to describe the interference one sound causes in the reception or processing of a second target sound. One sound may mask another sound by reducing the probability that it will be detected by the observer. It may also cause masking by interfering with the extraction of important properties of the sound such as its pitch or timbre, by diminishing the ability to identify the source or accurately estimate its location, or by interfering with the recognition of any special information emanating from the source such as spoken language.

In recent years, there has been an increasing body of

evidence to support the theory that auditory masking consists of two separate components that originate at different physiological levels. These physiological levels are not precisely determined, but may roughly be divided into the categories of ''peripheral'' and ''central.'' Peripheral masking is usually thought of as overlapping patterns of excitation on the basilar membrane and in the auditory nerve. Performance is limited by signal-to-noise constraints in the frequency regions surrounding the signal. Because in general there is a good correspondence between the amount of energy in the masker falling near the signal frequency and the amount of masking that occurs, peripheral masking is often called ''energetic masking'' (cf. Watson, 1987) and the ''critical-band energy-detector model'' is broadly successful in predicting peripheral masking (e.g., Green and Swets, 1974). In contrast, central masking[1] occurs despite a sufficiently robust representation in the auditory nerve for successfully accomplishing the task. Thus, in peripheral masking, performance is limited because there is insufficient information available to a central processor to solve the task, while in central masking sufficient information is available, but it is the processing itself, or nonoptimal decision rules applied to the processed stimulus, that lead to failure in the task. Stimulus manipulations and listening conditions that create high observer uncertainty have been shown to produce large amounts of masking that appear to be central in origin. This type of masking has been termed ''informational masking'' (Pollack, 1975; Watson *et al.*, 1975, 1976).

Most studies of informational masking have measured the detection of the presence of a tone (e.g., Neff and Green, 1987; Leek and Watson, 1984; Neff, 1995; Kidd *et al.*, 1994; Oh and Lutfi, 1998) or discrimination of a difference in the frequency or intensity of a tone (e.g., Watson *et al.*, 1975, 1976; Howard *et al.*, 1984; Kidd *et al.*, 1986; Leek *et al.*, 1991; Neff and Jesteadt, 1996). Detecting the presence of a pure tone or discriminating a change in the frequency or intensity of a pure tone are tasks that are particularly well-suited for the study of informational masking because it is possible to create masking by placing masker energy at frequencies remote from the signal. This minimizes energetic masking and provides support for the interpretation that the interference in performance is due to central factors.

While it is of great interest to study informational masking for detection or discrimination tasks for pure tones, a more complete understanding of the factors affecting multisource listening includes study of more sophisticated and demanding tasks of the observer using complex sounds. However, for some tasks—such as the recognition of speech from one talker among several talkers—it is difficult to control for energetic masking because the information is distributed over a broad range of frequencies. This complicates the determination of whether the masking that is observed is due to peripheral or to central mechanisms. Recently, we have used a nonspeech pattern identification task to study binaural processing and informational masking in complex, multisource environments (Kidd *et al.*, 1995a, 1998). In those experiments the task of the listener was to choose which of a set of previously learned narrow-band patterns was present on a given experimental trial. The presence of informational maskers—typically sequences of randomly drawn multitone complexes played synchronously with the signal—can greatly interfere with performance, causing shifts in the identification-level functions of 50 dB or more. This procedure has the advantage that it presumably requires more complex judgments than do pure-tone detection or difference limen tasks, but maintains the stimulus control necessary to produce large amounts of informational masking with concomitantly small amounts of energetic masking. Further, because the stimuli are presented well above detection threshold, the task bears a closer correspondence to the ''cocktail party problem,'' where the challenge is to select among and extract information from audible sources.

Although there have been many studies of informational masking in the past decade, the phenomenon is still not well understood and little is known about what causes interference in various tasks, especially for suprathreshold listening situations. In this study we examined the hypothesis that informational masking for suprathreshold identification tasks could be manipulated by varying the degree to which the masker and signal share similar properties. As a means for evaluating this hypothesis, we used the nonspeech pattern identification task mentioned above (cf. Kidd *et al.*, 1998). The basic idea was that the greater the similarity between the signals and the maskers (or parts of the maskers, as discussed in following sections) along a relevant perceptual dimension, the greater the uncertainty they would create for the listener which consequently would be reflected in larger amounts of informational masking.

In the following experiments, the identification of nonspeech patterns was measured in quiet and in the presence of four maskers. One masker was Gaussian noise that produces primarily peripherally based energetic masking and was included as a reference against which the results from the informational maskers could be compared. Along the postulated continuum of similarity, the Gaussian noise masker is the least similar to the signal set. The three informational maskers were comprised of sequences of multitone complexes played synchronously with the elements of the signal patterns. They differed in the rules used to generate the frequencies of the complexes in a manner that was expected to vary the degree of similarity between the maskers, or portions of the maskers, and the set of signals. If similarity is important in producing informational masking, and if our assumption is correct that these maskers vary in their degree of similarity to the signal set, then we should expect that the three informational maskers would produce different amounts of masking. It should be noted at the outset, however, that our notion of varying the degree of similarity of the informational maskers was only partially successful, both according to the experimental results and the results of a similarity metric applied to the stimuli, and an alternative explanation based on a limitation in the capacity to process multiple co-occurring auditory ''streams'' is also considered in the discussion.

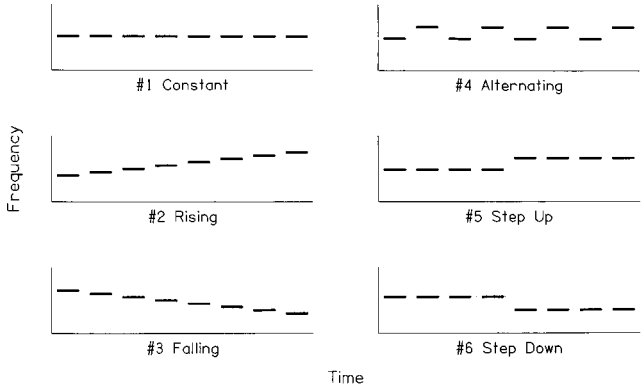


FIG. 1. Schematic of the six signal patterns in sound spectrogram form.

II. METHODS

A. Subjects

Three young-adult university students with normal hearing served as listeners. All three listeners had prior experience participating in psychoacoustic experiments and were trained on the experimental procedures and stimuli for several hundred trials prior to collection of the data reported here. The subjects were trained until stable performance was observed and the range of levels necessary to map out the performance-level functions for each subject in each condition was determined. The training phase was identical to the test phase described below, the only difference being that the practice data are not reported here. The subjects were paid for their participation.

B. Stimuli

All sounds were generated digitally and played through 16-bit digital-to-analog converters (TDT) at a rate of 20 000 samples, then low-pass filtered at 7500 Hz and attenuated by programmable attenuators. The conversion from digital amplitude to voltage was verified prior to each experimental session and the acoustical output of the headphone was calibrated using a standard coupler and sound-level meter. Also, the algorithms for combining tones into complexes were verified prior to conducting the experiment. The signals were sequences of eight contiguous tone bursts arranged in six spectro-temporal patterns. These signal patterns have been used previously in studies of nonspeech sound identification (Kidd *et al.*, 1995a, 1998; see also Weber, 1977). Each burst was 60 ms in duration including 10-ms cosine-squared rise/decay times. The six frequency patterns, illustrated in Fig. 1, were: (1) constant; (2) rising; (3) falling; (4) alternating; (5) step up; and (6) step down.

The signal patterns (and two of the informational maskers, as described below) were generated from 16 frequency bands spaced equally on a logarithmic scale. The lowest and highest frequencies of the bands were 200 and 6540 Hz. Within each band there were eight frequencies from which the patterns were constructed with each frequency separated by 2% of the band's center frequency. The width of each frequency band was thus 14% of the center frequency. On each trial one of the 16 possible bands was

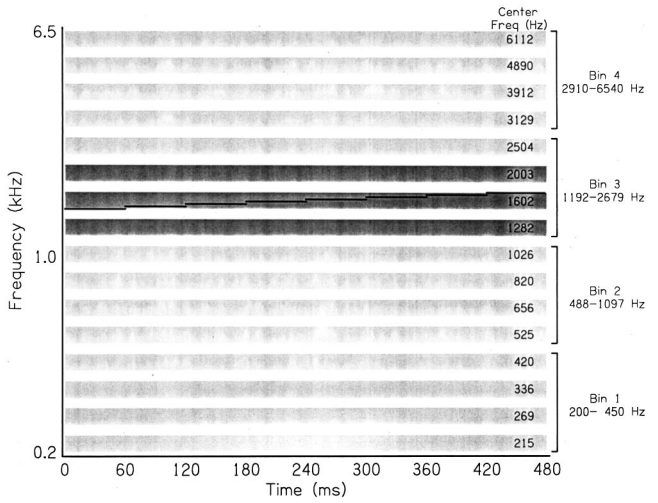


FIG. 2. Schematic of the 16 bands (shaded regions) from which the signal patterns and multitone maskers for the MBDN and FMOD conditions were constructed. A signal pattern (#2) is shown in heavy lines. The three darker-shaded bands indicate the ''protected region'' containing and adjacent to the signal. The bins used for pooling data for *post hoc* analysis for the different signal frequencies are shown to the right of the bands.

selected at random to contain a signal pattern. The signal—one of the six shown in Fig. 1—was then generated from the eight frequencies available within that band. For signals #4 (alternating), #5 (step up), and #6 (step down), the patterns were formed from the highest and lowest frequencies in the band. The constant frequency of signal #1 was chosen randomly from the eight possible frequencies in the selected band. Signal level was equal for tones in a pattern. When the signal was masked by an informational masker, a ''protected region'' which excluded masker tones was established around the signal. The protected region was approximately equal to 35% of the center frequency of the pattern. Figure 2 illustrates the 16 narrow bands and shows signal pattern #2 (rising) located in one of the bands with the two protected bands flanking the signal.

Also shown in Fig. 2 are the four subsets of the 16 frequency bands used for pooling the results for *post hoc* analysis discussed in a later section. Signal levels were chosen in 10-dB steps (5-dB steps when presented in the Gaussian noise masker) over a sufficient range to estimate the performance-level functions for that particular condition. There were 60 trials in each block of trials.

There were two types of maskers: Gaussian noise, intended to produce primarily energetic masking; and randomized multitone complexes, intended to produce primarily informational masking. The maskers were played in sequences of eight 60-ms bursts gated synchronously with the signal bursts.

The Gaussian noise masker was comprised of a series of eight independent bursts with 10-ms rise–fall times bandpass filtered from 200–6540 Hz and presented at 70 dB SPL. There was no protected region surrounding the signal as was the case for the informational maskers. Figure 3 (top left panel) is a schematic of the Gaussian noise masker illustrated along with signal #2.

The informational maskers were sequences of multitone bursts comprised of tones that were generated according to

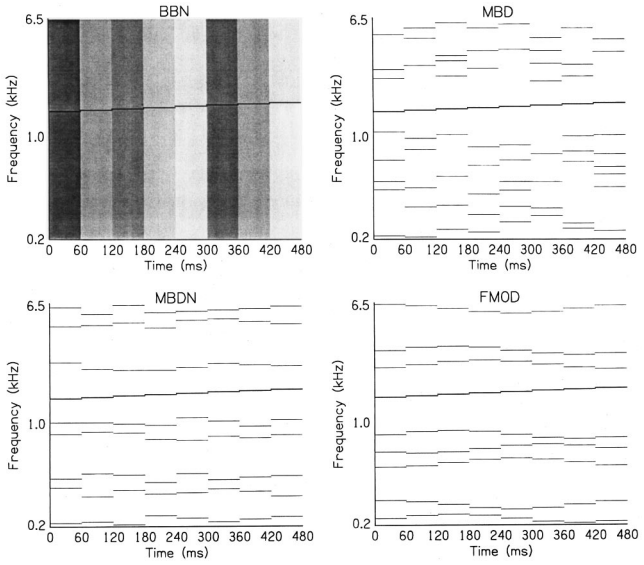


FIG. 3. Schematic representations of the four maskers. The upper left panel shows the broadband noise masker (BBN), the upper right panel shows the multiple-bursts different (MBD) masker, the lower left panel shows the multiple-bursts different narrowband (MBDN) masker, and the lower right panel shows the frequency-modulated (FMOD) masker. Note that the different shadings for the BBN masker indicate independent masker bursts. See the text for more details.

three different randomization schemes—one scheme for each type of informational masker. For all three informational maskers there were eight tones in each burst of the sequence and the levels of the tones were randomly chosen from a uniform distribution of levels ranging from 50–70-dB SPL. Thus, the overall level of a given masker burst depends on the specific levels of the tones chosen on that particular burst, but would always fall between approximately 59 and 79 dB SPL. In all cases the frequencies of the tones within a given burst were spaced farther apart than 1/duration (i.e., 1/0.06 s=16.7 Hz). In one masker, called ''multiple-bursts different'' (MBD), the eight frequencies in each burst were chosen randomly from the frequency range from 200–6540 Hz, excluding the protected region, and each masker tone level was a random sample from the range of 50–70 dB SPL. The MBD masker is illustrated in Fig. 3 (upper right panel) along with signal #2.

The other two informational maskers contained frequencies drawn from the 16 frequency bands described above for the signal. One masker is designated ''multiple-bursts different, narrow-band'' (MBDN). On each trial the band containing the signal and the two adjacent bands were excluded from selection as masker bands. Eight of the 13 remaining bands were randomly chosen to contain masker tones. On each burst, one masker tone fell within each of the selected bands. Within each masker band, the frequency was a random sample on each burst from among the eight possible frequencies within the band. This resulted in a masker comprised of eight narrow-band (over time) tone sequences. The level of the first tone in a given masker band was chosen at random on each presentation from a uniform range from 50–70 dB SPL but was held constant for subsequent bursts in that band for that presentation. The MBDN masker is shown in Fig. 3 (lower left panel) paired with signal #2.

The third informational masker was also created from the same set of 16 bands of 8 frequencies as were used to generate the MBDN maskers and on each presentation, 8 of the 13 available bands (excluding signal and flankers) were chosen at random to contain maskers. However, for this masker the frequency in the first burst in a given band was chosen at random from the eight possible frequencies, but subsequent frequencies in the burst sequence followed an orderly sinusoidal pattern. The variation in frequencies in a given masker band thus roughly approximated a frequency-modulated sine wave with random starting phase (among the eight possible frequencies) and completed one cycle of modulation during the eight bursts in a sequence. This masker is designated ''FMOD'' and is illustrated in Fig. 3 (lower right panel) also paired with signal #2. As with the MBDN masker, the level of the first burst in each band was chosen randomly from a uniform distribution ranging from 50–70 dB SPL and the subsequent tones in that band were played at that level.

C. Procedures

The experimental task was one-interval, six-alternative, forced-choice (1I-6AFC) with response feedback provided after every trial. On each trial, the listener was presented with one signal pattern chosen at random from the set of six patterns, played synchronously with a masker. The task of the listener was to choose which of the six signal patterns had been presented. The trials were grouped in blocks spanning a range of signal levels sufficient to map out the performance-level functions. Within each block, the signal levels tested were randomized.

The subjects listened monaurally through a TDH-50 earphone. The sounds were played to each listener's preferred ear. The listeners were seated in a double-walled IAC booth with an LCD computer terminal that contained a schematic illustration similar to Fig. 1 mounted above the response keys corresponding to each pattern. After the stimulus presentation, the listeners were permitted to take as much time as they liked to indicate their response on the keyboard. Interval timing and response feedback were displayed on the monitor on every trial.

Five conditions were tested: signal only (QUIET), Gaussian noise (BBN), and the MBD, MBDN, and FMOD multitone maskers. All of the data plotted for individual listeners represent percent-correct identification computed over approximately 200 trials per point. The listeners participated in 2-h time blocks two or three times per week and were allowed to take ''breaks'' as needed during each session.

III. RESULTS

Figure 4 contains the data plotted on coordinates of percent correct identification versus signal level. The data were pooled across signal frequency within the four bins corresponding to bands 1–4, 5–8, 9–12, and 13–16 as indicated from bottom to top in Fig. 2. The different frequency bins are plotted in the four rows in Fig. 4. Each column of Fig. 4 contains the data from one subject. The data points were fit[2] with logistic functions having the form

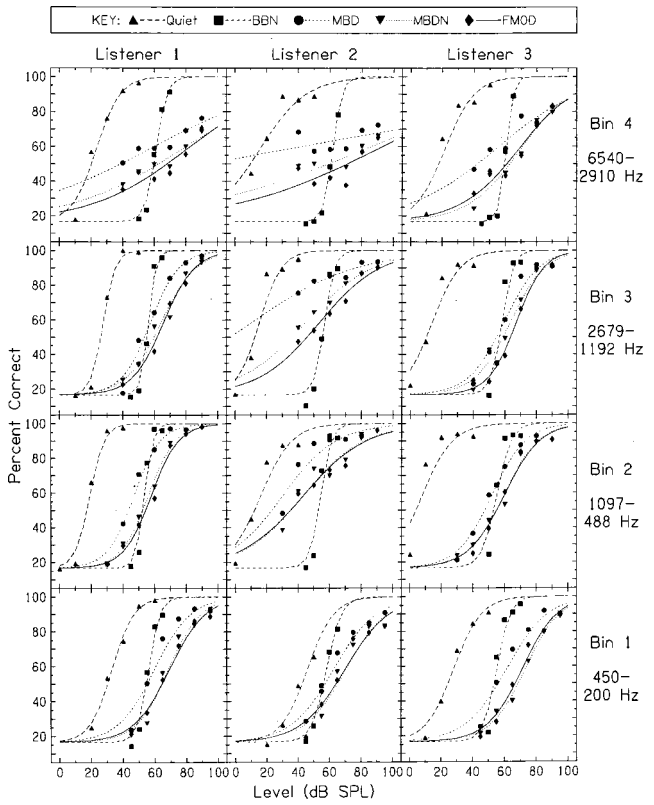


FIG. 4. Identification-level functions for individual subjects for the four frequency bins. The triangles are for the signal-only (QUIET) condition, squares are for the Gaussian noise masker (BBN), circles are for the MBD masker, inverted triangles are for the MBDN masker, and diamonds are for the FMOD masker. The lines are best-fitting logistic functions.

$$p(c) = \alpha + \frac{1.0 - \alpha}{1.0 + e^{-k(m-x)}},$$

where $\alpha = 0.167$ (chance performance, 1 of 6), m is the level corresponding to the midpoint of the function (about 58.5% correct), and k is the slope. The slopes, k, and midpoints, m, of the fitted logistic functions, along with estimates of goodness of fit,[3] are contained in Table I for each subject and averaged across the group of subjects.

First, both for the QUIET condition and all four masked conditions, performance increased with increasing signal level. The slopes of the performance-level functions vary across maskers and frequency bins. In the QUIET condition, the group average slopes varied from about 0.10 to 0.17. For the BBN condition, the slopes of the functions increased considerably relative to QUIET and, on average, ranged from 0.24 at the lowest frequency bin to 0.36 at the highest frequency bin. A similar steepening of slope of the performance-level functions in Gaussian noise relative to quiet has been noted previously by Kidd *et al.* (1995a). Because QUIET and BBN psychometric functions are not parallel, the amount of masking was greater at low signal-to-masker ratios (S/M) and less at high S/M. As is apparent from inspection of Fig. 4 and Table I, the group mean slopes for all of the informational masking conditions are much shallower than for the QUIET or, especially, for the BBN condition. This again means that the amount of masking (relative to QUIET) depends greatly on S/M. For MBD the group average slopes ranged from 0.02 to 0.09, for MBDN the range was from 0.04 to 0.09, and for FMOD the range was from 0.04 to 0.09 and, for all three maskers, the slopes were shallowest for the highest frequency bin. Figure 5 contains plots of the slope values.

An analysis of variance of the slopes of all five conditions indicated that masker type was a significant main effect ($F[4,8] = 98.7$; $p < 0.001$) but that frequency was not ($F[3,6] = 1.33$; $p = 0.348$). There was a significant interaction between masker type and frequency ($F[12,24] = 3.43$; $p < 0.005$). *Post hoc* analyses (Duncan's multiple range test) indicated that the three informational maskers were not significantly different from each other but that all were significantly shallower than for the BBN masker or for QUIET,

TABLE I. The midpoint, m, and slope, k, of the fitted functions for all five conditions (columns) and four frequency bins (rows). Also given is the proportion of the variance, PV, accounted for by the fit. Within each row the individual values are shown for all three subjects along with the intersubject means and standard errors.

		CONDITION														
Frequency bins (range in Hz)	Listener	QUIET			BBN			MBD			MBDN			FMOD		
		m	k	PV	m	k	PV	m	k	PV	m	k	PV	m	k	PV
4	1	23.4	0.13	0.96	61.2	0.28	0.99	56.6	0.02	0.88	77.2	0.03	0.92	80.6	0.03	0.83
	2	15.7	0.07	0.88	61.8	0.35	0.99	32.6	0.01	0.20	81.4	0.02	0.56	89.5	0.02	0.49
(6540–2910)	3	21.7	0.11	0.89	60.4	0.44	0.99	53.6	0.04	0.91	69.5	0.06	0.95	68.5	0.05	0.93
	Avg(se)	20.3(2.3)	0.10 (0.02)	0.91 (0.02)	61.2 (0.4)	0.36 (0.04)	0.99 (0.00)	47.6 (7.6)	0.02 (0.01)	0.66 (0.23)	76.0 (3.5)	0.04 (0.01)	0.81 (0.13)	79.5 (6.1)	0.04 (0.01)	0.75 (0.13)
3	1	27.8	0.26	0.99	56.3	0.42	0.99	57.5	0.11	0.97	64.2	0.09	0.97	65.9	0.10	0.99
	2	15.5	0.14	0.93	56.8	0.31	0.96	10.4	0.03	0.90	44.5	0.05	0.99	53.5	0.05	0.97
(2679–1192)	3	14.0	0.11	0.93	58.2	0.28	0.94	59.3	0.09	0.97	63.2	0.09	0.97	67.1	0.11	0.98
	Avg(se)	19.1(4.4)	0.17 (0.05)	0.95 (0.02)	57.1 (0.6)	0.34 (0.04)	0.96 (0.02)	42.4 (16.0)	0.08 (0.02)	0.95 (0.02)	57.3 (6.4)	0.08 (0.01)	0.99 (0.00)	62.2 (4.3)	0.09 (0.02)	0.98 (0.01)
2	1	19.2	0.22	0.99	53.6	0.36	0.97	47.0	0.12	0.98	55.8	0.12	0.99	57.9	0.12	0.98
	2	16.2	0.10	0.92	54.5	0.29	0.94	30.2	0.06	0.85	43.1	0.05	0.88	43.3	0.05	0.91
(1097–488)	3	7.8	0.10	0.82	54.9	0.22	0.89	51.8	0.09	0.99	59.1	0.09	0.98	59.6	0.09	0.97
	Avg(se)	14.4(3.4)	0.14 (0.04)	0.91 (0.05)	54.3 (0.4)	0.29 (0.04)	0.93 (0.02)	43.0 (6.6)	0.09 (0.02)	0.94 (0.05)	52.7 (4.9)	0.09 (0.02)	0.95 (0.04)	53.6 (5.2)	0.08 (0.02)	0.96 (0.02)
1	1	33.1	0.15	0.99	56.2	0.28	0.97	59.7	0.09	0.94	68.4	0.10	0.97	69.3	0.09	0.98
	2	46.0	0.12	0.96	58.2	0.21	0.98	62.8	0.07	0.97	70.3	0.08	0.94	70.0	0.07	0.94
(450–200)	3	28.1	0.12	0.97	55.3	0.24	0.92	61.4	0.08	0.95	73.1	0.09	0.99	71.3	0.09	0.98
	Avg(se)	35.7(5.4)	0.13 (0.01)	0.98 (0.01)	56.6 (0.9)	0.24 (0.02)	0.96 (0.02)	61.3 (0.9)	0.08 (0.00)	0.95 (0.01)	70.6 (1.3)	0.09 (0.01)	0.97 (0.02)	70.2 (0.6)	0.08 (0.01)	0.97 (0.02)
Average (s.d.)		22.4(9.3)	0.14 (0.03)	0.94 (0.03)	57.3 (2.8)	0.31 (0.05)	0.96 (0.02)	48.6 (8.8)	0.07 (0.03)	0.88 (0.14)	64.1 (11.0)	0.07 (0.03)	0.93 (0.08)	66.4(11.1)	0.07 (0.03)	0.91 (0.11)

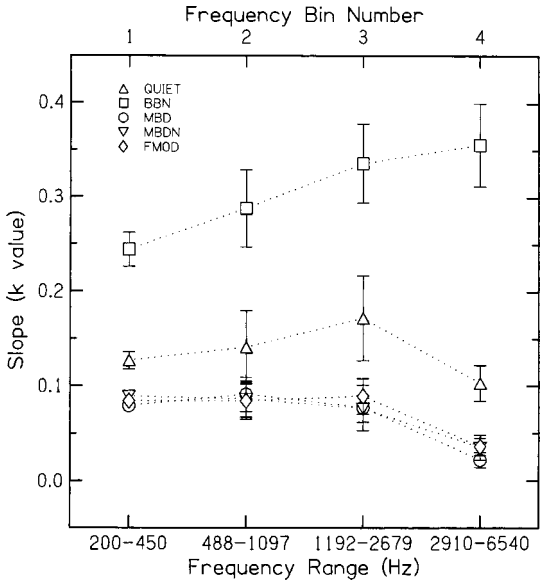


FIG. 5. The intersubject average slopes (*k*'s from logistic functions) of the fitted functions from Table I for the quiet condition and the four masked conditions. The error bars are standard errors of the means.

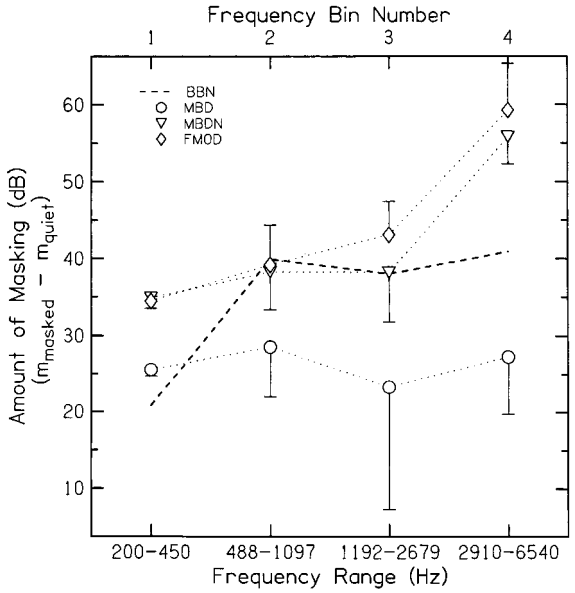


FIG. 6. The difference in decibels between the midpoints of the fitted identification-level functions for the quiet condition and the four masked conditions averaged across listeners. The error bars are standard errors of the means.

which were also significantly different from each other ($p < 0.05$).

At moderately high levels, all three subjects' performance in QUIET exceeded 90% correct. The midpoints of the functions varied across frequency. On average, the highest overall midpoint was 35.7 dB SPL, which occurred for the lowest frequency bin, and the lowest overall midpoint was 14.4 dB SPL, which was found for the 488–1097-Hz bin. This higher SPL required to identify the low-frequency signals is due to the normal increase in the audibility curve at low frequencies. Relative to QUIET, the BBN masker (squares) shifted the functions to the right along the abscissa. The level of the BBN masker was chosen—somewhat arbitrarily—to produce an amount of masking that was roughly comparable to that produced by the informational maskers. We may consider the shift in midpoints as an estimate of the amount of masking produced, recognizing that differences in slope affect the amount of masking estimated in this way. Thus, on average, the BBN masker shifted performance at the midpoints of the functions by about 20 dB for the lowest frequency bin, and about 40 dB for the other three bins. The smaller effect at the lower frequency bin is again due to the higher quiet threshold resulting from the shape of the audibility curve; the noise, as expected, shifted the masked values to similar levels across frequency.

The performance-level functions for the three informational maskers were also shifted along the abscissa relative to the QUIET condition. The amount of masking, as reflected by the shift in the midpoints of the functions, ranged from about 23–29 dB for all four bins of the MBD masker to 35 to 60 dB for both the MBDN and FMOD maskers in the highest frequency bin. It is noteworthy that larger differences between subjects were apparent here than for the BBN masker and the fits to the functions were sometimes rather poor (e.g., Listener 2 at the highest frequency bin, as indicated in Table I). This finding is typical for informational versus energetic masking in general and this task in particular (cf. Kidd *et al.*, 1998). The amount of masking as reflected by the shift in the midpoints of the fitted functions is plotted in Fig. 6.

Because the amount of informational masking that is produced is of primary interest in the study, and because the slopes of the informational maskers were so similar and so different than BBN or QUIET, we next conducted an analysis of variance (ANOVA) of the midpoints of the functions for the three informational maskers. This ANOVA indicated that masker type was a significant main effect ($F[2,4] = 8.51$; $p = 0.036$), whereas frequency was not. There was no significant interaction between masker and frequency. *Post hoc* analyses (Duncan multiple range test) indicated that the MBDN and FMOD maskers produced significantly greater amounts of masking than did the MBD masker ($p < 0.05$). However, the slightly greater amount of masking observed for FMOD was not significantly different than that produced by MBDN.

It is of interest to determine whether changes in the structure of the maskers would cause greater interference in the identification of some signals than others. It is possible, for example, that there is an interaction between masker type and the probability of correct identification of specific signals. Although our previous work using this signal set (Kidd *et al.*, 1995a, 1998) suggested that the signals are roughly equally identifiable, we did not use all of the maskers employed in the current study. Figure 7 is a histogram showing the percent correct obtained for each of the six signals in the five conditions tested here.

To obtain the values plotted in Fig. 7, the data were combined across frequency bins and levels, then averaged across subjects. In general, the patterns of correct identification of the signal set are remarkably similar, especially for the three informational maskers. For those maskers, signal #3 (rising) was the signal most likely to be correctly identi-

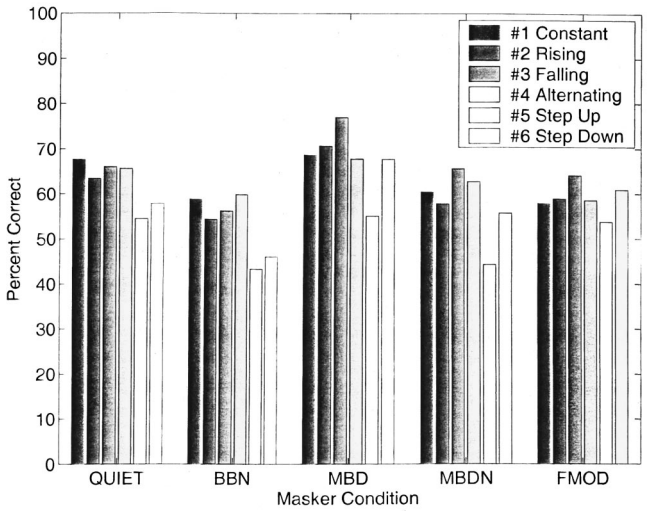


FIG. 7. Percent-correct identification for each of the six signals in all five listening conditions. The signal patterns are: #1 (constant), #2 (rising), #3 (falling), #4 (alternating), #5 (step up), and #6 (step down). The data were averaged across, subjects, frequencies, and levels (see the text).

fied and signal #5 (step up) was the least likely. The other four signals were roughly equally identifiable and fell within a range of about 5%. For QUIET and BBN, signals #5 and #6 (step down) were the least identifiable, with the other four signals being nearly the same. An analysis of variance was performed on these data after the percent-correct values were converted using an arcsine transform. The results indicated that neither masker type ($F[4,8]=2.28$; $p=0.15$) nor signal type ($F[5,10]=1.6$; $p=0.25$) was statistically significant. Thus, there do not appear to be any strong interactions between the type of masker and the relative ability to identify the members of this signal set.

IV. DISCUSSION

The principal finding is that the amount of masking produced by the MBDN and FMOD maskers was much greater than that produced by the MBD masker. Further, the slopes of the performance-level functions were nearly identical for these three maskers and were significantly shallower than were found for quiet or for the primarily energetic masker, BBN. The main difference between the MBD masker and the MBDN and FMOD maskers was that the masker elements for MBDN and FMOD were constrained to fall in narrow frequency regions throughout the burst sequence. In contrast, the frequencies in the MBD masker were randomized over the entire frequency range on each burst. This meant that the MBDN and FMOD maskers could form multiple, co-occurring perceptual ''streams.'' The difference in the randomization of level (for MBD the masker tones were chosen randomly for every component on every burst, while for MBDN and FMOD the level of the first tone in a band was chosen randomly and the successive tones were held constant at the level) probably contributed to the perceptual coherence of the streams in MBDN and FMOD as well. Because the signal set also is comprised of narrow bands of tones, the MBDN and FMOD maskers may be viewed as more similar to the set of signals than the MBD masker. Thus, the results provide some support for the motivating hypothesis that increasing similarity between signal and masker—or more specifically, portions of the maskers—will increase the amount of informational masking. However, these data do not conclusively prove that similarity, *per se*, is the important property of the maskers in producing the greater amounts of informational masking, and alternative explanations are considered below.

It is important to note that, up to this point, our use of the term ''similarity'' is entirely subjective and the hypothetically relevant stimulus dimension separating the maskers—the extent to which they resemble the signal patterns—does not have units that we can use to compute distances. However, because pattern identity is conveyed by the variation (or, for signal #1, the lack of variation) in frequency of the pattern elements, a reasonable measure of similarity[4] would be to determine how close the masker frequencies would come to matching the signals. This is based on the assumption that errors might be made because the listener could ''hear out'' a narrow-band portion of a masker and mistake it for one of the signals because it sounded sufficiently similar to that signal. In order to estimate this, we made the following computations: first, for a given masker type, we generated a masker sample using the same rules for stimulus generation as were used in the experiment (without a signal or ''protected region''). Then, each of the signal patterns was ''swept'' in frequency over the entire range of masker frequencies (200–6540 Hz) in 1-Hz steps, with the signal patterns scaled proportional to center frequency just as in the experiment. At each step, the difference in frequency was calculated between each tone in the signal sequence and the closest (in frequency) corresponding masker tone. Then, the root-mean-square (rms) difference was computed for the set of eight frequency differences from all eight bursts. For example, suppose the masker was MBD. One sample of that masker would be generated containing eight sets of eight tones—one set for each burst. We would begin with signal pattern 1 at 200 Hz. The difference in frequency between the frequency of the signal in burst 1 (200 Hz) and the nearest masker frequency would be computed. Then, the same would be done for burst 2,3,...,8 and the rms computed for all eight differences. That value would be stored as a percent error relative to the frequency (or center frequency) of the signal tones. Now, signal #1 would be incremented to 201 Hz and the procedure repeated. This would be done for the entire frequency range and all six signals. For any given masker sample, the closest match (smallest rms deviation) between any of the signals and the masker was found. Thus, for one masker sample, one rms deviation was found that represented the closest match between that masker and any of the six signals. Also, because we had the smallest rms deviation for each of the six signals given that masker, the average deviation across the signal set was computed. Distributions of rms deviations were accumulated across 2000 masker samples for each of the three multitone maskers. The distributions obtained in this manner varied considerably in form for the different maskers. For MBD, the mean value of the best match was about 10.6% of the center frequency and the distribution was skewed well above the mean. For MBDN,

the distribution was approximately Gaussian, with a mean closest match having an rms deviation of about 2%. For the FMOD masker, the distribution took only two values with the mean deviation falling at about 1.9%. When the average for the set of six signals was computed, the rms deviations increased only slightly to 11.5%, 2.8%, and 3.0%, respectively. Thus, as we would expect, both MBDN and FMOD maskers more closely approximated actual signal patterns than did MBD, and had about the same rms deviation values. To that extent, then, MBDN and FMOD maskers were quantitatively about the same in similarity to the signals and were both more similar to the signals than MBD.

There are relevant examples in both the visual and auditory psychophysical literature suggesting that similarity affects the ability to perceptually segregate or extract some property of visual or auditory images. And, there is ample evidence that perceptually segregating a signal from a masker can improve performance in a variety of tasks (e.g., Neff, 1995; Kidd *et al.*, 1994, 1998). Although there does not appear to be an exact parallel in the vision literature to the current experiment, there are areas of vision perception that seem related. For example, visual ''pop-out'' (e.g., Treisman and Gelade, 1980; see also the recent review by Nakayama and Joseph, 1998) occurs when a particular visual object has some distinctive feature that causes it to segregate perceptually from background images. This phenomenon is usually measured by determining the time it takes to search for, locate, and identify the target or, if search time is limited, the accuracy that is achieved during the visual search (e.g., Palmer, 1996). Generally, as the similarity of the target and masker set is increased, the longer the search times obtained or the poorer the accuracy in time-limited search. Other examples taken from the auditory literature are those of the identification of concurrently spoken vowels and of speech recognition in the case of multiple simultaneous talkers. For concurrent vowels, differences in the fundamental frequencies of the talkers or frequency modulating one voice while holding the other constant can significantly improve discriminability (e.g., Culling and Summerfield, 1995; de Cheveigne, 1997). Although, as mentioned in the Introduction, it is difficult to determine informational masking for two or more simultaneous talkers, there have been two recent studies of speech-masking speech that have specifically examined the issue. Freyman *et al.* (1999) measured the recognition of a speech signal in the presence of a speech masker from the same two loudspeakers in a free field, both leading in time to the front. When the masking speech sound lead to the right and the signal lead to the front, speech recognition improved relative to the former case. The interpretation of this improvement was that the identical, but right-leading masker pulled the masker image away from the speech signal image, perceptually segregating it from the masker and decreasing informational masking. When the signal and masker talkers were the same sex—thus presumably more similar in certaining occurred than when the two talkers were different sexes. Brungart's (2001) work suggests an even stronger effect of similarity in informational masking of speech by speech. His experiments used the coordinate response measure speech-recognition test (Bolia *et al.*, 2000) in which the target speech signal is identified by a call sign (e.g., in the phrase ''Ready Baron go to green three now'' the word ''Baron'' is the call sign, indicating that is the target talker) in a phrase and performance is measured for recognition of subsequent test words in the target utterance. As in the current study, he found evidence for two types of masking: energetic masking when the masker was noise and informational masking when the masker was another talker. Evidence for informational masking was found in the form of nonmonotonicities of the performance-level functions. The nonmonotonicities, in which performance actually improved as signal-to-noise ratio decreased, were apparently due to perceptual segregation of the signal from the masker by differences in level, reducing the informational component of masking. The error patterns were also nonrandom such that the responses tended to be from the masker sentence more frequently than could be accounted for by chance. The informational component also increased as the similarity between the talkers increased. The least amount of masking due to a competing talker occurred when the two talkers were different sexes, more masking was produced when the talkers were the same sex, and the greatest amount of masking occurred when the target and interfering sounds were produced by the same talker.

It is interesting to note that, in the present study, causing the elements to vary in frequency within each band in an orderly way (FMOD) did not significantly increase the amount of masking over random variation in frequency within the narrow masker bands (MBDN).[5] Although the similarity metric described above did not reveal any differences between the two maskers, we initially felt that FMOD might produce more masking because orderly variation in frequency within a narrow band might sound more plausible as a ''pattern'' possibly drawing attention away from the signal or leading to confusions. It is possible that the lack of a significant difference between the MBDN and FMOD maskers was because the MBDN masker caused more informational masking than anticipated. This could occur if the randomly varying frequencies within a narrow band sometimes approximated an actual pattern. Although it would be extremely rare for an MBDN masker stream to actually form one of the signal patterns, it could occasionally form part of a pattern or approximate an actual pattern sufficiently to cause confusions. If we assume that there is some internal noise in the coding of the frequencies of the narrow-band tone sequences (e.g., Kidd *et al.*, 1995b), then inexact matches to ''templates'' of the patterns stored in memory could lead to misidentifications.

While similarity between the maskers, or portions of the maskers, and the signal set may in fact be the determining factor in the magnitude of the informational masking found here, there are other plausible explanations that do not depend on similarity. It is possible that the three informational masking conditions put different demands on the ''processing capacity'' of the listener. The extremely shallow slopes of the performance-level functions found for the informational maskers—much shallower than those for the Gaussian noise,

for example—imply that considerable attentional resources were being devoted toward solving the task (e.g., Norman and Bobrow, 1975). If we assume that the listener can hear individual parts of the MBDN and FMOD maskers as separate perceptual streams, then it is reasonable to assume that these streams would demand more processing than if no streams were formed and the masker frequencies varied randomly, as with MBD, rarely lining up in frequency over time. Thus, the better performance in the MBD masker could be due to the listener having fewer streams to evaluate than in the MBDN and FMOD maskers.

A somewhat related argument that does not depend on similarity or on the listener perceiving multiple co-occurring streams in the maskers is based on the idea that the listener might improve performance by focusing an ''attentional filter'' on the correct frequency region during the burst sequence. Consider the set of eight masker tones and one signal tone comprising the initial burst of the burst sequence for any of the informational maskers. Because each of the nine tones could be the first element of any of the six patterns, the listener initially is faced with 54 equally plausible hypotheses. For the MBD masker, many of these hypotheses may rapidly be discarded because the frequencies in subsequent bursts are not sufficiently close to that in the first burst to plausibly be one of the signals. This might allow the shifting or focusing of attention on the correct frequency region. There is considerable evidence that attentional filtering in the frequency domain can improve performance in tasks where the signal frequency is uncertain. Several studies using the ''probe-signal'' method (Greenberg and Larkin, 1968) have demonstrated significant improvements in performance in attended frequency regions relative to unattended frequency regions (e.g., MacMillan and Schwartz, 1975; Scharf *et al.*, 1987; Hafter *et al.*, 1993) and similar results have been found in the temporal (Dai and Wright, 1995) and spatial (Arbogast and Kidd, 2000) dimensions.

It should be pointed out that there is always the possibility of some energetic masking influencing the results of informational masking experiments. Even with the relatively wide protected region surrounding the signal, some excitation from the more proximal maskers (recall that the masker tones were chosen randomly from a uniform distribution ranging from 50–70 dB SPL) would occasionally spread into the frequency region surrounding the signal. However, in this case, because of the rules used to draw the masker frequencies and levels and the constant width of the protected region, the expected amount of energetic masking produced by the MBD, MBDN, and FMOD maskers is about the same (that is, the nearest masker frequencies and the expected distributions of levels are the same for the three maskers). Also, the signals were presented at a level well above masked detection threshold, which lies near the chance point (i.e., about 17% correct; cf. Kidd *et al.*, 1995a) on the performance-level functions. The fact that the three informational masker slopes are significantly shallower than QUIET means that the amount of masking (level difference at each percent-correct point between masked and QUIET psychometric functions) varies markedly with S/M. At low S/M there is less masking, while at high S/M there is more masking. Thus, as the signals became more audible, the amount of interference caused by the maskers, relative to quiet, increased. In some cases here the chance points were difficult to determine from the logistic fits because the slopes were so shallow and, if extrapolated to lower signal levels than those tested, would unreasonably predict performance better than that obtained in quiet. However, even without a more precise estimate of threshold it is clear that the differences in performance between the three informational maskers cannot be due to differences in the small amounts of energetic masking that may have been present.

V. SUMMARY

(1) All masked conditions produced significant masking as evidenced by shifts in the identification-level functions to higher levels relative to those obtained in quiet.
(2) The MBDN and FMOD informational maskers, which were comprised of sets of narrow-band tone sequences, produced significantly more masking than did the MBD informational masker in which masker frequencies varied randomly throughout the burst sequence.
(3) The slopes of the identification-level functions were much shallower for the three informational maskers than for the energetic masker or for no masker. Further, the slopes tended to decrease with increasing frequency for the informational maskers and increase with increasing frequency for the energetic masker.
(4) A simple ''template-matching'' measure of similarity was sufficient to capture the qualitative trends found in the data: MBD was much less similar to the signal set than MBDN and FMOD, which were about the same.

These results provide qualified support for the proposition that the amount of informational masking observed in a nonspeech pattern identification task can be made greater by increasing the similarity between the masker and the signal set. However, that conclusion must be tempered by the fact that similarity here was manipulated arbitrarily and the assumption that similarity is the relevant dimension causing the informational masking is just that—an assumption. Further, alternative explanations, such as that invoking limitations in the capacity to process multiple co-occurring perceptual streams, can just as plausibly account for the results. Future study to examine these issues is necessary before it is possible to decide among these alternative explanations. It seems clear, though, that the important characteristic of the more effective informational maskers in interfering with the identification of narrow-band nonspeech patterns was that the sequential elements of the maskers were confined to narrow frequency ranges, just as was the case for the set of signal patterns.

ACKNOWLEDGMENTS

Work supported by NIH/NIDCD Grants Nos. DC04545 and DC00100, the ONR-managed MURI Grant No. Z883402, the Dudley Allen Sargent and Sargent Accelerated Research Funds, and the Boston University Hearing Research Center. The authors wish to express their appreciation

to Ervin Hafter and the Department of Psychology, University of California at Berkeley, for support during the preparation of this manuscript while the first and second authors were visitors on leave. We also are grateful to Donna L. Neff, Virginia M. Richards, and an anonymous reviewer for comments on an earlier version of the manuscript.

[1]The term ''central masking'' has also been used in the clinical audiology literature to describe the elevation in threshold that occurs in signal detection in the test ear by the presence of a contralateral masking noise used to prevent cross-ear hearing (e.g., Goldstein and Newman, 1985).

[2]The logistic fits were computed using the FMINS routine of the MATLAB software package (Mathworks Inc., Natick, MA, 2000).

[3]The proportion of variance (PV) accounted for by the fits was computed using the equation: $PV = 1 - [\Sigma_{i=1}^{n}(X_i - \hat{X}_i)^2 / \Sigma_{i=1}^{n}(X_i - \bar{X})^2]$, where X_i is the observed proportion correct value corresponding to level i, $\hat{X}_i$ indicates the predicted value of X_i, $\bar{X}$ is the mean of the Xi's, and n is the number of levels tested.

[4]Other measures of similarity could be employed. For example, subjects could listen to pairs of maskers and rate them according to their perceived degree of similarity. However, the computational template-matching approach used here was based on the assumption that the interference in pattern identification caused by the informational maskers would be related to how closely they approximated the actual signal patterns leading to misidentifications.

[5]MBDN and FMOD maskers were clearly different; i.e., an experienced listener could indicate whether masker samples were MBDN or were FMOD at close to 100% accuracy. This was true regardless of whether the entire masker was played or only single narrow-band sequences from the maskers were played. The frequency variation in FMOD had a characteristic sound that was obviously different than the narrow-band but random MBDN masker.

Arbogast, T. L., and Kidd, Jr., G. (**2000**). ''Evidence for spatial tuning in informational masking using the probe-signal method,'' J. Acoust. Soc. Am. **108**, 1803–1810.

Bollia, R., Nelson, W., Ericson, M., and Simpson, B. (**2000**). ''A speech corpus for multitalker communications research,'' J. Acoust. Soc. Am. **107**, 1065–1066.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT Press, Cambridge, MA).

Brungardt, D. S. (**2001**). ''Informational and energetic masking effects in the perception of two simultaneous talkers,'' J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., and Simpson, B. D. (**2001**). ''Contralateral masking effects in dichotic listening with two competing talkers in the target ear,'' J. Acoust. Soc. Am. **109**, 2486.

Cherry, E. C. (**1953**). ''Some experiments on the recognition of speech with one and two ears,'' J. Acoust. Soc. Am. **25**, 975–979.

Culling, J., and Summerfield, Q. (**1995**). ''The role of frequency modulation in the perceptual segregation of concurrent vowels,'' J. Acoust. Soc. Am. **98**, 837–846.

Dai, H., and Wright, B. A. (**1995**). ''Detecting signals of unexpected or uncertain durations,'' J. Acoust. Soc. Am. **98**, 798–806.

Darwin, C. J., and Carlyon, R. P. (**1995**). ''Auditory grouping,'' in *Hearing: From Handbook of Auditory Perception*, 2nd ed., edited by B. C. J. Moore (Academic, London), pp. 387–424.

de Cheveigne, A. (**1997**). ''Concurrent vowel identification. I. Effects of relative amplitude and F_0 difference,'' J. Acoust. Soc. Am. **101**, 2839–2847.

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (**1999**). ''The role of perceived spatial separation in the unmasking off speech,'' J. Acoust. Soc. Am. **106**, 3578–3588.

Goldstein, B. A., and Newman, C. W. (**1985**). ''Clinical masking: A decision-making process,'' in *Handbook of Clinical Audiology*, 3rd ed., edited by J. Katz (Williams and Wilkens, Baltimore), pp. 170–201.

Green, D. M., and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Krieger, New York) (originally published by Wiley, New York, 1966).

Greenberg, G. Z., and Larkin, W. D. (**1968**). ''Frequency-response characteristic of auditory observers detecting signals of a single frequency in noise: The probe-signal method,'' J. Acoust. Soc. Am. **44**, 1513–1523.

Hafter, E. R., Schlauch, R. S., and Tang, J. (**1993**). ''Attending to auditory filters that were not stimulated directly,'' J. Acoust. Soc. Am. **94**, 743–747.

Hawley, M., Litovsky, R., and Colburn, H. S. (**1999**). ''Speech intelligibility and localization in a multisource environment,'' J. Acoust. Soc. Am. **105**, 3436–3438.

Howard, J. H., O'Toole, A. J., Parasuraman, R., and Bennett, K. B. (**1984**). ''Pattern-directed attention in uncertain-frequency detection,'' Percept. Psychophys. **35**, 256–264.

Kidd, Jr., G., Mason, C. R., and Green, D. M. (**1986**). ''Auditory profile analysis of irregular sound spectra,'' J. Acoust. Soc. Am. **79**, 1045–1053.

Kidd, Jr., G., 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.

Kidd, Jr., G., Mason, C. R., and Rohtla, T. L. (**1995a**). ''Binaural advantage for sound pattern identification,'' J. Acoust. Soc. Am. **98**, 1977–1986.

Kidd, Jr., G., Mason, C. R., and Dai, H. (**1995b**). ''Discriminating coherence in spectro-temporal patterns,'' J. Acoust. Soc. Am. **97**, 3782–3790.

Kidd, Jr., G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (**1998**). ''Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns,'' J. Acoust. Soc. Am. **104**, 422–431.

Leek, M. R., and Watson, C. S. (**1984**). ''Learning to detect auditory pattern components,'' J. Acoust. Soc. Am. **76**, 1037–1044.

Leek, M. R., Brown, M. E., and Dorman, M. F. (**1991**). ''Informational masking and auditory attention,'' Percept. Psychophys. **50**, 205–214.

Macmillan, N. A., and Schwartz, M. (**1975**). ''A probe-signal investigation of uncertain frequency detection,'' J. Acoust. Soc. Am. **58**, 1051–1058.

Nakayama, K., and Joseph, J. S. (**1998**), ''Attention Pattern Recognition and Pop-out in Visual Search'' in *The Attentive Brain*, edited by R. Parasuraman (Bradford-MIT Press, Cambridge, MA), pp. 279–298.

Neff, D. L. (**1995**). ''Signal properties that reduce masking by simultaneous random-frequency maskers,'' J. Acoust. Soc. Am. **98**, 1909–1920.

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.

Norman, D. A., and Bobrow, D. G. (**1975**). ''On data-limited and resource-limited processes,'' Cognit Psychol. **7**, 44–64.

Oh, E., and Lutfi, R. A. (**1998**). ''Nonmonotonicity of informational masking,'' J. Acoust. Soc. Am. **104**, 3489–3499.

Palmer, J. (**1996**). ''Attentional effects in visual search accuracy and search time,'' in *Visual Attention*, edited by R. Wright (Oxford University Press, New York), pp. 348–388.

Pollack, I. (**1975**). ''Auditory informational masking,'' J. Acoust. Soc. Am. Suppl. 1 **57**, S5.

Pollack, I., and Pickett, J. M. (**1958**). ''Stereophonic listening and speech intelligibility against voice babble,'' J. Acoust. Soc. Am. **30**, 131–133.

Scharf, B., Quigley, S., Aoki, C., Peachey, N., and Reeves, A. (**1987**). ''Focused auditory attention and frequency selectivity,'' Percept. Psychophys. **42**, 215–223.

Treisman, A., and Gelade, G. (**1980**). ''A feature-integration theory of attention,'' Cognit Psychol. **12**, 97–136.

Watson, C. S. (**1987**). ''Uncertainty, informational masking and the capacity of immediate auditory memory,'' in *Auditory Processing of Complex Sounds*, edited by W. A. Yost and C. S. Watson (Erlbaum, Hillsdale, NJ), pp. 267–277.

Watson, C. S., Kelly, W. J., and Wroton, H. W. (**1976**). ''Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of uncertainty,'' J. Acoust. Soc. Am. **60**, 1176–1186.

Watson, C. S., Wroton, H. W., Kelly, W. J., and Benbasset, C. A. (**1975**). ''Factors in the discrimination of tonal patterns. I. Component frequency, temporal position and silent intervals,'' J. Acoust. Soc. Am. **57**, 1175–1185.

Weber, D. L. (**1977**). ''Detection and recognition of auditory patterns,'' Percept. Psychophys. **46**, 1–8.

Yost, W. A. (**1991**). ''Auditory image perception and analysis,'' Hear. Res. **56**, 8–18.

Yost, W. A., Dye, Jr., R. H., and Sheft, S. (**1996**). ''A simulated 'Cocktail Party' with up to three sound sources,'' Percept. Psychophys. **58**, 1026–1036.

Yost, W. A. (**1997**). ''The cocktail party problem: Forty years later,'' in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. A. Gilkey and T. R. Anderson (Erlbaum, Hillsdale, NJ), pp. 329–348.

Learning to perceive pitch differences[a)]

Laurent Demany[b)] and Catherine Semal
Laboratoire de Neurophysiologie, Université Victor Segalen and CNRS (UMR 5543), 146 rue Léo-Saignat, F-33076 Bordeaux, France

(Received 23 January 2001; revised 11 July 2001; accepted 28 November 2001)

This paper reports two experiments concerning the stimulus specificity of pitch discrimination learning. In experiment 1, listeners were initially trained, during ten sessions (about 11 000 trials), to discriminate a monaural pure tone of 3000 Hz from ipsilateral pure tones with slightly different frequencies. The resulting perceptual learning (improvement in discrimination thresholds) appeared to be frequency-specific since, in subsequent sessions, new learning was observed when the 3000-Hz standard tone was replaced by a standard tone of 1200 Hz, or 6500 Hz. By contrast, a subsequent presentation of the initial tones to the contralateral ear showed that the initial learning was not, or was only weakly, ear-specific. In experiment 2, training in pitch discrimination was initially provided using complex tones that consisted of harmonics 3–7 of a missing fundamental (near 100 Hz for some listeners, 500 Hz for others). Subsequently, the standard complex was replaced by a standard pure tone with a frequency which could be either equal to the standard complex's missing fundamental or remote from it. In the former case, the two standard stimuli were matched in pitch. However, this perceptual relationship did not appear to favor the transfer of learning. Therefore, the results indicated that pitch discrimination learning is, at least to some extent, timbre-specific, and cannot be viewed as a reduction of an internal noise which would affect directly the output of a neural device extracting pitch from both pure tones and complex tones including low-rank harmonics. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1445791]

PACS numbers: 43.66.Fe, 43.66.Hg [MRL]

I. INTRODUCTION

When repeated measurements of some perceptual discrimination threshold are made in an initially naive subject, it is generally found that the subject's performance gets better and better before stabilizing: the measured thresholds decrease, at first rapidly and then more and more slowly. Of course, if the subject had never participated previously in any psychophysical experiment, it is not so surprising to find that his or her performance is not immediately optimal: the response demands of the task must be learned and some attentional adaptation to the experimental situation is needed; in other words, an initial period of ''procedural learning'' (Robinson and Summerfield, 1996) is necessary. In many cases, however, performance is still suboptimal after thousands of trials, even though the response demands of the task are very simple (see, e.g., Leek and Watson, 1984; Wright *et al.*, 1997; Cansino and Williamson, 1997). Thus, in addition to a presumably rapid process of procedural learning, a more extended process of genuinely perceptual learning (or ''stimulus learning,'' in the terminology proposed by Robinson and Summerfield, 1996) certainly takes place. Indeed, it has often been reported—especially with regard to vision—that subjects trained for a long time to discriminate a given standard stimulus from neighboring stimuli did not transfer their learning to the discrimination of other stimuli, even though only the stimuli were changed, not the procedural aspects of the task. This reveals that a large part of the learning process was specific to the standard stimulus employed during the training period.

How can one interpret the fact that perceptual learning concerning stimulus A does not generalize to stimulus B? For two stimuli A and B which activate separate groups of neurons in some sensory map, one possible hypothesis is that training focused on A selectively modifies the response characteristics of those neurons that respond to A, and/or increases the number of neurons that respond to A but not B. Learning would thus be due to local modifications in a sensory map. This hypothesis has been supported by Recanzone *et al.* (1993). They trained monkeys in a pitch discrimination task, using pure-tone stimuli, and examined the neural correlates of the animals' perceptual learning at the primary auditory cortex level. They found that training focused on a narrow frequency region resulted in an expansion of the cortical area representing that frequency region. Other examples of learning-induced local changes in auditory maps are given by Weinberger (1995) and Edeline (1999).

However, some specific discrimination learning effects seem to call for a different interpretation. Instead of local changes in sensory maps, they seem to reflect specificities of perceptual *mechanisms*. In the domain of pitch, effects of this kind were reported by Demany (1985) and Grimault *et al.* (in press). We shall describe these two studies in some detail since the research reported in the present paper is closely related to them.

The aim of Demany (1985) was to determine to what extent, for human listeners, frequency discrimination learning is frequency-specific. In his experiment, frequency discrimination thresholds for a standard pure tone of 200 Hz

[a)]This work has been presented at the 140th meeting of the Acoustical Society of America, Newport Beach, 2000.
[b)]Electronic mail: Laurent.Demany@psyac.u-bordeaux2.fr

were measured in a ''pretest'' and a ''post-test,'' using an adaptive forced-choice procedure. The pretest and the post-test, consisting each of 50 trials, were separated by a training phase in which 10 blocks of 70 trials were run. During the training phase, frequency discrimination thresholds were again measured, with an adaptive procedure, but the frequency of the standard pure tone varied across subjects. In four groups of subjects, this frequency was respectively 200, 360, 2500, and 6000 Hz. It appeared that the effect of training on the threshold measured in the post-test (standard frequency: 200 Hz) was very similar for the groups trained at 200, 360, and 2500 Hz: For these three groups, the improvement observed from pretest to post-test was roughly the same. However, the improvement was markedly smaller for the group trained at 6000 Hz.

Therefore, this experiment suggested that, in humans, frequency discrimination learning is not *strongly* frequency-specific. How can one explain that, nonetheless, the benefit of training abruptly decreased when the standard frequency used in the training phase varied from 2500 to 6000 Hz? Several other psychophysical studies (not concerned with perceptual learning) have indicated that the pitch of pure tones changes in quality, rather abruptly, in the vicinity of 5000 Hz (see especially Attneave and Olson, 1971 and Semal and Demany, 1990): Above 5000 Hz, pitch becomes ''amusical'' and hiss-like. Perhaps not coincidentally, there are convincing reasons to believe that, in the periphery of the human auditory system, tone frequencies that lie respectively below and above 5000 Hz are coded differently: up to 5000 Hz, a temporal coding of frequency is likely to be available (cf., e.g., Rose *et al.*, 1967) and effective (Moore, 1973; Moore and Glasberg, 1989); beyond 5000 Hz, by contrast, it seems that frequency can only be coded by means of tonotopic cues. Demany (1985) pointed out that his results might reflect this duality of frequency coding mechanisms. He hypothesized that the groups trained at 360 and 2500 Hz completely transferred their perceptual learning to 200 Hz because these three frequencies are coded in the same manner—by means of temporal cues—in the auditory periphery. For the group trained at 6000 Hz, the explanation of the weaker transfer would be that subjects learned to use purely tonotopic cues, i.e., cues which are not the most efficient ones at 200 Hz.

This was an appealing idea, but an alternative hypothesis must also be considered: It could be that the perceptual learning that took place at 6000 Hz was not transferred to 200 Hz simply because the interval formed by these two frequencies exceeded some limit, corresponding for instance to a critical distance within a tonotopically organized neural map. Actually, the *complete* transfer of learning that occurred from 2500 Hz to 200 Hz is somewhat puzzling since these two frequencies were already separated by more than 3.5 octaves. It may be relevant to note, in this regard, that Demany's subjects were not trained very extensively: as mentioned above, only 700 trials were performed during the training phase of the experiment. In the domain of vision, there is some evidence that the selectivity of perceptual learning on a given condition increases with the amount of learning (Karni and Sagi, 1993; Ahissar and Hochstein, 1997). A similar phenomenon perhaps takes place in the case of pitch. Indeed, in a recent study on the frequency specificity of frequency discrimination learning, using a training program which was markedly more extensive than Demany's, Irvine *et al.* (2000) found that perceptual learning achieved at 5000 Hz did not completely transfer to 8000 Hz, and vice versa; these two frequencies are only 0.7 octave apart. Analogous results were reported by Wright (1998) for two frequencies that were 2 octaves apart (1000 and 4000 Hz).

One goal of the first experiment to be reported here was to clarify Demany's (1985) results and to try to confirm that the frequency specificity of frequency discrimination learning can reveal the existence of two frequency coding mechanisms, respectively operative below and above 5000 Hz. To this end, we used a methodology which differed in several respects from that of Demany. Importantly, the subjects' training was much longer in the new experiment. Our new experiment was also intended to answer the following question: when training in a given frequency discrimination task is given monaurally, is the corresponding perceptual learning specific to the trained ear or does it generalize to the other ear? To our knowledge, this question had never been asked before.

In the second experiment to be reported here, subjects learned to discriminate from each other periodic sounds which were no longer pure tones but complex tones, with a spectrum consisting of consecutive harmonics of a missing fundamental. It is well-known that a complex tone typically evokes a single and very precise pitch sensation which is equivalent to that evoked by a pure tone at the fundamental frequency ($F0$), even if there is no energy at this frequency in the spectrum of the complex. The corresponding pitch, that we shall call ''low pitch'' following Plomp (1976), is subjectively more salient than the pitch of any particular harmonic as long as the spectrum consists of more than two consecutive (and roughly isointense) harmonics (Moore and Glasberg, 1990; Laguitton *et al.*, 1998). This phenomenon has fascinated many psychoacousticians since the middle of the 19th century, but there is still no consensual explanation of it (see Houtsma, 1995, for a recent review). An important question is: Does the mechanism of low-pitch extraction depend on the resolvability of the harmonics in the auditory periphery? In other words, is the low pitch of a complex consisting of resolved pure tones extracted in the same manner as the low pitch of a set of unresolved pure tones? Grimault *et al.* (in press) tackled this question with a discrimination learning paradigm. They reasoned that if there were two distinct mechanisms of low-pitch extraction—one for resolved spectral components and the other for unresolved components—then subjects who have learned to make low-pitch discriminations for complexes of a given type (resolved or unresolved) might transfer this learning to other complexes of the same type, but not to complexes of the other type. By contrast, a unitary model of low-pitch extraction predicted that such selectivity of learning should not be observed. Grimault *et al.* did observe selective learning phenomena which were consistent with the ''dual'' model. Their study thus provides a strong argument against any unitary model of low-pitch extraction.

The most elaborate unitary model of low-pitch extraction is the model that was developed by Meddis and his co-workers (Meddis and Hewitt, 1991a, 1991b; Meddis and O'Mard, 1997). Its basic idea (inspired by Licklider, 1951) is that the central auditory system extracts low pitch by first computing the autocorrelation function of the neural spike train elicited at the output of each peripheral auditory filter; and then summing these autocorrelation functions across filters. The delay for which the resulting sum is maximal would provide an estimation of the stimulus periodicity and would thus correspond to the low pitch perceived. This model is elegant and powerful. It can account for numerous psychophysical phenomena concerning pitch in general (especially in the frequency range for which spike trains in the auditory nerve convey precise temporal information, that is below about 5000 Hz). However, the model seems unable to account for the results of Grimault *et al.* that we just mentioned. In addition, Carlyon and Shackleton (1994) and Carlyon (1998) have raised other objections against the idea that low pitch would be extracted in the same manner for complex tones made up of resolved and unresolved harmonics. Moreover, a study by Kaernbach and Demany (1998) suggests that, *for unresolved harmonics*, the mechanism of low-pitch extraction is not akin to an autocorrelation algorithm (because the auditory system appears to be sensitive only to first-order temporal regularities in an amplitude envelope). One way to solve these problems is to suppose that the auditory system does use an autocorrelation mechanism to extract low pitch, but only for sets of harmonics that are resolved in the auditory periphery. Note that it would then be natural to assume that, below 5000 Hz, the pitch of a single pure tone is extracted exactly like the low pitch of a resolved complex, i.e., by an autocorrelation mechanism. According to this scheme, therefore, although pitch discrimination learning focused on resolved complexes should not be largely transferred to unresolved complexes (in agreement with the results of Grimault *et al.*, in press), the same learning could be largely transferred to pure tones.

In the second experiment reported here, we wished to determine whether such a transfer does indeed occur. Each subject was initially trained to detect small differences in period between resolvable complex tones, using a fixed standard complex, and the transfer of this learning to pure tones was assessed as a function of the relation between the standard complex and the standard pure tone. In one condition, these two standard stimuli had identical periods and were thus matched in pitch; a large transfer of learning could then be expected, although the two standard stimuli had no common spectral component and, of course, had quite different timbres. In another condition, by contrast, the standard pure tone had a pitch which differed from the low pitch of the standard complex, but the frequency of this pure tone was equal to that of one spectral component of the complex (so that some transfer could also be expected here, under certain assumptions). In a third condition, finally, the two standard stimuli were unrelated to each other. The results obtained in these three conditions will be discussed in light of two theories of low-pitch extraction: the autocorrelation model mentioned above and the theory provided by Terhardt (1974, 1979).

Although our two experiments were intended to answer distinct questions, they were closely related from the methodological point of view. In particular, the subjects' tasks were *formally* the same (only the stimuli differed), and the initial training sessions were organized similarly. This allowed us to compare the time courses of discrimination learning for two kinds of stimuli: pure tones and resolved complex tones. Grimault *et al.* (in press) suggested that learning to perceive differences in low pitch (up to an asymptotical level of performance) takes less time for unresolved complexes than for resolved complexes. In our case, it was interesting to determine whether learning would be more rapid for pure tones than for resolved complexes.

II. EXPERIMENT 1

A. Overview

In this experiment, subjects learned to discriminate a monaural pure tone of 3000 Hz from ipsilateral pure tones with slightly different frequencies. We then assessed the transfer of this perceptual learning to the frequency discrimination of: (1) ipsilateral pure tones near 1200 Hz; (2) ipsilateral pure tones near 6500 Hz; (3) contralateral pure tones near 3000 Hz. Note that 1200 Hz and 6500 Hz are approximately equidistant from 3000 Hz on a logarithmic frequency scale, as well as on the ''ERB'' scale derived by Glasberg and Moore (1990) from measurements of the auditory filters' bandwidths.[1] On this basis, it could be expected that the amount of learning transfer would be similar in the first two conditions. On the other hand, while two frequencies such as 1200 and 3000 Hz lie within the domain of ''musical'' pitch and are likely to be coded in the same manner (i.e., by temporal cues) in the periphery of the human auditory system, it is reasonable to believe that a 6500-Hz frequency, lying outside the domain of musical pitch, is coded differently (only by means of spatial, tonotopic cues). From this point of view, and given Demany's (1985) study described in the previous section, it could be expected that the amount of learning transfer would be larger in the first condition than in the second one. The third condition was used in order to test the hypothesis that frequency discrimination learning is at least partly reflected by changes in neural responses to sound at a peripheral level of the auditory system (i.e., at a level where there is still no binaural convergence). If this hypothesis were true, we ought to find that a monaural discrimination learning does not completely transfer to the contralateral ear. If the hypothesis were wrong, a complete transfer might be observed.

B. Method

1. Measurement of thresholds

On each trial, the subject was presented with three successive pure tones, separated by 250-ms pauses. Each tone had a total duration of 250 ms and was gated on and off with 10-ms cosinusoidal amplitude ramps. The first tone was defined as the standard. One of the two subsequent tones, selected at random, was a repetition of the standard. The re-

maining tone, defined as the target, differed from the standard by a positive frequency shift. The subject's task was to identify the position of the target by pressing one of two buttons (respectively, labeled ''2'' and ''3''), on a response box. Visual feedback was provided immediately: following each correct response, an LED located in front of the appropriate button was switched on for 300 ms; no LED was switched on if the response was wrong. Any response initiated the next trial after a delay of 500 ms.

Frequency discrimination thresholds were measured with an adaptive procedure (Kaernbach, 1991), in two types of blocks of trials that we respectively called the ''*3000-Hz* blocks'' and the ''*mixed* blocks.'' In each *3000-Hz* block, the standard tone was at 3000 Hz and 110 trials were run. The frequency shift of the target was initially set to 50 cents (2.93%). During the first 10 trials, the shift (in cents) was multiplied by 2.3 after any wrong response, and divided by $(2.3)^{1/3}$ after any correct response. During the following 100 trials, the shift was multiplied by 1.5 after any wrong response and divided by $(1.5)^{1/3}$ after any correct response. The median of the shifts used in these 100 final trials served as an estimation of the shift for which the probability of a correct response was 0.75, and was taken as the subject's discrimination threshold.[2] Each *mixed* block consisted of three interleaved blocks of 110 trials for which the standard tone was, respectively, at 1200, 3000, and 6500 Hz; the standard varied regularly—in a saw-tooth manner—from trial to trial, and a threshold was measured for each standard separately. In each of the three interleaved blocks, the frequency shift of the target was manipulated according to the same rules as in the *3000-Hz* blocks.

Subjects were tested in a double-walled soundproof booth (Gisol, Bordeaux). The tones were monaurally presented by means of a Sennheiser HD265 earphone (the same earphone for the two ears), in a continuous background of ipsilateral pink noise. The pink noise was bandpass filtered between 500 and 11000 Hz (Stanford Research, SR640 and SR645). Its SPL was 46 dB. Each tone also had a nominal SPL of 46 dB, which corresponded to a sensation level of about 20 dB in the noise background. Given that the sensation level of each tone was determined by the masking effect of the noise, possible local irregularities in the frequency response of the earphone, or the subject's ear, were unlikely to affect this sensation level and thus to provide loudness cues in the discrimination task: the irregularities in question had no effect on the signal-to-noise ratio in a given frequency region. The tones and the noise were generated in real time, at a sampling rate of 25 000 Hz, via separate 16-bit digital-to-analog converters (Oros AU22).

2. Experimental sessions

For each subject, the experiment was carried out in 16 test sessions, preceded by one brief preliminary session. Each test session lasted about 1 h. Consecutive sessions were separated by at least one night (cf. Karni *et al.*, 1994) and at most five days.

The preliminary session began with a measurement of the subject's absolute detection threshold, at each ear, for tones of 1200, 3000, and 6500 Hz. These measurements were made using a Békésy tracking procedure. Any potential subject for whom one of the six measured thresholds exceeded 15 dB HL was dismissed at this point. For those who met the audiometric criterion, explanations were then given about the procedure used to measure frequency discrimination thresholds. Finally, in order to familiarize the subject with the response box and to check that the task was understood, 20 dummy trials were run with visual stimuli (strings of letters briefly presented on a monitor screen) instead of sounds.

In the first ten test sessions (sessions 1–10), frequency discrimination thresholds were measured for a fixed ear: the left ear for half of the subjects and the right ear for the other half. Session 1 began with one *mixed* block and then included ten *3000-Hz* blocks. In sessions 2–10, only *3000-Hz* blocks were run: ten blocks each time. At the end of session 10, therefore, 11 110 trials had been run with a standard tone fixed in frequency (3000 Hz) and laterality; only 220 trials had been run with different standard tones (in the *mixed* block of session 1).

In sessions 11–16, by contrast, half of the blocks were *mixed* blocks, during which the stimuli were presented to the same ear as in sessions 1–10. The other half consisted of *3000-Hz* blocks in which all stimuli were presented to the contralateral ear. Each of these six sessions included three *mixed* blocks and three *3000-Hz* blocks, presented alternately. The first block of each session was a *mixed* block for half of the subjects, and a *3000-Hz* block for the other half.

3. Subjects

The data reported below were obtained from eight students, paid for their services. None of them had previously taken part in a psychophysical experiment. Two additional students were dismissed at the end of session 1 because, for both of them, the mean threshold measured in the *3000-Hz* blocks of this session was so low (less than 5 cents) that we did not expect to find a significant improvement in the subsequent sessions.

C. Results and discussion

1. Learning during sessions 1–10

Figure 1 (panel a) shows how the thresholds measured in the *3000-Hz* blocks of sessions 1–10 varied from session to session. For a given session and subject, the computed statistic was the *geometric* mean of the ten measured thresholds. The eight thin curves display the individual results and the thick curve displays their geometric means. From session 1 to session 10, thresholds improved by a mean factor of 2.4. However, it can be seen that this improvement took place almost entirely in the first four or five sessions: the geometric means obtained in sessions 5–10 differed by at most 12% from each other. The mean obtained in session 10 was 5.8 cents; this is quite close to 6.2 cents, the threshold predicted by Wier *et al.*'s (1977) formulas describing frequency discrimination performance as a function of frequency and sensation level for ''expert'' listeners. The general equation proposed by Nelson *et al.* (1983) predicted a higher threshold:

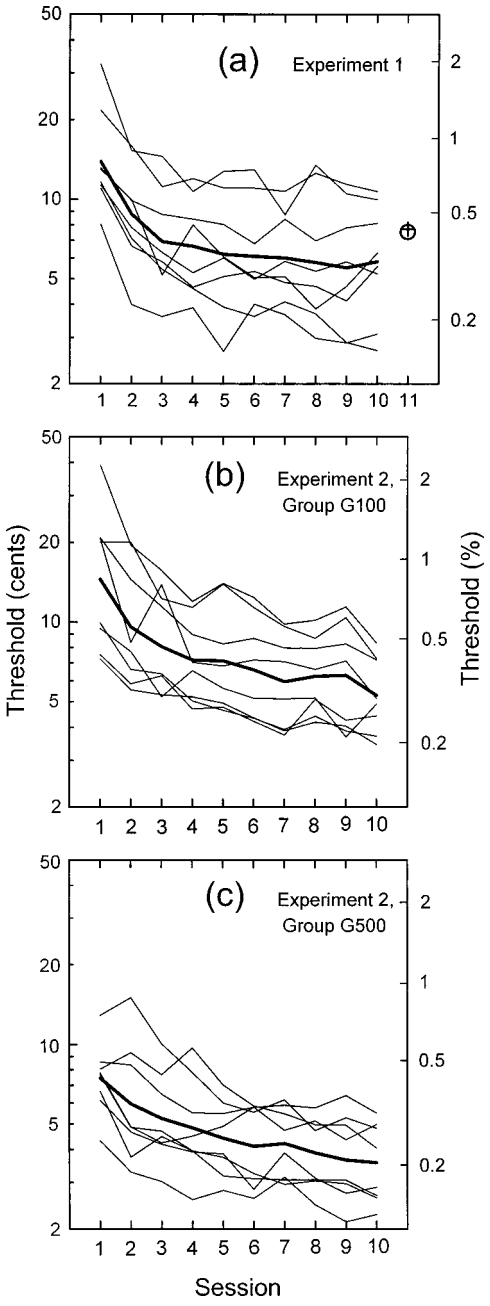


FIG. 1. Frequency discrimination thresholds measured in sessions 1–10 of experiment 1 (panel a) and experiment 2 (panel b for group G100, panel c for group G500). Thresholds are expressed in cents on the left-hand ordinate axis and as relative frequency differences—in %—on the right-hand ordinate axis. For the present data, these two units are approximately equivalent (cf. footnote 2). The left-hand scale is logarithmic and therefore the right-hand scale is almost so. Thin curves display the individual results and thick curves display their geometric means. In panel (a), the cross and the circle plotted for session 11 represent the mean thresholds obtained for 3000-Hz standard tones presented, respectively, in the contralateral *3000-Hz* blocks and the ipsilateral *mixed* blocks.

10.8 cents. So, it seems that sessions 1–10 were sufficient to provide maximum perceptual learning for the discrimination of monaural pure tones near 3000 Hz.

2. Frequency specificity

Our experimental procedure allowed us to assess in two ways the transference of the learning analyzed above to ipsilateral pure tones near 1200 and 6500 Hz. First, it was interesting to compare the thresholds measured in the *mixed* block run at the start of session 1 to those measured in the next *mixed* block, i.e., the first *mixed* block of session 11. The corresponding data are plotted in Fig. 2. A repeated-measures analysis of variance (ANOVA) performed on the logarithms of thresholds revealed highly significant effects of the "session" factor [$F(1,7)=56.1$, $P<0.001$] and the "standard frequency" factor [$F(2,14)=33.5$, $P<0.001$], but no significant interaction between these two factors [$F(2,14)<1$]. Thus, the improvement obtained at 3000 Hz was not significantly larger than those obtained at 1200 and 6500 Hz, as if the learning that occurred during the 100 *3000-Hz* blocks of sessions 1–10 had no frequency specificity. However, given that the *mixed* block of session 1 was run at the very start of this session and that subjects had never taken part previously in any psychophysical experiment, it is reasonable to think that most of the improvement observed in the next *mixed* block reflects procedural rather than perceptual learning (cf. our Introduction).

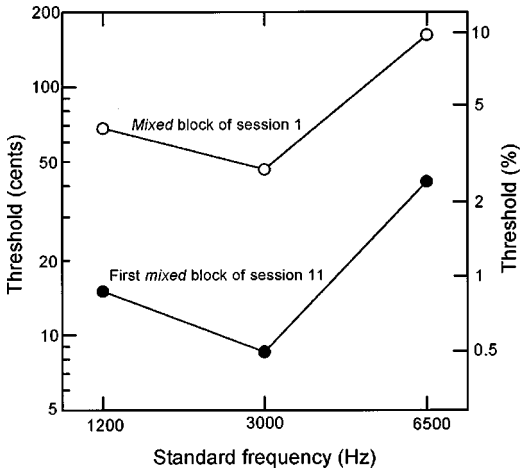


FIG. 2. Thresholds obtained in experiment 1 for the *mixed* block of session 1 and the first *mixed* block of session 11. As in Fig. 1, thresholds are expressed both in cents (on a log scale, left-hand ordinate) and as relative frequency differences, in % (right-hand ordinate).

Indeed, the learning that took place in sessions 1–10 appears to be strongly frequency-specific when its transfer is assessed only from the data obtained in sessions 11–16. The three solid curves in panel (a) of Fig. 3 show how the thresholds measured in the *mixed* blocks of these six sessions varied from session to session. Since we were mainly interested in the effect of standard frequency on thresholds' time course, the (geometric) means of the thresholds obtained at a given frequency in each session were divided by the mean obtained in session 11 for the same frequency, and Fig. 3(a) displays the resulting ratios. It is clear that the mean thresholds decreased markedly more at 1200 Hz (downward triangles) and 6500 Hz (upward triangles) than at 3000 Hz (circles).[3] Therefore, the transfer of the training received in the ten previous sessions appeared to be larger when the standard frequency was the same (3000 Hz) than when it was different. In order to evaluate the frequency effect more precisely, we computed for each subject the slope of the regression line summarizing the time course of the logarithms of thresholds, at each frequency. These slopes were then sub-

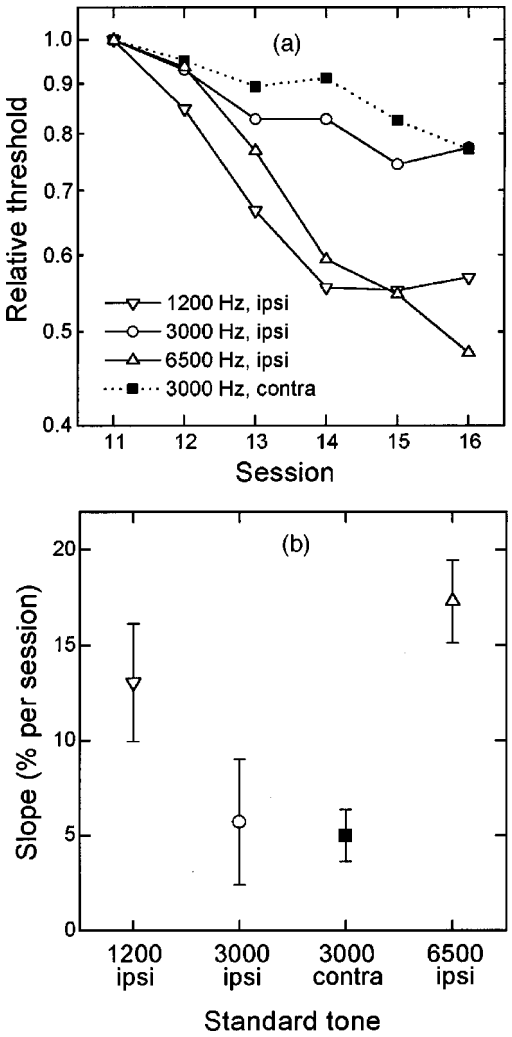


FIG. 3. (a) Time course of the thresholds measured in sessions 11–16 of experiment 1; for each session and each of the four conditions, the geometric mean of the thresholds measured in a given session was divided by the geometric mean of the thresholds measured in session 11; the ordinate scale is logarithmic; in session 16, the geometric means obtained in conditions 1200-ipsi, 3000-ipsi, 6500-ipsi, and 3000-contra were, respectively: 5.8, 5.8, 19.9, and 5.9 cents. (b) Means of the slopes of the regression functions summarizing thresholds' time course (on a log scale) for each standard tone and each subject; vertical bars represent standard errors.

mitted to a repeated-measures ANOVA. The overall effect of frequency was significant [$F(2,14)=5.69$, $P=0.016$]. A planned comparison revealed that the slopes were significantly shallower at 3000 Hz than at the other two frequencies [$t(14)=3.16$, $P=0.003$, one-tailed test]. The complementary planned comparison indicated that, contrary to our hypothesis, the slopes were not significantly steeper at 6500 Hz than at 1200 Hz [$t(14)=1.18$, $P=0.13$, one-tailed test]. The mean slopes (expressed as percentages of threshold decrease across consecutive sessions) are displayed in panel (b) of Fig. 3 (open symbols, conditions "1200-ipsi," "3000-ipsi," and "6500-ipsi"). It is worthy to note that whereas the mean slopes measured at 1200 and 6500 Hz were significantly different from 0 [$t(7)\geqslant 4.03$, $P\leqslant 0.002$, one-tailed tests], this was not the case for the mean slope measured at 3000 Hz [$t(7)=1.72$, $P=0.065$].

One must conclude from this analysis of the slopes that its outcome conflicts with a suggestion offered by Demany (1985) with regard to the frequency specificity of frequency discrimination learning. He suggested that: (1) the across-frequency transfer of learning is complete, even over intervals as large as 3.6 octaves, when the tested frequencies remain in the "musical" range; (2) the transfer is markedly weaker across the upper boundary of this frequency range, corresponding to approximately 5000 Hz. Point (1) is contradicted by our finding of an incomplete—and indeed weak—transfer from 3000 to 1200 Hz: both of these frequencies clearly fall within the musical range (see, e.g., Semal and Demany, 1990) and they are separated by only 1.3 octave. On the other hand, our data are not inconsistent with the hypothesis that 5000 Hz represents a kind of barrier for the transfer of perceptual learning, although we failed to confirm that something special occurs in the vicinity of this frequency.

In order to understand why we contradicted Demany's point (1), it is relevant to recall that Demany compared thresholds measured in an initial pretest to thresholds measured in a later post-test. Given that the initial pretest was quite short (50 trials), a significant part of the improvement that he observed was probably due to a process of procedural learning, without any stimulus specificity, rather than to perceptual learning *per se* (as pointed out by Irvine *et al.*, 2000). We performed a similar comparison in the present experiment and, as mentioned above, its outcome provided no evidence for a frequency specificity of frequency discrimination learning. However, this cannot be the whole story since Demany did obtain evidence for a specificity of that kind. So, in order to account for the apparent discrepancy between the two studies, one must probably take into account the fact that Demany's subjects were not trained very extensively following the pretest, even though this training certainly resulted in genuine perceptual learning. More precisely, our suggestion is that, in a frequency discrimination task using a constant standard stimulus, the selectivity of *perceptual* learning (as opposed to procedural learning) increases with the amount of practice and learning. Analogous ideas have been expressed with regard to perceptual learning in vision (Karni and Sagi, 1993; Ahissar and Hochstein, 1997).

3. Ear specificity

In panel (a) of Fig. 3, the dotted curve with filled squares shows the time course over sessions 11–16 of the mean thresholds measured during the *3000-Hz* blocks, which were run contralaterally to those of sessions 1–10. The mean of the slopes of the individual regression functions is plotted in panel (b) (closed square, condition "3000-contra"), as well as its standard error. It is legitimate to compare this mean slope to the other mean slopes displayed in panel (b) since the number of trials (and threshold measurements) per session was the same in each case. If the learning that took place in sessions 1–10 had no ear specificity at all, we should have found that the mean slope obtained in condition 3000-contra was (1) not significantly larger than that obtained in condition 3000-ipsi, and (2) not significantly larger than 0. The first prediction was verified—indeed, it was in condition 3000-contra that the mean slope was smallest—but the second prediction was disproved [$t(7)=3.62$,

$P = 0.004$, one-tailed test]. This pattern of results is somewhat paradoxical since, as mentioned in the previous section, the mean slope obtained in condition 3000-ipsi was not significantly larger than 0. The explanation is of course that the variance of the slopes was smaller in condition 3000-contra than in condition 3000-ipsi.

These results are therefore ambiguous. Another way to determine if learning was ear-specific is to compare the mean threshold obtained in the *3000-Hz* blocks of session 11 to the means obtained in the previous sessions. In Fig. 1 (panel a), the mean obtained in session 11 is plotted as a cross. Notice that this mean was higher than those obtained in each of the eight previous sessions. Such a rise can be taken as evidence for an ear specificity of learning. On the other hand, Fig. 1 also shows that the thresholds measured at 3000 Hz during session 11 were not significantly larger in the *3000-Hz* blocks (condition 3000-contra) than in the *mixed* blocks [condition 3000-ipsi; the corresponding mean threshold is plotted as a circle (surrounding the cross)]. The reason why thresholds were somewhat elevated in the *mixed* blocks is probably that in these blocks the standard tone changed from trial to trial instead of being fixed. Nonetheless, the evidence for an ear specificity of learning is again ambiguous.

Overall, we can conclude that the ear specificity of learning was at most weak. Fig. 3 clearly indicates that it was weaker than the *frequency* specificity of learning.

In the domain of vision, it has been found that some perceptual learning phenomena are at least partly monocular: they transfer incompletely, or even weakly, from one eye to the other (Karni and Sagi, 1991; Fahle *et al.*, 1995). This is quite striking since, for normal observers, a given image presented only to the left eye is hard to discriminate, in any respect, from the same image presented only to the right eye. In the domain of hearing, by contrast, a stimulus presented to the left ear is inevitably easy to discriminate from the same stimulus presented to the right ear, on the basis of a difference in subjective lateralization. The strong monocularity of some visual learning phenomena has been taken as evidence that these phenomena originate from local, experience-dependent modifications of the neuronal connections between cells in the *primary* visual vortex (Karni and Sagi, 1991). However, other visual learning phenomena, though stimulus-specific, transfer completely from one eye to the other (e.g., Fiorentini and Berardi, 1991). Similarly, tactile learning can strongly transfer across hands (Sathian and Zangaladze, 1997; Spengler *et al.*, 1997; Harris *et al.*, 2001). The fact that, in the present experiment, the ear specificity of frequency discrimination learning appeared to be weak, at most, implies that this learning cannot be mainly due to neuronal modifications at a peripheral level of the auditory system—before binaural convergence, i.e., in the cochlear nuclei or even more peripherally. We cannot completely rule out that such peripheral modifications exist, assuming that the ear specificity exists. However, note that even if a strong ear specificity had been found, an additional study would be needed to test the hypothesis of a peripheral (monaural) origin of learning: its ear specificity might instead rest upon lateralization phenomena involving the binaural system.

III. EXPERIMENT II

A. Overview

As indicated in the Introduction, experiment 2 was concerned with the transfer of pitch discrimination learning from a complex tone to pure tones. Each subject was initially trained to discriminate a fixed standard complex, consisting of harmonics 3–7 of a given $F0$ (f), from complexes with the same spectral structure (i.e., harmonics 3–7) but slightly higher $F0$s. Given that the maximum harmonic rank was equal to 7 and that our subjects had normal hearing, the complexes were unambiguously of the ''resolved'' type (see Shackleton and Carlyon, 1994). Following this training phase—which was similar to that used in experiment 1, except for the stimuli—subjects were required to detect frequency differences between pure tones and we assessed the transfer of the initial learning in three conditions. In condition ''FUNDAM,'' the frequency of the standard pure tone (g) was equal to f; thus, this standard pure tone was matched in pitch to the standard complex that was employed during the initial training phase. In condition ''SPECT,'' g was equal to $5 \cdot f$, so that the standard pure tone was identical in frequency to the median spectral component of the standard complex. In condition ''NOVEL,'' finally, g was remote—namely, at least 1.8 octave away—from both the $F0$ of the standard complex (f) and the frequencies of its spectral components. We found in experiment 1 that discrimination learning at a given frequency was poorly transferred to frequencies less than 1.8 octave away. On this basis, it could be predicted that the transfer of learning would be poor, or at most incomplete, in the NOVEL condition of the present experiment. Our main question was, therefore: will the transfer be larger in the other two conditions, especially in the FUNDAM condition?

B. Method

1. Stimuli

The standard complex used in the initial training phase had an $F0$ of 100 Hz for one-half of the subjects (Group G100), and 500 Hz for the other half (group G500). Thus, the median harmonic of this standard complex had a frequency of 500 Hz for group G100, and 2500 Hz for group G500. For both groups, the three standard pure tones used subsequently had the following frequencies: (1) 100 Hz (FUNDAM condition for G100, NOVEL condition for G500); (2) 500 Hz (SPECT condition for G100, FUNDAM condition for G500); (3) 2500 Hz (NOVEL condition for G100, SPECT condition for G500).

The harmonics of each complex had equal amplitudes. Their relative phases were set as suggested by Pressnitzer and Patterson (2001) in order to minimize the amplitude of a potential combination tone at $F0$. Namely, the waveform of a complex being

$$s(t) = \sum_{i=3}^{i=7} [\cos(2\pi \cdot i \cdot f \cdot t + \varphi_i)], \tag{1}$$

we had

$$\varphi_i = \pi(i-1) \cdot i/4. \tag{2}$$

All stimuli were presented monaurally, to the subject's preferred ear, with the equipment already used in experiment 1. Their temporal parameters were the same as those adopted in experiment 1, and they were again presented in a continuous background of ipsilateral pink noise. The pink noise was low-pass filtered at 4900 Hz and had an SPL of 55 dB (which corresponded to 45 dB A). Each standard complex was set to an SPL giving a nominal sensation level of 20 dB in the noise background, as determined by preliminary measurements made on two listeners with normal audiograms. On the basis of similar measurements, each standard pure tone was set to an SPL giving a nominal sensation level of 25 dB. Employing a noise background was advantageous for the reason mentioned in Sec. II B 1, but also because, in the present case, this noise served to mask a potential combination tone at f or $2 \cdot f$.

2. Procedure

The procedure was basically similar to that of experiment 1, but simpler. Again, 16 test sessions were run following a brief preliminary session. This time, the preliminary session took place on the same day as session 1 and did not include dummy trials using visual stimuli. It was devoted to measurements of the subject's absolute detection threshold for tones of 100, 500, 2500, and 4000 Hz, presented to the relevant ear. As before, potential subjects for whom any threshold exceeded 15 dB HL were dismissed.

In the test sessions, discrimination thresholds were measured with exactly the same method as before. Sessions 1–10 were organized identically. Each of them consisted of ten blocks of 110 trials in which the standard stimulus was a fixed complex (with an $F0$ of either 100 or 500 Hz, as stated above). Sessions 11–16 were also organized identically. Each of them consisted of three blocks of 330 trials, which were designed exactly like the *mixed* blocks of experiment 1 except that, in the present case, the frequencies of the three standard pure tones were, respectively, 100, 500, and 2500 Hz. In contrast to the procedure used in experiment 1, sessions 1–10 did not include any block of the *mixed* type, and sessions 11–16 consisted *only* of such blocks.

3. Subjects

Both of the experimental groups (G100 and G500) contained eight subjects, who were recruited in the same student population as the subjects of experiment 1 and who were paid for their services. None of them had previously taken part in a psychophysical experiment. Whereas in experiment 1 two potential subjects had been dismissed at the end of session 1, no such selection took place in the present case.

C. Results

1. Sessions 1–10

The two lower panels of Fig. 1 show how thresholds varied during sessions 1–10, for group G100 (panel b) and group G500 (panel c). From session 1 to session 10, thresholds improved by a mean factor of 2.8 for G100 and 2.1 for G500. These two factors are roughly similar to the factor of 2.4 obtained in experiment 1, for stimuli which were pure tones rather than complex tones. However, an inspection of Fig. 1 suggests that the curves obtained for G100 and G500 were somewhat different in *shape* from the curve obtained in experiment 1. In our previous experiment, the thresholds' improvement took place almost entirely in the first four of five sessions, and the following part of the curve was essentially a plateau. Here, by contrast, the initial improvement was less abrupt and it did not seem that subjects had finished learning after the tenth session. For each subject of the two experiments, we computed the slope of the regression function summarizing the thresholds' time course (on a log scale) during (1) sessions 1–3 and (2) sessions 8–10; the difference between these two slopes was then submitted to an ANOVA. The overall effect of the standard stimulus (experiment 1 vs G100 vs G500) was significant [$F(2,21)=4.93$, $P=0.018$]. Planned comparisons revealed a significant difference between the two experiments [$|t(21)|=2.93$ $P=0.008$], but no significant difference between G100 and G500 [$|t(21)|=1.12$, $P=0.28$].[4]

Our results thus suggest that frequency discrimination learning is a slower process when the standard stimulus is a resolved complex tone than when it is a single pure tone. Admittedly, the sample of subjects used in experiment 1 was slightly biased insofar as we rejected two potential subjects whose thresholds were ''too good'' in session 1. However, if their thresholds had been essentially constant throughout sessions 1–10, as we guessed, the rejection of these two subjects would not account for the fact that the regression functions summarizing the thresholds' time course from session 5 to session 10 were significantly flatter in experiment 1 than in experiment 2 [$|t(22)|=2.13$, $P=0.04$]. Another small bias stemmed from the presence, in experiment 1 but not experiment 2, of one *mixed* block at the beginning of session 1. However, this bias tended to *reduce*, rather than to increase, the contrast between the learning curves: The initial *mixed* block provided an extra opportunity for learning; in its absence, therefore, the data points corresponding to session 1 of experiment 1 would have been somewhat higher, which would have increased the initial slope of the mean curve, making it even more different from the mean curves obtained in experiment 2.

2. Sessions 11–16

Analyses of regression functions indicated that, over the last three sessions of the initial training phase (i.e., sessions 8–10), thresholds decreased with a mean slope of 6.5% per session. By contrast, over the next three sessions (sessions 11–13), the mean slope was equal to 15.7% per session. This increase in slope was statistically significant [$t(15)=2.87$, $P=0.003$, one-tailed test]. We can conclude from it that the transfer of perceptual learning from the complex tones to the pure tones was, at most, incomplete.[5]

The dependency of learning transfer on the experimental condition (FUNDAM, SPECT, or NOVEL) was assessed by considering the slopes obtained throughout sessions 11–16 (as in experiment 1). Figure 4 displays the mean values of these slopes, and the geometric means of the thresholds themselves, for each group and each standard frequency. Since a large slope reflected a small transfer, and a small

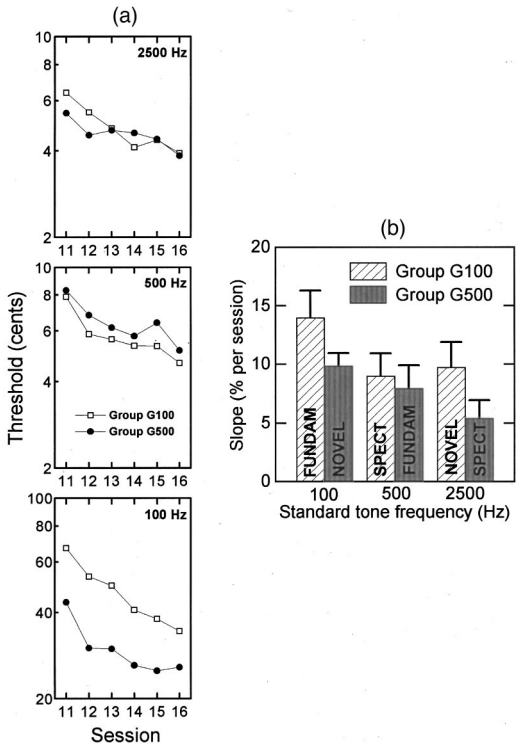


FIG. 4. (a) Thresholds measured in sessions 11–16 of experiment 2; each panel displays the results obtained in each group of subjects for a standard tone of a given frequency; the ordinate scales are logarithmic. (b) Mean slopes of the regression lines summarizing the thresholds' time course (on a log scale); vertical bars represent standard errors.

slope a large transfer, it was reasonable to predict that the slopes would be largest in the NOVEL condition and smallest in the FUNDAM condition. But this prediction was not verified. For the 100-Hz standard pure tone, surprisingly, the mean slope obtained in the FUNDAM condition (i.e., for group G100) was larger than the mean slope obtained in the NOVEL condition (i.e., for group G500). However, the difference in question was not statistically significant [$t(14) = 1.45$, $P = 0.169$]. Figure 4(b) shows that, for each group, it was in the SPECT condition that the mean slope was smallest. At 2500 Hz, however, the slopes were not significantly different in the SPECT condition (group G500) and the NOVEL condition (group G100) [$t(14) = 1.55$, $P = 0.142$]. At 500 Hz, similarly, the slopes were not significantly different in the SPECT (group G100) and FUNDAM (group G500) conditions [$t(14) < 1$]. Note that if instead of comparing the slopes, one compares only the thresholds obtained in session 11, *t*-tests also fail to demonstrate significant differences between the three conditions [$t(14) \leqslant 2.11$, $P \geqslant 0.053$].

D. Discussion

In this experiment, subjects had to detect differences in $F0$ between complexes made up of harmonics with identical ranks. Since the harmonics had identical ranks, the task could conceivably be performed by detecting frequency changes in individual harmonics—i.e., spectral pitch changes—rather than changes in low pitch. However, Moore and Glasberg (1990, 1991; see also Moore *et al.*, 1992) presented convincing evidence that, for resolvable complexes consisting of many harmonics (six or more), low pitch is the cue that subjects use in order to detect changes in $F0$. For the complexes used here, which comprised five harmonics, we felt that this was also true: to our ears, the low pitch was much more salient than the spectral pitch of any individual harmonic. This impression is consistent with experimental results reported by Laguitton *et al.* (1998). In their study, subjects were presented with various pairs of successive complexes and had to judge, for each pair, whether pitch rose or fell from the first to the second complex. The paired complexes were such that a rise in $F0$ was associated with a fall in the frequencies of spectral components, and vice versa; therefore, the perceived direction of pitch change revealed which aspect of pitch—spectral pitch or low pitch—was most salient. The results indicated that spectral pitch is generally more salient than low pitch for complexes consisting of only two harmonics (in agreement with previous results of Smoorenburg, 1970, and Houtsma and Fleuren, 1991), but that low pitch becomes the most salient percept as soon as the number of harmonics exceeds two. It should also be recalled, in this context, that our complexes were presented in a background of noise rather than in quiet. This tends to increase the salience of low pitch relative to that of purely spectral features (Moore and Glasberg, 1991; see also Hall and Peters, 1981).

Let us suppose, therefore, that what subjects learned to perceive more and more accurately, during sessions 1–10, was not the spectral component of the complexes but only the output of a neural ''periodicity detector'' such as an autocorrelator (cf. the Introduction). Let us assume more precisely that the device in question identifies both the low pitch of resolved complexes, like those used here, and the pitch of isolated pure tones (with periods in the same range); for a neural autocorrelator, each of these pitches would correspond to the most prominent peak in a ''summary autocorrelation function'' (SACF) equal to the sum of autocorrelation functions computed from the outputs of all the auditory filters excited by the stimulus (Meddis and Hewitt, 1991a, 1991b; Meddis and O'Mard, 1997). Under this assumption, we should have found a complete transfer of learning in the FUNDAM condition. Moreover, on the basis of the results obtained in experiment 1, a smaller transfer was expected in the NOVEL condition since in the latter condition the period of the standard pure tone was separated by at least 2.3 octaves from the period of the standard complex. Given that these two predictions were disproved, the hypothesis from which they derive must be wrong; apparently, during sessions 1–10, discrimination learning was not due to a reduction in an internal noise directly affecting the output of a pitch extractor which would essentially be a periodicity detector and would work identically for resolved complexes and isolated pure tones.

An alternative hypothesis, consistent with a suggestion of Meddis and Hewitt (1991b) and Meddis and O'Mard (1997), is that subjects learned to process the entire SACF of the stimuli rather than just the peak corresponding to their period. If this were true, the magnitude of learning transfer should have depended on the similarity between the SACF of the standard complex and the SACFs of the standard pure

TABLE I. Correlation (r) between the SACF of the standard complex and the SACF of the standard pure tone, for each group of subjects and experimental condition.

	Condition		
	FUNDAM	NOVEL	SPECT
Group G100	0.05	0.00	0.29
Group G500	0.06	0.00	0.25

tones which were presented subsequently. The correlations between those SACFs are displayed in Table I.[6] This table indicates that the SACFs were essentially noncorrelated in the FUNDAM condition, as well as the NOVEL condition, and only weakly correlated in the SPECT condition. This tallies with the experimental results, especially the fact that the transfer of learning was not larger in the FUNDAM condition than in the NOVEL condition. However, since the hypothesis under consideration here is that the entire SACF is relevant for *both* complex tones and pure tones, it does not immediately account for the finding of slower learning for complex tones (sessions 1–10 of experiment 2) than for pure tones (sessions 1–10 of experiment 1). Actually, this finding is also difficult to explain under the hypothesis that we considered previously.

A third hypothesis, which can account for the finding in question, is that learning was due to modifications in the tonotopically distributed temporal information provided from an auditory filter bank to a neural periodicity detector such as an autocorrelator.[7] In other words, the subjects' training would have reduced an internal noise affecting not the *output* of the periodicity detector but its *input*, in spite of the fact that, when the stimulus was a complex tone, its low pitch was much easier to perceive consciously than spectral pitches. If, in the course of learning, the information provided by each of the excited auditory filters is modified more or less independently of that provided by the other filters, learning might be slower for a resolved complex than for a pure tone simply because, in the former case, a larger number of auditory filters are excited and the temporal information coming from these multiple filters is more variable across filters. Since the modifications that supposedly occur for a given standard stimulus should primarily depend on its power spectrum, the corresponding learning should have some *spectral* specificity. Therefore, the hypothesis that we consider here predicted, like the previous one, that the transfer of learning would be weak in the FUNDAM and NOVEL conditions, but larger in the SPECT condition. The actually observed pattern of results was not radically different from this prediction, but the hypothesis is not satisfactorily verified since the transfer was not significantly larger in the SPECT condition than in the other two conditions. One possible way to explain the absence of a significant advantage of the SPECT condition is to argue that, in this condition, the transfer of learning was tested using a standard pure tone which was only one of the standard complex's five spectral components; the information provided by the four other components was perhaps modified to a larger extent during the learning process.

Up to this point, we have discussed the results in the framework of a single pitch theory. However, what we found is in fact consistent with any theory positing that the low pitch of a resolved complex and the pitch of an isolated pure tone are extracted by basically different mechanisms. Such a point of view was previously supported by Hall and Soderquist (1978) in experiments concerning the adaptation effect produced by a pure or complex tone on the perception of low pitch in a subsequently presented complex. Other experimental arguments for a duality of mechanisms, and against the notion of a common ''periodicity detector,'' were presented by McFadden (1988).

In order to make this dual hypothesis more precise, one may simply assume that temporal information is used for the extraction of low pitch whereas only place information (i.e., a tonotopic code) is used for the extraction of pitch from a pure tone. It does not seem likely, however, that only place matters in the case of a pure tone with a frequency as low as 100 Hz (Moore, 1973; Hartmann, 1997, Chap. 12). Yet, this should be admitted to account for the results obtained in group G100.

The so-called ''pattern recognition theories'' of low-pitch perception (see Houtsma, 1995, for a recent review) posit that the extraction of low pitch from a resolved complex begins with a measurement of the resolved components' frequencies (or pitches), but then rests upon mechanisms which do not come into play for the extraction of pitch from an isolated pure tone. Interestingly, one of these theories, developed by Terhardt (1974, 1979; see also Whitfield, 1970, for related ideas) assumes that the perception of low pitch originates from an associative learning process taking place during infancy (or even *in utero*). In the human acoustic environment, complex tones consisting of a *complete* series of resolvable harmonics abound and are biologically important (e.g., the vowels of speech). According to Terhardt, this leads to the formation, in some long-term auditory memory, of associative links between the pitch of a pure tone and the pitches of its subharmonics. The associations, once established, would then be used to extract a low pitch from stimuli containing resolved pure tones. Thus, a missing fundamental that one ''hears'' in spite of its absence in the stimulus would actually be more a recollection than a genuine sensation. Note that, if Terhardt is right, what our subjects perceived more and more accurately in the missing-fundamental complexes presented during sessions 1–10 was not their low pitches, since these low pitches were determined by associations which were established before the experiment and could not be refined within it. For an improvement in the precision of low-pitch perception, each complex should have possessed a spectral component at its fundamental frequency. From Terhardt's point of view, therefore, the thresholds' decrease during sessions 1–10 must have reflected an improvement in the encoding of *spectral* information. On this basis, one could make sense of an effect of spectral complexity on the speed of learning, as mentioned above. It was again predictable that the transfer of learning would be larger in our SPECT condition than in the other two conditions, and this was not verified. However, the most crucial prediction of Terhardt's theory was that the transfer would not be larger in

the FUNDAM condition than in the NOVEL condition, and this turned out to be true.

IV. CONCLUSIONS AND FINAL COMMENTS

Experiment 1, in which only pure tones were used, showed that human pitch discrimination learning can be strongly pitch-specific within the *musical* pitch domain (i.e., below 4 or 5 kHz). A previous study (Demany, 1985) suggested that this is not the case, but the discrepancy may stem from the fact that subjects were trained much more extensively in the present experiment: we suppose that, when a fixed standard tone is used in the training, the specificity of perceptual learning increases in the course of learning.

Experiment 1 also showed that pitch discrimination learning is, at most, only weakly ear-specific: monaural learning is largely transferred to the contralateral ear. This result implies that learning cannot be *mainly* due to changes in the response characteristics of neurons connected to only one cochlea. It is interesting to note that, nevertheless, partial cochlear lesions appear to be able to induce remapping phenomena in such neurons, at least for the cat (Rajan *et al.*, 1993).

In experiment 2, subjects were initially trained to detect pitch differences between missing-fundamental complex tones with resolved spectral components, and we assessed the transfer of this learning to the detection of pitch differences between pure tones. We obtained no evidence for learning transfer from a given complex to a pure tone with the same pitch; clearly, such a similarity in pitch did not, *per se*, favor the transfer. Thus, our data indicated that pitch discrimination learning is, at least to some extent, ''timbre-specific'' (or ''spectrum-specific''). This timbre specificity may appear counterintuitive in light of other psychophysical facts concerning pitch. For instance, Semal and Demany (1991, 1993) and Semal *et al.* (1996) provided evidence that, in short-term auditory memory, the sensory trace of the pitch of a sound (renewed on each experimental trial) is completely dissociated from the sound's timbre.

Our data also suggested that pitch discrimination learning is slower for a resolved complex than for a pure tone. This result can be seen as congruent with the recent report by Grimault *et al.* (in press) that pitch discrimination learning is slower for a resolved complex than for an unresolved complex. The temporal fine structure of auditory filters' responses to a resolved complex may vary greatly from one filter to another; this is indeed the case for filters such as the auditory-nerve fibers (cf., e.g., Sachs and Young, 1980). However, the variation across filters will be smaller if, instead of a resolved complex, the stimulus is either a pure tone or an unresolved complex. If, for these three types of stimuli, pitch discrimination learning basically reflects an improvement in the processing of the filters' temporal responses, then it makes sense to find that learning is slowest for a resolved complex.

The present work provides new information about auditory perceptual learning. However, contrary to our expectations, it did not clarify the mechanisms of pitch perception. Experiment 1 failed to confirm (but did not disprove, either) the idea that, for pure tones, the upper limit of ''musical'' pitch, near 5 kHz, is a boundary between two types of frequency coding mechanisms. Experiment 2 demonstrated that, if humans possess a pitch extractor which is essentially a periodicity detector and works identically for resolved complexes and isolated pure tones, then pitch discrimination learning is not a reduction of an internal noise affecting directly the *output* of this periodicity detector; otherwise, we should have found an essentially complete transfer of learning from a resolved complex to a pure tone with the same period. However, our results are compatible with the hypothesis that this periodicity detector exists and that the internal noise which is reduced in the learning process affects its input rather than its output. Our results are also compatible with the pitch theory proposed by Terhardt (1974, 1979), although the experiment had the potential to provide clear evidence against its main assumption—namely, the assumption that the extraction of low pitch from a resolved complex is in itself the product of a learning process.

ACKNOWLEDGMENTS

The authors wish to thank Sylvain Clément, Christophe Micheyl, Daniel Pressnitzer, and Beverly Wright for fruitful discussions, as well as Cecilia Maubaret and Odon Noblia for their active participation in the experiments.

[1]An equation put forward by Glasberg and Moore (1990) implies that there are 7.6 ERBs between 1200 and 3000 Hz, and 6.8 ERBs between 3000 and 6500 Hz.

[2]In this experiment, as well as the subsequent one, thresholds were measured in cents. The cent is a *logarithmic* unit since it corresponds to a given frequency *ratio* (equal to $2^{\pm 1/1200}$). However, because most of the thresholds that we measured were small (smaller than 100 cents), the size of these thresholds in cents was very nearly proportional to their size expressed as a relative frequency *difference*: Frequency shifts of, e.g., 1 cent and 100 cents, which differ by a factor of 100 in terms of cents, differ by a factor of about 102.9—which is close to 100—in terms of relative frequency differences; for frequency shifts smaller than 100 cents, the departure from proportionality is even smaller.

[3]Yet, in sessions 11–16, consecutive *mixed* blocks were separated by one *3000-Hz* block (with contralateral stimuli). The potential bias resulting from this aspect of our procedure, when thresholds' time course at each frequency is assessed only from the data obtained in the *mixed* blocks, is an artifactual *amplification* of thresholds' decrease at 3000 Hz (since part of this decrease could actually be due to a selective transfer of learning from the intervening *3000-Hz* blocks).

[4]A similar ANOVA performed on the data from sessions 1–2 and 9–10 (instead of 1–3 and 8–10) led to the same conclusions.

[5]One might argue that the increase in slope was partly due to the use of *mixed* blocks of trials in sessions 11–16: hypothetically, in these sessions, subjects learned to deal with the fact that the standard stimulus changed from trial to trial. However, the fact that the standard stimuli had never been presented in the previous sessions was quite probably a more important factor since, in sessions 11–16 of experiment 1, there was no significant improvement with time of the thresholds measured in condition 3000-ipsi (cf. Sec. II C 2 and Fig. 3). As Fig. 4 will show, markedly larger threshold improvements were found in sessions 11–16 of experiment 2.

[6]The SACFs were obtained from the ''AMS'' simulation of auditory processing, which is available on the Internet (ftp://ftp.essex.ac.uk/pub/omard/dsam/). We found that they were very similar in shape to the mathematical autocorrelation functions of the stimuli themselves.

[7]The filter bank in question should not be identified as the auditory nerve. Its site should be more central since, according to the results of experiment 1, frequency discrimination learning is unlikely to be mainly due to modifications in the behavior of ''monaural'' neurons.

Ahissar, M., and Hochstein, S. (**1997**). ''Task difficulty and the specificity of perceptual learning,'' Nature (London) **387**, 401–406.

Attneave, F., and Olson, R. K. (**1971**). ''Pitch as a medium: A new approach to psychophysical scaling,'' Am. J. Psychol. **84**, 147–166.
Cansino, S., and Williamson, S. J. (**1997**). ''Neuromagnetic fields reveal cortical plasticity when learning an auditory discrimination task,'' Brain Res. **764**, 53–66.
Carlyon, R. P. (**1998**). ''Comments on 'A unitary model of pitch perception' [J. Acoust. Soc. Am. **102**, 1811–1820 (1997)],'' J. Acoust. Soc. Am. **104**, 1118–1121.
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.
Demany, L. (**1985**). ''Perceptual learning in frequency discrimination,'' J. Acoust. Soc. Am. **78**, 1118–1120.
Edeline, J. M. (**1999**). ''Learning-induced physiological plasticity in the thalamo-cortical sensory systems: A critical evaluation of receptive field plasticity, map changes and their potential mechanisms,'' Prog. Neurobiol. **57**, 165–224.
Fahle, M., Edelman, S., and Poggio, T. (**1995**). ''Fast perceptual learning in hyperacuity,'' Vision Res. **35**, 3003–3013.
Fiorentini, A., and Berardi, N. (**1981**). ''Learning in grating waveform discrimination: Specificity for orientation and spatial frequency,'' Vision Res. **21**, 1149–1158.
Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Grimault, N., Micheyl, C., Carlyon, R. P., and Collet, L. (in press). ''Evidence for two pitch encoding mechanisms using a selective auditory training paradigm,'' Percept. Psychophys. .
Hall, J. W., and Peters, R. W. (**1981**). ''Pitch for nonsimultaneous successive harmonics in quiet and noise,'' J. Acoust. Soc. Am. **69**, 509–513.
Hall, J. W., and Soderquist, D. R. (**1978**). ''Adaptation of residue pitch,'' J. Acoust. Soc. Am. **63**, 883–893.
Harris, J. A., Harris, I. M., and Diamond, M. E. (**2001**). ''The topography of tactile learning in humans,'' J. Neurosci. **21**, 1056–1061.
Hartmann, W. M. (**1997**). *Signals, Sound, and Sensation* (AIP, Woodbury, NY).
Houtsma, A. J. M. (**1995**). ''Pitch perception,'' in *Hearing*, edited by B. C. J. Moore (Academic, San Diego), pp. 267–295.
Houtsma, A. J. M., and Fleuren, J. F. M. (**1991**). ''Analytic and synthetic pitch of two-tone complexes,'' J. Acoust. Soc. Am. **90**, 1674–1676.
Irvine, D. R. F., Martin, R., Klimkeit, E., and Smith, R. (**2000**). ''Specificity of perceptual learning in a frequency discrimination task,'' J. Acoust. Soc. Am. **108**, 2964–2968.
Kaernbach, C. (**1991**). ''Simple adaptive testing with the weighted up–down method,'' Percept. Psychophys. **49**, 227–229.
Kaernbach, C., and Demany, L. (**1998**). ''Psychophysical evidence against the autocorrelation theory of auditory temporal processing,'' J. Acoust. Soc. Am. **104**, 2298–2306.
Karni, A., and Sagi, D. (**1991**). ''Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity,'' Proc. Natl. Acad. Sci. U.S.A. **88**, 4966–4972.
Karni, A., and Sagi, D. (**1993**). ''The time course of learning a visual skill,'' Nature (London) **365**, 250–252.
Karni, A., Tanne, D., Rubenstein, B. S., Askenasy, J. J. M., and Sagi, D. (**1994**). ''Dependence on REM sleep of overnight improvement of a perceptual skill,'' Science **265**, 679–682.
Laguitton, V., Demany, L., Semal, C., and Liégeois-Chauvel, C. (**1998**). ''Pitch perception: A difference between right- and left-handed listeners,'' Neuropsychologia **36**, 201–207.
Leek, M. R., and Watson, C. S. (**1984**). ''Learning to detect auditory pattern components,'' J. Acoust. Soc. Am. **76**, 1037–1044.
Licklider, J. C. R. (**1951**). ''A duplex theory of pitch perception,'' Experientia **7**, 128–134.
McFadden, D. (**1988**). ''Failure of a missing-fundamental complex to interact with masked and unmasked pure tones at its fundamental frequency,'' Hear. Res. **32**, 23–40.
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.
Moore, B. C. J. (**1973**). ''Frequency difference limens for short-duration tones,'' J. Acoust. Soc. Am. **54**, 610–619.
Moore, B. C. J., and Glasberg, B. R. (**1989**). ''Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation,'' J. Acoust. Soc. Am. **86**, 1722–1732.
Moore, B. C. J., and Glasberg, B. R. (**1990**). ''Frequency discrimination of complex tones with overlapping and nonoverlapping harmonics,'' J. Acoust. Soc. Am. **87**, 2163–2177.
Moore, B. C. J., and Glasberg, B. R. (**1991**). ''Effects of signal-to-noise ratio on the frequency discrimination of complex tones with overlapping or nonoverlapping harmonics,'' J. Acoust. Soc. Am. **89**, 2858–2865.
Moore, B. C. J., Glasberg, B. R., and Proctor, G. M. (**1992**). ''Accuracy of pitch matching for pure tones and for complex tones with overlapping or nonoverlapping harmonics,'' J. Acoust. Soc. Am. **91**, 3443–3450.
Nelson, D. A., Stanton, M. E., and Freyman, R. L. (**1983**). ''A general equation describing frequency discrimination as a function of frequency and sensation level,'' J. Acoust. Soc. Am. **73**, 2117–2123.
Plomp, R. (**1976**). *Aspects of Tone Sensation* (Academic, London).
Pressnitzer, D., and Patterson, R. D. (**2001**). ''Distortion products and the pitch of harmonic complex tones,'' 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, The Netherlands), pp. 97–104.
Rajan, R., Irvine, D. R. F., Wise, L. Z., and Heil, P. (**1993**). ''Effect of unilateral partial cochlear lesions in adult cats on the representation of lesioned and unlesioned cochleas in primary auditory cortex,'' J. Comp. Neurol. **338**, 17–49.
Recanzone, G. H., Schreiner, C. E., and Merzenich, M. M. (**1993**). ''Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys,'' J. Neurosci. **13**, 87–103.
Robinson, K., and Summerfield, A. Q. (**1996**). ''Adult auditory learning and training,'' Ear Hear. **17**, 51S–65S.
Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). ''Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey,'' J. Neurophysiol. **30**, 769–793.
Sachs, M. B., and Young, E. D. (**1980**). ''Effects of nonlinearities on speech encoding in the auditory nerve,'' J. Acoust. Soc. Am. **68**, 858–875.
Sathian, K., and Zangaladze, A. (**1997**). ''Tactile learning is task specific but transfers between fingers,'' Percept. Psychophys. **59**, 119–128.
Semal, C., and Demany, L. (**1990**). ''The upper limit of 'musical' pitch,'' Music Percept. **8**, 165–176.
Semal, C., and Demany, L. (**1991**). ''Dissociation of pitch from timbre in auditory short-term memory,'' J. Acoust. Soc. Am. **89**, 2404–2410.
Semal, C., and Demany, L. (**1993**). ''Further evidence for an autonomous processing of pitch in auditory short-term memory,'' J. Acoust. Soc. Am. **94**, 1315–1322.
Semal, C., Demany, L., Ueda, K., and Hallé, P. A. (**1996**). ''Speech versus nonspeech in pitch memory,'' J. Acoust. Soc. Am. **100**, 1132–1140.
Shackleton, T. M., and Carlyon, R. P. (**1994**). ''The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination,'' J. Acoust. Soc. Am. **95**, 3529–3540.
Smoorenburg, G. F. (**1970**). ''Pitch perception of two-frequency stimuli,'' J. Acoust. Soc. Am. **48**, 924–942.
Spengler, F., Roberts, T. P. L., Poeppel, D., Byl, N., Wang, X., Rowley, H. A., and Merzenich, M. M. (**1997**). ''Learning transfer and neuronal plasticity in humans trained in tactile discrimination,'' Neurosci. Lett. **232**, 151–154.
Terhardt, E. (**1974**). ''Pitch, consonance, and harmony,'' J. Acoust. Soc. Am. **55**, 1061–1069.
Terhardt, E. (**1979**). ''Calculating virtual pitch,'' Hear. Res. **1**, 155–182.
Weinberger, N. M. (**1995**). ''Dynamic regulation of receptive fields and maps in the sensory cortex,'' Annu. Rev. Neurosci. **18**, 129–158.
Whitfield, I. C. (**1970**). ''Central nervous processing in relation to spatiotemporal discrimination of auditory patterns,'' in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden, The Netherlands), pp. 136–147.
Wier, C. C., Jesteadt, W., and Green, D. M. (**1977**). ''Frequency discrimination as a function of frequency and sensation level,'' J. Acoust. Soc. Am. **61**, 178–184.
Wright, B. A. (**1998**). ''Generalization of auditory-discrimination learning,'' Assoc. Res. Otolaryngol. Abs., Abs. **413**, 104.
Wright, B. A., Buonomano, D. V., Mahncke, H. W., and Merzenich, M. M. (**1997**). ''Learning and generalization of auditory temporal-interval discrimination in humans,'' J. Neurosci. **17**, 3956–3963.

Decision strategies of hearing-impaired listeners in spectral shape discrimination

Jennifer J. Lentz[a] and Marjorie R. Leek
Army Audiology and Speech Center, Walter Reed Army Medical Center, Washington, DC 20307

(Received 20 July 2001; revised 17 October 2001; accepted 19 December 2001)

The ability to discriminate between sounds with different spectral shapes was evaluated for normal-hearing and hearing-impaired listeners. Listeners detected a 920-Hz tone added in phase to a single component of a standard consisting of the sum of five tones spaced equally on a logarithmic frequency scale ranging from 200 to 4200 Hz. An overall level randomization of 10 dB was either present or absent. In one subset of conditions, the no-perturbation conditions, the standard stimulus was the sum of equal-amplitude tones. In the perturbation conditions, the amplitudes of the components within a stimulus were randomly altered on every presentation. For both perturbation and no-perturbation conditions, thresholds for the detection of the 920-Hz tone were measured to compare sensitivity to changes in spectral shape between normal-hearing and hearing-impaired listeners. To assess whether hearing-impaired listeners relied on different regions of the spectrum to discriminate between sounds, spectral weights were estimated from the perturbed standards by correlating the listener's responses with the level differences per component across two intervals of a two-alternative forced-choice task. Results showed that hearing-impaired and normal-hearing listeners had similar sensitivity to changes in spectral shape. On average, across-frequency correlation functions also were similar for both groups of listeners, suggesting that as long as all components are audible and well separated in frequency, hearing-impaired listeners can use information across frequency as well as normal-hearing listeners. Analysis of the individual data revealed, however, that normal-hearing listeners may be better able to adopt optimal weighting schemes. This conclusion is only tentative, as differences in internal noise may need to be considered to interpret the results obtained from weighting studies between normal-hearing and hearing-impaired listeners. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1451066]

PACS numbers: 43.66.Fe, 43.66.Sr, 43.66.Ba [SPB]

I. INTRODUCTION

It is well established that listeners with healthy auditory systems are able to distinguish sounds based on spectral shape (Green, 1988). However, the effect of hearing impairment on the ability to discriminate between sounds with different spectral shapes has received limited study. Whether hearing loss alters the ability to discriminate such sounds is an important issue, given that spectral cues play a significant role in speech perception. Due to the nature of many sensorineural hearing losses, which often result in different degrees of hearing loss at different frequencies, there is a possibility that hearing-impaired listeners cannot use across-frequency level information to the same degree as normal-hearing listeners. For example, it is unknown whether differences in loudness growth across frequency (e.g., Buus *et al.*, 1995) would lead to a diminished ability to compare levels across frequency. This study evaluates whether hearing-impaired listeners show deficits in their ability to detect differences in spectral shape and investigates whether hearing-impaired listeners rely on spectral regions different from normal-hearing listeners when discriminating complex sounds.

Spectral shape discrimination, often termed "profile analysis," is accomplished by making level comparisons across different frequency regions (see Green, 1988, for a review). In a typical profile-analysis task, listeners discriminate between a standard stimulus that is the sum of equal-amplitude tones and a signal stimulus having a different spectral shape. The component tones usually are well separated in frequency to eliminate masking of adjacent spectral components. In early profile analysis experiments, the signal stimulus had a tone added in-phase to a single component of the standard (cf. Spiegel *et al.*, 1981; Green *et al.*, 1983). For such a task, normal-hearing listeners have exquisite sensitivity to changes in the power spectra of sounds, often superior to that measured for intensity discrimination of single tones (Green and Mason, 1985; Dai and Green, 1992). One concern in all profile analysis experiments is that listeners may discriminate between sounds on the basis of the absolute change in level of a single spectral component. In order to reduce the strength of this single-channel cue, experimenters randomly vary the overall level of the stimuli on every presentation, typically over a range of 20 to 40 dB (Green, 1988). Level variation randomizes the absolute level of each component while retaining the overall spectral shape of the stimulus.

The likely reason that spectral shape discrimination has not been frequently tested in hearing-impaired listeners is that the deficits associated with hearing loss lead to many

[a]Present address: Department of Speech and Hearing Sciences, 200 South Jordan Ave., Indiana University, Bloomington, IN 47405-7002. Electronic mail: jjlentz@yahoo.com

potential experimental confounds. Experiments must be performed at high stimulus levels to ensure audibility. Configurations of the hearing loss also may interact with roving levels. For example, when the level of the stimulus is selected to be low, a sloping hearing loss may act as a low-pass filter resulting in a stimulus where high-frequency components are inaudible to a hearing-impaired listener (Kidd and Mason, 1992). At high stimulus intensities, sounds may be uncomfortably loud. Reduced dynamic ranges complicate the application of the generally large level variation distributions used to decrease the prevalence of single-channel cues. However, the role that abnormal loudness growth might play in the ability to discriminate sounds with different spectral shapes is unclear, as hearing loss has little effect on intensity discrimination (Florentine *et al.*, 1993; Buus *et al.*, 1995). Stimuli also must be constructed with care to eliminate effects of reduced frequency selectivity that impoverish the spectral contrast available to a hearing-impaired listener (Glasberg and Moore, 1986; Summers and Leek, 1994).

Methods have been developed to evaluate which portions of the power spectrum are used by listeners when discriminating between sounds with different spectral shapes. These include Conditional On a Signal Stimulus (COSS) methods, developed by Berg (1989, 1990), and correlation methods described by Richards and Zhu (1994) and Lutfi (1995). Such methods can determine whether listeners with hearing loss adopt different decision rules than normal-hearing listeners when comparing levels across a wide frequency range. An experimenter can estimate the relative importance (or relative weight) of each frequency region in the discrimination of sounds with different spectral shapes. In a two-interval forced-choice task, the levels of the individual components are altered randomly and independently every time the stimuli are presented to the listener. The listener decides which interval was the signal, and the listener's response is correlated with the level difference across the two intervals for each component to estimate that component's contribution to the listener's decision.

Applying similar methods to various spectral shape discrimination tasks, Berg and Green (1990) and Dai *et al.* (1996) reported nearly optimal weighting strategies for normal-hearing listeners detecting a level increment in one component of a multi-tone sound. That is, listeners gave the signal component the greatest weight and used the level of additional stimulus components as comparisons, giving weights of $-1/(n-1)$ where n is the total number of components. In contrast, Berg and Green (1992) and Matiasek and Richards (1999) found that, when listeners detected a signal which involved changing the level of all components, they did not always adopt the optimal decision strategy defined by the linear model.

The correlation technique assumes that the decision variable is a linear combination of the random variables generated by altering the component levels, an assumption which may be valid only to a first approximation. Matiasek and Richards (1999) showed that nonlinearities in the transformation from stimulus variables to a decision variable can lead to misestimations of the weighting strategy. Theoretically, for a fixed-frequency signal, listeners should apply the same weights to signal-present and signal-absent stimuli (Dai *et al.*, 1996). However, Matiasek and Richards showed that differences in weighting strategy for signal-present and signal-absent stimuli are prominent for spectrally dense stimuli but are much less prevalent when the stimuli have fewer components widely spaced in frequency.

Correlation techniques have been applied to evaluate the abilities of hearing-impaired listeners to compare levels across frequency in intensity discrimination tasks. Doherty and Lutfi (1999) asked whether hearing-impaired listeners could discriminate a level change of a single component of a harmonic complex, ranging in frequency from 250 to 8000 Hz, as well as normal-hearing listeners. The optimal result would be that a listener ignores all components (weights=0) except the signal component (weight=1). Doherty and Lutfi showed that both normal-hearing and hearing-impaired listeners weighted the signal component most heavily. In instances where the signal frequency was 250 and 1000 Hz, both groups of listeners were equally able to ignore the nonsignal components. However, when the signal component was at 4000 Hz (i.e., near the edge of their listeners' hearing losses), hearing-impaired listeners may have been better able to ignore the nonsignal components. In another study, Doherty and Lutfi (1996) evaluated spectral weights for a task in which listeners detected a level increment added to all components of a multi-tone sound, which also spanned frequencies from 250 to 8000 Hz. In contrast to normal-hearing listeners, who tended to weight all components equally, hearing-impaired listeners usually weighted some of the high-frequency components more heavily. This finding suggested that listeners placed greater importance on components near the edge of their hearing loss.

In the following experiment, a correlation technique was used to investigate decision strategies adopted by normal-hearing and hearing-impaired listeners when discriminating sounds with different spectral shapes. In contrast to the work reported by Doherty and Lutfi (1996, 1999), this experiment evaluated decision strategies underlying spectral shape discrimination, rather than intensity discrimination. Overall level variation was used in a set of conditions to determine whether changes in the perceived loudness of sound generated by level variation led to difficulty in comparing sounds with different spectral shapes for hearing-impaired listeners. Thresholds and weighting functions were compared between normal-hearing and hearing-impaired listeners to determine whether hearing loss leads to a diminished ability to distinguish between sounds with different spectral shapes.

II. METHODS

A. Listener characteristics

Six normal-hearing listeners (NH1–6), ranging in age from 28 to 68 years, and six hearing-impaired listeners (HI1–6), ranging from 65 to 76 years, participated. All normal-hearing listeners had pure-tone-audiometric thresholds within 20 dB of the 1989 ANSI standard between 250 and 4000 Hz. Hearing-impaired listeners were selected to have a pure-tone-average at 2, 3, and 4 kHz greater than 35 dB HL. All hearing-impaired listeners had hearing loss for at

TABLE I. Audiometric thresholds (dB HL *re*: ANSI 1989) of the test ear for normal-hearing (NH1–NH6) and hearing-impaired listeners (HI1–HI6).

			Frequency (Hz)						
Listener	Age	Test ear	250	500	1000	2000	3000	4000	8000
NH1	28	R	5	10	5	15	15	0	25
NH2	52	R	5	5	15	10	5	5	10
NH3	68	L	15	10	10	5	15	20	30
NH4	46	R	5	5	5	0	0	0	15
NH5	31	R	0	0	5	5	5	0	0
NH6	42	R	5	5	5	10	10	10	25
HI1	76	L	40	50	50	55	65	65	85
HI2	67	R	25	20	40	55	65	65	70
HI3	65	R	55	60	65	70		70	60
HI4	70	R	25	30	25	20	30	65	60
HI5	77	L	30	30	50	55	60	55	70
HI6	71	L	10	15	25	35	55	50	50

least one of the frequencies present in the experimental stimuli, and most had losses throughout the range of stimulus frequencies. Hearing losses were moderate and bilateral; the site of lesion was presumed to be cochlear based on air- and bone-conduction thresholds and immittance audiometry. For the hearing-impaired listeners, the ear with better audiometric thresholds was the test ear. For normal-hearing listeners, the right ear always was tested, except for NH3 who had a mild hearing loss in the right ear, and therefore her left ear was used. The audiometric configurations for all listeners are reported in Table I. All listeners had prior experience with psychoacoustic experiments. Listeners NH1 and NH2 are the first and second authors, respectively.

B. Stimuli

The standard was the sum of five tones ranging from 200 to 4200 Hz spaced equidistantly on a logarithmic scale. On each stimulus presentation, the starting phases of the component tones were selected randomly and independently from a uniform distribution ranging from 0 to 2π rad. For all normal-hearing listeners, the mean level per component was 80 dB SPL. Three of the hearing-impaired listeners were tested at the same level as the NH listeners (HI1, HI5, and HI6), and the other three were tested at 85 dB SPL. The signal stimulus had a tone added in-phase to the 920-Hz (central) component of the standard. The signal strength is described as the amplitude of the added tone relative to the mean amplitude of the component to which it was added (i.e., sig re stan, in dB). The stimuli were generated digitally, played by a two-channel 16-bit digital-to-analog converter (TDT DA1) at a sampling rate of 20 kHz, low-pass filtered at 8.5 kHz (TDT FLT3; approx. 105 dB/oct falloff with a maximum attenuation of 65 dB), and fed into one phone of a TDH-49 headset. The stimuli were 300 ms in duration, including 30-ms cosine-squared rise/decay ramps. Two experimental variables were varied during data collection: (a) component amplitudes and (b) overall levels.

1. Component-by-component perturbation

In the no-perturbation conditions, the standard was the sum of equal-amplitude tones. In the perturbation conditions, the component levels were altered from an equal-amplitude standard by adding randomly and independently drawn deviates from a Gaussian distribution with a mean of zero and a standard deviation of 2.5 dB on each stimulus presentation. A new selection of component levels was presented to the listener in each interval, so that the same stimulus was not presented twice.

2. Overall level randomization

Measurements also were obtained both with and without across-interval level randomization. These conditions will be referred to as rove and no-rove, respectively. For the rove conditions, the overall levels of the stimuli were varied on every presentation through the use of an external attenuator. The value of the attenuator was randomly altered based on draws from a uniform distribution with a 10-dB range and a 0.1-dB gradation. This small range of level variation was used to ensure that no component was inaudible to a hearing-impaired listener.

C. Procedure

Thresholds were estimated in a two-alternative forced-choice task, with trial-by-trial signal levels chosen according to an adaptive tracking procedure. Listeners were seated in a sound-attenuating room and heard two sounds, the signal and standard stimuli, separated by 300 ms. Whether the signal or standard was presented first was determined randomly with equal likelihood. Listeners indicated which interval contained the signal stimulus by touching the appropriate area on a touch-screen terminal. Correct answer feedback was provided to the listener following each trial.

Two interleaved two-down, one-up tracking procedures of 70 trials each were used to estimate a threshold of 71% correct detections (Levitt, 1971). One tracked threshold for the no-perturbation stimuli and one for the perturbation stimuli. Whether a trial consisted of no-perturbation or perturbation stimuli was chosen randomly. At the beginning of every track, the mean level of the signal tone was set to about 15 dB above an estimate of the listener's final threshold in the no-perturbation condition. The initial step size of the tracking procedure was 4 dB, and after three reversals the

step size was reduced to 1 dB. The mean of the masker levels at the last even number of reversal points, excluding the first three, was taken as threshold.

Data first were collected for the no-rove conditions (two interleaved tracks-no-perturbation and perturbation) and then collected for rove conditions. Nine threshold estimates were obtained for each of the four experimental conditions (no-perturbation/no-rove, perturbation/no-rove, no-perturbation/rove, and perturbation/rove). The final threshold reflects the average of the last eight threshold estimates. Only one listener (NH5) showed significant practice effects in the no-rove conditions, and therefore she repeated nine threshold estimates in the no-rove conditions after finishing the rove conditions. Those repeated thresholds are reported here.

D. Weighting analysis

A weighting analysis describing the relative contribution of each component to a decision statistic was based on data obtained in the perturbation conditions. The weighting analysis was applied separately to the rove and no-rove data using the perturbation tracks only. Application of the weighting analysis to psychophysical experiments requires several key assumptions. First, the weights are assumed to be constant across all trials for both the no-perturbation and perturbation tracks—that is, the decision process is the same for these two types of stimuli which have been randomly interleaved during data collection. Second, the decision variable, Z, is assumed to be a weighted linear combination of internally represented random variables which are obtained from the level differences per component (interval 1−interval 2), or X_i. The X_i are normally distributed with a mean μ_i and standard deviation σ_i, which includes external as well as any internal noise. The decision variable is given by

$$Z=\sum_{i=1}^{n} a_i X_i + C,$$

where a_i is the weight for the ith component, n is the total number of components, and C is a criterion or response bias. For the nonsignal components, $\mu_i=0$. A response variable, T, is defined as

$$T=\begin{cases}1, & \text{if } Z>0\\ 0, & \text{if } Z\leqslant 0\end{cases}.$$

A listener responds that the signal was present in interval 1 ($T=1$) if $Z>0$ and responds that the signal was present in interval 2 ($T=0$) if $Z\leqslant 0$.

Richards and Zhu (1994) showed that the correlation between the internally represented stimulus variables X_i and the response variable T, $\rho_{X_i,T}$, can be used to estimate a_i. Richards and Zhu showed that $\rho_{X_i,T}$ is proportional to a_i multiplied by the standard deviation, σ_i. The $\rho_{X_i,T}$ then must be "adjusted" by $1/\sigma_i$ to estimate a_i. If the X_i have equal variance, this adjustment is not needed. Whether the X_i are of equal or unequal variance may be crucial to the interpretation of the function relating weight to component frequency, as each correlation coefficient must be evaluated in relation to the noise associated with the X_i. Experimenters often assume that for normal-hearing listeners, the X_i are equal variance. However, for hearing-impaired listeners, it is unclear that this assumption is valid. This point will be addressed in more detail in Sec. IV.

In order to estimate weights, the individual trials of the last eight adaptive tracks were sorted by signal level. Dai *et al.* (1996) showed that correlations are dependent on signal level, with the lowest correlations at high signal levels. To increase the probability of measuring differences in the correlation between the two subject groups, 60 trials with the highest signal levels were excluded from further analysis. As a result, 500 trials (70 trials per track×8 tracks=560 trials−60 trials) were used in the analysis for each subject. Although, other studies in the literature used a greater number of trials for the correlation analysis [Doherty and Lutfi (1996) used 1000+ trials], 500 was deemed appropriate based on the study of Richards and Zhu (1994) who used 400 trials in their simulations.

Using methods similar to those described by Richards and Zhu (1994) and Lutfi (1995), responses for each trial were coded as 0 or 1, depending on whether the listener responded interval 1 (coded as 1) or interval 2 (coded as 0). For both groups the level difference in dB (interval 1−interval 2) was determined for each component. These level differences were computed only for the randomly selected perturbation values (i.e., the signal level was not included for the 920-Hz component). These 500 trials were blocked into 10 subgroups with 50 discriminations in each group. For all subgroups, the product moment correlation coefficient between the level differences and the responses was computed. Averaged correlation coefficients were obtained by first converting the correlation coefficients to Fisher Z values, averaging the Z scores, and then using the inverse Fisher Z transform to get an average correlation coefficient. Standard errors of the mean were computed on the Z scores as well. For convenience, the function relating correlation coefficient to component frequency will be referred to as a weighting function. Following Dai *et al.* (1996), correlations also were calculated between the listener's response and the difference of attenuator values between intervals 1 and 2 for the perturbation/rove condition (ρ_{rove}).

III. RESULTS

A. Sensitivity

Figure 1 plots average thresholds as sig re stan in dB for the no-rove (unfilled bars) and rove (hatched bars) conditions for no-perturbation and perturbation conditions. Data obtained from normal-hearing listeners are plotted to the left, and data obtained from hearing-impaired listeners are plotted to the right. Error bars indicate standard errors of the mean across the six listeners in each group. Using a repeated-measures ANOVA, a significant effect of overall level variation on thresholds is present in that thresholds are higher in the rove conditions than the no-rove conditions [$F(1,10)=10.5$, $p<0.01$]. Thresholds also are higher in the perturbation conditions relative to the no-perturbation conditions [$F(1,10)=53.1$, $p<0.001$]. The interaction between perturbation and level randomization was significant as well, with

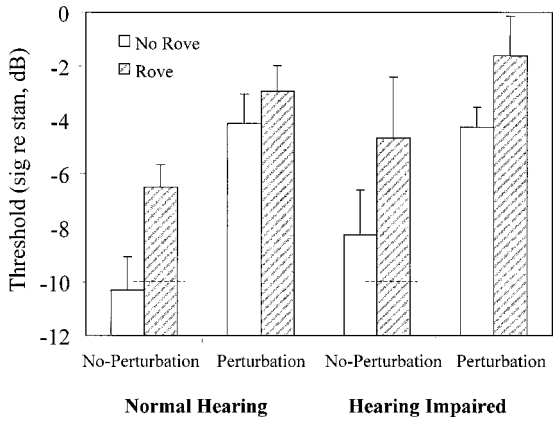


FIG. 1. Average thresholds are plotted for normal-hearing listeners to the left and hearing-impaired listeners to the right for the no-perturbation and perturbation conditions. No-rove and rove conditions are denoted by the unfilled and hatched bars, respectively. The dashed line at −10 dB indicates threshold for an ideal observer in the no-perturbation/rove condition. Error bars indicate standard errors of the mean across six listeners.

a smaller increase in threshold between rove and no-rove conditions when perturbation was present [$F(1,10)=7.1$, $p<0.03$]. On average, the presence of component-by-component perturbation degraded performance by approximately 5.1 dB in the no-rove condition and 3.3 dB in the rove condition. Overall level randomization degraded performance by approximately 3.7 dB in the no-perturbation condition and by about 2 dB in the perturbation condition. There was no main effect of subject group [$F(1,10)=0.6$, $p=0.449$], nor was group involved in any significant interaction. Thresholds averaged across all conditions were −6.0 for normal-hearing listeners and −4.7 for hearing-impaired listeners. This small difference in threshold is outweighed by the large variability across listeners.

Data in the rove conditions can be compared with data obtained in past experiments, though large interlistener variability is apparent across multiple studies. Thresholds in the current experiment are somewhat higher than reported in the past. Green *et al.* (1983) tested spectral shape discrimination thresholds using a five-component stimulus spanning 314–3178 Hz and reported thresholds near −12 dB. Our average thresholds in the no-perturbation/rove conditions of −6.5 dB for normal-hearing listeners and −4.7 dB for hearing-impaired listeners are higher than the thresholds reported by Green *et al.*, but other studies report thresholds similar to the current estimates for five-tone stimuli with narrower frequency ranges of 520–1900 Hz (e.g., Green *et al.*, 1984; Henn and Turner, 1990; Kidd, 1993). In profile-analysis experiments, thresholds tend to decrease with more practice (Kidd *et al.*, 1991), and in many of the previous studies, listeners had longer practice sessions (e.g., 5–10 h) than used here. In addition, the interleaving of fixed-amplitude stimuli amid random-amplitude stimuli increases the uncertainty in the task and probably increased thresholds in the No-Perturbation conditions.

Thresholds in the perturbation/rove condition also are somewhat higher than reported previously. Berg and Green (1990) measured thresholds for a five-component stimulus ranging from 200 to 5000 Hz and had 2 dB of level perturbation added. Their threshold of −5 dB was somewhat better than our −2.9 dB for normal-hearing listeners and −1.6 dB for hearing-impaired listeners. The current study used a slightly greater level of perturbation, which probably contributed to the higher thresholds observed.

The dashed lines in Fig. 1 indicate predicted threshold for an ideal, noiseless observer who uses the change in level of a single component to make discriminations for the no-perturbation/rove condition [approximately −10 dB; for a derivation see Green (1988), pp. 19–21]. Because thresholds are higher than predicted by the energy rule when overall level randomization is present, there remains a possibility that listeners were using a single-channel intensity cue, rather than a comparison of levels across frequency to discriminate between these sounds. Thus, the possible interpretations are listeners are comparing levels across frequency, listeners are performing simple intensity discrimination of the signal component, or some combination of the two. The results from the weighting analysis will provide further insight as to whether listeners indeed use only the signal frequency comparison as the basis of their discriminations.

B. Weights

The individual variability in this task is summarized in Fig. 2 which shows weights as a function of component frequency for each listener. Recall that the weights were obtained using only the trials in the perturbation conditions. The left panels show data obtained from normal-hearing listeners and the right panels show data from hearing-impaired listeners. Filled and unfilled symbols denote rove and no-rove data, respectively. Error bars indicate standard errors across ten estimates of the correlation coefficients. The shape of the functions relating correlation coefficient (or weight) to component frequency are similar across listeners, but the magnitude and sign of the weights for the nonsignal components show a good deal of variability. Figure 3 plots averaged data for the normal-hearing and hearing-impaired listeners in the left and right panels, respectively. Filled symbols denote rove data, and unfilled symbols denote no-rove data. Here, error bars indicate the standard errors of the mean across the six listeners in each group.

First, looking at the averaged data (Fig. 3), both groups of listeners show similar patterns in the functions relating weight to component frequency. A repeated-measures ANOVA indicated significant main effects for component frequency [$F(4,40)=54.4$; $p<0.01$] and overall level randomization [$F(1,10)=13.7$; $p<0.005$]. The largest correlation coefficient was measured for the component at the signal frequency, 920 Hz, and much smaller correlation coefficients were measured for the non-signal components. Averaged across frequency, the correlations are somewhat lower for the rove conditions than for the no-rove conditions. Level variation also reduced the correlation of the signal-bearing component. A strong correlation coefficient in the no-rove condition can be weakened somewhat by the presence of level variation, suggesting that, in part, listeners used the level of the signal tone as the basis of their discriminations. An interaction between component frequency and overall level randomization also was present [$F(4,40)=2.75$; $p<0.05$].

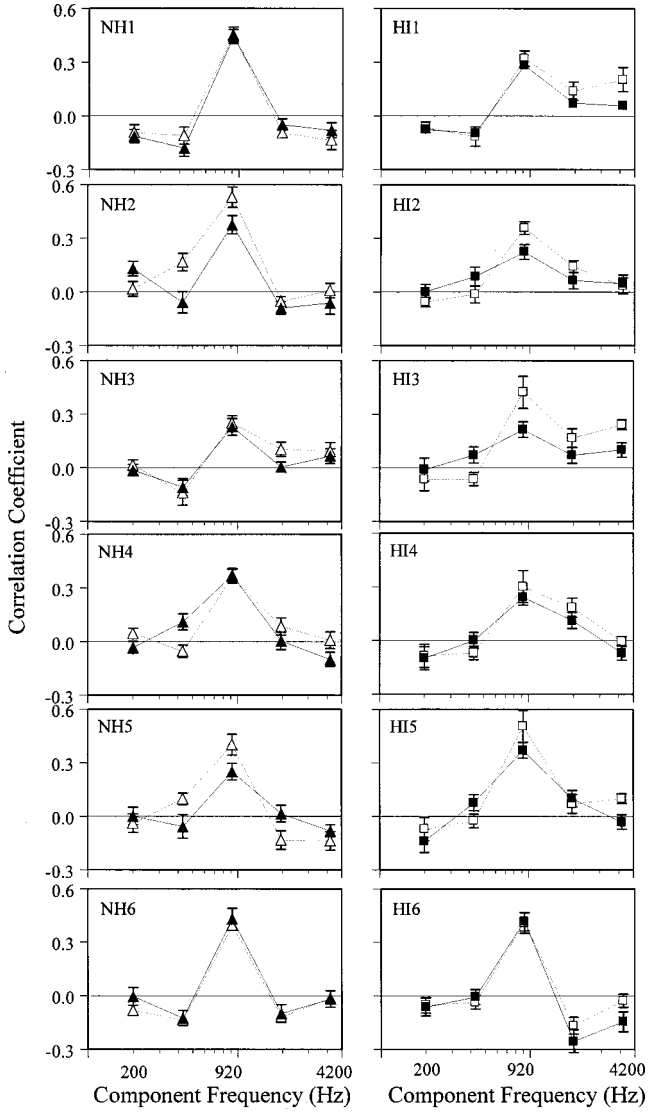


FIG. 2. Individual weighting functions (correlation coefficients as a function of component frequency) are plotted for the normal-hearing listeners in the left panels and hearing-impaired listeners in the right panels. The filled and unfilled symbols denote rove and no-rove data, and the error bars represent standard errors of the mean across ten replicate estimates of the correlation coefficient.

To some degree, the functions have different shapes between rove and no-rove conditions. This is most evident in the hearing-impaired data. No significant main effect of group was present [$F(1,10)=1.24$; $p=0.29$], but there was a significant three-way interaction among group, component frequency, and overall level variation [$F(4,40)=2.81$; $p<0.04$]. The differences between the rove and no-rove functions are more pronounced for the hearing-impaired listeners than for the normal-hearing listeners.

For each function, the weights across frequency are correlated to some degree due to the addition of the signal tone (Dai *et al.*, 1996). Therefore, a nonparametric Wilcoxon signed ranks test was used to determine whether the coefficients were different from zero. The p values reported have not been corrected for multiple comparisons to emphasize the result that even without taking multiple comparisons into account, few nonsignal correlations are different from zero.

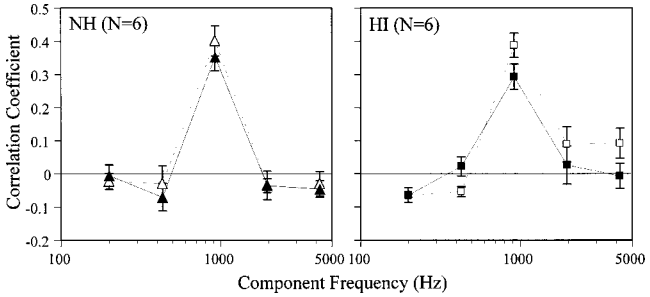


FIG. 3. Averaged weighting functions are shown for the normal-hearing (left panel) and hearing-impaired (right panel) listeners for rove (filled symbols) and no-rove (unfilled symbols) conditions. Error bars are standard errors of the mean across six listeners.

With respect to the averaged data, all correlation coefficients at the signal frequency were significantly different from zero ($p<0.03$), but only hearing-impaired listeners showed averaged weights for nonsignal components that were different from zero. These components were the two low-frequency components in the no-rove condition (200 Hz: $p<0.03$; 420 Hz: $p<0.05$) and the lowest-frequency component in the rove condition (200 Hz: $p<0.03$). The weights were negative, indicating that the hearing-impaired listeners were able to compare levels across frequency at least in the low frequencies. That all other correlations were not significantly different from zero suggests that, on average, those nonsignal components contribute little to the detection of the signal tone.

The analysis of the data as averages, however, swamps differences that may occur across listeners, which are evident in Fig. 2. For the normal-hearing listeners, of the 48 correlations obtained for the nonsignal components (6 listeners×4 components×2 conditions—rove/no-rove), only 14 had p values less than 0.05 (29%). Similarly, hearing-impaired listeners had 15 correlations that were significantly different from zero (31%). Across the two conditions, each listener had at least one component with a correlation coefficient significantly different from zero, though, statistically, the probability of this result is not negligible. To some degree, listeners may have compared levels across frequency, but primarily used the level change of the signal component to make discriminations.

For the no-rove condition (unfilled symbols), the optimal decision rule would be to ignore all stimulus components, except the signal-bearing component. Because only the signal-bearing component reliably indicates a change between signal and standard stimuli, the correlation coefficient for the 920-Hz component should be high, and the nonsignal components should not be correlated with the response. For both groups of listeners, most weights were not significantly different from zero, but this result is more pronounced for the normal-hearing listeners than the hearing-impaired listeners. Of the 24 possible comparisons in the no-rove condition, normal-hearing listeners had five weights different from zero (21%) in the no-rove condition, whereas hearing-impaired listeners had ten (42%). This analysis of individual data gives some support for an interpretation that hearing-impaired listeners did not adopt the optimal decision rule as readily as the normal-hearing listeners.

TABLE II. Correlations between listener's response and level of stimulus (ρ_{rove}).

	Listener no.					
	1	2	3	4	5	6
NH	0.06	0.40	0.28	0.35	0.11	−0.01
HI	0.24	0.48	0.43	0.26	0.44	0.14

In the rove condition (filled symbols), correlation coefficients for the nonsignal components again are close to zero. In the presence of level randomization, an optimal detector would predict a correlation function with negative (nonzero) weights for the nonsignal components and positive weights for the signal components (Berg and Green, 1990). That is, listeners would make across-frequency comparisons of level, using the nonsignal components as a reference to the change in level at the signal frequency. Analyzing the data from individual listeners indicates that in some cases, listeners may have used the nonsignal components as a reference. In the rove condition, normal-hearing listeners had nine weights different from zero (38%), whereas hearing-impaired listeners had only five (21%). Again, there is weak support for some differences between normal-hearing and hearing-impaired listeners. In general, however, across frequency-information was used only on a limited basis by both groups of listeners. Neither group of listeners adopted the optimal decision rule.

For both rove and no-rove conditions, normal-hearing listeners may have been slightly better at adopting the optimal decision rules. It is difficult to make this conclusion, however, due to the small numbers of correlations that were significantly different from zero. The small number of listeners tested and the variability across listeners undermine any strong conclusions regarding whether listeners performed simple intensity discrimination, spectral shape discrimination, or some combination of the two. A greater range of level variation might have encouraged listeners to switch decision strategies from using only the level change at the signal frequency. Differences between the rove and no-rove functions are apparent, as reflected in the results of the ANOVA.

To further test whether listeners based their decisions on overall intensity or spectral cues, correlation coefficients were calculated relating a listener's response to the absolute level difference between the two intervals (signal and standard) of a trial. These correlations were obtained only for the rove conditions, where overall level was explicitly varied through the attenuator. Table II shows these correlations for each listener. The correlations between the overall level differences and the listeners' responses (ρ_{rove}) were positive, except for NH6, indicating that the presence of level variation contributed to most listeners' responses. For the normal-hearing listeners, the correlation ranged between −0.01 to 0.40, with a mean of 0.21, and for the hearing-impaired listeners, the correlation coefficients ranged between 0.14 and 0.48, with a mean of 0.35. If listeners completely ignored level variation, the correlation coefficients would be near zero, which was the case for only two of the normal-hearing listeners, NH1 and NH6. Any positive correlation between the level variation and a listener's response indicates that responses are based partially on overall level, rather than a change in spectral shape across frequency. A two-tailed *t*-test indicates no significant differences between the normal-hearing and hearing-impaired listeners' ρ_{rove} distributions ($t = 0.08$). However, ρ_{rove} is positively correlated with age ($r = 0.59$). The older a listener, the greater the ρ_{rove}, suggesting that changes in stimulus level may be more difficult for older listeners to ignore.

IV. DISCUSSION

In this experiment, the stimuli were constructed so that potential limitations in audibility and/or reduced frequency selectivity would be minimized for hearing-impaired listeners. The mean levels of the stimuli were selected to eliminate the possibility that random selections of component level and overall stimulus level would result in an inaudible component. The component spacing was selected to be sufficiently wide (components spaced greater than an octave apart) so that within-channel masking would not affect sensitivity. When these precautions were taken, the results indicate that NH and HI listeners show similar sensitivity to changes in spectral shape and similar correlation coefficients across frequency.

A. Sensitivity

Normal-hearing and hearing-impaired listeners had similar thresholds when detecting a tone added in-phase to a component of a background sound. Thresholds were not lower than would be expected for a single-channel decision rule, which suggests that listeners may have based their decisions on the change in level of the 920-Hz component rather than using across-frequency cues. Data obtained from the weighting analysis support this interpretation to some extent.

Because these data suggest that, in part, listeners were using intensity cues to discriminate sounds, these data should be considered in light of the intensity discrimination literature on hearing impairment. Florentine *et al.* (1993) and Buus *et al.* (1995) measured pure-tone intensity discrimination at different sound pressure levels across a wide range of hearing loss configurations. Threshold values for intensity discrimination (ΔL) of about 1.9 dB (Florentine *et al.*, 1993) and 0.9 dB (Buus *et al.*, 1995) were reported for listeners with audiometric configurations similar to ours, and there was also greater variability in measurements for hearing-impaired listeners than for normal-hearing listeners. Thresholds in the no-perturbation/no-rove condition form the best comparison to these studies. After converting from sig re stan to ΔL, our normal-hearing listeners had thresholds of about 2.3 dB and our hearing-impaired listeners had thresholds of about 2.8 dB. These results are somewhat poorer than measured by Florentine *et al.* and Buus *et al.*, and the differences probably reflect uncertainty introduced by interleaving no-perturbation and perturbation tracks and the small influence of the nonsignal components. As with their studies, we observed some across-listener variability. Overall, this study

agrees with their findings that hearing-impaired listeners exhibit similar sensitivity to that of normal-hearing listeners in intensity discrimination tasks.

The experimental design employed here may have had some role in encouraging listeners to use intensity cues rather than spectral shape cues. In experimental manipulations where perturbation is added across-trial (rather than across-interval as used here), Kidd *et al.* (1988) were able to change the way memory was used in spectral shape discrimination tasks. It is thought that listeners use the context-coding mode of memory to detect changes in the spectral shape of sounds, as the listener compares each sound to a remembered reference and recalls a categorical representation of the result of that comparison (Kidd *et al.*, 1988). Tasks which depend on the context code of memory, such as profile analysis, do not show large effects of interstimulus interval duration (Durlach and Braida, 1969; Green *et al.*, 1983). In contrast, in the sensory-trace mode, a listener maintains representation of a stimulus by rehearsing an auditory image in short-term memory. These tasks, including intensity discrimination, are largely dependent on interstimulus interval duration, as the representation of the image decays with time (Durlach and Braida, 1969; Berliner and Durlach, 1973; Green *et al.*, 1983). Kidd *et al.* (1988) showed that, for profile analysis, increasing the across-trial component-by-component perturbation shifted the dependence on context coding to the sensory-trace mode of memory. The methods applied in the current experiment do not directly compare with Kidd *et al.*'s study because a no-perturbation track was interleaved with a perturbation track, but the interleaving of these tracks may have led listeners to rely more on sensory-trace coding than context coding.

In this study, where the stimuli were the sum of tones, component-by-component perturbation did not differentially affect thresholds for hearing-impaired and normal-hearing listeners. For both groups of listeners, thresholds were elevated in the perturbation conditions compared to the no-perturbation conditions. This result replicates that of Kidd *et al.* (1991) who found that normal-hearing listeners' thresholds increased with increasing levels of component-by-component perturbation (see also Lentz and Richards, 1998). The presence of level variation also increased thresholds. Level variation should affect thresholds in instances where either the change in the overall level of the stimulus cues the signal or where only the change in a single component cues the signal. The change in threshold again suggests that spectral shape discrimination is not the sole process contributing to these data.

The 10-dB level variation did not significantly change thresholds for the hearing-impaired listeners compared to the normal-hearing listeners. Because of recruitment in the impaired ears, their perception of the level variation may have been that the loudness of the sounds varied over a greater range than for the normal-hearing listeners. The large differences in sensitivity across listeners, which often occur in profile-analysis experiments, make it difficult to determine whether adding perturbation to the stimulus has differential effects for hearing-impaired listeners.

B. Weights

Overall, hearing-impaired listeners showed weights similar to normal-hearing listeners in this task. In Fig. 3, it appears that hearing-impaired listeners had a tendency to weight high-frequency components to a greater extent than normal-hearing listeners, especially in the no-rove condition. Looking at the individual hearing-impaired data, of the ten weights that were significantly different from zero in the no-rove conditions, eight of them were for the 1900- and 4200-Hz components. The three-way interaction between group, component frequency, and overall level variation suggests that hearing-impaired listeners may adopt different weights than normal-hearing listeners depending on whether rove was present or absent. This conclusion is tentative, however, as in the weighting analysis, the X_i are assumed to have equal variance. For hearing-impaired listeners, there is a strong possibility that the variance in noise across channels is not equal, especially for the listeners used in this study who have sloping hearing losses. The variance of the ''internal'' representation of the noise may be greater in regions of greater hearing loss (e.g., Humes *et al.*, 1988). Four of the hearing-impaired listeners have near-normal hearing in the low frequencies and moderate losses at the high frequencies. Thus, the high-frequency channels may be noisier than the low-frequency channels and the correlation coefficients may need to be adjusted by some value to equate them to the low-frequency channels (e.g., Richards and Zhu, 1994). In this analysis, the correlation coefficients for the high-frequency channels may be somewhat overestimated due to the greater proportion of noise in those channels. That the noisier channels would lead to overly high correlation coefficients may seem counterintuitive. However, if we assume that a decision variable, Z, is an unweighted sum of multiple random variables, X_i, and one of the channels, X_j, has a much larger variance than all other X_i, then the value of X_j will dominate the value of Z. Thus, the decision variable will be highly correlated with X_j and correlated to a much lesser degree with the X_i that have little variance. In this way, channels with greater variance lead to greater correlation coefficients. Richards and Zhu (1994) derived this argument mathematically.

This issue also may be relevant when addressing the results of Doherty and Lutfi (1996) who found weighting strategies to be different for hearing-impaired and normal-hearing listeners. Doherty and Lutfi (1996) measured weighting functions for an intensity discrimination task in which listeners detected an overall change in intensity of a six-tone complex. Their listeners tended to have sloping audiograms, with greater hearing loss in the high frequencies. Weights were higher for components near the edges of their hearing loss, but normal-hearing listeners tended to weight all components equally. Their analysis was based on the assumptions of the Conditional-On-a-Single-Stimulus (COSS) analysis, in which the weights should be unaffected by differences in internal noise across channels (Berg, 1989; Doherty and Lutfi, 1996). The weights estimated using the correlation analysis proposed by Richards and Zhu (1994) are not so immune to internal noise differences. The current study used the correlation method rather than the COSS

analysis to evaluate whether a different methodology would lead to the same conclusions as Doherty and Lutfi. Additionally, the COSS analysis requires a constant signal level be used, and a tracking procedure was used here to jointly obtain sensitivity measures and weights.

The current study may have found only slight differences in the weighting strategy between normal-hearing and hearing-impaired listeners because the correlations for the nonsignal components were estimated to be near zero. If the weights of the high-frequency components in the Doherty and Lutfi (1996) study and the 4000-Hz data of Doherty and Lutfi (1999) are overestimated due to the presence of additional noise in the region of one's hearing loss, the result that hearing-impaired listeners place greater weight on the high-frequency components may be called into question. Spectral shape changes which would lead to correlations different from zero can be used to evaluate whether different groups of listeners have different weighting strategies or whether noise varies across channels for a single group of listeners. It might be more fruitful to use a more complex spectral shape change, but differences in internal noise across frequency must be kept in mind when interpreting the weighting analysis. Another technique that may be useful is to perform a weighting analysis on catch trials that have no signal component, as the correlations would be the largest there (see Dai *et al.*, 1996).[1] Further experiments are necessary to determine whether an additional (internal) noise source may account for the differences seen among these studies.

When linear models are applied to data obtained from hearing-impaired listeners, and normal-hearing listeners are used as a comparison, another consideration is the loss of cochlear nonlinearity resulting from hearing loss (e.g., Moore, 1995). For listeners with sloping audiograms, high-frequency channels may be treated as linear, but the lower-frequency channels may have nonlinear processing intact. If variability in the external stimulus is compressed, or reduced, in the healthy channels and not compressed in the other channels, the effective noise in the system would be greater for the damaged channels than for the unimpaired channels. The loss of compression, therefore, may effectively be treated as greater noise in regions of damage.

If the weights for the hearing-impaired listeners in the current study do not need to be adjusted by a correction factor, then a very tentative conclusion would be that hearing-impaired listeners did not adopt optimal decision strategies as well as normal-hearing listeners. It is difficult to make such a case with the evidence presented here, however, as the variability across listeners in both groups in the weighting functions is rather large.

In the current study, normal-hearing and hearing-impaired listeners possibly discriminated between sounds based on the change in intensity of the signal component to some degree. This interpretation was further supported by the correlation between the listeners' responses and the level of the attenuator. The differences between groups were small; however, there was a tendency for older listeners to select stimuli with greater intensities at the signal.

These results place rather stringent limitations on models of impaired auditory processing, in that peripheral hearing loss may not lead to changes in decision strategies attributable to central processing or to changes in the ability to use the information presented to the central auditory system. These data suggest that hearing-impaired listeners are not distracted by nonsignal sounds in the region of hearing loss. At least in the relatively simple spectral shape discrimination task used here, long-term hearing loss may not alter decision strategies. It is possible that listeners do not rely more on low-frequency hearing, as might be expected in a system where high-frequency information is not represented well. The stimuli were designed in this experiment to provide equally ''good'' information to all frequency regions present in the stimuli. The impaired auditory systems were just as competent in using this ''good'' information as the healthy systems of the normal-hearing listeners.

V. SUMMARY AND CONCLUSIONS

(1) Sensitivity to simple changes in spectral shape is similar for normal-hearing and hearing-impaired listeners. Both groups of listeners were affected by external variation (component-by-component perturbation and level randomization) to nearly the same extent. However, large individual differences were present in both groups of listeners.
(2) Weights across frequency were relatively similar for normal-hearing and hearing-impaired listeners. On average, hearing-impaired listeners did not rely on energy in the regions of their hearing loss to a greater extent than more normal frequency regions.
(3) Hearing-impaired listeners may not have adopted the optimal weighting strategy as well as normal-hearing listeners. However, differences in internal noise may need to be considered to interpret the results obtained from weighting studies across normal-hearing and hearing-impaired listeners. Further studies of spectral shape discrimination are needed to determine whether differences in internal noise across frequency are present in impaired ears, and whether changes in the noise alter decision processing for complex sounds.

ACKNOWLEDGMENTS

This work was supported by Grant No. DC 00626 from the National Institute on Deafness and Communication Disorders and by the Department of Clinical Investigation, Walter Reed Army Medical Center under Work Unit #2587. The authors thank Laura Dreisbach and Lina Kubli for assistance during data collection. We also would like to thank Ken Grant, Virginia Richards, and Van Summers for their helpful comments on this manuscript. All subjects participated voluntarily and provided written, informed consent. 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]In this study, the correlations for the nonsignal components are expected to be small, but adding a tone to a stimulus further decreases the correlations of all components, leading to nonindependence of the weights across frequency (Dai *et al.*, 1996). To determine whether the large signal levels

impacted the analyses reported here, the data were reevaluated using 200 trials which had the lowest signal levels. The averaged correlation coefficients changed by no more than 0.04, but at times weights obtained by individual listeners varied greatly. Twenty-two of the 96 nonsignal weights varied by 0.06 or greater. Approximately half (55%) of these weights were estimated to have larger magnitudes when only the 200 low-signal level trials were included. Thus, including all trials from the track has little impact on the interpretation of the current results.

Berg, B. G. (**1989**). ''Analysis of weights in multiple observation tasks,'' J. Acoust. Soc. Am. **86**, 1743–1746.

Berg, B. G. (**1990**). ''Observer efficiency and weights in a multiple observation task,'' J. Acoust. Soc. Am. **88**, 149–158.

Berg, B. G., and Green, D. M. (**1990**). ''Spectral weights in profile listening,'' J. Acoust. Soc. Am. **88**, 758–766.

Berg, B. G., and Green, D. M. (**1992**). ''Discrimination of Complex Spectra: Spectral Weights and Performance Efficiency,'' in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, London), pp. 373–379.

Berliner, J. E., and Durlach, N. I. (**1973**). ''Intensity perception. IV. Resolution in roving-level discrimination,'' J. Acoust. Soc. Am. **53**, 1270–1287.

Buus, S., Florentine, M., and Zwicker, T. (**1995**). ''Psychometric functions for level discrimination in cochlearly impaired and normal listeners with equivalent-threshold masking,'' J. Acoust. Soc. Am. **98**, 853–861.

Dai, H., and Green, D. M. (**1992**). ''Auditory intensity perception: Successive versus simultaneous, across-channel discriminations,'' J. Acoust. Soc. Am. **91**, 2845–2854.

Dai, H., Nguyen, Q., and Green, D. M. (**1996**). ''Decision rules of listeners in spectral-shape discrimination with or without signal-frequency uncertainty,'' J. Acoust. Soc. Am. **99**, 2298–2306.

Doherty, K. A., and Lutfi, R. A. (**1996**). ''Spectral weights for overall level discrimination in listeners with sensorineural hearing loss,'' J. Acoust. Soc. Am. **99**, 1053–1058.

Doherty, K. A., and Lutfi, R. A. (**1999**). ''Level discrimination of single tones in a multitone complex by normal-hearing and hearing-impaired listeners,'' J. Acoust. Soc. Am. **105**, 1831–1840.

Durlach, N. I., and Braida, L. D. (**1969**). ''Intensity perception I: Preliminary theory of intensity resolution,'' J. Acoust. Soc. Am. **46**, 372–383.

Florentine, M., Reed, C. M., Rabinowitz, W. M., Braida, L. D., Durlach, N. I., and Buus, S. (**1993**). ''Intensity perception. XIV. Intensity discrimination in listeners with sensorineural hearing loss,'' J. Acoust. Soc. Am. **94**, 2575–2586.

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.

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford U.P., New York).

Green, D. M., Kidd, G., Jr., and Picardi, M. C. (**1983**). ''Successive versus simultaneous comparison in auditory intensity discrimination,'' J. Acoust. Soc. Am. **73**, 639–643.

Green, D. M., and Mason, C. R. (**1985**). ''Auditory profile analysis: Frequency, phase and Weber's law,'' J. Acoust. Soc. Am. **77**, 1155–1161.

Green, D. M., Mason, C. R., and Kidd, G., Jr. (**1984**). ''Profile analysis: Critical bands and duration,'' J. Acoust. Soc. Am. **75**, 1163–1167.

Henn, C. C., and Turner, C. W. (**1990**). ''Pure-tone increment detection in harmonic and inharmonic backgrounds,'' J. Acoust. Soc. Am. **88**, 126–131.

Humes, L. E., Espinoza-Varas, B., and Watson, C. S. (**1988**). ''Modeling sensorineural hearing loss. I. Model and retrospective evaluation,'' J. Acoust. Soc. Am. **83**, 188–202.

Kidd, G., Jr. (**1993**). ''Individual differences in the improvement in spectral shape discrimination due to increasing number of nonsignal tones,'' J. Acoust. Soc. Am. **93**, 992–996.

Kidd, G., Jr., and Mason, C. R. (**1992**). ''A new technique for measuring spectral shape discrimination,'' J. Acoust. Soc. Am. **91**, 2855–2864.

Kidd, G., Jr., Mason, C. R., and Hanna, T. E. (**1988**). ''Evidence for sensory-trace comparisons in spectral shape discrimination,'' J. Acoust. Soc. Am. **84**, 144–149.

Kidd, G., Jr., Mason, C. R., Uchanski, R. M., Brantley, M. A., and Shah, P. (**1991**). ''Evaluation of simple models of auditory profile analysis using random reference spectra,'' J. Acoust. Soc. Am. **90**, 1340–1354.

Lentz, J. J., and Richards, V. M. (**1998**). ''The effects of amplitude perturbation and increasing numbers of components in profile analysis,'' J. Acoust. Soc. Am. **103**, 535–541.

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

Lutfi, R. A. (**1995**). ''Correlation coefficients and correlation ratios as estimates of observer weights in multiple-observation tasks,'' J. Acoust. Soc. Am. **97**, 1333–1334.

Matiasek, M. M., and Richards, V. M. (**1999**). ''Relative weights for three different psychophysical tasks,'' J. Acoust. Soc. Am. **106**, 2209(A).

Moore, B. C. J. (**1995**). *Perceptual Consequences of Cochlear Damage* (Oxford U.P., New York).

Richards, V. M., and Zhu, S. (**1994**). ''Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients,'' J. Acoust. Soc. Am. **95**, 423–434.

Spiegel, M. F., Picardi, M. C., and Green, D. M. (**1981**). ''Signal and masker uncertainty in intensity discrimination,'' J. Acoust. Soc. Am. **70**, 1015–1019.

Summers, V., and Leek, M. R. (**1994**). ''The internal representation of spectral contrast in hearing-impaired listeners,'' J. Acoust. Soc. Am. **95**, 3518–3528.

Maximum speed of pitch change and how it may relate to speech[a)]

Yi Xu[b)] and Xuejing Sun
Department of Communication Sciences and Disorders, 2299 North Campus Drive, Northwestern University, Evanston, Illinois 60208

(Received 23 May 2001; revised 13 November 2001; accepted 29 November 2001)

How fast speakers can change pitch voluntarily is potentially an important articulatory constraint for speech production. Previous attempts at assessing the maximum speed of pitch change have helped improve understanding of certain aspects of pitch production in speech. However, since only ''response time''—time needed to complete the middle 75% of a pitch shift—was measured in previous studies, direct comparisons with speech data have been difficult. In the present study, a new experimental paradigm was adopted in which subjects produced rapid successions of pitch shifts by imitating synthesized model pitch undulation patterns. This permitted the measurement of the duration of entire pitch shifts. Native speakers of English and Mandarin participated as subjects. The speed of pitch change was measured both in terms of response time and excursion time—time needed to complete the entire pitch shift. Results show that excursion time is nearly twice as long as response time. This suggests that physiological limitation on the speed of pitch movement is greater than has been recognized. Also, it is found that the maximum speed of pitch change varies quite linearly with excursion size, and that it is different for pitch rises and falls. Comparisons of present data with data on speed of pitch change from studies of real speech found them to be largely comparable. This suggests that the maximum speed of pitch change is often approached in speech, and that the role of physiological constraints in determining the shape and alignment of F_0 contours in speech is probably greater than has been appreciated. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1445789]

PACS numbers: 43.70.Aj, 43.70.Bk, 43.70.Jt [AL]

I. INTRODUCTION

Speech is produced by a biomechanical system that has various inherent limitations. Many of these limitations may play a role in shaping the acoustic signal generated in speech production. One of them is the maximum speed at which speakers can change their pitch voluntarily. The importance of this limitation has not been widely recognized among those who are interested in the patterns of fundamental frequency variations in speech, however. Part of the reason for this lack of appreciation is probably a general feeling that in speech we approach our biomechanic limits only occasionally. In recent years, however, there has been accumulating evidence that ''time pressure'' may play a part in determining the shape and alignment of certain F_0 contours in speech (Caspers and Heuven, 1993; Ladd, Mennen, and Schepman, 2000; Xu, 1998, 2001). Unless some kind of limit is reached or approached, of course, time pressure should not make much difference. Nevertheless, the exact role the maximum speed of pitch change may play in speech is far from clear. This is partly because the data we have obtained about the speed of pitch change are far from complete, and partly because we have yet to pinpoint any direct link between maximum speed of pitch change and actual variations of F_0 contours in speech.

The maximum speed of pitch change was studied in the 70's by Ohala and Ewan (1973) and Sundberg (1979). Both studies used similar methods. Subjects were asked to shift from one pitch level to another as fast as possible upon command (Sundberg, 1979).[1] Then, the speed of pitch change was assessed by measuring the response time—time used to complete the fastest portion (the middle 75%) of a pitch shift, as illustrated in Fig. 1. The term ''response'' was first used by Ohala and Ewan (1973) in the phrase ''the response characteristics of the larynx in voluntary pitch change.'' The definition of the measurement, namely, time corresponding to the middle 75% of the pitch change, was also first used by Ohala and Ewan (1973). Sundberg (1979) used response time in his paper to refer to this measurement. Response time therefore provides a measurement of the time it takes the subject to respond to the command (hand waving or light flash as described in his paper) for making the pitch shift. Several findings of these early studies are potentially important for speech. First, the average response time was found to be around 79, 85, and 101 ms for pitch rises of 4, 7, and 12 semitones, respectively, and about 68, 70, and 73 ms for pitch drops of 4, 7, and 12 semitones. Thus, the speed of pitch change seems to increase as the interval becomes larger. Second, as also indicated by the measured response time, pitch lowering was faster than pitch elevation. Third, female subjects was found to have shorter response times than male subjects.

The estimates obtained by Ohala and Ewan (1973) and Sundberg (1979) constitute a very important step toward a

[a)]Part of the preliminary results of this study were presented at the 6th International Conference on Spoken Language Processing, Beijing, 2000 (Xu and Sun, 2000).
[b)]Electronic mail: xuyi@northwestern.edu

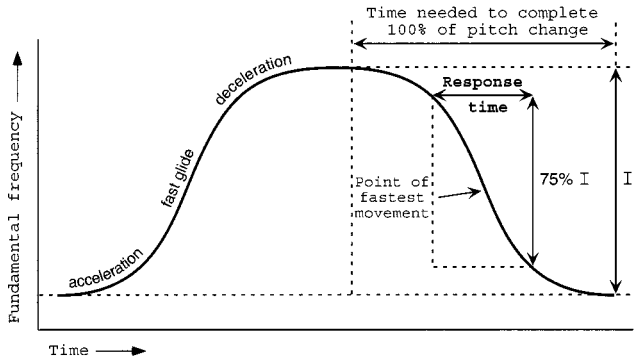


FIG. 1. First, illustration of *Response time*—time used to complete 75% of a pitch change, as defined by Ohala and Ewan (1973) and Sundberg (1979) (partially adapted from Sundberg, 1979). Second, illustration of the time needed to complete 100% of a pitch change. Third, conceptual division of a pitch movement into three phases: acceleration, fast glide, and deceleration.

clear understanding of the physiological limits on the speed of pitch change. However, those data are incomplete in two respects. First, as can be seen in Fig. 1, response time does not fully reflect the fastest instantaneous pitch movement possible, which should occur somewhere in the middle of the rising and falling ramps in the pitch change curve. Second, and more importantly, by definition, response time does not tell us how much time it takes for the speaker to complete 100% of a pitch shift, which is potentially much longer, as Fig. 1 clearly suggests.

The time needed to complete 100% of a pitch shift is potentially important for our understanding of pitch production in speech. In tone languages, for example, utterances often contain alternate high and low pitches. As observed in Xu (1997, 1999), the transitions between high and low pitches take a considerable amount of time, and the duration of the transition affects the shape of the F_0 contours. Furthermore, the minimum duration of the pitch transition may also limit how closely adjacent F_0 peaks or valleys can follow each other. Such limits may play a role in determining the alignment of F_0 peaks and valleys relative to segmental units such as syllables (Xu, 1999, 2001; Caspers and van Heuven, 1993).

It is also possible that articulatory limits on the speed of pitch change are never reached in speech, and that, instead, it is certain perceptual constraints that limit how fast pitch shifts are made in speech. This has been suggested by 't Hart, Collier, and Cohen (1990). In an effort to understand the rate of pitch change observed in speech, they considered both production and perception as possible contributing factors. Eventually, they dismissed production in favor of perception as the determining factor, based mainly on their interpretation of the data reported by Ohala and Ewan (1973) and Sundberg (1979). They first computed the maximum speed of pitch change using the response time for the 12-st condition as reported by Sundberg (1979), and came up with the value of 120 st/s. They then looked for the fastest pitch movement in Dutch and found it to be only 50 st/s. Based on this comparison, they concluded that articulatory limits simply could not have been responsible for the observed rate of pitch change in speech. Instead, they theorized, it must have been listeners' limited perceptual ability to distinguish between different rates of pitch change that has forced speakers to use slower pitch change rates in speech (pp. 71–75). A similar interpretation of Sundberg's (1979) data was also adopted by Caspers and van Heuven (1993).

Such interpretation of the data reported by Ohala and Ewan (1973) and Sundberg (1979) seems to have exaggerated the maximum speed of pitch change in two ways. First, the data from both studies indicate that the rate of pitch change is faster for a larger pitch interval than for a smaller one. So, unless a particular pitch movement actually spans 12 st, it is inappropriate to use the data for that interval as the direct indicator of speed of pitch change at other intervals. Second, response time, by definition (Sundberg, 1979), measures the time it takes to complete only 75% of the pitch change. Although it is not yet clear at this point how long it will take to complete 100% of a pitch change, it should certainly take longer than response time, as indicated by Fig. 1. Thus, it is inappropriate to treat response time as the ultimate indicator of the physiological limits on the speed of pitch change, because it corresponds to only part of the pitch shift.

A better way to estimate the actual physiological limits on the speed of pitch change would be to examine how long it takes to complete 100% of a pitch change at various pitch intervals. It probably would have been difficult, however, for Ohala and Ewan (1973) and Sundberg (1979) to measure the time interval of complete pitch changes with the experimental paradigm they employed, even if that was what they had in mind initially. In Fig. 1, at the beginning of a pitch shift, the F_0 movement seems to continually accelerate: changing very slowly at first, and gradually reaching full speed. Near the end of the pitch shift, the speed of movement decelerates, and the curve levels off gradually as the target pitch is being reached. To measure the time needed to complete 100% of the pitch shift as shown in Fig. 1, they would have to locate the exact points in time when the shift began and when it ended. However, the asymptotes near the onset and offset of the pitch shift make it hard to determine these points. To lessen the uncertainty in determining the end points of a pitch shift, therefore, it is necessary to minimize the duration of these asymptotes. One way to do that is to have the speaker produce a very quick succession of high and low pitches. This would generate a pitch undulation pattern that goes up and down rapidly, in which the lingering time on each pitch should be reduced to a minimum.

In the present study, we adopted a new paradigm to assess the maximum speed of pitch change by having subjects produce rapid pitch undulation patterns through imitation of rapidly alternating high–low pitch sequences. The maximum speed of pitch change is then estimated by measuring, among other things, the time used to complete 100% of each pitch shift. To see if these measurements can provide more direct indications as to whether and how physiological limits may play a critical role in shaping the F_0 contours of speech, we compared our data with data on speed of pitch change in real speech obtained in previous studies.

In addition, we also examined the effects of language, gender, pitch carrier (sustained schwa versus /malamalama/), and pitch shift direction on various measurements. To check

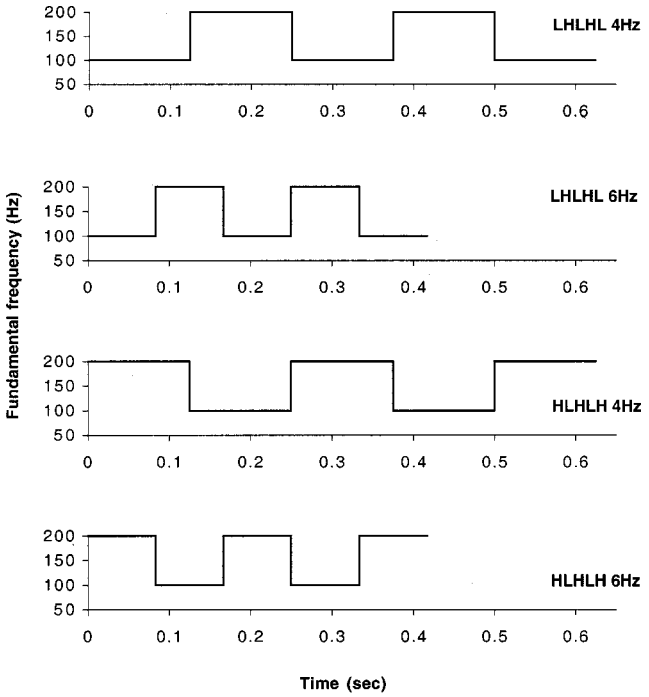


FIG. 2. Schematic representations of model pitch undulation patterns used as stimuli in the experiment, where 4 and 6 Hz refer to 4 and 6 undulation cycles per s, respectively. These examples have the base frequency of 100 Hz.

for possible differences due to language, we used native speakers of Mandarin and English as subjects. Mandarin has lexical tones (pitch patterns that can differentiate words) while English only has pitch accents related to intonation and word stress. It is possible that because successive high- and low-tone sequences frequently occur in Mandarin, native speakers of Mandarin can make pitch undulations faster than native speakers of English.[2] In addition, because pitch contours are usually carried by the vocalic portion of syllables in speech, it is possible that the syllable structure may either hinder the production of pitch undulation or facilitate it. To test which is the case, we used both CV syllables and simple vowels as the pitch carriers. To verify the gender difference as reported by Sundberg (1979), we also compared data from male and female speakers.

II. METHOD

A. Stimuli

The stimuli were model pitch undulation patterns to be imitated by subjects, as illustrated in Fig. 2. As can be seen in Fig. 2, the pitch undulation patterns are either HLHLH or LHLHL, where H and L represent relatively high and low pitches, respectively. The model pitch undulation patterns were built by modifying naturally produced voice samples. First, schwa-like vowels with relatively steady fundamental frequencies were recorded by a male and a female speaker. The duration of the vowel was approximately 1 s. Based on these original voice samples, a series of pitch manipulations was carried out to generate stimuli with desired pitch variation patterns.

Using the TD-PSOLA resynthesis function of the PRAAT program,[3] the fundamental frequencies of the original vowels were modified to generate a number of steady-state vowels. Based on the male voice, vowels with the base frequencies of 50, 60, 70, 80, 90, 100, and 115 Hz were generated. Based on the female voice, vowels with the frequencies of 115, 130, 145, 165, 185, 205, and 230 Hz were generated. These two ranges of frequencies correspond to the lower limits of male and female voices.

For each vowel at a particular fundamental frequency, 12 model pitch undulation patterns were generated. These pitch patterns differed in three ways. In terms of pitch variation pattern, they were either /HLHLH/ or /LHLHL/. In terms of pitch variation interval, the difference between H and L was, following Sundberg (1979), 4, 7, or 12 semitones. In terms of pitch variation rate, the duration of each HLH or LHL cycle in a pattern was either 1/4 or 1/6 s. In other words, the undulation frequency (i.e., the number of HLH or LHL cycles per second) of a model pitch pattern was either 4 or 6 Hz. The 4-Hz condition was to help warm up the subject during each trial, while the 6-Hz condition was to elicit the fastest pitch changes possible. It has been shown that 6 Hz is at the slower end of the involuntary vibrato rate in singing (Prame, 1994; Dejonckere, Hirano, and Sundberg, 1995). Presumably, *voluntary* pitch undulations are unlikely to exceed 6 Hz.

B. Subjects

Nineteen native speakers of American English (10 females and 9 males) and 22 native speakers of Mandarin Chinese (12 females and 10 males) between the ages of 18 and 45, recruited from Northwestern University campus, participated in the experiment. Subjects all reported having normal hearing, vision, and language ability. While some of the subjects had musical or voice training of some kind, none of them was a professional singer or involved in a professional singing group. Those with professional voice training were excluded so as to avoid the effect of extensive voice training on pitch undulation rate (Sundberg, 1979). The tasks of the experiment turned out to be too difficult for several subjects. Some of them were not able to produce the desired pitch shifts in the right order. Those subjects were all native speakers of English. The others produced pitch ranges smaller than two semitones in many trials. Those were both Mandarin and English subjects. As a result, only 36 subjects generated data suitable for analysis. Of the remaining subjects, 16 are English speakers (8 females and 8 males) and 20 are Chinese speakers (11 females and 9 males).

C. Procedure

The experiment was conducted in the Speech Acoustics Laboratory at Northwestern University. The subjects imitated the model pitch patterns using both a schwa and a syllable sequence (/malamalama/) as the pitch carriers. In total, each subject produced 240 trials (3 pitch intervals×2 carriers×2 patterns×2 undulation rates×2 sessions×5 repetitions).

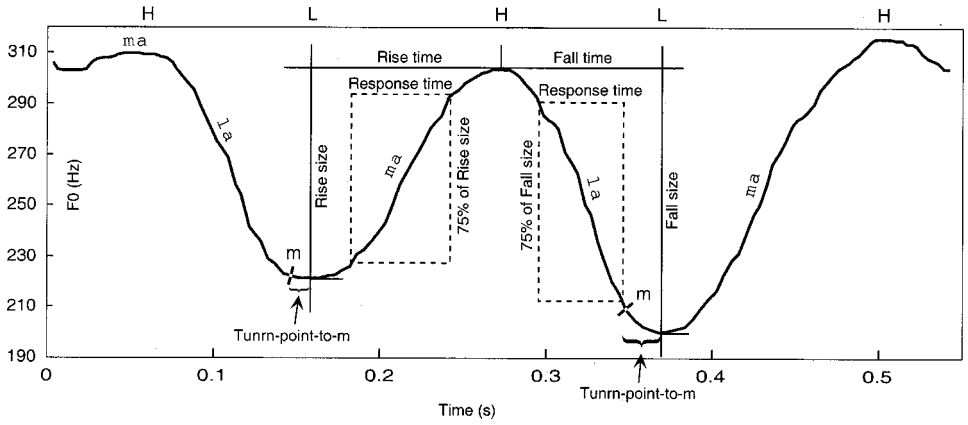


FIG. 3. Illustration of measurement of rise and fall excursion time, rise and fall response time, and turn-point-to-*m* in an HLHLH trial spoken with /malamalama/. The two cross points on the curve labeled "*m*" are the onsets of second and third /m/ in the trial. The fact that both turning points occur after "*m*" means that both turn-point-to-*m* values are negative.

During the experiment, the subject was seated in a sound-treated booth in front of a computer monitor. The experiment procedure was controlled by a set of HTML files, which were displayed by Netscape Navigator (Netscape Communications Corp.). A condenser microphone was used for the recording, and the vocalization was directly digitized onto a computer hard disk in a Macintosh G4 computer using SOUNDEDIT (Macromedia Inc.). For each subject, a comfortable pitch level was first determined before the start of the practice trials by choosing from a range of prerecorded voice samples played by the first HTML page. (Among the female subjects, seven selected the base frequency of 185 Hz, eight selected 205 Hz, and four selected 230 Hz. Among the male speakers, one selected the base frequency of 80 Hz, two selected 90 Hz, eight selected 100 Hz, three selected 115 Hz, and three selected 130 Hz.) The experimental stimuli were organized into three HTML pages, each containing model undulation patterns with the pitch intervals of 4, 7, or 12 semitones, respectively. On each page the undulation models are divided into two patterns—HLHLH and LHLHL, and two rates—4 and 6 Hz. The subject selected one of the stimuli each time by clicking on the corresponding button. The model pattern was then played through the loudspeaker. The subject was instructed to imitate the stimuli five times in each session, and as accurately as possible in terms of both pitch interval and undulation frequency.

The experimenter sat outside the recording booth, watching another computer screen showing the same display as seen by the subject, and listening to the subject's vocalization through a pair of headphones. The experimenter monitored the subject's performance and gave instructions when necessary. Since the task was somewhat difficult for some subjects, an intensive practice session was held for them before the real trials.

The whole recording process consisted of four sessions. In the first and third sessions, the subject imitated the pitch models with a schwa, while in the second and fourth sessions, the subject imitated the models with the syllable sequence /malamalama/. This particular syllable sequence was found in pilot tests to be the fastest vowel–sonorant sequence one could produce. Between sessions, the subject was given the chance to take a break. During each trial, i.e., for each model, the subject was allowed to replay the model pattern as many times as they wished. The experimenter would ask the subject to repeat a trial if it was felt to be necessary.

D. F_0 extraction and measurement

The F_0 extraction was done using a procedure similar to the ones used in Xu (1997, 1998, 1999, 2001). The procedure combines custom computer programming with ESPS/WAVES+ (Entropic Inc.). The digitized signals were transferred to a Dell workstation running on the LINUX platform. The *epochs* program was then used to mark every pitch period in each undulation sequence, and the labels were saved into a text file. After that, the waveform, the period labels, and the spectrogram of the signal were displayed in *xwaves*. The period labels were examined carefully for spurious vocal pulse markings such as double labeling and period skipping. Apparent errors were corrected manually.

While checking and correcting the vocal period labels, segmentation labels were also added at the onset and offset of all vocalizations and at the boundaries between /m/ and adjacent vowels for the /malamalama/ trials.

The vocal period and segment labels for each trial were saved in a text file. All the text files were then processed by a set of custom-written C programs. The programs converted the duration of pitch periods into F_0 values, and then smoothed the resulting F_0 curves using a *trimming algorithm* that eliminates sharp bumps and edges (Xu, 1999).[4] The trimmed F_0 curves were then subjected to further analysis using a set of custom-written MATLAB procedures. The following measurements were taken by the MATLAB procedures, most of which are illustrated in Fig. 3.

Excursion size (*rise* or *fall*)—pitch difference (in st) between adjacent F_0 minimum and maximum in the middle undulation cycle. Excursion size is expressed in semitone in order to make the data from individual speakers, especially across genders, more comparable.[5]

Excursion time (*rise* or *fall*)—time interval between adjacent F_0 maximum and minimum in the middle undulation cycle.

Excursion speed=excursion size/excursion time.

"*Response time*"—time interval corresponding to the middle 75% of *excursion size* (in Hz), as defined by Ohala and Ewan (1973) and Sundberg (1979).[6]

TABLE I. Mean values of various measurements under the effects of language, gender, direction (of pitch change), (pitch) carrier, and interval (of pitch change), together with probability values resulting from five-factor mixed-measure ANOVAs. Significant p values are printed in boldface.

	Language		Gender		Direction		Carrier		Interval		
	Chinese	English	Female	Male	Rise	Fall	Mala	Schwa	4	7	12
Excursion size (st)	4.4	5.8	5.0	5.0	4.8	5.3	5.1	5.0	3.8	4.7	6.6
	$p=$ **0.0037**		$p=0.8773$		$p<$ **0.0001**		$p=0.1396$			$p<$ **0.0001**	
Excursion time (ms)	125.3	141.2	128.7	136.4	132.5	132.2	133.6	131.2	125.7	128.2	143.2
	$p=$ **0.0219**		$p=0.2467$		$p=0.6987$		$p=0.2052$			$p<$ **0.0001**	
Response time (ms)	69.6	75.6	75.6	68.4	71.7	72.7	73.6	70.8	70.7	70.6	75.4
	$p=$ **0.0431**		$p=$ **0.0308**		$p=0.4608$		$p=$ **0.0208**			$p=$ **0.0005**	
Excursion speed (st/s)	35.9	42.1	40.1	37.0	36.5	40.8	38.6	38.7	30.8	37.5	47.6
	$p=0.1195$		$p=0.3674$		$p<$ **0.0001**		$p=0.8541$			$p<$ **0.0001**	
Maximum velocity (st/s)	60.8	72.4	65.0	67.1	61.3	70.6	66.4	65.6	50.3	62.4	85.1
	$p=0.0749$		$p=0.8323$		$p<$ **0.0001**		$p=0.3625$			$p<$ **0.0001**	

Maximum velocity—positive and negative extrema in the velocity curve corresponding to the rising and falling ramps in the middle undulation cycle. Velocity curves were computed by taking the first derivative of the F_0 curves after they were further smoothed by a five-point median filter and a seven-point (for male speakers) or 17-point (for female speakers) Hanning window.

For /malamalama/ files, the following measurements were also taken:

Peak-to-m—average time interval between the second and third F_0 maxima and onset of the second and third /m/ in LHLHL.

Valley-to-m—average time interval between the second and third F_0 minima and onset of the second and third /m/ in HLHLH.

Note that the value of *peak-to-m* or *valley-to-m* is negative if the peak or valley occurs after the onset of /m/, as illustrated in Fig. 3.

In the analyses, only data meeting the following criteria are included:

(a) Excursion size (rise or fall) >1 st;
(b) Excursion size (rise or fall) <2 standard deviations about the mean;
(c) Excursion time (rise or fall) <2 standard deviations about the mean.

Also, since the study is investigating the fastest speed of pitch change possible, only trials in the 6-Hz undulation frequency condition were processed for analysis. After applying these criteria to all trials in the 6-Hz condition, 3553 of the 4320 data points (82%) remained for further analysis. Of the excluded data points, 226 failed criterion (a), and 541 failed criteria (b) or (c).

III. ANALYSES

A. Effects of language, gender, direction of pitch change, pitch carrier, and interval of pitch change

Table I displays the mean *excursion size, excursion time, response time, excursion speed, and maximum velocity* broken down according to language (Chinese/English), gender (female/male), direction of pitch change (rise/fall), pitch carrier (malamalama/schwa), and interval of pitch change (4/7/12 st). Also displayed in the table are the probability values resulting from five-factor mixed-measure ANOVAs performed on the five measurements. Of the independent variables, language and gender are between-group factors, and the rest are within-group factors.

From Table I it can be seen that the effect of interval is significant for all measurements. Also, a set of Student–Newman–Keuls *post hoc* tests found the differences between all pairs of the three intervals to be significant at the 0.05 level, with the exception of excursion time and response time between 4 and 7 st. This indicates that (a) subjects managed to produce different excursion sizes for the three pitch-shift intervals, and (b) the speed of pitch change varied across the intervals. The mean interval sizes achieved by the subjects, however, are not quite what we had hoped for. In particular, for the 12-st condition, the mean interval achieved was only 6.5 st. Interestingly, the English subjects achieved greater excursion sizes than the Mandarin subjects.

In addition to excursion size, the effect of language is also significant for excursion time and response time. For both of them, the English subjects had greater means than the Mandarin subjects. This does not mean, however, that native English speakers are slower in making pitch changes. In fact, their excursion speed and maximum velocity are both somewhat faster than those of the Mandarin subjects, although neither difference reaches significance. It could be the case that the larger excursion size of the English subjects actually gave rise to the faster speed. This is partially verified by Table II, which shows that excursion speed and maximum velocity are highly correlated with excursion size, but not with excursion time and response time, despite the fact that time is actually in the equation for computing excursion speed. The fact that the English subjects produced larger pitch excursions and hence faster pitch changes than the Mandarin subjects is somewhat surprising to us, because presumably, speakers of a tone language should have better ability to make local pitch changes.[7]

As shown in Table I, the effect of direction is significant for excursion size, excursion speed, and maximum velocity. It is not significant, however, for excursion time and response time. While this is somewhat different from Sundberg (1979), where response time was found to be different for pitch lowering and pitch elevation, falls in the present data

TABLE II. Correlation (r) of various factors (computed from 432 mean values of each measurement: 2 directions×2 carriers×3 intervals×36 subjects=432 means).

	Excursion size	Excursion time	Excursion speed	Maximum velocity	Response time
Excursion size	1.000	0.384	0.859	0.920	0.212
Excursion time		1.000	−0.103	0.084	0.845
Excursion speed			1.000	0.956	−0.194
Maximum velocity				1.000	−0.106
Response time					1.000

are nevertheless consistently faster than rises.

There are also significant interactions between direction and interval for excursion size, excursion time, and response time. For excursion size, the interaction is largely due to greater differences between rise size and fall size at smaller intervals ($\Delta f = 0.6$ st and $\Delta f = 0.5$ st) than at the largest interval ($\Delta f = 0.2$ st). For excursion time and response time, on the other hand, the interaction is largely due to the lack of differences at the intervals of 4 and 7 st. In contrast, both excursion time and response time are longer when the interval is 12 st.

The main effect of carrier is significant only for response time, as shown in Table I. However, a number of interactions involving carrier reached or approached significance level. Interestingly, it is excursion size, excursion time, and response time that have significant or near-significant interactions, as can be seen in Fig. 4. English subjects show much larger differences between /malamalama/ and the schwa than Mandarin subjects for excursion size, excursion time, and response time. Furthermore, Mandarin speakers' excursion size is smaller when the carrier is /malamalama/ than when it is the schwa, whereas the difference with English speakers is reversed. It is possible that, for the Mandarin speakers, it is less natural to change pitch repeatedly within a sustained vowel than to associate each pitch value with a syllable, because the latter is similar to what they do in speaking their native language. If this is the case, in performing the task they may tend to use the usual pitch range for lexical tones which has been found to require only a small portion (up to 6 st, cf. Xu, 1999, in press) of a speaker's pitch range (up to 2 octaves, cf. Fairbanks, 1959). For the English subjects, in contrast, maybe associating a different pitch with each of the successive syllables is quite unnatural and consequently they had to use more effort in performing the task, resulting in a larger pitch range.[8]

Also, male subjects show larger differences between /malamalama/ and the schwa than female subjects for all three measurements (though only near-significance level for response time). The differences may seem to indicate that, with respect to these interactions, female speakers overall behave more like Mandarin speakers than like English speakers. However, the two probability levels do not seem high enough to warrant a clear conclusion about the gender effect at this point.

Somewhat surprisingly, the main effect of gender was not significant for any of the measurements except response time. But, response time is longer for female subjects than for male subjects, which is just the opposite of what is suggested by Sundberg's (1979) data. Nor was there any significant interaction between gender and direction as can be observed in Sundberg's data. As it turns out, however, there are other gender differences that are actually quite robust, as we will discuss next.

B. Excursion time versus response time

As suggested in Fig. 1, a complete pitch shift probably consists of three phases: acceleration, rapid glide, and deceleration.[9] Response time, as defined by Ohala and Ewan (1973) and Sundberg (1979), is the amount of time it takes the speaker to complete the middle 75% of a pitch change. Conceptually, therefore, pitch movement during response time corresponds largely to the fast glide phase of the pitch shift, and the rest of the excursion time to the initial acceleration and final deceleration. A comparison of excursion time with response time may thus let us see the distribution of time between the fast glide and acceleration–deceleration phases of a pitch shift. Table III displays excursion size, excursion time, response time, and the ratio of excursion time to response time broken down according to interval, direction, and gender.

As can be seen in the table, the ratio of excursion time to response time ranges from 1.62 to 2.07, and the mean ratio is 1.87. Thus, excursion time is always much longer than response time, and in many cases even more than twice as long. This indicates that a large portion of excursion time is missing when estimating the maximum speed of pitch change with response time alone.

Table III also reveals an interesting gender difference in terms of the ratio of excursion to response time. A five-factor (language, gender, direction, carrier, and interval) mixed ANOVA finds the main effect of gender highly significant for this ratio: $f(1,32) = 187.26$, $p < 0.0001$. Also highly significant are the main effect of interval [$f(1,32) = 24.24$, $p < 0.0001$] and the interaction between gender and interval [$f(1,32) = 26.63$, $p < 0.0001$]. As can be seen in the last row of Table III, for male speakers the excursion/response ratio remains fairly constant around 2. But, for female speakers the ratio increases steadily from 4 to 12 st, and is smaller than that of the males even for the largest intervals. Looking closer for the source of this difference, we notice in Table III that at each interval in both directions, female speakers have shorter excursion time but longer response time than male speakers. All this seems to indicate that female speakers use less time than male speakers in the acceleration and deceleration phases of the pitch shift. There are two possible ex-

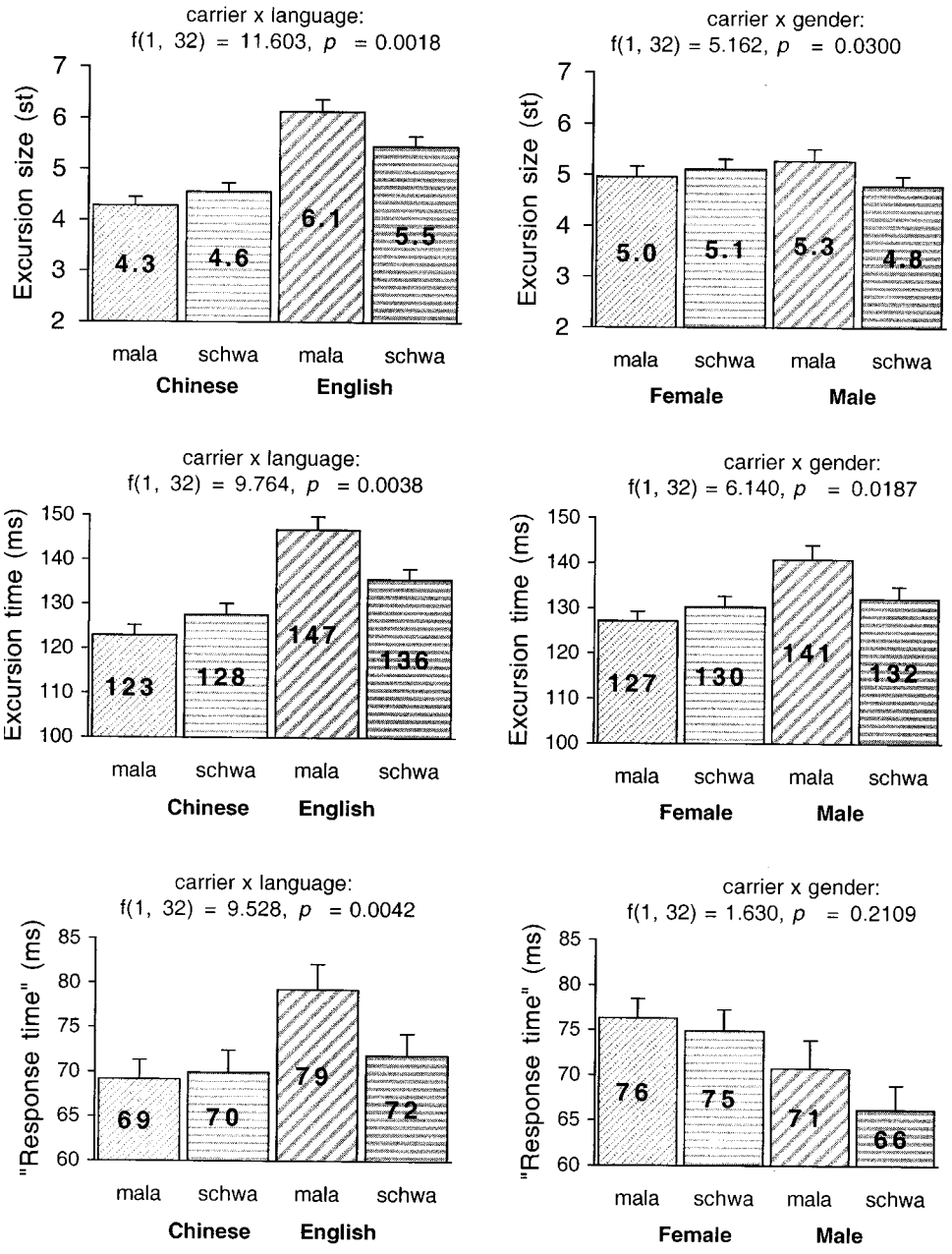


FIG. 4. Interactions of carrier with language and gender on excursion size, excursion time, and response time. Results of the five-factor mixed-measure ANOVAs for each interaction are displayed on top of the respective graph.

planations for this finding. The first is that female speakers have more powerful laryngeal muscles than male speakers so that they can start and stop a pitch shift faster than male speakers can. The second explanation, which we think is more plausible, is that female speakers have less laryngeal mass and hence less laryngeal inertia, thus needing less time than male speakers to initiate and end a pitch shift.

The main effects of direction and carrier also reach significance level [$f(1,32)=6.088, p=0.0191$; $f(1,32)=5.475, p=0.0257$]. But, the differences in the means are

TABLE III. Excursion size, excursion time, response time, and ratio of excursion time to response time. The ratios are means of individual ratios from all subjects.

Direction	Rise						Fall						
Interval	4		7		12		4		7		12		
Gender	f	m	f	m	f	m	f	m	f	m	f	m	Mean
Excursion size (st)	3.3	3.7	4.4	4.5	6.8	6.1	3.9	4.3	4.9	5.1	6.8	6.5	5.0
Excursion time (ms)	120	129	123	133	144	148	122	132	126	133	137	144	133
Response time (ms)	75	64	72	65	80	73	76	67	75	69	76	73	72
Excursion/response	1.62	2.07	1.70	2.07	1.82	2.04	1.62	2.00	1.69	1.96	1.83	2.02	1.87

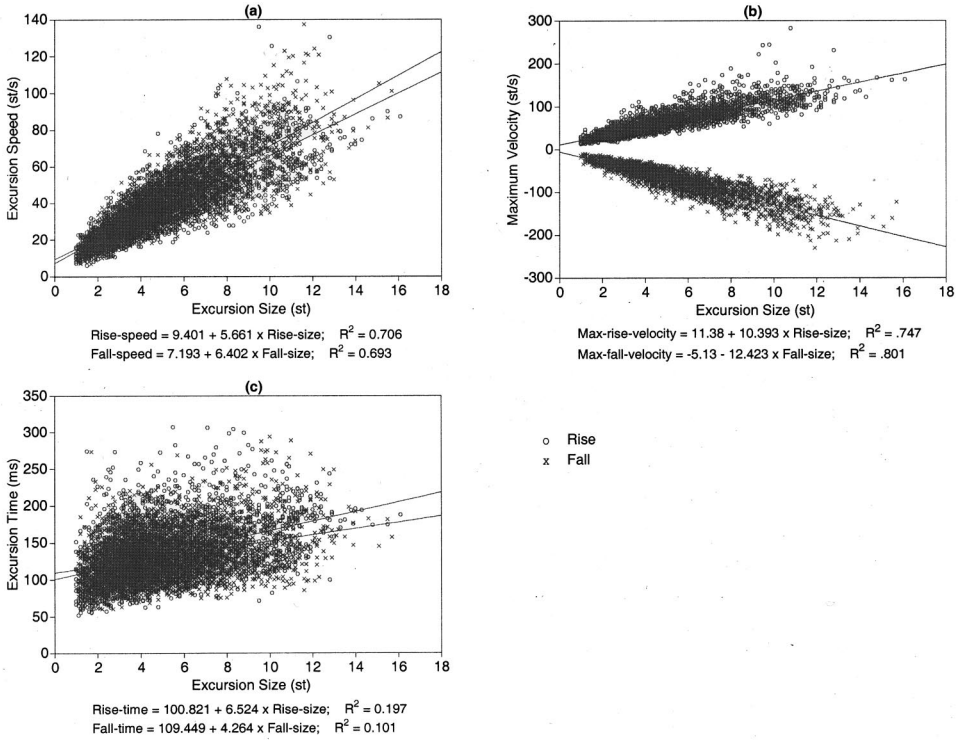


FIG. 5. Scatter plots of excursion speed, maximum velocity, and excursion time produced by all subjects as functions of excursion size. Only data points meeting the criteria listed in Sec. II D "F_0 extraction and measurement" are included. Each plot thus consists of 3553 data points.

so small (1.878 vs 1.844 for rise and fall, and 1.847 vs 1.876 for /malamalama/ and schwa) that we do not want to attribute them much importance.

C. Excursion time and excursion speed as functions of excursion size

We can see in Table I that excursion time, response time, excursion speed, and maximum velocity all vary significantly with pitch change interval. To observe their relation with pitch change interval in more detail, excursion time, excursion speed, and maximum velocity are plotted as functions of excursion size in all conditions for all the subjects in Fig. 5. As can be seen in the figure, in general both excursion speed and maximum velocity change fairly linearly with excursion size. Based on this observation, simple linear regressions were performed for each subject on both pitch rises and falls, with excursion size as independent variable and excursion time, excursion speed and maximum velocity as dependent variables.

Table IV shows the coefficients (intercept and slope, i.e., b and a in $z=b+ax$) of simple linear regressions of excursion time over excursion size, as well as excursion time computed with these regression parameters for the excursion sizes of 4, 7, and 12 st. Also shown in the table are the maximum, minimum, and standard deviation of the coefficients and computed excursion time across subjects. As can be seen, the deviation of the computed excursion time is quite large, especially when the excursion size is large. To reduce the deviation, three subjects (Nos. 11, 15, 36) whose computed excursion time at any of the six excursion sizes (3 sizes$\times$2 directions) deviates more than 2 standard deviations about the mean were taken out, and a new set of regression coefficients and computed excursion time values was obtained, as shown in Table V.

Regression coefficients for excursion speed and maximum velocity were also computed and are displayed in Tables VI and VII together with the predicted values at excursion sizes of 4, 7, and 12 st. Consistent with Fig. 5, for both excursion speed and maximum velocity, the mean R^2 values for pitch rise (0.723 and 0.734) and pitch fall (0.708 and 0.753) are quite high in Tables VI and VII. This further suggests that the relationship between excursion size and excursion speed is quite linear. The linear equations displayed in Tables VI and VII can therefore be used to predict the

TABLE IV. Coefficients (intercept and slope) of simple linear regressions of excursion time over excursion size, excursion time computed with these coefficients for the excursion sizes of 4, 7, and 12 st, and R^2 of the regression analyses. Data rows 1–4 display the mean, maximum, minimum, and standard deviation, respectively. In the Max and Min rows, all the values except R^2 are taken from the subject whose average rise time or fall time across the three intervals is the largest among all subjects. This way the maximum and minimum rise time can be computed with the corresponding intercept and slope values.

	Rise time (ms)						Fall time (ms)					
	Intercept	Slope	4 st	7 st	12 st	R^2	Intercept	Slope	4 st	7 st	12 st	R^2
Mean	90.4	9.7	129	158	206	0.355	100.9	6.3	126	145	177	0.201
Max	146.7	32.3	216	313	475	0.714	138.9	22.5	171	209	321	0.672
Min	48.7	3.1	91	105	128	0.053	51.3	−0.5	95	99	97	0.008
Std dev.	22.0	6.0	25.6	39.1	66.4	0.172	18.7	4.4	17.9	26.5	45.9	0.167

TABLE V. Coefficients (intercept and slope) of simple linear regressions of excursion time over excursion size, excursion time computed with these coefficients for the excursion sizes of 4, 7, and 12 st, and R^2 of the regression analyses (data from 33 subjects). Data rows 1–4 display the mean, maximum, minimum, and standard deviation, respectively. In the Max and Min rows, all the values except R^2 are taken from the subject whose average rise time or fall time across the three intervals is the largest among all subjects. This way the maximum and minimum rise time can be computed with the corresponding intercept and slope values in the table.

	Rise time (ms)						Fall time (ms)					
	Intercept	Slope	4 st	7 st	12 st	R^2	Intercept	Slope	4 st	7 st	12 st	R^2
Mean	89.6	8.7	124	150	194	0.356	100.4	5.8	124	141	170	0.193
Max	75.8	20.6	158	220	323	0.714	104.9	11.7	152	187	245	0.672
Min	73.1	4.6	91	105	128	0.053	102.1	−0.5	100	99	97	0.008
Std dev.	20.3	4.4	18.8	26.6	45.4	0.179	15.4	3.5	15.9	23.0	38.5	0.158

TABLE VI. Coefficients (intercept and slope) of simple linear regressions of excursion speed over excursion size, excursion speed computed with these coefficients for the excursion sizes of 4, 7, and 12 st, and R^2 of the regression analyses (data from 33 subjects). Data rows 1–4 display the mean, maximum, minimum, and standard deviation, respectively. In the Max and Min rows, all the values except R^2 are taken from the subject whose average rise time or fall time across the three intervals is the largest among all subjects. This way the maximum and minimum rise speed can be computed with the corresponding intercept and slope values in the table.

	Rise speed (st/s)						Fall speed (st/s)					
	Intercept	Slope	4 st	7 st	12 st	R^2	Intercept	Slope	4 st	7 st	12 st	R^2
Mean	10.8	5.6	33	50	78	0.723	8.9	6.2	34	52	83	0.708
Max	9.0	8.6	43	69	112	0.914	−0.6	10.3	41	72	123	0.906
Min	10.2	3.6	25	35	54	0.325	13.0	3.4	27	37	54	0.371
Std dev.	4.3	1.3	4.8	8.0	14.2	0.139	5.0	1.6	4.6	8.4	15.8	1.46

TABLE VII. Coefficients (intercept and slope) of simple linear regressions of maximum velocity over excursion size, maximum velocity computed with these coefficients for the excursion sizes of 4, 7, and 12 st, and R^2 of the regression analyses (data from 33 subjects). Data rows 1–4 display the mean, maximum, minimum, and standard deviation, respectively. In the Max and Min rows, all the values except R^2 are taken from the subject whose average rise time or fall time across the three intervals is the largest among all subjects. This way the maximum and minimum values can be computed with the corresponding intercept and slope values in the table.

	Max rise velocity (st/s)						Max fall velocity (st/s)					
	Intercept	Slope	4 st	7 st	12 st	R^2	Intercept	Slope	4 st	7 st	12 st	R^2
Mean	12.4	10.5	54	86	139	0.734	−6.8	−12.1	−55	−92	−152	0.753
Max	0.2	18.0	72	126	216	0.927	−11.8	−8.7	−46	−72	−116	0.941
Min	20.8	6.2	46	64	95	0.270	13.9	−18.2	−59	−113	−204	0.410
Std dev.	7.7	2.5	7.9	13.9	25.6	0.137	9.0	2.1	6.1	10.0	19.7	0.126

TABLE VIII. Mean values of turn-point-to-*m* (ms) under the effects of language, gender, turn type, and interval, together with probability values resulting from four-factor mixed-measure ANOVAs.

Language ($p = 0.0021$)	Chinese				English			
Gender ($p = 0.0241$)	Female		Male		Female		Male	
Turn type ($p = 0.0012$)	Valley	Peak	Valley	Peak	Valley	Peak	Valley	Peak
	−3.7	5.8	−0.2	21.9	12.9	18.1	24.6	29.8

TABLE IX. Comparison of response time measured by Sundberg (1979) and computed response time based on the present data.

		Rise response time			Fall response time		
	Gender	4 st	7 st	12 st	4 st	7 st	12 st
Sundberg (1979)	Female	75	81	98	62	68	70
	Male	83	89	104	73	72	75
Present study	Female	73	80	91	75	78	84
	Male	63	78	104	65	73	86

maximum speed of pitch change at different pitch change intervals. The data on maximum velocity as of yet do not have any real speech data with which to compare. However, in situations where the onset or offset of a pitch change is hard to determine, and maximum instantaneous velocity is the only measurable speed indicator, data in Table VII may be used as useful reference.

D. Alignment of peak and valley with syllable

For trials with /malamalama/ as the carrier, analysis was also done on the alignment of F_0 peaks and valleys relative to syllable boundaries. Table VIII displays the mean values of turn-point-to-*m* (including both peak-to-*m* and valley-to-*m*), broken down according to language, gender, and turn type. Also displayed in the table are the probability values resulting from four-factor mixed-measure ANOVAs. Of the independent variables, language and gender are between-group factors, and turn type and interval are within-group factors. The effect of interval is not significant and thus is not included in the table. As illustrated in Fig. 3, a positive value of turn-point-to-*m* means that the F_0 peak or valley corresponding to a syllable is realized *before* the end of the syllable; a negative value means that the peak or valley is realized *after* the syllable offset; and a small value, whether positive or negative, means that the peak or valley is realized close to the end of the syllable. The values in Table VIII indicate that Mandarin subjects produced F_0 peaks and valleys closer to the syllable offset than English subjects, and that there is a greater tendency for their peaks or valleys to occur *after* the end of the syllable. The same tendencies can also be seen in female versus male subjects, and in pitch falls versus pitch rises (i.e., valley versus peak).

IV. DISCUSSION

The main goal of the present study was to assess the maximum speed of pitch change in such a way that the assessment is more relevant to our understanding of F_0 contour variations in speech. By measuring not only response time as done in earlier studies (Ohala and Ewan, 1973; Sundberg, 1979) but also excursion time—time used to complete 100% of a pitch shift, the data so obtained can be more readily compared to data collected in studies of F_0 contours in real speech, as will be done next. As will be discussed subsequently, these comparisons make it possible for us to consider, in more realistic terms than before, implications of the maximum speed of pitch change on our understanding of F_0 contour production in speech in general. Finally, there are still inadequacies in the present study which can be improved in future research, as will be considered briefly.

A. Comparison with previous studies

There are a number of studies that have collected data to which the current data can be compared. In particular, Sundberg (1979), Caspers and van Heuven (1993), Ladd *et al.* (1999), Ladd *et al.* (2000), and Xu (1999) will be considered.

1. Sundberg (1979)

As mentioned earlier, Sundberg (1979), following Ohala and Ewan (1973), measured the response time of pitch shift, defined as the time to complete 75% of the pitch shift interval. To compare his data with ours, we estimated the mean response time from Fig. 3 of Sundberg (1979) (because no actual numbers were reported in the paper) and displayed them in the upper two rows in Table IX. The lower two rows of Table IX display the mean response time computed from our current data for males and females, respectively.

As can be seen in Table IX, while the computed response time values for pitch falls from the present study are slightly longer than those of Sundberg's, they are somewhat shorter than Sundberg's for pitch rises. Overall, the computed response time in the present study is comparable to that of Sundberg's. This suggests that the excursion time for Sundberg's subjects, had it been possible to measure it, could have been similar to that of the present study.

Also, the difference in speed between pitch rises and falls reported by Sundberg (1979) is largely confirmed. The mechanism behind this difference, however, is still unclear. Judging from the fact that the difference seems to increase with the size of pitch change, it is possible that it is due to the different muscles involved in pitch lowering and elevation. The former probably mainly involves the cricothyroid and thyroarytenoid, and the latter the infrahyoid strap muscles (sternohyoid, thyrohyoid, and sternothyroid) (Erickson, 1993; Hallé, 1994; Honda, 1995), which are more powerful, but slower at small ranges (Honda, 1995).

2. Caspers and van Heuven (1993)

Caspers and van Heuven (1993) examined the effect of time pressure on the realization of pitch rises and falls related to accents in Dutch. They measured the excursion size and F_0 slope of pitch rises and falls associated with prenuclear accents. Table X displays the fastest speed of pitch rises and falls for the female speaker and male speaker reported by Caspers and van Heuven. Also displayed in Table X are the excursion speeds computed from our current data for the average and fastest male and female speakers, respectively. The fastest speaker is the one whose computed excursion speed averaged across 4, 7, and 12 semitone is the greatest among all female or male speakers for pitch rises or falls. As can be seen in the table, in making pitch rises, the female speaker in their study is a bit faster than even the fastest female subject in the present study; the male speaker is faster than our average male speaker but slower than our fastest male speaker. In making pitch falls, their female speaker is faster than our average female speaker but slower than our fastest female speaker; their male speaker is a bit slower than our fastest male subject but about as fast as our average male subject. Overall, therefore, their speakers are faster than our average speakers but somewhat slower than our fastest speakers. This indicates that their two speakers were probably speaking near the speed limit when making the fastest pitch changes.

As mentioned in the Introduction, however, because they used response time for the 12-st interval reported by Sundberg (1979) as the actual limit of speed of pitch change, Caspers and Heuven concluded that the slope of pitch rises and falls was well within the articulatory limits. Comparisons of their data with ours as shown in Table X and the

TABLE X. Comparison of fastest excursion speed in Caspers and van Heuven (1993) and computed excursion speed based on the present data. The excursion sizes are as listed in Table 3 of their paper for the female and male speakers, respectively. The excursion speed from our data was computed using the intercepts and slopes of the fastest as well as the average male and female subjects in the present study at the same excursion sizes as those in Table 3 of Caspers and van Heuven.

		Rise		Fall	
	Subject	Size (st)	Slope/speed (st/s)	Size (st)	Slope/speed (st/s)
Caspers and van Heuven (1993)	Female	6.7	72	10.1	77
	Male	7.8	66	9.3	59
Present study	Average female	6.7	49	10.1	69
	Fastest female	6.7	67	10.1	96
	Average male	7.8	53	9.3	61
	Fastest male	7.8	72	9.3	84

newly-understood Sundberg data both suggest that the speakers examined by Caspers and van Heuven (1993) probably approached their maximum speed of pitch change quite frequently.

3. Xu (1999)

This study examined F_0 variations in Mandarin under the effects of tone and focus. The F_0 analysis in the study was done with a similar procedure as used in the present study. However, no data on the speed of pitch change were reported in the published paper. To extract data on the speed of pitch change from the raw data obtained in the study, we wrote a new C program to make the measurements. The program locates the F_0 peaks and valleys at the edges of pitch movements the same way as in the present study, and then measures the excursion time and excursion speed. The individual values of excursion speed were regressed over excursion size as in the present study for each of the eight Mandarin subjects. The regression equations were then used to compute excursion speed at the excursion intervals of 4, 7, and 12 st.

Table XI displays the mean excursion speed computed for the three intervals for different tones together with the excursion speed values as shown in Table VI and the mean excursion speed values from the Mandarin subjects alone. The values of speed of pitch change were divided into two groups: those associated with static tones and those with dynamic tones. Static tones refer to the H (High) and L (Low) tones. F_0 movements occur in this group when the tones of two adjacent syllables differ at the syllable boundary, such as in HL, LH, etc. Dynamic tones refer to the R (Rising) and F (Falling) tones, to which pitch movements are presumably intrinsic (Xu and Wang, 2001).

As can be seen in Table XI, although excursion speed associated with the static tones in Xu (1999) is much slower than the maximum excursion speed obtained in the present study, the excursion speed associated with the dynamic tones is fairly comparable with the present data both from all the subjects and from the Mandarin subjects alone. This comparison is quite interesting, because it shows that in speech, the maximum speed of pitch change is approached only when there is a strong demand for a fast pitch change.

4. Ladd et al. (1999) and Ladd, Mennen, and Schepman (2000)

These two studies investigated the alignment of F_0 peaks and valleys in the prenuclear rising accent in English (Ladd *et al.*, 1999) and Dutch (Ladd *et al.*, 2000). We computed the speed of pitch change (st/s) from the data reported in these studies and listed them in Table XII. Also listed in the table is the speed of pitch change in the present study estimated using the coefficients in Table VI for the same excursion sizes.

From Table XII it can be seen that in both cases, the speed of pitch change in those two studies is somewhat slower than the estimated speed obtained in the present study. This is despite the fact that both studies include conditions where time pressure is potentially applied to the realization of pitch movements. We do notice one thing that is different about these two studies when compared to the studies discussed earlier, however. That is, the prenuclear accents

TABLE XI. Comparison of excursion speed estimated from Xu (1999) and from the present data (same as in the first row of Table VI).

		Rise speed			Fall speed		
		4 st	7 st	12 st	4 st	7 st	12 st
Xu (1999)	Static tone	24	37	58	21	35	57
	Dynamic tone	31	51	83	29	49	81
All speakers in present study		33	50	78	34	52	83
Mandarin speakers in present study		34	52	82	34	53	84

TABLE XII. Comparison of excursion speed in Ladd *et al.* (1999) and Ladd *et al.* (2000) with computed excursion speed based on our data (Table VI). For Ladd *et al.* (1999), the excursion sizes are computed from data in Appendix B of their paper, the excursion time is obtained from Fig. 3 and Table 2 of their paper. For Ladd *et al.* (2000), for their experiment 1 the rise size was estimated based on their Table I, and rise time was computed from data in Table I as well as in the text. For their experiment 2, the rise size is from endnote 2 and rise time from both their Table II and endnote 2. The rise speeds in both studies were calculated by dividing rise size with rise time.

	Rise size (st)	Rise speed (st/s)
Ladd *et al.*, 1999: Experiment 1 fast	3.7	21
Present study	3.7	31
Ladd *et al.*, 1999: Experiment 2 fast	3.4	21
Present study	3.4	29
Ladd *et al.*, 2000: Experiment 1 fast	5.4	23
Present study	5.4	40
Ladd *et al.*, 2000: Experiment 2	6.5	31
Present study	6.5	46

in these studies always occur on a syllable that is followed by an unstressed syllable. If we assume that an unstressed syllable either does not have a pitch target of its own, or carries only a rather weak pitch target, then it is possible that the rise in a prenuclear accent that precedes an unstressed syllable is not implemented under the greatest time pressure. In other words, they are somewhat similar to the situation of the static tones in Mandarin, whose implementation also does not seem to require maximum speed of pitch change, as has been shown in Table XI. Naturally, the validity of this interpretation awaits closer examination in future studies.

B. Implications

As suggested earlier, a rapid pitch shift should consist of three phases: acceleration, rapid glide, and deceleration. Previous studies of the maximum speed of pitch change seem to have focused mainly on the second phase, i.e., the rapid glide (Ohala and Ewan, 1973; Sundberg, 1979). The present study takes all three phases into consideration when estimating the speed of pitch change. As it turns out, excursion time is nearly twice as long as response time. Furthermore, it is found that the speed of pitch change varies quite linearly with the size of pitch change, and that it varies also with the direction of pitch change. These findings have many implications for our understanding of pitch contours in speech as well as other aspects of speech production. In the following, we will discuss just a few of these implications.

1. How often is the maximum speed of pitch change reached in speech?

The role of articulatory constraints has been widely recognized in the phonetics and phonology literature. However, rarely do we see serious discussion on whether limits on the speed of articulatory movements are actually reached. Perhaps this is because of the general belief that human beings as biological systems would not allow their physiological limits to be approached very often when performing a task as routine as speech. Instead, more consideration is given to the economy of effort, as defined by Lindblom (1982), as the ultimate constraint in speech production. Economy of effort implies that the speaker is capable of making a more extreme articulation but chooses not to. A physiological limit, on the other hand, is a threshold that the speaker simply cannot cross. One study that does seriously consider the possible role of articulatory limit on the speed of pitch change in determining various aspects of F_0 contours in speech is 't Hart *et al.* (1990). As discussed in the Introduction, however, their interpretation of the data reported by Sundberg (1979) underestimated the actual articulatory limits on the speed of pitch change. The comparison of present data with those of Xu (1999) discussed earlier suggests that pitch change speed comparable to that obtained with a paradigm as demanding as that employed in the present study can be easily observed in real speech in Mandarin. For Dutch, the full excursion size found by 't Hart *et al.* (1990) is around 6 st (p. 53). At this interval, the speed of 50 st/s they reported is also comparable to the excursion speed in Table VI at the same interval. Also, as shown in Table X, the fastest pitch change speed reported by Caspers and van Heuven (1993) is comparable to the maximum speed of pitch change at similar pitch shift intervals. Furthermore, as mentioned by 't Hart *et al.*, in English, full-size rises and falls can span an octave and the rate of change can reach 75 st/s (p. 49). This again is comparable to the computed mean excursion speed for the 12-st interval shown in Table VI. These comparisons seem to suggest that the maximum speed of pitch change is probably approached or even reached more often than we have realized.

Note that this does not mean that the maximum speed of pitch change is reached all the time. Rather, there are many situations in which the thresholds are not likely approached. For example, the production of the static tones (H, L) in Mandarin probably does not often require maximum speed of pitch change, as Table XI seems to suggest. Also, the production of the prenuclear accent in English and Dutch probably does not call for maximum speed of pitch change if the stressed syllable is followed by an unstressed syllable, as Table XII appears to suggest.[10] What seems critical is that in each specific situation we need to try to recognize if a particular biomechanical limit may be approached and whether a condition exists that necessitates the approximation of that limit. In the following we will discuss a number of such situations and examine how maximum speed of pitch change may play a role in shaping certain F_0 contours in speech.

2. How may contextual tonal variations relate to maximum speed of pitch change?

In a series of studies on contextual tonal variations in Mandarin (Xu, 1994, 1997, 1999), it was found that the F_0 contour of a tone varies extensively with the offset F_0 of the preceding tone, especially when there is no voiceless consonant separating the vowels. The H tone in Mandarin, for example, is produced with an apparent rising contour when it follows the L tone. Likewise, the L tone is produced with an apparent falling contour when preceded by the H tone. Xu and Wang (2001) suggest that these seemingly long transitions are due to the fact that it takes time to make the required pitch change when shifting from one tone to another. What was not known, however, was how much time it would take for such transitions to complete. As shown in Table V, it

would take on average 142 ms (computed with the mean intercept and slope for rise time in the table) for a Mandarin speaker to complete a 6-st pitch rise. This means that in a syllable with an average duration of 180 ms (Xu, 1999), the greater half of the F_0 contour in the syllable would have to be used for completing the pitch rise from the L tone to the H tone even if the maximum speed of pitch has been achieved. The long transitions found in Xu (1997, 1999) now seem to have a clearly plausible articulatory explanation: speakers probably have no way of avoiding them, given the limit of their laryngeal physiology.

In many African tone languages, e.g., Yoruba, the H tone is said to change into a rising tone when preceded by the L tone, and the L tone is said to change into a falling tone when preceded by the H tone (Hyman and Schuh, 1974). While there are various phonological accounts of this kind of tonal variation (e.g., Hyman and Schuh, 1974; Goldsmith, 1990; Manfredi, 1993), it is not yet clear if such changes are due to speaker's intentional change of the articulatory target associated with the tone. From the limited duration data that can be obtained from Laniran (1992) and Akinlabi and Liberman (1995), it seems that the average syllable duration in Yoruba is no longer than that in Mandarin. This suggests that these dynamic F_0 patterns in Yoruba probably have much to do with speakers' articulatory limitations. If the maximum speed of pitch change found in the present study is universal, the rise in the H tone and fall in the L tone are probably inevitable whenever they are preceded by a tone with a very different offset pitch.

3. What are the linguistically meaningful pitch targets and how are they realized in speech?

As has been observed in a number of recent studies, certain F_0 events such as F_0 peaks and valleys maintain a relatively stable alignment with the onset or offset of the syllable (Arvaniti, Ladd, and Mennen, 1998; Caspers and van Heuven, 1993; Kim, 1999; Ladd *et al.*, 2000; Prieto, v. Santen, and Hirshberg, 1995; Xu, 1998, 1999, 2001). There are disagreements, however, over the interpretation of these alignment patterns. In particular, Ladd and his colleagues argue that these patterns indicate that F_0 turning points are linguistically meaningful targets and are ''anchored'' by speakers at the onset or offset of the syllable, and that observed F_0 shapes are merely interpolations between these targets (Arvaniti *et al.*, 1998; Ladd *et al.*, 1999; Ladd *et al.*, 2000). An alternative view recently offered by Xu and Wang (2001) and Xu (in press) contends that observed F_0 events such as peaks and valleys are not necessarily the underlying functional units *per se*. Rather, they are mostly products of the interaction between underlying pitch targets and various articulatory constraints. For example, the H, L, R, and F tones in Mandarin probably have the underlying pitch targets [high], [low], [rise], and [fall]. In speech production, these targets are synchronously implemented with the syllables that carry them due to, presumably, the constraints of coordinated movements (Kelso, 1984; Schmidt, Carello, and Turvey, 1990). Due to the constraint of the maximum speed of pitch change, however, the realization of these targets in contexts often deviates much from their realization in isolation, resulting in contextual tonal variations as discussed in the previous section. Furthermore, according to this view, the occurrence of F_0 peaks and valleys requires the right sequence of pitch targets, and the alignment of the peaks and valleys depends on the properties of the pitch targets involved. For example, in an LHL sequence, F_0 has to rise from the first L tone to realize the [high] of the H tone, and then fall to realize the [low] of the second L tone. This will result in a rising F_0 contour during the H-carrying syllable, a falling contour during the second L-carrying syllable, and a peak near the boundary between the second and third syllables.

What the findings of the present study tell us is that, no matter what form the linguistically meaningful targets take, implementing them takes time. If, for example, the speaker's task is to anchor an F_0 minimum at the onset of a syllable-initial consonant, as suggested by Ladd and colleagues, an average speaker would have to start the F_0 movement toward this low point at least 124 ms earlier, even if the range of the movement is just 4 st (cf. Table V). Furthermore, the speaker would have to adjust the onset of a pitch movement according to the size of the F_0 excursion toward that low point. This has yet to be confirmed by empirical data.

If, instead, the speaker's task is to implement a pitch target such as [high] in synchrony with a syllable, as suggested by Xu and Wang (2001), there would be no need to anticipate the size of the F_0 movement toward this target. Rather, the speaker just needs to start the implementation of the pitch target at the onset of the syllable and end the implementation at the offset of the syllable. Because it takes at least 124 ms to raise or lower pitch by 4 st according to the present data (Table V), much of the earlier F_0 contour during a syllable would form a transition from the preceding pitch target toward the current target. Furthermore, the shape of the transition and the height and slope of the F_0 contour near the end of the syllable would all vary depending on the magnitude and direction of the difference between the two adjacent targets and the duration of the syllable. Additionally, depending on the duration of a syllable, an F_0 peak associated with it would occur either before or after its offset. All of these have indeed been found in Mandarin (Xu, 1997, 1998, 1999, 2001).

The present data therefore seem to provide some support for the view that, at least in Mandarin, underlying pitch targets such as [high], [low], [rise], and [fall] constitute part of the meaningful linguistic units and they are produced synchronously with their associated syllables. It is possible; however, languages like English and Dutch are very different from Mandarin in terms of underlying pitch targets and their alignment, and this difference may explain the contrast between the aforementioned views. Further studies are needed to resolve this issue.

C. Limitations, caveats, and future directions

Despite the significance of the data obtained in the present study, we are aware of their limitations, and we want to also point out a few caveats and identify possible future directions. First, from Tables IV–VII it is apparent that there are large variations across subjects. Part of the variability

may be due to subjects' different abilities to perform the arbitrary task of the experiment. For example, compared to Sundberg's (1979) nonsinger subjects who were all taking a musical class at the time of the experiment, ours had rather diverse musical backgrounds. Although we did not find any contribution of musical training, the lack of musical training of some subjects may have contributed to the difficulty they experienced while trying to perform the task.

Second, the 6-Hz undulation rate that we used for the stimuli is probably a bit too fast, since even our fastest subjects did not achieve that rate. We used 6 Hz to ensure that we get the fastest speed possible. But, it may also have contributed to the difficulty some subjects experienced during the experiment. The other source of the variability may be each individual subject's true idiosyncratic speed of pitch change. It would be interesting in future studies to examine whether individual speakers' F_0 contour patterns are directly linked to their own maximum speed of pitch change.

Third, the significant effects of language, gender, and turn type on turn-point-to-*m*, as shown in Table VIII, are somewhat puzzling to us. In the /malamalama/ condition, what the subjects were asked to do is to produce the pitch undulation patterns together with the syllable sequences. Although there were no explicit instructions as to how precisely they should align the two, the implied requirement is that they produce the two synchronously. The patterns shown in Table VIII, however, do not seem to fit what one might predict from previous data. We know that female speakers might have faster speed of pitch change according to Sundberg's (1979) data, or they may have faster F_0 movement acceleration and deceleration as suggested by our new data discussed above. We also know that Mandarin probably requires higher precision of pitch target alignment than English because virtually every syllable is specified with a lexical tone. Finally, we know from data reported by Ohala and Ewan (1973), Sundberg (1979), and the current study that lowering pitch is faster than raising it. As speculated by Ohala (1978), ''since they can be accomplished quicker, they [falling tones] might be less likely than rising tones to 'spill over' onto the next syllable'' (p. 31). All these seem to suggest that the turning points should more likely occur after rather than before the end of the pitch-associated syllable for male than for female speakers, for Mandarin than for English speakers, and for rises than for falls. The fact that the opposite of all of these was found in the present data might suggest that there is some subtle but critical difference between subjects' tasks in the present study and the task of producing linguistically meaningful words and phrases.

Finally, this study did not look directly into the physiology of pitch production. And, the few physiological studies we reviewed could not provide us with direct explanations about the speed of pitch change observed in this study. So, it is not yet clear to us what, at the muscular level, makes changing pitch take as much time as found in the present study. This, again, can be resolved only by future studies.

V. CONCLUSIONS

The goal of the present study was to assess the maximum speed of pitch change in such a way that the results can be directly compared to data from real speech. This was motivated by our realization that previous attempts at measuring the speed of pitch change have provided only a partial picture. They obtained data only on response time—time needed to complete 75% of a pitch shift. As shown in Fig. 1, that corresponds only to the fast glide phase of a pitch shift. It became apparent to us that the initial acceleration and final deceleration should also be taken into consideration before the data on the speed of pitch change can be fully useful for speech research. In this study, we adopted a new experimental paradigm in which subjects produced rapid pitch shifts by imitating a series of model pitch undulation patterns. This enabled us to measure the complete duration of each pitch shift as well as that of the response time as defined in previous studies.

As it turns out, it takes nearly twice as long to complete an entire pitch shift as it takes to execute the middle 75% of the shift. This finding indicates that physiological limitation on the speed of pitch movement is probably much greater than has been recognized. More interestingly, we find that the maximum speed of pitch change obtained in this study is comparable to the maximum speed of pitch change observed in a number of existing studies on real speech. This suggests that the role of physiological constraints in determining the shape and alignment of F_0 contours in speech is probably more important than has been appreciated. While it is likely that very often articulatory movements do not reach maximum speed due to speakers' choice of not using excessive effort, in many other occasions, as the new data show, the absolute limit may indeed have been approached or even reached. This means that, for pitch contour production at least, absolute articulatory limits, just as the economy of efforts as suggested by Lindblom (1982), probably constitute a major articulatory constraint that helps to shape the acoustic signal in speech.

Our data also demonstrate more clearly than before the linear relations between the *size* and *speed* and between *size* and *peak velocity* of pitch change. Similar linear relations have been found between peak velocity and movement amplitude in both speech and nonspeech movements (Hertrich and Ackermann, 1997; Löfqvist and Gracco, 1997). This suggests that F_0 movements are probably quite similar to other speech and nonspeech movements. The linear relation between the size and speed of pitch movement also suggests that it is imperative to know the pitch change interval when determining whether the maximum speed of pitch change is approached in a given case. Linear regression coefficients displayed in Tables V–VII can be used in future studies as references for determining for each pitch variation pattern in a language, how much of it is explainable in terms of articulatory constraints, and how much of it has to be explained by other mechanisms such as language specific phonological rules. The method of inducing rapid pitch undulation patterns from human subjects as used in the present study can also be employed in future studies to test various speaker groups as well as individual speakers to see whether and how their speech patterns are related to their idiosyncratic limits on the speed of pitch change.

Finally, although our findings are about the speed of

changing pitch, they also raise the question of whether the maximum speed of other articulatory movements is also approached in speech more frequently than has been recognized. Future efforts in finding answers to this question may help further our understanding of speech production and speech perception in general.

ACKNOWLEDGMENTS

This work is supported in part by NIH Grant No. DC03902. We would like to thank Bob Ladd, two anonymous reviewers, and Yiya Chen for their valuable comments and suggestions on earlier versions of this paper.

[1]No information was given as to what vowel(s) the subjects used to sing the pitch notes in the study.

[2]One reviewer cautioned us that English may actually produce faster pitch change than Chinese, because prominence is realized in English as F_0 contrast between adjacent syllables. As found in Xu (1999), however, this kind of contrast is used just as readily in Mandarin. In any case, as it turned out, we found no significant difference in speed of pitch change between speakers of these two languages.

[3]We thank Paul Boersma and the Praat project at Institute of Phonetic Sciences, University of Amsterdam, for making their program freely available to speech researchers.

[4]As described in Xu (1999), the trimming algorithm removes only local bumps in an F_0 curve involving adjacent vocal periods. As shown in Fig. 2 in Xu (1999), the ''trimming'' makes sure that sharp local bumps are not mistaken as true extreme pitch points by subsequent algorithms that locate peaks and valleys in an F_0 curve.

[5]We also considered the ERB scale (Hermes and van Gestel, 1991; Hermes and Rump, 1994), but did not find it appropriate for our purpose, because it reduces the speaker differences too heavily.

[6]Although it is conceptually more appropriate to calculate response time using excursion size expressed in semitone, numerically the difference between the two ways of measurement is quite small. Besides, it is critical that our data are in a form that is directly comparable to that of the previous studies.

[7]We also examined the possible contribution of musical training, but did not find any.

[8]It is also possible that it is the phonetic difference in [1] in the two languages that caused the difference in excursion time and response time. The American English [1] is known to be ''dark,'' i.e., with the back of the tongue as well as the tongue tip raised, whereas the Mandarin [1] is quite ''light.'' The back of the tongue is bulkier than the tip of the tongue. Raising it may require more time and it may have lengthened the entire syllable.

[9]We are not suggesting, however, that there are clear boundaries between the phases. Rather, the F_0 contour throughout a pitch shift is continuous, and the division among the three phases is conceptual and arbitrary.

[10]The same can also be seen in the rise data in Table 3 in Caspers and van Heuven (1993).

Akinlabi, A., and Liberman, M. (**1995**). ''On the phonetic interpretation of the Yoruba tonal system,'' Proceedings of The 13th International Congress of Phonetic Sciences, Stockholm, pp. 42–45.

Arvaniti, A., Ladd, D. R., and Mennen, I. (**1998**). ''Stability of tonal alignment: The case of Greek prenuclear accents,'' J. Phonetics **36**, 3–25.

Caspers, J., and van Heuven, V. J. (**1993**). ''Effects of time pressure on the phonetic realization of the Dutch accent-lending pitch rise and fall,'' Phonetica **50**, 161–171.

Dejonckere, P. H., Hirano, M., and Sundberg, J. (**1995**). *Vibrato* (Singular, San Diego).

Erickson, D. (**1993**). ''Laryngeal muscle activity in connection with Thai tones,'' Ann. Bull. Res. Inst. Logoped. Phoniatr. **27**, 135–149.

Fairbanks, G. (**1959**). *Voice and Articulation Drillbook* (Harper & Row, New York).

Goldsmith, J. A. (**1990**). *Autosegmental and Metrical Phonology* (Blackwell, Oxford).

Hallé, P. A. (**1994**). ''Evidence for tone-specific activity of the sternohyoid muscle in Modern Standard Chinese,'' Lang. Speech **37**, 103–123.

Hermes, D., and van Gestel, J. (**1991**). ''The frequency scale of speech intonation,'' J. Acoust. Soc. Am. **90**, 97–102.

Hermes, D. J., and Rump, H. H. (**1994**). ''Perception of prominence in speech intonation induced by rising and falling pitch movements,'' J. Acoust. Soc. Am. **96**, 83–92.

Hertrich, I., and Ackermann, H. (**1997**). ''Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures,'' J. Acoust. Soc. Am. **102**, 523–536.

Honda, K. (**1995**). ''Laryngeal and extra-laryngeal mechanisms of *F*0 control,'' in *Producing Speech: Contemporary Issues—for Katherine Safford Harris*, edited by F. Bell-Berti and L. Raphael (AIP, New York), pp. 215–245.

Hyman, L., and Schuh, R. (**1974**). ''Universals of tone rules,'' Ling. Inq. **5**, 81–115.

Kelso, J. A. S. (**1984**). ''Phase transitions and critical behavior in human bimanual coordination,'' Am. J. Physiol.: Regulatory, Intergrative Comp. **246**, R1000–R1004.

Kim, S.-A. (**1999**). ''Positional effect on tonal alternation in Chichewa: Phonological rule vs phonetic timing,'' Proceedings of Annual Meeting of Chicago Linguistic Society, Chicago, pp. 245–257.

Ladd, D. R., Faulkner, D., Faulkner, H., and Schepman, A. (**1999**). ''Constant 'segmental anchoring' of *F*0 movements under changes in speech rate,'' J. Acoust. Soc. Am. **106**, 1543–1554.

Ladd, D. R., Mennen, I., and Schepman, A. (**2000**). ''Phonological conditioning of peak alignment in rising pitch accents in Dutch,'' J. Acoust. Soc. Am. **107**, 2685–2696.

Laniran, Y. (**1992**). ''Intonation in Tone Languages: The Phonetic Implementation of Tones in Yorùbá,'' Ph.D. dissertation, Cornell University.

Lindblom, B. (**1982**). ''Economy of speech gestures,'' in *The Production of Speech*, edited by P. F. MacNeilage (Springer, New York), pp. 217–245.

Löfqvist, A., and Gracco, V. (**1997**). ''Lip and jaw kinematics in bilabial stop consonant production,'' J. Speech Hear. Res. **40**, 877–893.

Manfredi, V. (**1993**). ''Spreading and downstep: Prosodic government in tone languages,'' in *The Phonology of Tone*, edited by H. v. d. Hulst and K. Snider (Mouton de Gruyter, New York), pp. 133–184.

Ohala, J. J. (**1978**). ''Production of tone,'' in *Tone: A Linguistic Survey*, edited by V. A. Fromkin (Academic, New York), pp. 5–39.

Ohala, J. J., and Ewan, W. G. (**1973**). ''Speed of pitch change,'' J. Acoust. Soc. Am. **53**, 345(A).

Prame, E. (**1994**). ''Measurements of the vibrato rate of ten singers,'' J. Acoust. Soc. Am. **96**, 1979–1984.

Prieto, P., v. Santen, J., and Hirshberg, J. (**1995**). ''Tonal alignment patterns in Spanish,'' J. Phonetics **23**, 429–451.

Schmidt, R. C., Carello, C., and Turvey, M. T. (**1990**). ''Phase transitions and critical fluctuations in the visual coordination of rhythmic movements between people,'' J. Exp. Psychol. Hum. Percept. Perform. **16**, 227–247.

Sundberg, J. (**1979**). ''Maximum speed of pitch changes in singers and untrained subjects,'' J. Phonetics **7**, 71–79.

'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).

Xu, Y. (**1994**). ''Production and perception of coarticulated tones,'' J. Acoust. Soc. Am. **95**, 2240–2253.

Xu, Y. (**1997**). ''Contextual tonal variations in Mandarin,'' J. Phonetics **25**, 61–83.

Xu, Y. (**1998**). ''Consistency of tone-syllable alignment across different syllable structures and speaking rates,'' Phonetica **55**, 179–203.

Xu, Y. (**1999**). ''Effects of tone and focus on the formation and alignment of F_0 contours,'' J. Phonetics **27**, 55–105.

Xu, Y. (**2001**). ''Fundamental frequency peak delay in Mandarin,'' Phonetica **58**, 26–52.

Xu, Y. (in press). ''Sources of tonal variations in connected speech,'' J. Chin. Ling.

Xu, Y., and Sun, X. (**2000**). ''How fast can we really change pitch? Maximum speed of pitch change revisited,'' Proceedings of The 6th International Conference on Spoken Language Processing, Beijing, pp. 666–669.

Xu, Y., and Wang, Q. E. (**2001**). ''Pitch targets and their realization: Evidence from Mandarin Chinese,'' Speech Commun. **33**, 319–337.

Electromagnetic articulograph based on a nonparametric representation of the magnetic field

Tokihiko Kaburagi
Department of Acoustic Design, Kyushu Institute of Design, 4-9-1, Shiobaru, Minami-ku, Fukuoka 815-8540, Japan and CREST, Japan Science and Technology Corporation, Japan

Masaaki Honda
Human and Information Science Laboratory, NTT Communication Science Laboratories, 3-1, Morinosato-Wakamiya, Atsugi, Kanagawa 243-0198, Japan and CREST, Japan Science and Technology Corporation, Japan

(Received 20 August 2001; accepted for publication 27 November 2001)

Electromagnetic articulograph (EMA) devices are capable of measuring movements of the articulatory organs inside and outside the vocal tract with fine spatial and temporal resolutions, thus providing useful articulatory data for investigating the speech production process. The position of the receiver coil is detected in the EMA device on the basis of a field function representing the spatial pattern of the magnetic field in relation to the relative positions of the transmitter and receiver coils. Therefore, the design and calibration of the field function are quite important and influence the accuracy of position detection. This paper presents a nonparametric method for representing the magnetic field, and also describes a method for determining the receiver position from the strength of the induced signal in the receiver coil. The field pattern in this method is expressed by using a multivariate spline as a function of the position in the device's coordinate system. Because of the piecewise property of the basis functions and the freedom in the selection of the rank and the number of the basis functions, the spline function has a superior ability to flexibly and accurately represent the field pattern, even when it suffers from fluctuations caused by the interference between the transmitting channels. The position of the receiver coil is determined by minimizing the difference between the measured strength of the received signal and the predicted one from the spline representation of the magnetic field. Experimental results show that the error in estimating the receiver position is less than 0.1 mm for a 14×14-cm measurement area, and this error can be further reduced by using a spline-smoothing technique. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1445785]

PACS numbers: 43.70.Aj, 43.70.Jt [AL]

I. INTRODUCTION

By observing the dynamically changing behavior of the speech articulators, we can deepen our understanding of the motor process in the production of speech utterances. Electromagnetic articulograph (EMA) devices have been studied as tools for providing useful information about the position and movement of fixed points inside and outside the vocal tract with a high degree of spatial and temporal resolution (Hixon, 1971; Sonoda, 1974; Schönle *et al.*, 1987; Schönle *et al.*, 1989; Tuller *et al.*, 1990; Perkell *et al.*, 1992; Kaburagi and Honda, 1994). Because the EMA device uses static or alternating magnetic fields, it is much less invasive than x-ray techniques (Fujimura *et al.*, 1973; Kiritani *et al.*, 1975) to obtain a large amount of articulatory data. Although the problem of invasiveness to the subject can be completely avoided with ultrasonic imaging techniques (Morrish *et al.*, 1984; Shawker *et al.*, 1985; Stone *et al.*, 1988; Stone, 1990; Stone and Lundberg, 1996), such techniques are limited by their relatively slow frame rate. The EMA device is also effective for studying dynamic models of articulatory movements (Kaburagi and Honda, 2001), articulatory-based speech synthesis (Kaburagi and Honda, 1998), and acoustic-to-articulatory inversion problems (Suzuki *et al.*, 1998).

The two-dimensional EMA device with alternating magnetic fields (Schönle *et al.*, 1987) uses three transmitter coils driven by currents of different frequencies. It simultaneously measures the electromagnetically induced currents in multiple receiver coils attached to the articulators and detects their positions in the midsagittal measurement plane. To obtain spatial information from the received signal, the spatial pattern of the magnetic field is represented as a voltage-to-distance function (VDF) incorporating the distance of a given receiver from any one of the three transmitter coils. Schönle *et al.* (1987) proposed a VDF in which the received signal was inversely proportional to the transmitter–receiver distance cubed. The position of the receiver coil is then determined as the crossing point of three circles, the center and radius of each given by the transmitter position and corresponding transmitter–receiver distance estimated using the VDF from the received signal.

On the other hand, Perkell *et al.* (1992) expressed the exponent part of the VDF as a polynomial function of the transmitter–receiver distance, and also described the importance of calibrating the VDF parameters in the local region in which the receiver coil would be located during the experiment. It has been recognized that the parameter calibration is of importance as well as the field representation problem

 0001-4966/2002/111(3)/1414/8/$19.00 © 2002 Acoustical Society of America

(Gracco and Nye, 1993; Hoole, 1993), and Kaburagi and Honda (1997) extended the idea of the local calibration to propose an adaptive method of the parameter calibration. This method was performed by selecting calibration data samples neighboring the receiver position during the experiment, which was *a priori* unknown, resulting in a higher measurement accuracy.

The effectiveness of local and adaptive calibration methods indicates that the global pattern of the actual magnetic field deviates from the theoretical field based on the assumption that a transmitter coil functions as a dipole source, and hence a VDF with fixed parameters only holds for a restricted region. A possible source of this pattern deviation is interference among the transmitting channels of the EMA device, which would affect the strength of the magnetic fields with each other. For example, the transmitter coils could be mutually inductive. Then, for any given transmitter coil, currents would be induced by neighboring transmitters so that the resulting reproduction of magnetic field attenuates the surrounding fields of the corresponding frequencies.

In this paper, we present a novel approach for representing the spatial pattern of the magnetic field and describe how the receiver position can be calculated using the proposed field function. Instead of using the voltage-to-distance function, the field pattern in our method is expressed by multivariate B-spline functions (Schumaker, 1981; Dierckx, 1993), which can depict a smooth curved surface as a linear combination of piecewise basis functions. The position and orientation of the receiver coil is then determined by minimizing the difference between the measured and predicted signal strengths.

The spline function has the following advantageous characteristics when used to represent the magnetic field. First, each basis function is defined locally and there are freedoms in the selection of the rank and the number of the basis functions. These flexibilities allow the spline function to accurately represent the desired curved surface. Second, the field pattern is expressed as a function of the x and y axes of the device, whereas the VDF only takes the transmittance–receiver distance into account. Therefore, our method explicitly accommodates the field function to the actual magnetic field even when it has location-dependent fluctuations. Third, the spline function provides a closed-form representation of the entire field pattern. Therefore, the proposed method circumvents the computationally inefficient pattern matching used in the adaptive calibration method to discover the local calibration data (Kaburagi and Honda, 1997). In calculating the receiver position, the optimal values of three unknown parameters, i.e., the positional variables x and y and the orientation θ, are simultaneously determined so that the prediction error with respect to the receiver signal is minimized. This problem becomes nonlinear, because the basis functions of the splines are piecewise polynomial functions. Therefore, a procedure based on the Gauss–Newton method is used to solve the nonlinear optimization problem.

This paper is organized as follows. Section II explains the spline-based representation of the magnetic field, and Sec. III presents the method for estimating the receiver position from the strength of the induced signal. Experimental results are then presented in Sec. IV to show the accuracy of the proposed method in representing the magnetic field and estimating the receiver position. The effect of smoothing the calibration data samples is also considered. Finally, Sec. V summarizes this work and gives our conclusions.

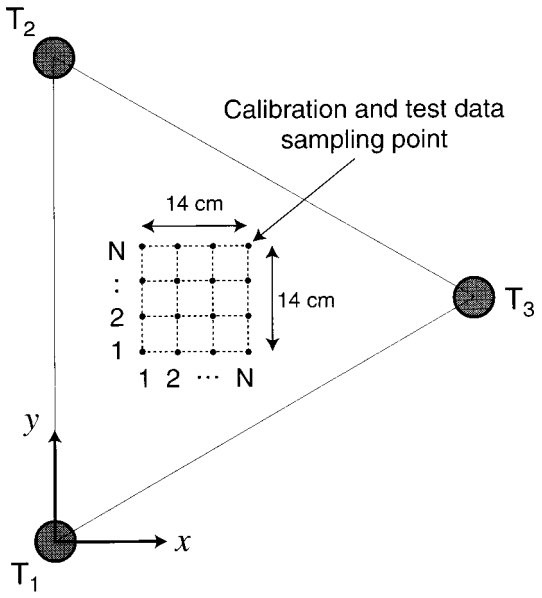


FIG. 1. Illustration of the spatial distribution of three transmitter coils and the rectangular measurement area specified for the observation of the articulatory organs. Calibration and test data samples are taken by using the receiver coil placed at known positions, i.e., the crossing points of the grid drawn to cover the entire measurement area, and by measuring the strength of the induced signal. The number of the data samples for each side of the rectangular is denoted as N.

II. MAGNETIC FIELD REPRESENTATION

This section explains the coordinate system and measurement area of the EMA device, and presents the method for representing the spatial pattern of the magnetic field using the multivariate spline functions. Then, it describes the calibration method to determine the values of weighing parameters included in the spline function. The explanations of the spline function given in this section are based on Schumaker (1981) and Dierckx (1993).

A. Hardware system

Figure 1 illustrates the two-dimensional EMA device used in this study (Carstens Articulograph AG100, Germany), which has three transmitter coils positioned at vertices T_1, T_2, and T_3 of a regular triangle with side lengths of 64.18 cm. They are driven at carrier frequencies of 12, 12.5, and 13 kHz, respectively, and generate alternating magnetic fields. The measurement area of articulatory movements is set inside triangle $T_1T_2T_3$ as a 14×14-cm rectangle so that it covers the entire range of measurement points on the articulators, i.e., jaw, lips, tongue, soft palate, and fixed maxillary references on the noise bridge and upper incisors. The position of the receiver coil is measured in a coordinate system

fixed to the transmitter coils. Note that the axes of the transmitter coils are parallel to one another and perpendicular to the measurement plane.

B. Spline-based representation of the magnetic field

The strength of the electromagnetically induced signal in the receiver coil can be expressed as $e_l = v_l(x,y)\cos\theta$, where $v_l(x,y)$ is the field function representing the strength of the magnetic field, and θ is the tilt angle of the receiver coil relative to the magnetic flux, identical for every transmitter channel. The induced signal is then reduced by a factor of $\cos\theta$. An index to the transmitters is represented as l.

We express the logarithm of the field function $v(x,y)$ using a multivariate B-spline function such that

$$\log\{v(x,y)\} = \sum_{p=1}^{P}\sum_{q=1}^{Q} c_{pq}N_p(x)N_q(y), \tag{1}$$

where $N_p(x)$ and $N_q(y)$, respectively, represent piecewise polynomial functions, the basis functions of the spline representation, with respect to the x- and y-axes. The rank of these basis functions is assumed to be m. P and Q indicate the number of basis functions to represent the entire field pattern, and c_{pq} denotes the weighting parameter in summing combinations of the basis functions.

Equation (1) forms a curved surface over the $x-y$ plane as the tensor product of the basis functions. For given data samples $v_{ij} = v(x_i, y_j)$ $(i = 1,2,...,N;\ j = 1,2,...,N)$, the weighting parameters c_{pq} can be determined so that the curved surface interpolates these samples. Another possible selection of the weighting parameters is the spline smoothing in which the curved surface is made to approximate the data samples by minimizing an error criterion. The basis function takes zero outside a certain interval along the axis. By virtue of this piecewise property, the spline function generally provides a good result when used for interpolating or smoothing given data samples and reproducing the overall shape of a curved surface. In addition, the smoothness of the surface can be controlled by parameter m, which determines the polynomial order of the basis function. From these properties, the spline function is applicable for a fine representation of the magnetic field.

C. Construction of the basis functions

By setting the number and the positions of the internal nodes, the basis functions $N_p(x)$ and $N_q(y)$ are uniquely determined regardless of the data samples to be interpolated or smoothed. Internal nodes of the two-variate spline function [Eq. (1)] can be arranged in a rectangular region $R = [a,b]\times[c,d]$ as follows:

$$a = \xi_{1-m} = \cdots = \xi_{-1} = \xi_0 < \xi_1 < \cdots < \xi_p < \cdots < \xi_{P-m+1} = \xi_{P-m+2} = \cdots = \xi_{P-1} = \xi_P = b,$$

and

$$c = \zeta_{1-m} = \cdots = \zeta_{-1} = \zeta_0 < \zeta_1 < \cdots < \zeta_q < \cdots < \zeta_{Q-m+1} = \zeta_{Q-m+2} = \cdots = \zeta_{Q-1} = \zeta_Q = d,$$

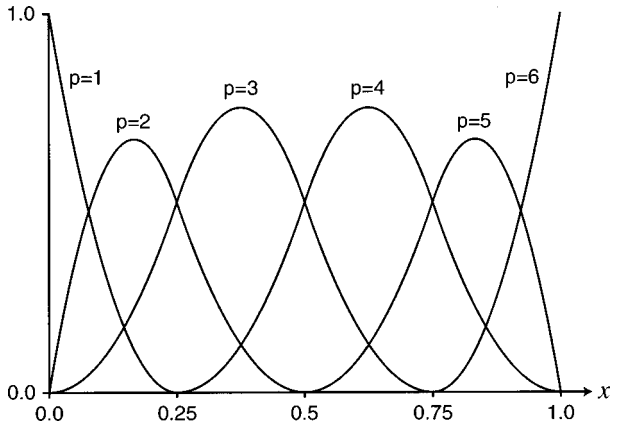


FIG. 2. Illustration of B-spline basis functions with the rank of three constructed by assigning nine internal nodes along the axis. Three nodes are overlapped on both sides and the number of nonoverlapping nodes is 3. The values of basis functions were computed at 256 axis positions within the interval from zero to one, by using de Boor's algorithm (de Boor, 1972) instead of directly evaluating divided differences of truncated power functions. The traces were then drawn for each basis function by linearly interpolating the obtained values.

where ξ_p and ζ_q represent the positions of the nodes along the x- and y-axes, respectively. Note that m nodes overlap on both sides, and the number of nonoverlapping nodes (n) equals $P-m$ and $Q-m$.

Then, the B-spline basis function can be constructed from the positions of successive $m+1$ nodes as

$$N_p(x) = (\xi_p - \xi_{p-m})M_p(x;\xi_{p-m},\ldots,\xi_p),$$

and

$$N_q(y) = (\zeta_q - \zeta_{q-m})M_q(y;\zeta_{q-m},\ldots,\zeta_q)$$

for $1 \leqslant p \leqslant P$ and $1 \leqslant q \leqslant Q$, where $M_p(x;\xi_{p-m},\ldots,\xi_p)$ and $M_q(y;\zeta_{q-m},\ldots,\zeta_q)$ are divided difference of truncated power function $M(x;z) = (x-z)_+^{m-1}$ with respect to $z = \xi_{p-m},\ldots,\xi_p$ and $z = \zeta_{q-m},\ldots,\zeta_q$, respectively. These basis functions have the following piecewise properties:

$$N_p(x)\begin{cases} >0, & \xi_{p-m} < x < \xi_p \\ =0, & \text{otherwise,} \end{cases}$$

and

$$N_q(y)\begin{cases} >0, & \zeta_{q-m} < y < \zeta_q \\ =0, & \text{otherwise,} \end{cases}$$

indicating that the basic functions takes a nonzero value only for a limited interval along the axis. The values of these basis functions can be effectively evaluated for a specific value of x or y using de Boor's iterative algorithm (de Boor, 1972). Figure 2 illustrates examples of the B-spline basis functions $N_p(x)$, where the rank is set at three and internal nodes are placed at 0.0, 0.25, 0.5, 0.75, and 1.0 along the x axis. Three nodes are overlapped on both sides at 0.0 and 1.0, and the number of nonoverlapping nodes is three. Six basis functions are constructed from this node assignment.

D. Determination of the weighting parameters

The weighting parameters in Eq. (1) can be determined so that the spline function forms a curved surface interpolat-

ing or smoothing a given set of calibration data samples. The field strength in a given location is obtained by positioning a receiver coil parallel to the magnetic field and then measuring the induced signal. For the purpose of obtaining a spline representation of the magnetic field, the calibration data are measured at the points of intersection of lines drawn at equal intervals in the horizontal and vertical directions, as shown in Fig. 1. Every calibration datum stores the measured field strength v_{ij} sampled at the position (x_i, y_j) for $1 \leqslant i \leqslant N$ and $1 \leqslant j \leqslant N$, where N gives the number of samples in each axis.

The equations specifying the spline interpolation condition can be derived by substituting the calibration data samples into Eq. (1) as

$$\log v_{ij} = \sum_{p=1}^{P} \sum_{q=1}^{Q} c_{pq} N_p(x_i) N_q(y_j). \tag{2}$$

The values of the unknown parameters c_{pq} are calculated by solving these $N \times N$ linear equations simultaneously. $N = m + n$ must be satisfied for the number of nodes (n) and the rank of the basis functions (m) so that the number of equations is coincident with that of the unknown parameters. On the other hand, the weights in the spline smoothing are determined under the condition $N > m + n$ so that a criterion C, the total square error between the measured and predicted field strengths, is minimized, where

$$C = \sum_{i=1}^{N} \sum_{j=1}^{N} \left\{ \log v_{ij} - \sum_{p=1}^{P} \sum_{q=1}^{Q} c_{pq} N_p(x_i) N_q(y_j) \right\}^2. \tag{3}$$

If the partial derivative of C with respect to c_{pq} is set to zero, the resulting normal equation can be used to determine the optimal values of the weighting parameters.

The procedure for forming the spline representation can be summarized as follows. For the spatial alignment of the internal nodes, the rectangular region R is taken to coincide with the measurement area, and the node position is selected at the intersecting points of grids with equal intervals, just like the sampling points of the calibration data. The basis functions with respect to x- and y axes are then constructed for given node positions under the condition $P = Q = n + m$. Finally, as described above, weights used in summing the basis functions are calculated from the two-dimensional calibration data depending on the interpolating or smoothing condition.

III. ESTIMATION METHOD OF THE RECEIVER POSITION

A. An error criterion of the signal prediction

This section describes the method for estimating the position of the receiver coil. The induced signal in the receiver coil is first separated in the frequency domain to obtain each component of the transmitting channels, which is expressed as e_l for $l = 1$, 2, and 3 below. On the other hand, the strength of the induced signal can be predicted as a function of the three unknown variables, positional variables x and y and the tilt angle θ, as

$$\hat{e}_l(x, y, \theta) = v(x, y) \cos \theta, \tag{4}$$

where the logarithm of the field strength is represented by the spline function as described in the previous section [Eq. (1)].

Therefore, the values of these variables can be determined so that the difference between the measured and predicted signal strengths is minimized. The criterion expressing the prediction error of the induced signal is defined as

$$S(x, y, \theta) = \sum_{l=1}^{3} \{\log e_l - \log \hat{e}_l\}^2. \tag{5}$$

This optimization problem becomes nonlinear since the magnetic field is represented by the piecewise polynomial functions, and we employ an iterative procedure based on the Gauss–Newton method (Dennis and Schnabel, 1983). The position of the receiver coil is estimated as the solution to this problem.

B. Derivation of the optimal solution

The values of the unknown parameters are incrementally optimized based on the Gauss–Newton method so that the prediction error S is minimized. This iterative step can be written as

$$\mathbf{x}^{k+1} = \mathbf{x}^k + \alpha \Delta \mathbf{x}^k, \tag{6}$$

where vectors $\mathbf{x}$ and $\Delta \mathbf{x}$, respectively, represent the unknown parameters and their updating values

$$\mathbf{x} = (x, y, \theta)^t, \tag{7}$$

and

$$\Delta \mathbf{x} = (\Delta x, \Delta y, \Delta \theta)^t. \tag{8}$$

Here, α is a reduction parameter $(0 < \alpha \leqslant 1)$, k indicates the index of iterations, and t denotes the transposition. The values of the updating vector $\Delta \mathbf{x}$ can be determined by solving the normal equation

$$A^t A \Delta \mathbf{x} = A^t (\mathbf{e} - \hat{\mathbf{e}}), \tag{9}$$

where A is the Jacobian matrix representing the sensitivity of the unknown parameters over the predicted signals

$$A = \begin{pmatrix} \partial \log \hat{e}_1 / \partial x & \partial \log \hat{e}_1 / \partial y & \partial \log \hat{e}_1 / \partial \theta \\ \partial \log \hat{e}_2 / \partial x & \partial \log \hat{e}_2 / \partial y & \partial \log \hat{e}_2 / \partial \theta \\ \partial \log \hat{e}_3 / \partial x & \partial \log \hat{e}_3 / \partial y & \partial \log \hat{e}_3 / \partial \theta \end{pmatrix}. \tag{10}$$

Vectors $\mathbf{e}$ and $\hat{\mathbf{e}}$, respectively, represent measured and predicted logarithmic signals as

$$\mathbf{e} = (\log e_1, \log e_2, \log e_3)^t, \tag{11}$$

and

$$\hat{\mathbf{e}} = (\log \hat{e}_1, \log \hat{e}_2, \log \hat{e}_3)^t. \tag{12}$$

The procedure for determining the optimal values of the unknown parameters can be summarized as follows:

(1) Given the measured signal $\mathbf{e}$ and the initial values of the unknown parameters $\mathbf{x}^0$, set the counter k at zero.
(2) Calculate the predicted signals $\hat{\mathbf{e}}$ and the value of each component of the Jacobian matrix A for $\mathbf{x}^k$. Then, solve the normal equation [Eq. (9)].

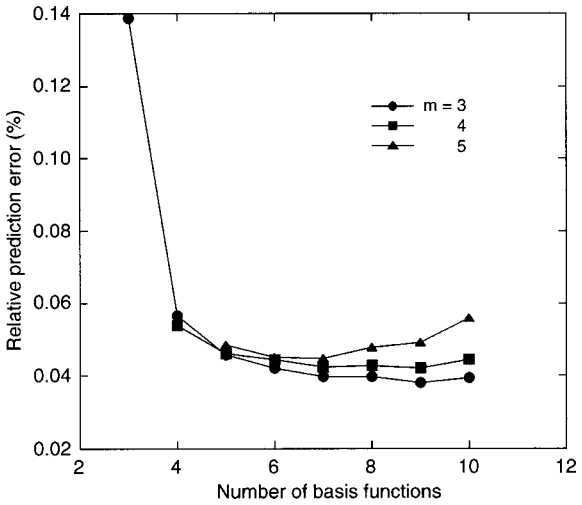


FIG. 3. Prediction error of the magnetic field strength as a function of the number of basis functions of the spline representation. The experimental condition specifying the rank of the basis function is denoted as *m*. The number of internal nodes and the weights are determined so that the spline function interpolates each set of calibration data.

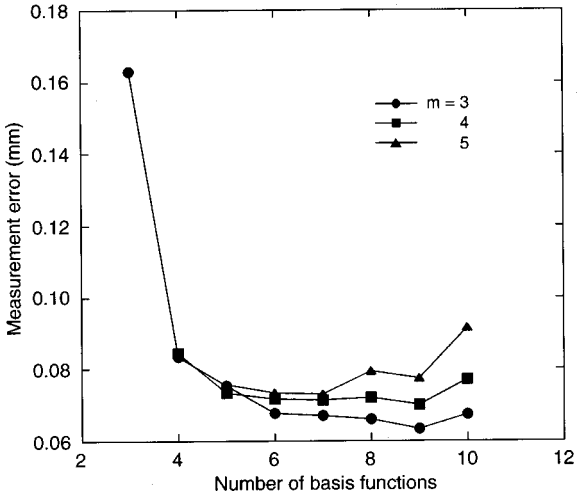


FIG. 4. Prediction error of the receiver position as a function of the number of basis functions. The rank of the basis function is denoted as *m*. The number of internal nodes and the weights are determined based on the interpolation condition.

(3) Update the values of the unknown parameters [Eq. (6)].
(4) When a convergence is obtained, quit the procedure. Otherwise, set the counter as $k=k+1$ and repeat from the second step.

In computing components of the Jacobian matrix, the values of the basis functions and their partial derivatives $\partial N_p(x)/\partial x$ and $\partial N_q(y)/\partial y$, should be evaluated for a specific value of $\mathbf{x}^k$. These calculations can be effectively performed based on de Boor's incremental algorithm (de Boor, 1972). In our study, the reduction parameter in Eq. (6) is expressed as $\alpha=0.1L$, where L is an integer ranging from 1 to 10. L is selected, at each iteration, so that the prediction error of the received signal [Eq. (5)] is minimized.

IV. EXPERIMENTAL RESULTS

This section presents the results of experiments conducted to examine the accuracy of the proposed method in representing the magnetic field and estimating the receiver position. To perform the experiments, samples of calibration and test data were measured, as shown in Fig. 1, at equally spaced lattice points in the measurement area. The sample number N of calibration data was selected in the range from 3 to 10 in eight steps. The sample number was fixed at 15 for the test data, so that the interval between adjacent sampling points was 1 cm, and three sets were measured. Note that the rotational angle of the receiver coil was set at zero both for calibration and test samples. Results are first shown for experiments in which a spline interpolation is used to represent the field pattern. Next, the effect of smoothing the calibration data samples is examined, and the convergence of the iterative procedure in predicting the receiver position is presented.

A. Accuracy for representing the magnetic field

Experimental results plotted in Fig. 3 show the error between the predicted and actual field strengths as a function of the number of the spline basis functions used to represent the magnetic field. The rank of the basis function (m) was set at three, four, and five, and the weights were determined so that the spline function formed an interpolated curved surface for each set of calibration data. Because the number of calibration data (N) must be equivalent to that of basis functions (P and Q), the number of nonoverlapping nodes (n) was set as $n=N-m$ for specified N and m. Then, the known position of each test sample (x_i,y_j) was substituted into the spline function, and the relative error between the predicted $v(x_i,y_j)$ and measured v_{ij} signal strengths was evaluated as $100\cdot|v(x_i,y_j)-v_{ij}|/v_{ij}(\%)$. The ordinate of the figure represents the mean error of three trials, where errors for 15×15 test samples were also averaged in each trial.

The figure clearly indicates that the error is less than 0.06% when the number of basis functions is greater than 3. The influence of the rank is relatively small, but an increase of the error is observed for $m=5$ when the sample number is greater than 7. The minimum error (0.04%) is obtained when P and Q are greater than 6 for $m=3$. These results suggest that the logarithm of the magnetic field is well represented by polynomial basis functions of the second order.

B. Accuracy in predicting the receiver position

Figure 4 shows the prediction accuracy of the receiver position. The spline interpolation of the calibration data was used to express the magnetic field as in the previous experiment, the receiver position (x,y) was calculated from the signal strength of each test sample, and the prediction error of the receiver position was evaluated as $\sqrt{(x-x_i)^2+(y-y_j)^2}$ for the actual position (x_i,y_j) included in the test sample. The iteration number of the position prediction procedure was fixed at 10. The mean of three trials is plotted along the ordinate.

The results indicate that the prediction error is less than 0.1 mm when the number of basis functions is more than 3. The influence of the rank of the basis functions is more apparent than in the previous experiment: the prediction may

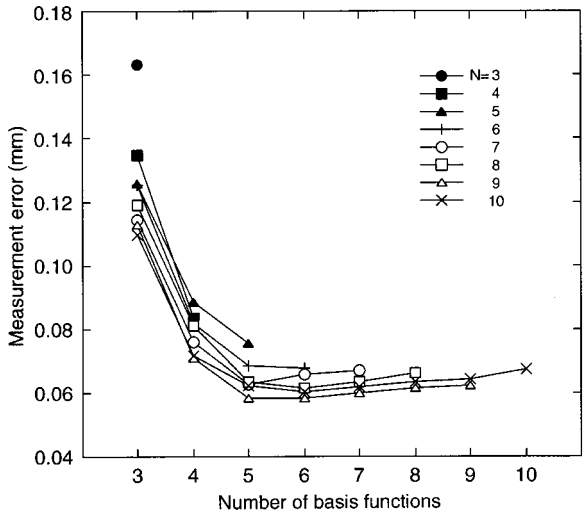


FIG. 5. Mean prediction error of the receiver position as a function of the number of basis functions. The number of calibration data samples along each axis is denoted as *N*. The number of internal nodes and the weights are determined so that the spline function smoothly approximates a given set of calibration data under the square-error criterion, while the rank of the basis function is fixed at 3.

be better performed for $m=3$ if the number of basis functions is the same. When *m* is set at 3 and *P* and *Q* are greater than 5, the error is less than 0.07 mm. The errors shown in Figs. 3 and 4 appear to have a similar tendency and indicate that an accurate representation of the magnetic field can result in precise *x* and *y* measurements.

C. Effect of smoothing the calibration data

The weights of the spline representation can be determined, as described in Sec. II, so that given calibration data samples are smoothly approximated by the spline function. Estimation errors of the receiver position shown in Figs. 5 and 6 were obtained by using the smoothing spline technique in representing the magnetic field. The figures, respectively, indicate mean and maximum errors of the test data samples, as a function of the number of basis functions (*P* and *Q*). The

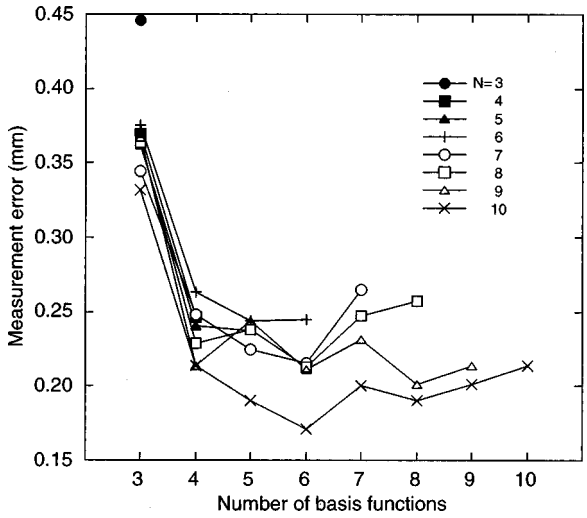


FIG. 6. Maximum prediction error of the receiver position as a function of the number of basis functions. Other experimental conditions are the same as described in Fig. 5.

rank (m) was fixed at 3, and the plotting symbols in these figures represent different numbers of calibration data samples (N), which ranged from 3 to 10. In addition, the symbols representing any given value of *N* are connected by straight lines. Together with the condition $P=Q=n+m$, $N=P=Q$ holds for the spline interpolation and $N>P$ and $N>Q$ for the smoothing. Therefore, the extreme right-hand symbol in a connected sequence corresponds to the interpolating condition, while the other symbols in the sequence correspond to smoothing.

It is clear from Fig. 5 that the overall shape of every sequence of symbols shows the same tendency regardless of the sample number of the calibration data: the minimum error is obtained when five or six basis functions are used. In addition, the prediction error generally decreases when the calibration sample number increases, even when the number of basis functions is the same, suggesting that the data-smoothing technique can reduce the influence of random error components that occasionally arise during the measurement of calibration data. By combining both results, and using eight to ten samples of calibration data in each axis, a minimum error of approximately 0.06 mm is obtained. Figure 6 indicates the maximum prediction error in the measurement area calculated using 225 test samples. The result shows almost the same tendency as the previous one with respect to the number of calibration data samples and basis functions. It can be concluded from the figure that the maximum error is below 0.25 mm if the number of basis functions is set at 5 or 6.

D. Other considerations

Because the basis function of the spline representation is determined by the spatial alignment of the internal nodes, the alignment pattern may influence the accuracy of the proposed method. An experiment was performed in which the optimal node alignment was found by a search procedure, such that the mean estimation error of the receiver position was minimized. The rank and the number of the basis function was selected as 3 and 5, respectively, and spline smoothing was applied to the calibration data. The number of non-overlapping nodes, for which the search method was applied, was set at 2, and the experiment was repeated while changing the number of calibration data from 5 to 10. The results showed that the difference between the prediction errors with and without optimization is very small (approximately 0.003 mm), suggesting that the internal nodes can be placed at equal intervals.

Finally, Fig. 7 shows a plot of the receiver position prediction error as a function of the number of iterations. It is clear from the figure that the Gauss–Newton method converges rapidly and that three or four iterations are sufficient to reach a minimum prediction error.

V. SUMMARY AND CONCLUSIONS

The representation of the magnetic field is one of the central issues in constructing an electromagnetic position sensing device. Commonly, the voltage-to-distance function has been derived from the assumption that each transmission

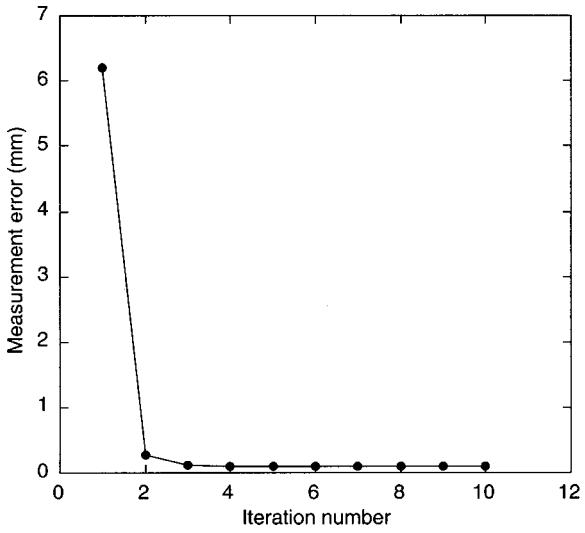


FIG. 7. Prediction error of the receiver position as a function of the number of iterations. The rank and the number of the basis functions are, respectively, 3 and 7. The number of calibration data samples along each axis is 7. Spline interpolation is used.

coil behaves like an electromagnetic dipole (Raab *et al.*, 1979; Schönle *et al.*, 1987, 1989) or closely approximates dipole behavior (Perkell *et al.*, 1992). This paper presented a novel method based on a nonparametric approach. A two-variate function is constructed by a linear sum of piecewise polynomial functions, i.e., B-spline basis functions, to represent the magnetic field of the two-dimensional articulographic device. The piecewise property of the basis functions makes it possible to accurately represent the spatial field pattern, even when it has local fluctuations. In addition, the spline function can provide a closed-form representation, which is computationally superior to the adaptive method we proposed before (Kaburagi and Honda, 1997) for calibrating unknown parameters included in the field function.

Based on a spline representation of the magnetic field, it is possible to predict the strength of the received signal for three unknown parameters of the receiver coil, i.e., the position and the tilt relative to the magnetic field, while incorporating the influence of the tilt angle. The position of the receiver coil can be then determined by minimizing the difference between the measured and predicted signal strengths. This framework produces a nonlinear optimization problem, which is solved here using a Gauss–Newton-based iterative procedure.

Experiments were performed to examine the accuracy of the proposed method. A receiver coil was placed at each crossing point of a grid drawn in a 14×14-cm measurement area of the magnetic device, and the received signal was measured to obtain several calibration and test data sets. Calibration data were used to determine the values of the weighting coefficients of the spline function, while test data were used to compute errors in predicting the strength of the magnetic field and the position of the receiver coil.

The main experimental results and related considerations are as follows:

(1) Experiments showed that the relative error in predicting the magnetic field strength is less than 0.06%, when the spline function is constructed to interpolate four or more samples of calibration data in each axis. The influence of the rank of the basis function is relatively small, but the results are generally better at 3 than at 4 or 5.
(2) If the same number of calibration data samples are used as in (1), the error in predicting the receiver position is less than 0.1 mm. It was revealed that this error can be further reduced by using the data smoothing technique. The error generally decreases when the number of calibration data samples increases for the same number of basis functions. The optimal number of basis functions is 5, and the minimum prediction error (approximately 0.06 mm) can be obtained by using more than six calibration data samples along each axis.
(3) About the spatial alignment of the internal nodes, the experimental results indicated that nodes can be placed at points separated by equal intervals on each side of a rectangle that covers the entire measurement area.
(4) The iterative procedure was revealed to converge quickly. Three or four iterations are enough to obtain an accurate prediction of the receiver position.

Kaburagi and Honda (1997) showed that the measurement accuracy of an articulograph device, the same one used in this study, is 0.230 mm when the voltage-to-distance function is calibrated globally, and this error decreases to 0.106 mm when a local or adaptive calibration is performed. Experimental results presented in this paper indicate that the spline function is slightly superior to the ordinal field representation method when an equivalent number of calibration data samples is used: the measurement error decreases from 0.106 to 0.084 mm for 4×4 calibration samples.

However, the two-dimensional articulograph device still has the problem of off-sagittal misalignment and rotation of the receiver coil that causes significant measurement errors. An informal experiment indicated that the influence of the off-sagittal misalignment can be compensated if the rotation angle is zero: the increase of the measurement error is approximately 0.1 mm as long as the misalignment is less than plus or minus 6 mm. However, the error rapidly increases when the rotation angle increases. It reaches about 3 mm when the rotation angles with respect to the *x* and *y* axes are both 20 deg and the misalignment is 6 mm.

To overcome this problem, the proposed method should be extended to measure the receiver coil position in three-dimensional space as well as the rotational angle relative to the three-dimensional pattern of the magnetic field. The principle and architectural design of the three-dimensional articulograph have been extensively investigated (Zierdt, 1993; Tillmann *et al.*, 1996; Zierdt *et al.*, 1999, 2000), resulting in the construction of an actual measurement system (Carstens Articulograph AG500, Germany). The spline function is expected to be useful for representing the field pattern of the three-dimensional system as well, because the magnetic fields are generated using six transmitting channels and hence the interference problem might be more apparent than in two-dimensional systems. To investigate the three-dimensional field pattern representation, experimental results obtained in the present study would be useful to determine

the values of several parameters included in the construction of the spline representation.

ACKNOWLEDGMENT

This research was partly supported by the Grants for Incentive Research from the Nissan Science Foundation.

de Boor, C. (**1972**). ''On calculating with B-splines,'' J. Approx. Theory **6**, 50–62.

Dennis, J. E., and Schnabel, R. B. (**1983**). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations* (Prentice-Hall, Englewood Cliffs, NJ).

Dierckx, P. (**1993**). *Curve and Surface Fitting with Splines* (Oxford University Press).

Fujimura, O., Ishida, Y., and Kiritani, S. (**1973**). ''Computer-controlled radiography for observation of movements of articulatory and other human organs,'' Comput. Biol. Med. **3**, 371–384.

Gracco, V. L., and Nye, P. W. (**1993**). ''Magnetometry in speech articulation research: Some misadventures on the road to enlightenment,'' Proceedings of the ACCOR Workshop on Electromagnetic Articulography in Phonetic Research (FIPKM **31**, der Universität München), pp. 91–104.

Hixon, T. J. (**1971**). ''An electromagnetic method for transducing jaw movements during speech,'' J. Acoust. Soc. Am. **49**, 603–606.

Hoole, P. (**1993**). ''Methodological considerations in the use of electromagnetic articulatory in phonetic research,'' Proceedings of the ACCOR Workshop on Electromagnetic Articulography in Phonetic Research (FIPKM **31**, der Universität München), pp. 43–64.

Kaburagi, T., and Honda, M. (**1994**). ''Determination of sagittal tongue shape from the positions of points on the tongue surface,'' J. Acoust. Soc. Am. **96**, 1356–1366.

Kaburagi, T., and Honda, M. (**1997**). ''Calibration methods of voltage-to-distance function for an electro-magnetic articulometer (EMA) system,'' J. Acoust. Soc. Am. **101**, 2391–2394.

Kaburagi, T., and Honda, M. (**1998**). ''Determination of the vocal tract spectrum from the articulatory movements based on the search of an articulatory-acoustic database,'' Proceedings of the International Conference on Spoken Language Processing (ICSLP98), 433–436.

Kaburagi, T., and Honda, M. (**2001**). ''Dynamic articulatory model based on multidimensional invariant-feature task representation,'' J. Acoust. Soc. Am. **110**, 441–452.

Kiritani, S., Itoh, K., and Fujimura, O. (**1975**). ''Tongue-pellet tracking by a computer-controlled x-ray microbeam system,'' J. Acoust. Soc. Am. **57**, 1516–1520.

Morrish, K., Stone, M., Sonies, B., Kurtz, D., and Shawker, T. (**1984**). ''Characterization of tongue shape,'' Ultrason. Imaging **6**, 37–47.

Perkell, J. S., Cohen, M. H., Svirsky, M. A., Matthies, M. L., Garabieta, I., and Jackson, M. T. T. (**1992**). ''Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements,'' J. Acoust. Soc. Am. **92**, 3078–3096.

Raab, F. H., Blood, E. B., Steiner, T. O., and Jones, H. R. (**1979**). ''Magnetic position and orientation tracking system'' IEEE Trans. Aerosp. Electron. Syst. **AES-15**, 709–717.

Schönle, P. W., Gräbe, K., Wenig, P., Höhne, J., Schrader, J., and Conrad, B. (**1987**). ''Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract,'' Brain Lang. **31**, 26–35.

Schönle, P. W., Müller, C., and Wenig, P. (**1989**). ''Echtzeitanalyse von orofacialen Bewegungen mit Hilfe der elektromagnetischen Artikulographie,'' Biomed. Tech. **34**, 126–130.

Schumaker, L. L. (**1981**). *Spline Functions: Basic Theory* (Wiley-Interscience, New York).

Shawker, T. H., Stone, M., and Sonies, B. C. (**1985**). ''Tongue pellet tracking by ultrasound: Development of a reverberation pellet,'' J. Phonetics **13**, 135–146.

Sonoda, Y. (**1974**). ''Observation of tongue movement employing magnetometer sensor,'' IEEE Trans. Magn. **MAG-10**, 954–957.

Stone, M., Shawker, T. H., Talbot, T. L., and Rich, A. H. (**1988**). ''Cross-sectional tongue shape during the production of vowels,'' J. Acoust. Soc. Am. **83**, 1586–1596.

Stone, M. (**1990**). ''A three-dimensional model of tongue movement based on ultrasound and x-ray microbeam data,'' J. Acoust. Soc. Am. **87**, 2207–2217.

Stone, M., and Lundberg, A. (**1996**). ''Three-dimensional tongue surface shapes of English consonants and vowels,'' J. Acoust. Soc. Am. **99**, 3728–3737.

Suzuki, S., Okadome, T., and Honda, M. (**1998**). ''Determination of articulatory positions from speech acoustics by applying dynamic articulatory constraints,'' Proceedings of the International Conference on Spoken Language Processing (ICSLP98), 2251–2254, 1998.

Tillmann, H. G., Zierdt, A., and Hoole, P. (**1996**). ''Towards a three-dimensional articulographic system,'' J. Acoust. Soc. Am. **100**, 2662.

Tuller, B., Shao, S., and Kelso, J. A. S. (**1990**). ''An evaluation of an alternating magnetic field device for monitoring tongue movements,'' J. Acoust. Soc. Am. **88**, 674–679.

Zierdt, A. (**1993**). ''Problems of electromagnetic position transduction for a three-dimensional articulographic measurement system,'' Proceedings of the ACCOR Workshop on Electromagnetic Articulography in Phonetic Research (FIPKM **31**, der Universität München), pp. 137–142.

Zierdt, A., Hoole, P., and Tillmann, H. G. (**1999**). ''Development of a system for three-dimensional fleshpoint measurement of speech movements,'' Proceedings of the International Congress of Phonetic Sciences (ICPhS99), pp. 73–76.

Zierdt, A., Hoole, P., Honda, M., Kaburagi, T., and Tillmann, H. G. (**2000**). ''Extracting tongues from moving heads,'' Proceedings of the 5th Seminar on Speech Production, pp. 313–316.

Intensity-importance functions for bandlimited monosyllabic words

Gerald A. Studebaker[a] and Robert L. Sherbecoe
Memphis Speech and Hearing Center, The University of Memphis, 807 Jefferson Avenue, Memphis, Tennessee 38105

(Received 6 June 2001; accepted for publication 29 November 2001)

A study was carried out to determine the relative importance to speech intelligibility of different intensities within the speech dynamic range. The functions that were derived are analogous to previous descriptions of the relative importance of different frequencies and are referred to here as intensity-importance functions (IIFs). They were obtained as follows. Sharply filtered bands of speech (NU6 monosyllabic words) were mixed with filtered noise and presented alone or in pairs at 19 signal-to-noise ratios (−25 to 41 dB). When paired bands were tested, the level and signal-to-noise ratio (SNR) of one band were held constant while the level and SNR of the other band were varied. The listeners were 100 normal hearers, organized into five 20-person groups. Each group provided speech recognition data for one of five frequency regions (141–562, 562–1122, 1122–1778, 1778–2818, and 2818–8913 Hz). Comparisons of the results for each group indicated that IIFs vary with frequency and SNR. Current methods for predicting intelligibility from physical measurements of speech audibility would need to be revised in order to take such findings into consideration. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1445788]

PACS numbers: 43.71.An, 43.71.Gv, 43.66.Dc [CWT]

I. INTRODUCTION

The effects of intensity on speech intelligibility have been of interest to speech and hearing scientists for almost a century. In view of this, surprisingly little attention has been given to the way in which the relative importance of the speech spectrum varies with intensity. Since the classic articulation index (AI) studies conducted by Bell Telephone Laboratories over 50 years ago (French and Steinberg, 1947; Fletcher and Galt, 1950), there have been few attempts to explore this topic. Some exceptions are the studies by Hirsh, Reynolds, and Joseph (1954), Pavlovic and Studebaker (1984), and Boothroyd (1990). In the first two studies, however, the importance of the speech intensity range was not the principal focus of the work, while in the third, the conclusions were based on a theoretical analysis of the shape of the performance-intensity (P-I) function rather than on actual experimental data.

To a large degree, lack of interest in the speech dynamic range appears to reflect the perception that it was adequately defined by the AI pioneers long ago. This perception was strongly reinforced when the American National Standards Institute AI writing group (ANSI S3.5-1969), and the European researchers who developed the Speech Transmission Index (Houtgast and Steeneken, 1973; Steeneken and Houtgast, 1980) adopted the now well-known 30-dB dynamic range.

This range has its origins in the work of Beranek (1947). His proposal for a uniformly weighted 30-dB range was based on his assessment of the short-term speech spectrum measurements reported by Dunn and White (1940). The strength of this idea, in comparison to the competing models of French and Steinberg (1947) and Fletcher and Galt (1950), was that it was simpler and yet provided satisfactory predictions for many listening conditions. However, the simplicity of the Beranek model is also its chief weakness. As Boothroyd observed in 1990, certain aspects of the model are implausible. In particular, he noted its assumption of abrupt transitions between zero and full importance at the limits of the dynamic range. And, as Levitt (1982) had pointed out earlier, Dunn and White were unable to measure very soft speech levels due to the noise limitations of their measuring equipment. Thus, 30 dB may be a conservative estimate for the dynamic range of speech.

The goal of this study was to investigate how the importance of added audibility for speech recognition changes with intensity and signal-to-noise ratio (SNR). The functions that describe the results are analogous to previous descriptions of the importance of different frequency regions for speech intelligibility and are therefore referred to using the name suggested by Boothroyd (1990), *intensity-importance functions*.

As conceptualized here, the intensity-importance function (IIF) for any given frequency band within the auditory spectrum has three major components. They are: (1) the width (in dB) of the effective[1] dynamic range of speech (EDRS); (2) the SNRs (in dB) that correspond to the upper and lower limits of the EDRS; and (3) the relative value of each decibel within the EDRS.

A survey of the literature reveals a number of conflicting ideas about each of these components of the IIF. Values proposed for the width of the speech dynamic range extend from as low as 30 dB to as high as 68 dB (ANSI S3.5-1997; Beranek, 1947; French and Steinberg, 1947; Fletcher and Galt, 1950; Studebaker *et al.*, 1999). Similarly, estimates for the upper limit of the range and the weights assigned to each

[a]Electronic mail: gstudbkr@memphis.edu

decibel within the range also differ. The upper limit, i.e., the speech peaks level, has been variously assumed to be either 12 or 15 dB above the long-term rms speech level (Beranek, 1947; Kryter, 1962; ANSI S3.5-1969; Steeneken and Houtgast, 1980; ANSI S3.5-1997) or to change in level with frequency (French and Steinberg, 1947; Studebaker and Sherbecoe, 1991; Studebaker *et al.*, 1993), while both linear (Beranek, 1947; Kryter, 1962; Steeneken and Houtgast, 1980) and nonlinear (French and Steinberg, 1947; Fletcher and Galt, 1950) weighting schemes have been applied to the interior of the range.

One of our own recent studies (Studebaker *et al.*, 1999) suggests that the 30-dB-wide IIF used in many audibility models is not accurate, at least for some high-level listening conditions. In that study, we attempted to predict the speech recognition scores of subjects who were tested at both very good (>15 dB) and very poor SNRs. We found that when the AI was adjusted so that it predicted a normal-hearing group's scores for the quiet conditions, it did not predict them well for the noisy conditions, or vice versa. This problem did not occur in the case of hearing-impaired subjects whose scores for the same conditions were similarly evaluated. Interestingly, these inconsistencies were substantially reduced, without causing other difficulties, by simply increasing the EDRS from 30 to 40 dB. This suggested that the problem is not fundamental to the AI itself, but may be due to the intensity-importance weighting function used in the calculations.

The results of other studies also reveal the need to reexamine the issue of the EDRS/IIF. In two recent reports, Rankovic (1997, 1998) concluded that the Fletcher and Galt (1950) method for predicting speech recognition performance is superior to the method described in ANSI S3.5-1969. While there are many differences between these two methods, one of the more significant differences is the IIF that each assumes. The Fletcher and Galt IIF is substantially broader than the ANSI IIF and presumes that the weight of each decibel in the EDRS decreases nonlinearly in relation to its maximum value. Thus, Rankovic's results at least do not support the 30-dB linearly weighted IIF used in ANSI S3.5-1969.

If the EDRS has characteristics that differ substantially from those long assumed by most researchers, this could have significant practical implications. For example, if the dynamic range of speech is larger than 30 dB, then the effects of noise at relatively good SNRs will be underestimated when the standard IIF is used. In this case the standard will provide misleading conclusions about the impact of relatively low-level noise on intelligibility. This would be especially noticeable under narrow-band listening conditions, such as those experienced by many hearing-impaired persons and hearing aid users. The problem may have been exacerbated with ANSI S3.5-1997, in which the entire EDRS was moved upward 3 dB relative to that specified in ANSI S3.5-1969.

Finally, many researchers have used audibility calculations to draw conclusions about various aspects of human auditory performance and the fundamental nature of speech itself (e.g., Kamm *et al.*, 1985; Humes *et al.*, 1986; Ludvigsen, 1987; Zurek and Delhorne, 1987; Dubno *et al.*, 1989; Ching *et al.*, 1998; Noordhoek *et al.*, 1999; Magnusson *et al.*, 2001). If the IIF that was assumed in those calculations was incorrect, then it may be necessary to revise at least some of the conclusions that were derived using them. For this reason, and those noted earlier, we concluded that additional study of the IIF was needed.

II. METHOD

A. Overview

The method used in this project consisted of three basic steps: (1) Sharply filtered bands of speech were presented to normal-hearing subjects over a range of SNRs. (2) Speech recognition scores for these bands were converted into AI^2 values. (3) The AI values were plotted as a function of SNR to form IIFs. A fundamental assumption underlying this approach is that the relative importance of an auditory region can be derived from changes in the speech recognition score that occur when the SNR or bandwidth of a speech stimulus is varied to include or exclude that region.

Testing was carried out in each of five contiguous frequency bands. Each band was presented by itself, and also in conjunction with a second band of the same speech located in a different frequency region. When the bands were presented in pairs, they were always spaced at least 1 octave apart. One of the bands, called the *pedestal band*, was kept at a fixed level, while the other band, called the *remote band*, was varied in level. To ensure that the entire dynamic range would be revealed, the widths of the bands were chosen so that virtually all of the subjects would still obtain scores below 100% when the pedestal and remote bands were heard together and the remote band was at its highest level.

Both one-band and two-band conditions were evaluated because each type of condition has certain disadvantages that could bias the derivation of the IIF. When they are used together, however, these disadvantages tend to offset one another.

The main problem with narrowly filtered one-band conditions is that they produce very low speech recognition scores. This causes several potential complications. First, the subjects may become frustrated with the difficulty of the task, with possible effects on their morale, concentration, and performance. Second, changes in audibility that may have a substantial effect on intelligibility when scores are near the middle of the performance range may have less effect, or even none at all, when scores are at or near the bottom of the range. Third, because the ends of the performance scale are nonlinear, small differences in the score will produce relatively large differences in the AI. This could cause large variations in the derived importance values.

For these reasons, all of the subjects were first tested under two-band conditions in which the pedestal band had a relatively low fixed level and a high SNR. This produced both an easier task and also a minimum level of performance to which the contributions of the variable-level remote band could be added. The bandwidths, levels, and SNRs of the pedestal bands were chosen so that when these bands were heard by themselves they had nominal AI values of about 0.2 and produced scores of about 20%.

TABLE I. Filter cutoff frequencies, slopes, and AIs for each band.

	Band number				
	1	2	3	4	5
Band limits (Hz)	141–562	562–1122	1122–1778	1778–2818	2818–8913
HP filter skirt (dB/oct)	116	610	1187	1368	1 622
LP filter skirt (dB/oct)	1377	2138	3033	3909	10 663
Nominal AI(%)	20.3	19.7	17.9	21.0	19.6

While the two-band approach addresses the problem of low scores, it has the disadvantage that the bands might interact. If speech information in the two bands were to add synergistically, the importance function would increase more rapidly with level. Conversely, if the energy from one band were to spread into and mask the energy in the other band, importance would change more slowly with level or might even decrease.

Therefore, after completing the two-band tests we also tested single bands so that the effects of interactions could be estimated. The same subjects were used and each received the same bands, SNRs, and intensity levels as in the two-band tests. To minimize learning effects, the two-band and one-band tests were conducted at least 1 week apart.

B. Subjects

One hundred adults (28 males, 72 females) between 18 and 35 years old (mean age: 23.6 years, age standard deviation: 3.3 years) were recruited from the Memphis area to serve as subjects. All subjects were either enrolled in college at the time they were tested or had completed at least 1 year of college education. None had formal academic training in speech and hearing science.

Each potential subject received a pure-tone hearing test, an auditory immittance test, and a series of practice monosyllabic word tests (50-item Auditec of St. Louis W-22 lists) in quiet and without filtering before the experiment. Only people with audiometric thresholds no worse than 15 dBHL (ANSI S3.6-1996) at the octave frequencies from 125 to 8000 Hz, normal tympanograms (± 100 daPa), and practice word scores of at least 90% were chosen.

C. Speech stimuli

The speech test stimuli were digitized copies of the Auditec of St. Louis recordings of the NU6 monosyllabic word test (Studebaker *et al.*, 1993). Prior to the experiment, these materials were software filtered into five frequency bands (MATLAB v5.21, The MathWorks, Inc.) Filtering was performed using high-order (>2500) equiripple finite-impulse-response (FIR) bandpass filters with transition bands of 74 Hz, nominal stop-band attenuation levels of approximately 200 dB, and passband ripple of less than 0.1 dB.

The filter cutoffs were chosen so that the bands would produce approximately equal intelligibility and have limits at the ''preferred'' frequencies specified in ANSI S1.6-1984, R1997. The cutoff frequencies used, the calculated filter skirt slopes, and the calculated AIs for each of the five bands (Studebaker *et al.*, 1993) are reported in Table I.

After filtering, the speech was digitally amplified using locally written software that preserved the original intensity relationships among the 200 words that compose the NU6 test, and also between the words and a speech calibration signal. (The calibration signal was a recording of the test words without their carrier phrases or the pauses between words.) The bands were then combined to produce two-channel recordings with the pedestal band on one channel and the remote band on the other. Finally, each signal was converted to 20-bit format (COOL EDIT PRO SPECIAL EDITION v1.1, Syntrillium, Inc.).

D. Noise stimuli

Two noises were used to restrict the audibility of the speech within each band. The noise used with the pedestal band was shaped to match the contour of the average normal-hearing threshold (ANSI S3.6-1996). The noise used with each remote band was shaped to match the speech-peaks spectrum of the talker. The limits of the threshold-shaped noise were 89 to 11 000 Hz. The limits of the talker-spectrum-matched (TSM) noise were always one-third of an octave wider, on either side, than the limits of the speech band with which it was combined.

E. Instrumentation

The speech stimuli were played at 25 kHz (their original sampling rate) using a personal computer (Dell Pentium 133 MHz) with a 20-bit multichannel sound card (Layla, Echo Audio Corp.). The level of the pedestal band was set using a programmable attenuator (PA4, Tucker-Davis Technologies [TDT]), while the remote band was first amplified an additional 10 dB (SS1, TDT) and then its level was also set with a programmable attenuator (PA4, TDT).

The pedestal and remote band speech signals were routed to a mixer (SM3, TDT), where the bands were recombined and each band was mixed with the desired level of masking noise. The mixed signals were then passed through a headphone buffer (HB6, TDT) to a wideband insert earphone (ER-1, Etymotic Research) designed to simulate sound-field listening conditions.

All of the subjects listened to the test stimuli monaurally while sitting in a double-walled, sound-treated room. Noise levels in the room were below those needed for testing audiometric thresholds using earphones. Measurements at the output of the earphone indicated that there was less than 0.10% harmonic distortion at the audiometric frequencies from 250 to 4000 Hz when the playback system was set to deliver the highest speech level presented. Subjects were in-

structed when to listen by messages displayed on a computer monitor. They responded by typing the words they heard on a computer keyboard.

F. Test conditions

The 100 subjects were assigned quasi-randomly (i.e., without replacement) to one of five 20-person groups. The subjects in each group were first tested with one of the pedestal band/remote band combinations and then with the same remote band alone. They also received two tests with just the pedestal band. One of these tests occurred during the two-band conditions and the other during the one-band conditions.

In the two-band tests, band 1, the lowest frequency band, and band 5, the highest frequency band, served as pedestals. When band 1 served as a pedestal, it was paired with bands 3, 4, and 5. When band 5 served as a pedestal, it was paired with bands 1 and 2.

The band 1 pedestal was always presented at 60 dB SPL while the band 5 pedestal was always presented at 54 dB SPL. These are the same SPLs that each band would have if the bands were part of an unfiltered speech signal presented at 63 dB SPL. The level of each remote band or single band was varied, in steps of 4 dB, to produce one of 19 SPLs ranging from 19 to 91 dB SPL.

The threshold-shaped noise was always presented at 33 dB SPL. This was equivalent to a masked threshold level of 0 dB HL. Masked pure-tone hearing tests were used to verify that the noise did not produce unwanted threshold elevation in either the pedestal band or the remote band. The level of the TSM noise varied, depending on its bandwidth, from about 45 to about 47 dB SPL. Within the passband of the chosen remote band, however, the TSM noise was always at 44 dB SPL.

All of the specified speech and noise levels were measured by repeatedly playing either the speech calibration signal or the masking noise itself through the testing system into a Zwislocki coupler and a sound-level meter (Larson-Davis 800B). The meter was set to linear frequency weighting, one-third-octave-band mode, integrate, and a 3-dB exchange rate. Output levels, based on a 2-min integration interval, were read directly from the meter's digital display.

SNRs for the five bands were derived from their respective speech and noise levels. These SNRs, when based on the SPLs of just the speech and noise within the frequency limits of the speech bands, had the following values. The two pedestal bands, 1 and 5, had SNRs of 33 and 25 dB, respectively. Each of the remote bands and the single bands had SNRs ranging, in 4-dB steps, from -25 to 47 dB.

Every subject received one 50-word NU6 list under each test condition presented. This corresponds to 40 NU6 lists per subject (19 lists$\times$2 presentations$+$1 list$\times$2 presentations). The order of the conditions was first randomized and then counterbalanced (Kirk, 1982) so that, for each group of 20 subjects, the four available NU6 lists were presented an equal number of times across the subjects and SNRs. Equal numbers of right and left ears were tested in each group. The subjects were paid for their participation.

G. Data processing procedures

1. Conversion of scores into importance values

The importance of an auditory area is determined from the AI change that occurs when the area is added to or deleted from what was previously audible. The AI change is based on the speech recognition scores obtained with and without the particular auditory area included in the presentation. Therefore, in order to assess the importance of an auditory area, it is first necessary to convert the speech scores into AI values.

In this study, AI values were derived from speech scores using the relative transfer function[3] reported by Studebaker *et al.* (1993). The importance, *I*, of an area was then calculated using Eq. (1). A_X and A_Y in this equation represent transforms of the scores obtained with and without the area of interest included and A_T represents the AI for the entire audible area

$$I=\frac{|A_X-A_Y|}{A_T}. \tag{1}$$

To simplify interpretation, each result was expressed as a percentage.

Other AI-type procedures, such as the speech intelligibility index (ANSI S3.5-1997) or the speech transmission index (Steeneken and Houtgast, 1980), assume that the importance assigned to each frequency band is uniformly distributed across the dynamic range and that the dynamic range is the same for all frequency bands. Neither of these assumptions was made here. Instead, the importance of each additional dB of SNR within any frequency band was viewed as a proportional part of the total importance of the entire auditory area, not just as a proportional part of an individual frequency band.

The A_T values (one for the single-band conditions and one for the two-band conditions) were determined by summing the contiguous positive importance values within each band and then combining the sums across bands. This approach has three advantages: (1) It permits the inclusion of negative importance values in the IIF but does not permit them to influence the function's shape or magnitude within the positive importance region. (2) It produces a set of cumulative functions for which the largest values in the positive importance region across bands add to 100%. (3) It produces an average cumulative curve that falls just short of 20%, and a total cumulative curve that falls just short of 100%, thereby correctly reflecting the fact that the curves for the five individual bands reach their peak values at different SNRs.

2. Filter-skirt effects

Scores on filtered speech materials may be affected by information in the filter skirts (Studebaker and Sherbecoe, 1993; Warren and Bashford, 1999; Warren, Bashford, and Lenz, 2000). Although the digital filters used in this study had very steep skirts (Table I), they were not infinitely steep. Therefore, it was expected that the bandwidth of the speech signal would increase with level and calculations were made to estimate the likely effect of this factor. The results indi-

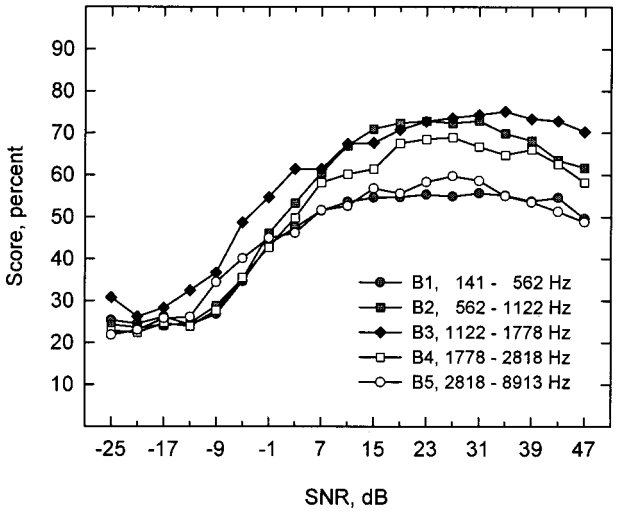


FIG. 1. Performance-intensity functions (in %) for the two-band conditions.

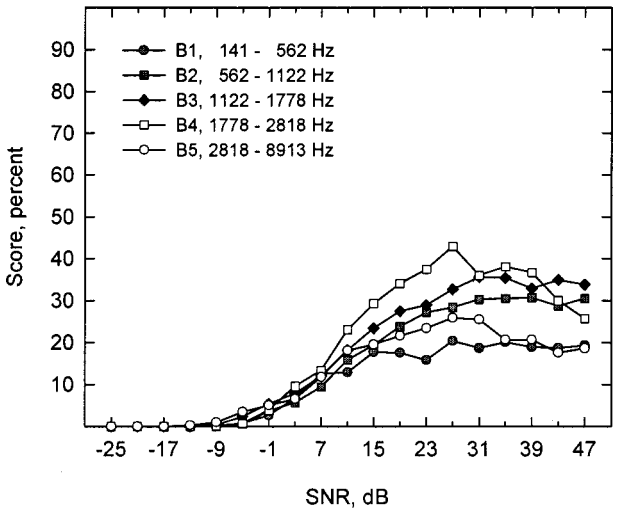


FIG. 2. Performance-intensity functions (in %) for the one-band conditions.

cated that at the highest listening level (91 dB SPL), the score for an average normal-hearing subject would increase by 1.1% to 5.5%, depending on the frequency band. The mean increase was 3.2%. At lower speech levels the effect would be smaller and, as the filter skirts have a nearly constant slope, uniformly distributed across the range of presentation levels used. It was concluded, therefore, that this factor would not significantly affect the results and correction for it was not included.

III. RESULTS

A. Experimental data

1. Two-band functions

Figure 1 shows the performance-intensity (P-I) functions (in percent) for the two-band conditions. Each function is based on data averaged across 20 subjects. It is apparent from this figure that the five bands did not produce identical P-I functions. Nevertheless, the functions have a number of similar characteristics.

One similarity is that all five bands produced relatively little change in the score below −13 dB SNR. Also, the absolute scores for all of the bands were close to the results for the pedestal bands alone (band 1 pedestal=21.9%, band 5 pedestal=24.5%). Both results suggest that the base of each P-I function was determined primarily by the pedestal band, as intended.

Between −13 and 3 dB SNR, the bands all produced curves that grow rapidly. The curves for bands 1, 2, and 4 increase similarly, while those for bands 3 and 5 do not. At 3 dB, each function's rate of growth begins to taper off, with some functions slowing more quickly than others. At about 15 to 19 dB SNR, growth slows further with some functions reaching a plateau and others continuing until about 31 to 35 dB SNR. Above this they all exhibit signs of rollover.

An unusual finding was that band 3, the central band on the frequency scale, produced noticeably better scores than the other bands below 7 dB SNR, particularly at −25 dB. It is unclear why that occurred, but it did not appear to have happened because of experimental error or because these subjects were overly proficient at the task. In any case, this result does not pose a significant problem because the critical factor in the analysis is the change in the score with SNR, not the score differences across bands at each SNR.

2. One-band functions

Figure 2 shows the P-I functions for the one-band conditions. Notice that without the pedestal, the scores for the five bands did not start to deviate from the baseline until −9 and −5 dB SNR. Then, they all rose at virtually the same rate with all bands producing very similar scores through 7 dB SNR.

Above 7 dB SNR, the growth rate slows and the P-I functions start to diverge. The low-to-midfrequency bands (1, 2, and 3) produced curves that reach a maximum at different levels but then remain nearly constant with level. In contrast, the two high-frequency bands (4 and 5) produced curves that reach a maximum at about the same SNR and then exhibit a fairly steady decline.

Just as in the two-band conditions, the bands with the highest and lowest frequency ranges produced the smallest absolute scores while the more centrally located bands produced the largest scores. In this case, however, band 4, rather than band 3, produced better overall performance.

3. Comparison of one-band and two-band functions

The overall data patterns in Figs. 1 and 2 are similar in a number of ways. One common feature is that the P-I functions for the five bands changed in a similar way at low SNRs, but in a distinctly differently way at high SNRs. The results for the low SNRs are consistent with the presumption that the bands have nearly equal frequency importance. The results at the high SNRs are not. Thus, both data sets support the conclusion that band importance changed with SNR. Both sets of curves also indicate that some bands produced more change in the score, and over a broader range of SNRs, than others.

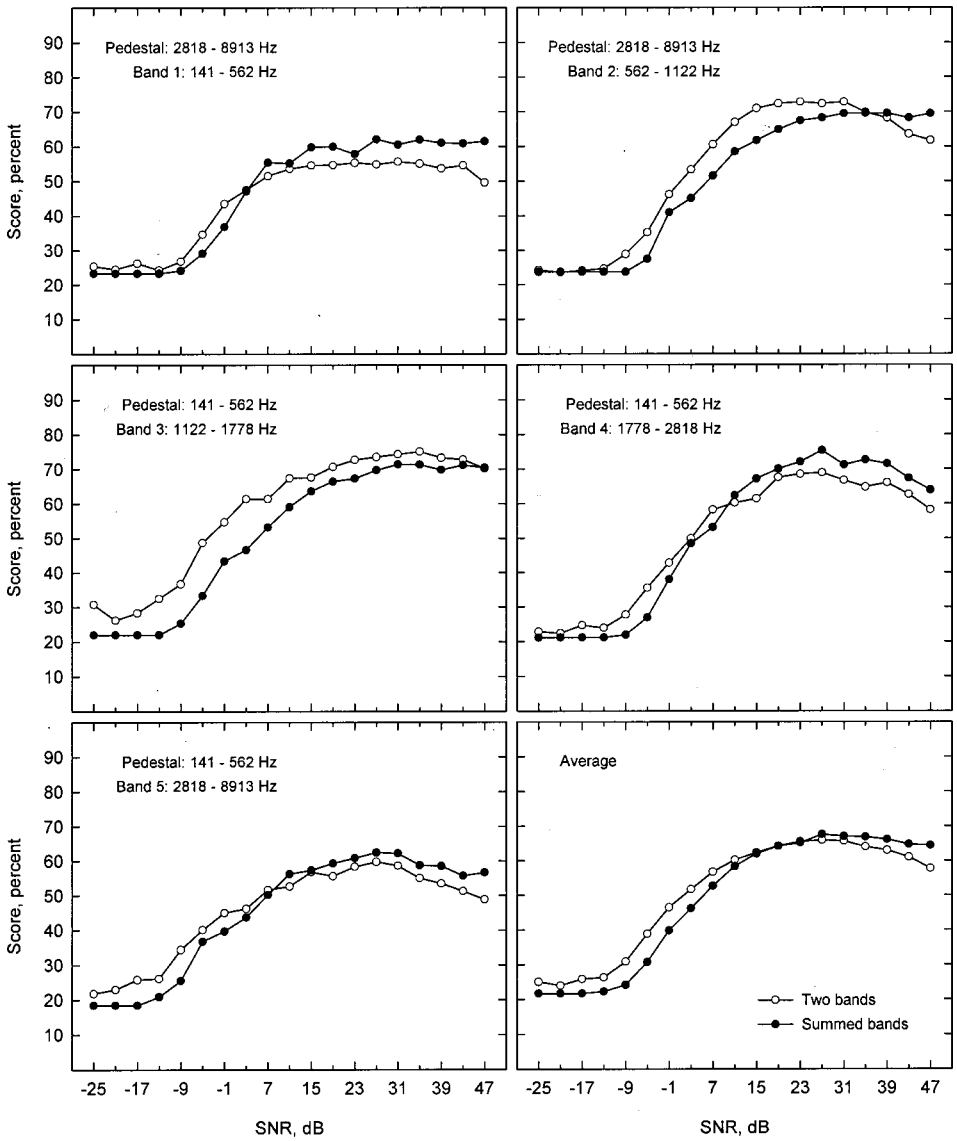


FIG. 3. Performance-intensity functions for the two-band and the comparable summed-band conditions. The open symbols are the functions from Fig. 1. The filled symbols are the functions predicted from the sum of the AIs for the pedestal band and one-band conditions. The last panel displays the average results for the five bands.

There are also notable differences between the one-and two-band functions. One is that the two-band functions rise off the baseline at lower SNRs. The change in score with SNR was also larger in the two-band conditions. However, this is just an artifact of the absolute score size. Another difference is that the functions for the five one-band conditions overlap more closely at low SNRs, while those for the two-band conditions exhibit more rollover at high SNRs. Finally, there are differences in the relationships among bands 2, 3, and 4 in the one-band and two-band conditions. These differences suggest that both positive and negative interactions occurred between the pedestal and remote bands and that the interactions varied in complex ways, depending on the speech levels and particular band combinations involved.

4. Comparison of two-band and summed-band functions

The difference in the absolute size of the scores for the one-band and two-band conditions makes it difficult to evaluate the interactions that may have occurred between the pedestal and remote bands. To deal with this problem the scores for the two-band conditions were compared with scores predicted from the sums of the AIs for the pedestal band and one-band conditions. This ''summed-band'' condition makes the assumption that when two bands are presented together, each band contributes independently to the total AI. The two-band results show what happened when the same bands were actually presented together.

The predicted scores for the summed bands were obtained by first converting the scores for the pedestal band and one-band condition that made up each two-band condition into AI values using the transfer function from Studebaker *et al.* (1993). These paired AIs were then added together and the sums converted back to scores using the same transfer function.

Figure 3 compares the summed-band and two-band functions for each of the five frequency ranges studied. The means for each function type are compared in the sixth panel. The different panels represent different groups of subjects. However, the two lines within each panel are based on data from the same 20 subjects. All 100 subjects are represented in the two mean lines.

Several patterns are evident. First, at low remote band levels, the two-band conditions produced higher scores than

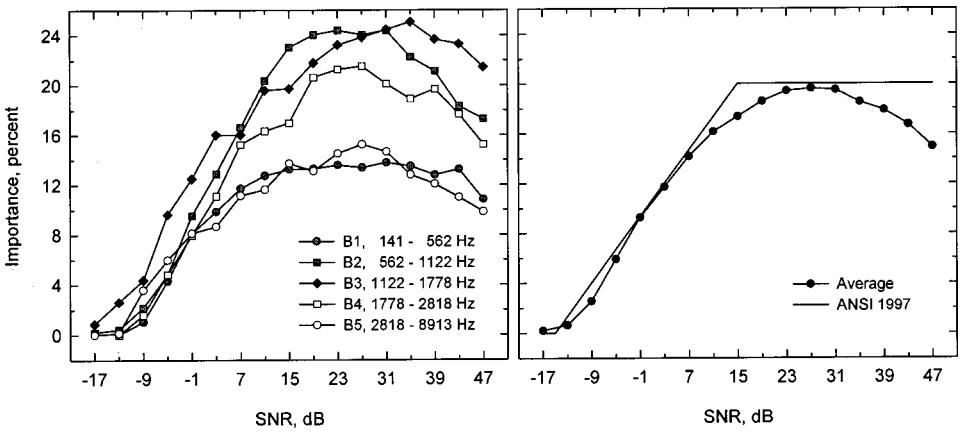


FIG. 4. Cumulative intensity-importance functions (in %) based on the two-band data. The right-hand panel shows the average results.

the summed bands. Also, in four of the five conditions, the two-band conditions produced higher scores than the summed bands at levels where the remote band alone produced 0% scores (see Fig. 2). At these very low SNRs, the two-band results were, on average, about 3.6% higher than the summed-band results. This difference cannot be attributed to learning because the subjects always received the two-band conditions before the one-band conditions.

At higher SNRs, the size of what appears to be a synergistic effect varied substantially with frequency. At the mid SNRs, bands 2 and 3 showed the largest effects. These bands were also the ones closest in frequency to their respective pedestal bands, though they were not adjacent. The difference between the two-band and summed-band conditions gradually diminished as the level of the remote band was increased, until, at the highest SNRs, most conditions reveal negative band interactions. The crossover point where the interactions reversed direction occurred, on average, at about 20 dB SNR (64 dB SPL).

It should be noted that the negative interactions in Fig. 3 do not indicate that one band was ever superior to two. As Figs. 1 and 2 illustrate, performance was always better when the pedestal band was present than when it was not. This was true even at the highest remote band levels where the negative interactions were largest.

There was no evidence that the data were significantly affected by a practice effect as a result of all subjects having received the one-band conditions after the two-band conditions. Such an effect, had it occurred, would have raised the scores for the one-band conditions and thereby produced higher scores for at least some of the summed-band conditions. Also, a comparison of each subject's two scores for the pedestal bands, one score obtained during the two-band tests and one during the one-band conditions, did not reveal significant changes in performance with practice. Mean differences between the two scores ranged from 4.2% to −0.4%, with an average difference of 1.4%.

B. Derived results

1. Cumulative functions

Cumulative intensity-importance functions describe the amount of importance that has accumulated up to a specified SNR. They are similar to P-I functions. Figures 4 and 5 show the cumulative IIFs derived from the two-band functions reported in Fig. 1 and the one-band functions reported in Fig. 2, respectively. Each figure also displays an average function, in the right-hand panel. A comparably scaled *de facto* IIF function for ANSI S3.5-1997 is included for comparison.

All of the IIFs are based directly on the original data. That is, neither these results, nor the original data, were smoothed or normalized in any way, with the exception that the importance values are expressed as percentages. When evaluating these functions, note that the lower-left area of each IIF is determined by the high-intensity parts of the speech signal and the upper-right area by the low-intensity parts.

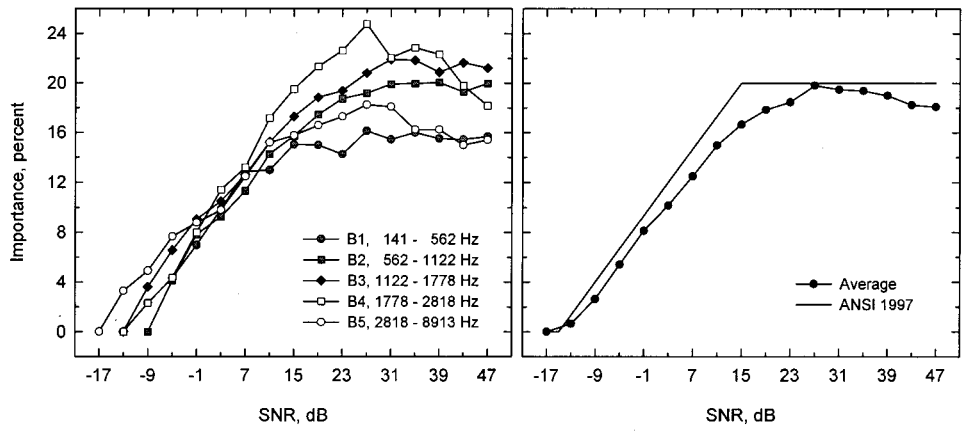


FIG. 5. Cumulative intensity-importance functions (in %) based on the one-band data. The right-hand panel shows the average results.

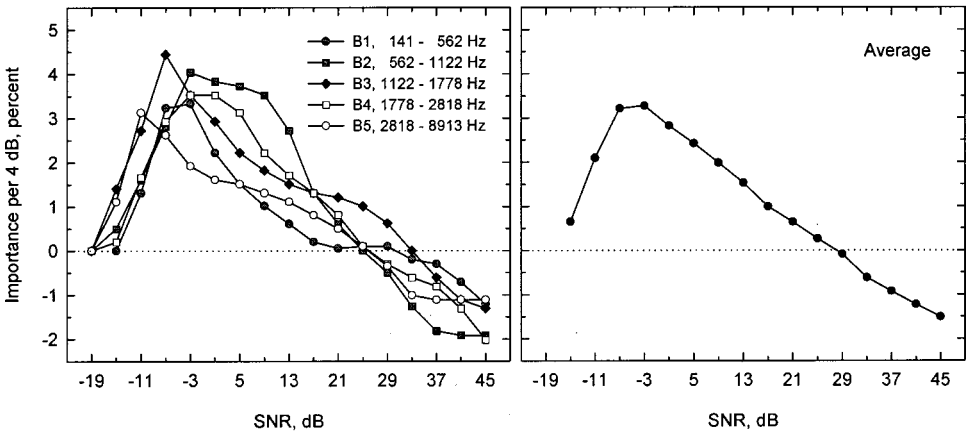


FIG. 6. Smoothed density IIFs based on the two-band data. The right-hand panel shows the average results.

Comparison of the IIFs in Figs. 4 and 5 with the P-I functions in Figs. 1 and 2 reveals that the inter-relationships among the bands have been well preserved. For example, at low SNRs the functions for the various bands are similar while at high SNRs there are substantial differences. Also, the one-band IIFs are more like each other than the two-band IIFs. Based on these and other similarities, it can be inferred that the IIFs accurately reflect the relative importance of the experimental conditions.

Some bands have peak values exceeding 20% (i.e., 1/5 of the total), while others have substantially smaller maximum values. Thus, the total importance of the five bands differed substantially even while their importance was similar at the poorer SNRs. Also, most of the IIFs have regions of negative importance reflecting the rollover in the P-I functions. Recall that the method used to create the functions made it possible to include these negative effects without influencing the positive part of the IIFs (i.e., the region to the left of the peak value).

2. Density functions

Although cumulative importance functions have the advantage that they can be readily related to the P-I functions upon which they were based, density importance functions have greater practical value because they provide a detailed view of the relative importance of each part of the intensity (or frequency) domain. Unfortunately, this view comes at the cost of a high degree of variability. This occurs because each value in a density function is based on a small subset of the data. In the present case, each density importance value is derived from the transforms of just two mean scores, each based on 1000 test words.

It is well known that speech recognition scores may include substantial chance effects (Thornton and Raffin, 1978) and it follows from statistical theory that the variability of the difference between two scores is still greater (Snodgrass, 1977, pp. 227–228). As importance function values are differences between transformed speech scores, they reflect the variability of the original scores. The same variability is also present in the cumulative functions, but is less evident due to the larger scale of the display.

The typical procedure to deal with local irregularities in functions is to use curve fitting or data smoothing. Unfortunately, all such procedures distort the data. This can be a particular problem near the ends of importance functions because these regions are often viewed as having special significance.[4] It was for this reason that data smoothing was not used up to this point. Nevertheless, due to relatively large point-to-point variability, smoothing is necessary to produce density importance functions that can be used in real-world applications.

The smoothing procedure used in this study worked as follows. A spreadsheet with two graphs was created for each band. One graph displayed the band's original cumulative IIF and an estimate of its smoothed cumulative IIF; the other graph displayed the band's original density function and an initial estimate of its smoothed density IIF. Both graphs were based on the same data. To smooth the data, one of us (G.A.S.) adjusted the values of the data points that alternated around the then-current smoothed density function while observing the relationship between the smoothed and unsmoothed data in both the cumulative and density graph displays. This was repeated until smoothed functions were obtained[5] that distorted the overall trends in the original data as little as possible in both displays simultaneously.

Figures 6 and 7 show the resulting smoothed density importance functions. The left panel of each figure displays the functions for each frequency band, while the right panel displays the average function for the five bands. The ordinate of each figure is scaled so that the importance values reflect the importance (in %) of the 4-dB intervals used to collect the original data. To determine importance per dB, it is necessary to divide the ordinate values by 4. Importance is expressed as a percentage of the total importance of all five bands combined. Thus, each band has a different total importance, but together, the sum of the positive importance values add to 100%. The sum of the positive values on the average function is slightly less than 20% because all of the bands did not reach maximum importance at the same SNR.

An aspect of Figs. 6 and 7 that requires mention is the nominal SNRs at which the importance values are plotted. In Figs. 1–3, the data are, of course, plotted at the actual SNRs used. The same practice is used in Figs. 4 and 5 because the curves depict the importance that has accumulated up to the SNR indicated on the abscissa. In contrast, in Figs. 6 and 7, the data are plotted like a histogram, i.e., the importance

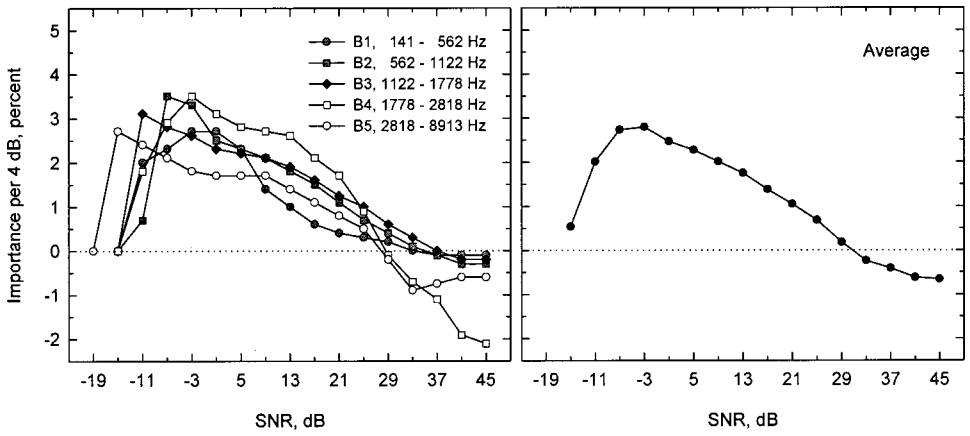


FIG. 7. Smoothed density IIFs based on the one-band data. The right-hand panel shows the average results.

values are placed at the center of the interval between the two SNRs that contribute to them. This practice is consistent with the methods used to generate and describe frequency-importance functions.

Notice that the one-band and two-band importance functions both show relatively large initial increases, jumping from 0% to 3% in only one or two 4-dB steps. This large increase was anticipated in the case of the one-band functions because the shape of the transfer function causes small changes in the score near 0% to produce large changes in AI. It was not expected in the case of the two-band functions. A comparison of the individual band results in Figs. 6 and 7 indicates that the two-band conditions did produce a somewhat more gradual initial rise in importance as SNR was increased. Generally, the transition to the peak value occurred over two 4-dB steps rather than one, as in the case of the one-band data. Nevertheless, the initial increase was still relatively large. Together, these results suggest that that some factor besides the transfer function shape contributes to the rapid initial increases in both the one-band and two-band functions.

As a related matter, note in Figs. 6 and 7 that the average density IIFs did not show as rapid an initial increase in importance with SNR. Inspection reveals that this occurred in large part because the different bands produced abrupt increases in the importance at different SNRs. This suggests that even while the importance of narrow bands may increase quickly with SNR, intensity importance for broadband speech may increase more slowly.

After reaching a peak value, importance gradually declined with SNR until it reached zero, somewhere between 25 and 30 dB SNR. The one-band IIFs are flatter and more alike in shape than the two-band functions. The greater variability of the two-band functions reflects the band interactions described earlier.

Figure 8 compares the two average density importance functions directly. Here, the ordinate values are multiplied by 5 so that the figure provides an estimate of the importance of the five bands combined. The overall configurations of the functions in Fig. 8 are similar. That is, as SNR is increased, importance rises relatively rapidly to a peak about 8 to 12 dB above the starting point before falling relatively gradually but steadily to 0% importance over another 30 dB or so. Although there are quantitative differences, the overall shape of the IIF does not depend strongly on whether it was determined using single bands or two-band combinations. The major difference between the two functions is that importance changes more with SNR for the two-band conditions than for the one-band conditions. This is another result of the positive and negative band interactions that occurred when two bands were presented together.

3. Smoothed cumulative functions

Figure 9 shows cumulative IIFs derived directly from the smoothed density functions in Fig. 8. Comparisons with the mean unsmoothed cumulative functions in Figs. 4 and 5 reveal that the two sets of curves are similar to each other. This shows that the smoothing operation did not substantively alter the overall trends of the data.

4. The ANSI S3.5 functions

Figure 8 also includes the *de facto* density intensity-importance function used in ANSI S3.5-1997. It has been plotted using increments of importance per 4 dB to make it comparable to the other two functions in the figure. The value at −15 dB SNR is only one-half as large as the others

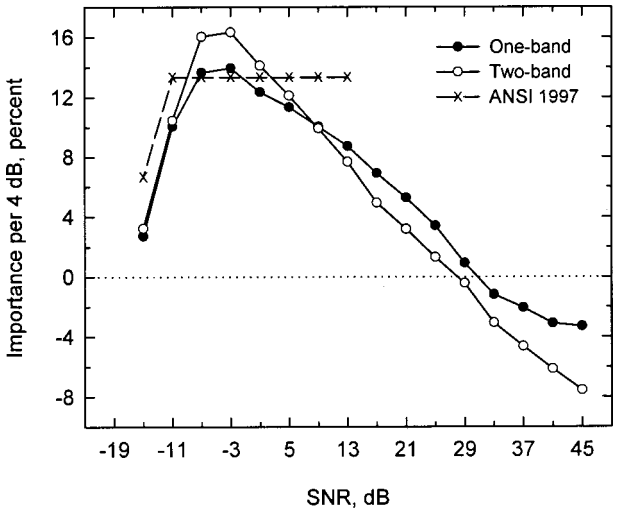


FIG. 8. The density IIFs obtained in this study, summed across the five frequency bands, versus the *de facto* IIF for ANSI S3.5-1997.

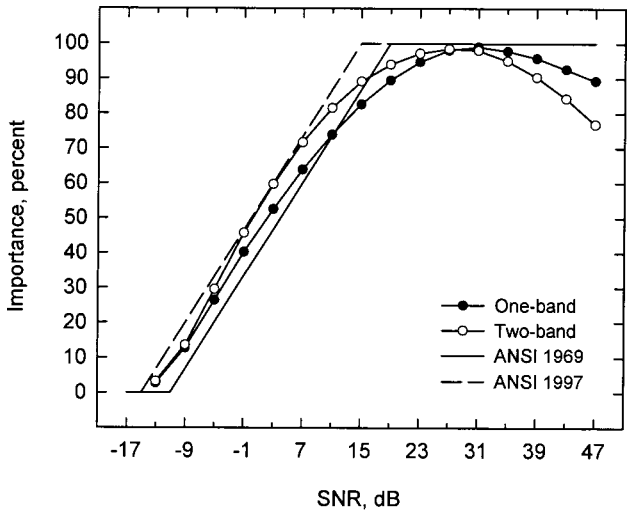


FIG. 9. The cumulative IIFs obtained in this study, summed across the five frequency bands, versus the *de facto* cumulative IIFs for ANSI S3.5-1997 and ANSI S3.5-1969. These functions are based on the same data reported in Fig. 8.

because it represents the importance of the half-interval (2 dB) that results when the ANSI function is mapped onto the SNRs used in this study.

Figure 9 includes the *de facto* cumulative IIFs used in ANSI standards S3.5-1997, as well as the one in S3.5-1969. The latter IIF is identical to the former except that it is shifted 3 dB to the right.

5. One-third-octave band functions

In order to use the results of this study in a conventional manner, it is necessary to convert them into IIFs for one-third-octave bands. Both two-band (Table II) and one-band (Table III) results are presented. The one-band results are more appropriate for single narrow-band conditions while the two-band results are more appropriate for broadband or multiple-band conditions where interactions across the frequency spectrum are more likely.

The values in Tables II and III were derived as follows. Each of the five bands was divided into its constituent one-third-octave bands. Each constituent band was assigned a relative importance value within its respective larger band using the frequency-importance data of Studebaker *et al.* (1993). These relative importance values were used to allocate importance proportionately among the one-third-octave bands within each of the five bands at each SNR. This procedure preserved the different rates of change in importance over SNR for each of the five bands. However, the relative importance across the one-third-octave bands within each of the five speech bands remains constant over SNR. The results in Tables II and III cover the same frequency range as the 1/3-octave-band method described in ANSI S3.5-1997.

In using Tables II and III, it is necessary to consider the fact that the values listed there represent the importance (in %) of a 4-dB change in SNR. As 1-dB steps are used in most calculations, a table with four times as many cells (and values one-fourth as large) would be required.

IV. DISCUSSION

The results of this study indicate that the importance of the speech intensity range has characteristics that are both qualitatively and quantitatively different from those of the *de facto* IIF now in wide use. These data are not consistent with a 30-dB-wide function with equal weighting from −15 to +15 dB. Nor are they consistent with the practice of using the same IIF in all frequency regions.

These conclusions are evident from the unprocessed percentage scores. Therefore, they do not depend upon the validity of the methods used to derive the IIFs. However, to compare our results with the literature and the ANSI standards, the data had to be transformed into IIFs. The following observations were made regarding these comparisons.

A. Comparisons with other IIFs

In contrast to the rectangular function used in ANSI S3.5-1997, this study revealed that the IIF is shaped more like an asymmetric peak function. It rises relatively quickly from the point where speech first becomes audible, reaches a peak within 8 to 12 dB, and then gradually declines to zero over the next 30 dB or so. It was also noted that the IIF varied substantially across the five frequency regions investigated.

Figure 8 shows that, in relation to the present data, the ANSI S3.5-1997 function assigns more weight to the ends of its EDRS and less weight at the center. Figure 9 shows that above 4 dB SNR, the slopes of the functions obtained in this study begin to decrease while the slopes of the two ANSI functions remain unchanged. According to the ANSI S3.5-1997 function, speech is completely audible at 15 dB SNR, while the ANSI 1969 function indicates that full audibility occurs at 18 dB SNR. In this study, complete audibility was not achieved until SNRs of 27 dB (two-band data) to 31 dB (one-band data) were reached.

The present results are also consistent with the ANSI functions in several ways. For example, the two-band data produced a cumulative IIF for the five combined bands that is similar to the ANSI S3.5-1997 function between −15 and 7 dB SNR. This suggests that using these results, instead of the ANSI S3.5-1997 IIF, is not likely to have a significant effect on predicted speech recognition performance at relatively poor to moderate SNRs. This would include many of the studies on the effect of noise on speech intelligibility found in the literature.

Comparisons with the IIFs reported in the classic studies of French and Steinberg (1947) and Fletcher and Galt (1950) are difficult because of numerous differences in assumptions and procedures. This is particularly true of the Fletcher and Galt study. Nevertheless, qualitatively, the IIF functions reported here are more like that of Fletcher and Galt than the function by French and Steinberg or either of the ANSI functions.

B. Speech peaks

The SNR where the intensity-importance function intercepts the *x* axis provides an estimate of the level of the effective speech peaks. These peaks are assumed to be 12 dB

TABLE II. Importance values (in %/4 dB) for one-third octave bands based on the two band data.

Band CF (Hz)	160	200	250	315	400	500	630	800	1000
Cutoff (Hz)	141–178	178–224	224–282	282–355	355–447	447–562	562–708	708–891	891–1122
Ctr. S/N									
45	−0.0565	−0.0774	−0.1256	−0.2047	−0.3077	−0.4386	−0.6415	−0.6278	−0.6473
41	−0.0330	−0.0451	−0.0733	−0.1194	−0.1795	−0.2559	−0.6415	−0.6278	−0.6473
37	−0.0141	−0.0193	−0.0314	−0.0512	−0.0769	−0.1097	−0.6077	−0.5948	−0.6133
33	−0.0094	−0.0129	−0.0209	−0.0341	−0.0513	−0.0731	−0.4220	−0.4131	−0.4259
29	0.0047	0.0064	0.0105	0.0171	0.0256	0.0366	−0.1688	−0.1652	−0.1704
25	0.0047	0.0064	0.0105	0.0171	0.0256	0.0366	0.0000	0.0000	0.0000
21	0.0024	0.0032	0.0052	0.0085	0.0128	0.0183	0.2026	0.1983	0.2044
17	0.0094	0.0129	0.0209	0.0341	0.0513	0.0731	0.4389	0.4296	0.4429
13	0.0283	0.0387	0.0628	0.1024	0.1538	0.2193	0.9116	0.8922	0.9199
9	0.0478	0.0654	0.1062	0.1731	0.2602	0.3709	1.1817	1.1566	1.1925
5	0.0707	0.0967	0.1570	0.2559	0.3846	0.5483	1.2492	1.2226	1.2606
1	0.1037	0.1418	0.2302	0.3753	0.5641	0.8041	1.2830	1.2557	1.2947
−3	0.1555	0.2128	0.3453	0.5630	0.8462	1.2062	1.3505	1.3218	1.3628
−7	0.1508	0.2063	0.3349	0.5459	0.8205	1.1697	0.9454	0.9252	0.9540
−11	0.0613	0.0838	0.1360	0.2218	0.3333	0.4752	0.5402	0.5287	0.5451
−15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.1688	0.1652	0.1704
Band CF (Hz)	1250	1600	2000	2500	3150	4000	5000	6300	8000
Cutoff (Hz)	1122–1412	1412–1778	1778–2239	2239–2818	2818–3548	3548–4467	4467–5623	5623–7079	7079–8913
Ctr. S/N									
45	−0.5879	−0.7235	−1.1234	−0.8941	−0.4426	−0.3177	−0.1905	−0.1001	−0.0588
41	−0.4975	−0.6122	−0.7302	−0.5812	−0.4426	−0.3177	−0.1905	−0.1001	−0.0588
37	−0.2713	−0.3339	−0.4494	−0.3577	−0.4426	−0.3177	−0.1905	−0.1001	−0.0588
33	0.0000	0.0000	−0.3370	−0.2682	−0.4024	−0.2888	−0.1732	−0.0910	−0.0534
29	0.2783	0.3425	−0.1685	−0.1341	−0.1408	−0.1011	−0.0606	−0.0318	−0.0187
25	0.4522	0.5565	0.0562	0.0447	0.0402	0.0289	0.0173	0.0091	0.0053
21	0.5427	0.6679	0.4494	0.3577	0.2012	0.1444	0.0866	0.0455	0.0267
17	0.5879	0.7235	0.7302	0.5812	0.3219	0.2310	0.1385	0.0728	0.0428
13	0.6783	0.8348	0.9549	0.7600	0.4426	0.3177	0.1905	0.1001	0.0588
9	0.8140	1.0018	1.2358	0.9835	0.5231	0.3755	0.2251	0.1182	0.0695
5	0.9949	1.2244	1.7413	1.3859	0.6036	0.4332	0.2598	0.1364	0.0802
1	1.3115	1.6140	1.9660	1.5647	0.6438	0.4621	0.2771	0.1455	0.0855
−3	1.5828	1.9479	1.9660	1.5647	0.7645	0.5487	0.3290	0.1728	0.1015
−7	1.9898	2.4488	1.6290	1.2965	1.0462	0.7509	0.4503	0.2365	0.1390
−11	1.2210	1.5027	0.9268	0.7377	1.2474	0.8953	0.5369	0.2820	0.1657
−15	0.6331	0.7792	0.1123	0.0894	0.4426	0.3177	0.1905	0.1001	0.0588

in ANSI standard S3.5-1969 and 15 dB in ANSI standard S3.5-1997. The 3-dB difference reflects the fact that the old and new versions of the standard are based on different data.[6] Because 4-dB steps were used in this study, a precise estimate of the speech peak level is not possible from the current results. However, the data in Fig. 9 tend to support the view that it is larger than 12 dB.

The data also suggest that the effective peaks may grow larger at higher frequencies. Again, the 4-dB step size used in this study leaves a degree of uncertainty. However, nonzero performance was observed to vary from SNRs of −9 to −17 dB (mean=−13) across frequency bands. Therefore, although the present results are not inconsistent with the 15-dB peak level assumed in ANSI S3.5-1997 for broadband speech, different values may be appropriate for bandlimited speech, depending on the frequency region remaining audible.

C. Effective dynamic range

It is difficult to assign a single number to the EDRS because the width of the IIF varies with frequency and importance gains decline asymptotically with SNR (Fig. 9). In general, however, the present data suggest that the EDRS is greater than the 30 dB value now widely assumed. For example, the difference between the first and last nonzero importance values on our functions range from 36 to 44 dB, with an average value of just over 40 dB.

It might be argued that these results overestimate the EDRS because the upper end of each IIF adds only a small amount of importance that is spread over a relatively broad range of SNRs. According to our data, 9% (two-band) to 16% (one-band) of the intensity importance lies outside the IIF specified in ANSI S3.5-1997. Under relatively broadband conditions, this extra importance would not matter because

TABLE III. Importance values (in %/4 dB) for one-third octave bands based on the one-band data.

Band CF (Hz)	160	200	250	315	400	500	630	800	1000
Cutoff (Hz)	141–178	178–224	224–282	282–355	355–447	447–562	562–708	708–891	891–1122
Ctr. S/N									
45	−0.0047	−0.0064	−0.0104	−0.0170	−0.0255	−0.0364	−0.1008	−0.0986	−0.1017
41	−0.0047	−0.0064	−0.0104	−0.0170	−0.0255	−0.0364	−0.1008	−0.0986	−0.1017
37	−0.0047	−0.0064	−0.0104	−0.0170	−0.0255	−0.0364	−0.0336	−0.0329	−0.0339
33	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0336	0.0329	0.0339
29	0.0094	0.0128	0.0208	0.0339	0.0510	0.0727	0.1343	0.1315	0.1356
25	0.0141	0.0192	0.0312	0.0509	0.0765	0.1091	0.2351	0.2301	0.2372
21	0.0187	0.0257	0.0416	0.0679	0.1020	0.1454	0.3695	0.3616	0.3728
17	0.0281	0.0385	0.0625	0.1018	0.1530	0.2182	0.5038	0.4931	0.5084
13	0.0469	0.0641	0.1041	0.1697	0.2551	0.3636	0.6046	0.5917	0.6101
9	0.0656	0.0898	0.1457	0.2376	0.3571	0.5091	0.7053	0.6903	0.7117
5	0.1078	0.1475	0.2394	0.3904	0.5867	0.8363	0.7725	0.7561	0.7795
1	0.1265	0.1732	0.2811	0.4582	0.6887	0.9818	0.8397	0.8218	0.8473
−3	0.1265	0.1732	0.2811	0.4582	0.6887	0.9818	1.1084	1.0848	1.1185
−7	0.1078	0.1475	0.2394	0.3904	0.5867	0.8363	1.1755	1.1505	1.1862
−11	0.0937	0.1283	0.2082	0.3394	0.5101	0.7272	0.2351	0.2301	0.2372
−15	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Band CF (Hz)	1250	1600	2000	2500	3150	4000	5000	6300	8000
Cutoff (Hz)	1122–1412	1412–1778	1778–2239	2239–2818	2818–3548	3548–4467	4467–5623	5623–7079	7079–8913
Ctr. S/N									
45	−0.0900	−0.1107	−1.1734	−0.9339	−0.2402	−0.1724	−0.1034	−0.0543	−0.0319
41	−0.0900	−0.1107	−1.0617	−0.8450	−0.2402	−0.1724	−0.1034	−0.0543	−0.0319
37	0.0000	0.0000	−0.6147	−0.4892	−0.3002	−0.2155	−0.1292	−0.0679	−0.0399
33	0.1350	0.1661	−0.3911	−0.3113	−0.3603	−0.2586	−0.1551	−0.0814	−0.0478
29	0.2699	0.3322	−0.0559	−0.0445	−0.0801	−0.0575	−0.0345	−0.0181	−0.0106
25	0.4499	0.5536	0.5029	0.4003	0.2001	0.1437	0.0861	0.0452	0.0266
21	0.5623	0.6921	0.9499	0.7560	0.3202	0.2298	0.1378	0.0724	0.0425
17	0.7198	0.8858	1.1734	0.9339	0.4403	0.3160	0.1895	0.0995	0.0585
13	0.8548	1.0519	1.4528	1.1563	0.5604	0.4022	0.2412	0.1267	0.0744
9	0.9447	1.1626	1.5087	1.2008	0.6805	0.4884	0.2929	0.1538	0.0904
5	0.9897	1.2180	1.5646	1.2453	0.6805	0.4884	0.2929	0.1538	0.0904
1	1.0347	1.2734	1.7322	1.3787	0.6805	0.4884	0.2929	0.1538	0.0904
−3	1.1697	1.4395	1.9557	1.5566	0.7205	0.5171	0.3101	0.1629	0.0957
−7	1.2596	1.5502	1.6205	1.2897	0.8406	0.6033	0.3618	0.1900	0.1116
−11	1.3946	1.7163	1.0058	0.8005	0.9607	0.6895	0.4135	0.2172	0.1276
−15	0.0000	0.0000	0.0000	0.0000	1.0808	0.7757	0.4652	0.2443	0.1435

intelligibility would be at or near 100% throughout the added range. However, under narrow-band conditions, it could influence monosyllabic word test performance by as much 15% to 25%. The effect would be larger for speech with steeper P-I functions (e.g., sentences). It should be noted that concerns about an exact value for the EDRS can be avoided by using tables of values, such as Tables II and III, that account for the variations in importance with SNR.

The results of this study suggest that low-level noise, under some conditions, may have a more detrimental effect on performance than previously thought. This effect may not be noticeable when the bandwidth is broad because then the SNRs needed to produce less than perfect performance are relatively poor. But, with limited bandwidth, such as in the cases of hearing loss and hearing aids, SNRs greater than 15 dB may be needed in order to achieve the best speech recognition possible. Such results may also partly explain why some frequently cited studies found that intelligibility predictions for narrow-band speech at high SNRs are not as good as those for broadband speech at low SNRs (Beranek, 1947; Hirsh *et al.*, 1954).

D. Frequency effects

The results for each frequency band are based on less data and are more variable than the mean results. Nevertheless, it is evident that frequency had an impact on all three aspects of the IIF. In the case of the speech peaks, the highest frequency band had the highest effective peaks. This result is consistent with previous studies that indicate the physical peaks vary with frequency (Dunn and White, 1940; Cox *et al.*, 1988; Sherbecoe *et al.*, 1993) and also with studies that have suggested that some number of dB should be added to these physical peaks to derive the effective peaks (Studebaker and Sherbecoe, 1991; Studebaker *et al.*, 1993). The most prominent frequency effect, however, was that some

bands accumulated importance faster, and over a broader range of SNRs, than others. As a result, both the dynamic range of speech and the relative importance of the bands varied with SNR.

E. Band interactions

In this study, both positive and negative band interactions occurred when more than one speech band was present. These interactions may be of interest because of their possible implications for listener performance with multiband hearing aids. While only tentative conclusions can be drawn about these phenomena from these data, nevertheless, some observations can be made.

At low speech levels, subject performance in the two-band conditions was often greater than predicted by the sum of the independent contributions of the bands. This result appears to be a demonstration of the synergistic effects seen in other studies (Pollack, 1948; Hirsh *et al.*, 1954; Kryter, 1960; Grant and Braida, 1991; Müsch and Buus, 2001a, 2001b). However, unlike previous studies which found that these effects were bigger when the bands were more widely spaced in frequency, we observed greater synergism when the bands were relatively close together (although they were separated by at least an octave).

The negative interactions that occurred at high speech levels are also interesting. Such interactions are often said to result from spread of masking. If this was the cause of the rollover observed in this study, then it had to be produced by speech masking speech because the noise levels used were relatively low and were not varied. Further, if that is correct, then the results for bands 1 and 2 would reflect high-frequency spread and the results for bands 3, 4, and 5 would reflect low-frequency spread.

Close analysis of the data revealed some unexpected results, given the classical descriptions of high- and low-frequency spread of masking. The first was that the data patterns for bands 1 and 2, where high-frequency spread was more likely, were not fundamentally different from those for bands 3 to 5, where low-frequency spread was more likely. A second curious finding was that the negative effects of intensity began at relatively low speech levels. For example, rollover occurred when band 1 exceeded 87 dB SPL (43 dB SNR) and band 2 exceeded 75 dB SPL (31 dB SNR). Although this pattern is consistent with high-frequency spread of masking, the relatively low speech levels and gradual nature of the reductions at those levels suggest that a different process may be involved.

More consistent with expectations is the observation that band 3 produced less rollover than bands 4 and 5. Band 3 was closer to the low-frequency pedestal band. An increase in the amount of masking with increased frequency separation is a characteristic of ''remote'' or downward spread of masking (Bilger and Hirsh, 1956; ANSI S3.5-1969). However, it is not possible to reach any firm conclusions about this from the results of this study.

F. The AI equation

Equation (2) is commonly used to define the articulation index and to illustrate the basic features of how it is calculated.

$$A = P \int_0^\infty I(f)\,W(f)\,df. \tag{2}$$

This equation states that the articulation index (A) is found by dividing the audible spectrum into bands, multiplying each band's intensity weight $W(f)$ by its frequency weight $I(f)$, and then summing the products. W equals the effective speech peaks-to-noise ratio (in dB) in each band f, divided by the dynamic range of speech (in dB).

It is virtually always assumed that the dynamic range of speech is constant over frequency. However, even if the EDRS were allowed to vary with frequency, Eq. (2) would be misleading. This is because the importance of a given decibel of audibility within a band would be inversely influenced by the size of the dynamic range assigned to that band as a whole. This problem can be avoided by expressing the importance of any given frequency band-by-SNR area in relation to total importance, as done here, rather than only to the importance of the band in which it is found. Procedurally, this means that in order to obtain A, importance values for each area of audibility need to be extracted from a lookup table, such as Table II, or derived from a multiple regression equation, and then added.

G. Frequency-importance functions (FIF)

Finally, it should be noted that the decision to interpret these data as revealing a change in the IIF over frequency is fundamentally an arbitrary choice. It is equally possible to interpret the data as revealing a change in the FIF over SNR. From a practical standpoint, it does not matter which way the data are viewed. Except for possible smoothing artifacts, the values in Tables II and III would be essentially the same from either perspective.

The general character of the results, when viewed as a change in the FIF with SNR, can be seen by inspection of either the raw data (Figs. 1 and 2) or of the IIFs that were derived from that data (Figs. 4–7). Quantitative examples can be generated by plotting the data in Tables II and III as a function of frequency.

V. CONCLUSIONS

(1) IIFs for monosyllabic words are shaped like asymmetric peak functions. They rise sharply from the point where speech first becomes audible, reach a peak within 8 to 12 dB, and then gradually decline to zero over approximately the next 30 dB.
(2) IIFs for monosyllabic words have widths that are larger than 30 dB but, because that width varies with frequency and importance changes more slowly at higher SNRs, it is not possible to describe the EDRS accurately with a single number.
(3) The effective peaks of monosyllabic words vary in size with frequency, becoming larger at higher frequencies.

Overall, however, the sizes of the peaks obtained here are not inconsistent with the 15 dB used in ANSI S3.5-1997.

(4) The equation often used to define the AI may be misleading because it implies that the IIF has the same characteristics at all frequencies. This study suggests those characteristics vary with frequency.

(5) The data from this study can also be interpreted as evidence that frequency-importance functions vary with SNR.

ACKNOWLEDGMENTS

This work was supported in part by Research Grant Number 5 R01 DC 00154 from the National Institute on Deafness and other Communication Disorders, National Institutes of Health, and by funds from The Center for Research Initiatives and Strategies for the Communicatively Impaired (CRISCI) at the University of Memphis. The authors would also like to thank Catherine Gwaltney and Ginger Gray for their help in collecting the data and Joseph Matesich for his help with software development.

[1]In this paper the adjective ''effective'' is used in conjunction with the phrase ''dynamic range of speech.'' This adjective signifies that the dynamic range is based on measurable changes in human speech recognition performance rather than on physical measurements of the speech signal obtained with electronic equipment.

[2]The designation AI is used as a generic reference to any importance weighted audibility index, such as the Articulation Index.

[3]The relative, rather than the absolute, transfer function was used because the relative function makes no *a priori* assumptions about the size of the speech peaks or dynamic range. The absolute transfer function does (Studebaker *et al.*, 1993). However, if the importance values are expressed as percentages, as done here, the same result is obtained whether the absolute or the relative transfer function is used.

[4]Data smoothing almost always has the effect of increasing the apparent size of the effective speech peaks.

[5]An inherent characteristic of density importance data is that successive values tend to alternate between being too large and too small. Also, there is generally a correlation between the size of the successive ''errors.'' These features make the general trends of the data relatively obvious and the smoothing operation easier and less subject to arbitrary judgment than might be expected.

[6]The speech peaks level in ANSI S3.5-1969 was based on the observation that, between 1000 and 1400 Hz, 1 percent of the short-term (125 ms) speech levels spoken by male talkers are at least 12 dB above the long-term average level (French and Steinberg, 1947; Beranek, 1947). The 1997 standard reflects both the results of speech recognition studies that presented speech in talker-spectrum-matched noise at different SNRs (e.g., Studebaker *et al.*, 1993) and the procedures that were devised to calculate the speech transmission index (Steeneken and Houtgast, 1980).

ANSI (**1969**). ANSI S3.5-1969, ''American National Standard Methods for the Calculation of the Articulation Index'' (American National Standards Institute, New York).

ANSI (**1984**). ANSI S1.6-1984 R1997, ''American National Standard Preferred Frequencies, Frequency Levels and Band Numbers for Acoustical Measurements'' (American National Standards Institute, New York).

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).

Beranek, L. L. (**1947**). ''The design of speech communication systems,'' Proc. Inst. Rad. Eng. **35**, 880–890.

Bilger, R. C., and Hirsh, I. J. (**1956**). ''Masking of tones by bands of noise,'' J. Acoust. Soc. Am. **28**, 623–630.

Boothroyd, A. (**1990**). ''Articulation index: Importance function in the intensity domain,'' J. Acoust. Soc. Am. Suppl. 1 **88** S31.

Ching, T. Y. C., 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.

Cox, R. M., Matesich, J. S., and Moore, J. N. (**1988**). ''Distribution of short-term rms levels in conversational speech,'' J. Acoust. Soc. Am. **84**, 1100–1104.

Dubno, J. R., Dirks, D. D., and Schaefer, A. B. (**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.

Dunn, H. K., and White, S. D. (**1940**). ''Statistical measurements on conversational speech,'' J. Acoust. Soc. Am. **11**, 278–288.

Fletcher, H., and Galt, R. H. (**1950**). ''The perception of speech and its relation to telephony,'' J. Acoust. Soc. Am. **22**, 89–151.

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

Grant, K. W., and Braida, L. D. (**1991**). ''Evaluating the articulation index for auditory-visual input,'' J. Acoust. Soc. Am. **89**, 2952–2960.

Hirsh, I. J., Reynolds, E. G., and Joseph, M. (**1954**). ''Intelligibility of different speech materials,'' J. Acoust. Soc. Am. **26**, 530–538.

Houtgast, T., and Steeneken, H. J. M. (**1973**). ''The modulation transfer function in room acoustics as a predictor of speech intelligibility,'' Acustica **28**, 66–73.

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.

Kamm, C. A., Dirks, D. D., and Bell, T. S. (**1985**). ''Speech recognition and the articulation index for normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **77**, 281–288.

Kirk, R. E. (**1982**). *Experimental Design: Procedures for the Behavioral Sciences*, 2nd ed. (Brooks/Cole, Belmont, MA).

Kryter, K. D. (**1960**). ''Speech bandwidth compression through spectrum selection,'' J. Acoust. Soc. Am. **32**, 547–556.

Kryter, K. D. (**1962**). ''Methods for the calculation and use of the articulation index,'' J. Acoust. Soc. Am. **34**, 1689–1697.

Levitt, H. (**1982**). ''Speech discrimination ability in the hearing impaired: Spectrum considerations,'' in *The Vanderbilt Hearing Aid Report, State of the Art-Research Needs*, edited by G. A. Studebaker and F. H. Bess (Monographs in Contemporary Audiology, Upper Darby, PA), pp. 32–43.

Ludvigsen, C. (**1987**). ''Prediction of speech intelligibility for normal-hearing and cochlearly hearing-impaired listeners,'' J. Acoust. Soc. Am. **82**, 1162–1171.

Magnusson, L., Karlsson, M., and Leijon, A. (**2001**). ''Predicted and measured speech recognition performance in noise with linear amplification,'' Ear Hear. **22**, 46–57.

Müsch, H., and Buus, S. (**2001a**). ''Using statistical decision theory to predict speech intelligibility. I. Model structure,'' J. Acoust. Soc. Am. **109**, 2896–2909.

Müsch, H., and Buus, S. (**2001b**). ''Using statistical decision theory to predict speech intelligibility. II. Measurement and prediction of consonant-discrimination performance,'' J. Acoust. Soc. Am. **109**, 2910–2920.

Noordhoek, I. M., Houtgast, T., and Festen, J. M. (**1999**). ''Measuring the threshold for speech reception by adaptive variation of the signal bandwidth. I. Normal-hearing listeners,'' J. Acoust. Soc. Am. **105**, 2895–2902.

Pavlovic, C. V., and Studebaker, G. A. (**1984**). ''An evaluation of some assumptions underlying the articulation index,'' J. Acoust. Soc. Am. **75**, 1606–1612.

Pollack, I. (**1948**). ''Effects of high pass and low pass filtering on the intelligibility of speech in noise,'' J. Acoust. Soc. Am. **82**, 413–422.

Rankovic, C. M. (**1997**). ''Prediction of speech reception for listeners with sensorineural hearing loss,'' in *Modeling Sensorineural Hearing Loss*, edited by W. Jesteadt (Erlbaum, Mahwah, NJ), pp. 421–431.

Rankovic, C. M. (**1998**). ''Factors governing speech reception benefits of adaptive linear filtering for listeners with sensorineural hearing loss,'' J. Acoust. Soc. Am. **103**, 1043–1057.

Sherbecoe, R. L., Studebaker, G. A., and Crawford, M. R. (**1993**). ''Speech spectra for six recorded monosyllabic word tests,'' Ear Hear. **14**, 104–111.

Snodgrass, J. G. (**1977**). *The Numbers Game: Statistics for Psychology* (Oxford University Press, New York).

Steeneken, H. J. M., and Houtgast, T. (**1980**). ''A physical method for measuring speech-transmission quality,'' J. Acoust. Soc. Am. **67**, 318–326.
Studebaker, G. A., and Sherbecoe, R. L. (**1991**). ''Frequency-importance and transfer functions for recorded CID W-22 word lists,'' J. Speech Hear. Res. **34**, 427–438.
Studebaker, G. A., and Sherbecoe, R. L. (**1993**). ''Frequency-importance functions for speech recognition,'' in *Acoustical Factors Affecting Hearing Aid Performance*, 2nd ed., edited by G. A. Studebaker and I. Hochberg (Allyn and Bacon, Needham, MA), pp. 185–204.
Studebaker, G. A., Sherbecoe, R. L., and Gilmore, C. (**1993**). ''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.
Thornton, A. R., and Raffin, M. J. M. (**1978**). ''Speech-discrimination scores modeled as a binomial variable,'' J. Speech Hear. Res. **21**, 507–518.
Warren, R. M., and Bashford, J. A. (**1999**). ''Intelligibility of 1/3-octave speech: Greater contribution of frequencies outside than inside the nominal passband,'' J. Acoust. Soc. Am. **106**, L47–52.
Warren, R. M., Bashford, J. A., and Lenz, P. W. (**2000**). ''Intelligibility of bandpass speech: Effects of truncation or removal of transition bands,'' J. Acoust. Soc. Am. **108**, 1264–1268.
Zurek, P. M., and Delhorne, L. A. (**1987**). ''Consonant reception in noise by listeners with mild and moderate sensorineural hearing impairment,'' J. Acoust. Soc. Am. **82**, 1548–1559.

Bell clapper impact dynamics and the voicing of a carillon

N. H. Fletcher[a)]
Research School of Physical Sciences and Engineering, Australian National University, Canberra 0200, Australia

W. T. McGee
61 Calder Crescent, Holder 2611, Australia

A. Z. Tarnopolsky
School of Aerospace and Mechanical Engineering, University College, The University of New South Wales, Australian Defence Force Academy, Canberra 2600, Australia

(Received 31 August 2001; revised 22 November 2001; accepted 13 December 2001)

The periodic re-voicing of the bell clappers of the Australian National Carillon in Canberra provided an opportunity for the study of the acoustic effects of this operation. After prolonged playing, the impact of the pear-shaped clapper on a bell produces a significant flat area on both the clapper and the inside surface of the bell. This deformation significantly decreases the duration of the impact event and has the effect of increasing the relative amplitude of higher modes in the bell sound, making it ''brighter'' or even ''clangy.'' This effect is studied by comparing the spectral envelope of the sounds of several bells before and after voicing. Theoretical analysis shows that the clapper actually strikes the bell and remains in contact with the bell surface until it is ejected by a displacement pulse that has traveled around the complete circumference of the bell. The contact time, typically about 1 ms, is therefore much longer than the effective impact time, which is only a few tenths of a millisecond. Both the impact time and the contact time are reduced by the presence of a flat on the clapper. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1448517]

PACS numbers: 43.75.Kk [ADP]

I. INTRODUCTION

The National Carillon in Canberra was presented to the people of Australia by Britain to mark the Golden Jubilee of the establishment of Canberra as Australia's national capital in 1913. The carillon was opened by the Queen in 1970, a year that also marked the two hundredth anniversary of James Cook's famous voyage in which the east coast of Australia was mapped for the first time. The carillon itself was designed and built by John Taylor and Company of Loughborough in England, and is housed in an elegant white tower on tiny Aspen Island in Canberra's central Lake Burley Griffin. It has a total of 53 bells and its compass is four and a half octaves from G_2 to C_7. The pitch of the carillon is set a semitone lower than standard pitch at about $A_4 = 415$ Hz.

The carillon is played from a keyboard and pedalboard at a lower level in the tower (where there is also an identical practice clavier using undercut metal slats and resonator tubes). An auxiliary set of hammers or clappers on a small number of bells is used, under automatic control, to sound out the hours and quarter hours with Westminster chimes.

Experience has shown that the sound of the bells changes progressively as they are used. Initially the sound improves ''as the clappers get to know the bells,'' as one authority picturesquely puts it. After a few years of regular use, however, the sound becomes ''brighter'' with greater development of higher partials. While this high development of upper partials is perhaps desirable in isolated bells, their inharmonicity can make the sound discordant in a carillon where bells are played in harmony.

Examination of the bells shows that the change of tone is associated with the development of elliptical flat areas on both bell and clapper in the impact region. As usage demands, therefore, a re-voicing operation is undertaken in which the clapper surface is ground or filed to restore its curved shape at the impact site. This re-voicing makes the sound ''mellow'' once again. No attempt is made to reshape the impact site on the bell itself, since this might affect the tuning of its overtones. One of us (W. T. McG.) is responsible for the maintenance (including re-voicing) of the Australian National Carillon, and this afforded the opportunity to study several mechanical and musical aspects of the operation. Conveniently for our study, re-voicing had not been carried out for several years, as more urgent mechanical matters required prior attention. Consequently, the clappers exhibited a level of damage larger than normally acceptable, making the results of re-voicing much more clear. No work at all was required on the bells themselves, so this study is complementary to the recent work of Boutillon and David,[1] whose study concentrated upon the reconditioning of the bells in a carillon.

Rather surprisingly, in this particular carillon, the position on the inside of the bell against which the clapper strikes is not initially smooth. Indeed the bell-tuning operation has apparently been carried out on a lathe with a quite broad cutting tool, with the result that the inside surface bears regular circumferential grooves each 1–2 mm wide and almost 1 mm deep, separated from their neighbors by flat areas also about 1–2 mm wide. After the 30 years of playing that this

a)Electronic mail: neville.fletcher@anu.edu.au

instrument has received, the impact site on the bell wall has been almost flattened over an elliptical area somewhat larger than the flat developed on the clapper, this exaggeration being due to small variations in the position of the impact site because of mechanical slack in the mechanism.

The mechanism itself consists of a pear-shaped clapper of cast iron, mounted on a hinged rod. The original mechanism was modified some years ago, and the weight of each clapper is now balanced by a spiral spring. The clapper is drawn into impact with the bell by a flexible steel cable typically about 30 cm long that connects it to the rods, levers, and cables leading up from the clavier.

It is not our purpose here to survey what is known of the acoustics of bells or of carillon design. The interested reader will find appropriate material in a book by Fletcher and Rossing[2] and in a collection of papers edited by Rossing,[3] included among which is an extensive monograph by Bigelow.[4]

II. MEASUREMENTS

To examine the effects of re-voicing the clappers, three bells were selected for study: Nos. 9 (E_3), 29 (C_5), and 47 ($F_6^\sharp$), effectively spanning the compass of the instrument. The results for the midrange bell (C_5) will be reported in some detail and then comparable results for the other bells quoted.

Bell 29 is about 43 cm in diameter and 36 cm in height to its shoulder and is cast from bronze. Its mass is about 59 kg. The spherical part of its pear-shaped cast-iron clapper has a diameter of about 95 mm. The estimated mass of the clapper is about 5 kg. The impact damage flat on the clapper was approximately elliptical, with dimensions 17 mm×14 mm, while that on the inside of the bell was slightly larger at 20 mm×14 mm. The grooves on the bell surface had been plastically deformed so that the surface was essentially flat, though traces of the groove pattern did remain.

The re-voicing itself was carried out by hand and eye, using a powered angle grinder and then a file for finishing. The resultant clapper surface was of approximately uniform spherical curvature but had some file marks across it of approximate depth 0.1 mm.

Since the plastic washers used in the bell mount isolated it electrically from the clapper, it was possible to examine the contact time τ simply by using the impact as a switch in a simple electrical circuit consisting of a 9 V cell and a 1000 Ω resistor in series and displaying the voltage across the resistor on a storage oscilloscope. When a simple impact was achieved—a point to be considered again later—its duration was about 0.6 ms. The same experimental arrangement was used to measure contact time after re-voicing, the typical duration then being about 1.0 ms, though becoming as long as 1.3 ms for particularly gentle strikes.

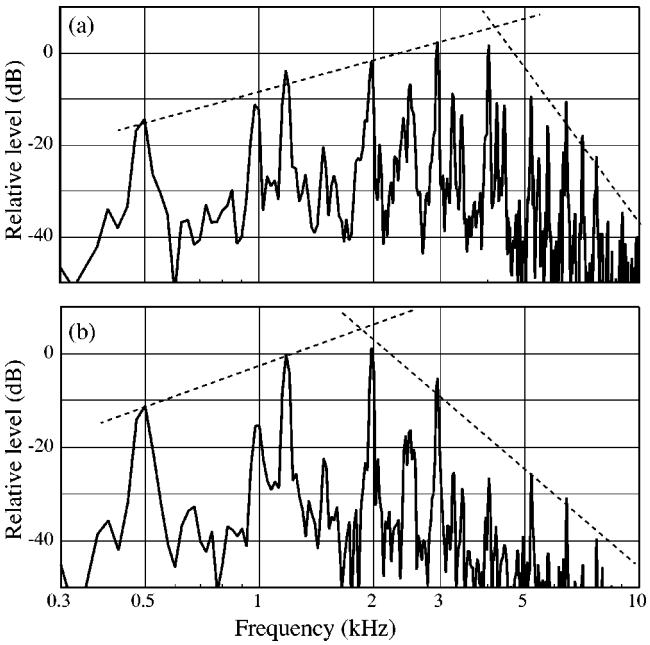


FIG. 1. (a) Sound spectrum of bell 29 before re-voicing; (b) sound spectrum of the same bell after re-voicing.

An interesting feature of this experiment, which will later be seen to be significant, is that, while impacts widely separated in time so that the bell has ceased vibrating were quite clean and reproducible, impacts in rather quick succession with only a few seconds separation displayed jitter in the form of repeated very brief contacts. This jitter can be attributed to the effects of residual vibration of the bell.

The radiated sound signal was recorded using a microphone and tape recorder located in the bell chamber about 4 m from the bell. The microphone position was the same for recordings before and after re-voicing so that comparison was possible irrespective of resonances or other effects within the bell chamber. Recorded signals from a number of bell strokes before and after re-voicing were later subject to frequency analysis using a fast Fourier transform technique, the sample being selected to commence 40 ms after the initiating impulse.

A typical pair of spectra for bell 29 is shown in Fig. 1. The spectral envelope can be adequately described in terms of the two straight lines shown, and the frequency f^* at the intersection of these lines gives a measure of the ''brightness'' of the sound. This frequency is about 4 kHz before re-voicing and about 2 kHz afterwards. The distribution of

TABLE I. Experimental results.

Bell	Bell diam (cm)	Bell mass (kg)	Clapper diam (mm)	Clapper flat diam (mm)	τ before (ms)	τ after (ms)	f^* before	f^* after
9(E_3)	120	950	200	18	1.5	3.5	900 Hz	600 Hz
29(C_5)	43	59	95	15	0.6	1.2	4 kHz	1.8 kHz
47($F_6^\sharp$)	20	7.7	56	7	0.6	0.8	>4 kHz[a]	>4 kHz[a]

[a]Uncertain because of wind noise and other problems.

modes within the spectral envelope is determined by the design and tuning of the bells[2,4] and need not be considered here though it is, of course, the primary concern of the designer and bell founder.

Similar analyses were carried out for the other two bells, and the complete results are summarized in Table I. The clapper flat diameter quoted is the geometric mean of the two diameters of the contact ellipse. It can be seen that the behavior scales quite consistently for all three bells.

III. THEORY

Two aspects of the impact behavior require theoretical investigation. The first is the impact itself, including contact time, behavior of the contact force, and the way in which these quantities depend upon the size of the damage flat and upon parameters such as impact speed. The second is the vibrational excitation and sound output to be expected from this type of impact.

The whole subject of the impact of one body upon another is of great practical importance, and has been the subject of much theoretical and experimental investigation, dating back to the classic work of Hertz in the nineteenth century. An excellent exposition with extensive references has been given by Goldsmith.[5]

A rigorous theory of impact in the present case necessarily involves the complex elastic deformation of the bell and clapper during the impact and also the vibration and bodily motion of the bell during that time. Such an analysis would be extremely complicated and is in any case unnecessary for the present problem. The analysis to be presented here therefore includes only enough detail to establish the effects of the major physical parameters and to give a reasonable approximation to the numerical values involved.

A. Hertzian impact theory

Most reasonably simple treatments of impact are based upon Hertz's assumption that the impact takes place sufficiently slowly that a static treatment of the elastic distortions is adequate. His theory covers the case of impact of spheres on spheres and of spheres on plane objects, but could reasonably be extended to the more complex geometry of a sphere with a contact flat. Following the treatment in Chapter 4 of Goldsmith,[5] we find that, for a sphere of mass m and radius R impacting on a very massive plate at a speed V, the Hertz theory gives a contact time

$$\tau_H \approx 4.5\left[\frac{(\delta_1+\delta_2)^2 m^2}{RV}\right]^{1/5}, \tag{1}$$

where

$$\delta_1=\frac{1-\mu_1^2}{\pi E_1}; \quad \delta_2=\frac{1-\mu_2^2}{\pi E_2} \tag{2}$$

and μ_1, μ_2 are the Poisson's ratios and E_1, E_2 the Young's moduli of the sphere and plate, respectively. During the impact, the force follows a curve like

$$F \approx F_{\max}\sin(\pi t/\tau_H), \tag{3}$$

where

$$F_{\max}\approx 0.44\frac{m^{3/5}R^{1/2}V^{6/5}}{(\delta_1+\delta_2)}. \tag{4}$$

Equations (1)–(4), which apply only for an undamaged clapper, serve as a check for the extended but less accurate theory to be developed in Sec. III B.

These quasistatic equations can apply only in the limit of very slow impacts which allow any elastic vibrations generated to dissipate completely in a negligible fraction of the impact time. For more rapid impacts it is necessary to consider the effects of wave generation and reflection from the boundaries of the impacting objects. This is difficult enough in the case of a small sphere impacting on a clamped rod[5] and becomes impossibly complicated for more complex geometries. For very rapid impacts, of the type exemplified by impact of a small sphere on a large object such as an extended plate, the situation is somewhat simplified, since the deformation of the sphere can be taken to be quasistatic, and the generation of elastic waves in the plate can be approximated by a simple wave impedance, provided there is not time during the impact for any reflections from the plate boundaries to return to the impact site. The latter assumption is a reasonable approximation for the case of clapper impact on a bell, and will be the basis of the theory developed in Sec. III B.

B. An extended impact theory

It is necessary now to formulate a simple theory from which predictions can be made of the impact behavior of a sphere with an initial flat upon the surface of a bell. The geometry assumed during the impact is as shown in Fig. 2. The coordinate of the center of the clapper, assumed spherical for simplicity, relative to a fixed plane is y_1, and that of the surface of the bell, assumed plane, is y_2, the coordinate of this surface before the impact being y_0. During the im-

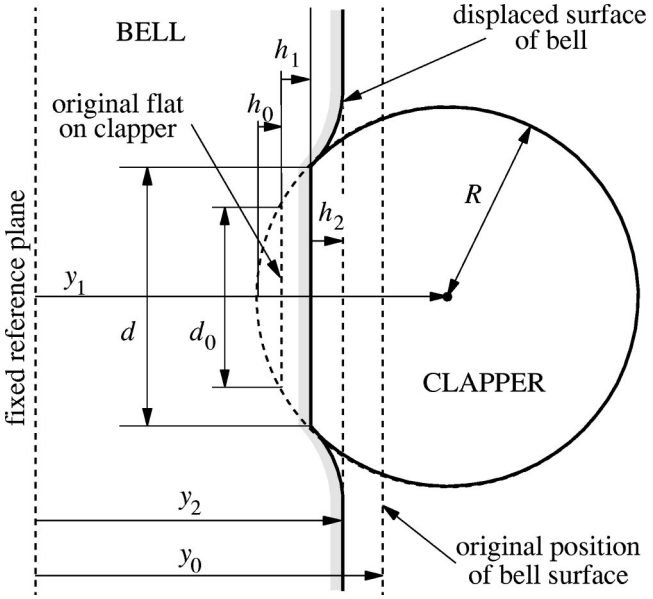


FIG. 2. Assumed geometry of the bell and clapper during impact. The bell is displaced from its initial position y_0 because of elastic wave generation, while the clapper deforms the surface elastically and is itself deformed, the diameter of the contact flat increasing from its initial value d_0 to d.

pact, the clapper penetrates the bell surface by a distance h_2 and the depth of the flat below the extension of the spherical surface increases from h_0 to h_0+h_1. At the same time the diameter of the contact flat increases from its initial magnitude d_0 to a new time-dependent magnitude d. It is assumed that all distortions are elastic rather than plastic, at least during the course of a single impact event.

The first approximation to be made is to assume that the bell is large enough and the impact short enough that there are no significant reflections during the contact time. Since the bell has axial symmetry, this assumption is equivalent to the proviso that there is not enough time during the impact for a wave to travel round the circumference of the bell and return to the impact point before the clapper rebounds, a matter that will be discussed later. The large-scale behavior of the bell during impact can now be approximated by

$$F=-Z\frac{dy_2}{dt}, \tag{5}$$

where F is the force imposed by the clapper impact, y_2 is the coordinate of the bell surface as shown in Fig. 2, and Z is the characteristic impedance for flexural waves on the bell. The total mass of the bell is sufficiently much larger than that of the clapper that it is reasonable to ignore its bodily displacement. Exact evaluation of Z is difficult, because of the complex geometry of the bell, but an approximation can be made by assuming it to behave like an infinite flat plate. An analysis has been given by Skudrzyk,[6] who shows that, for the case of a plate or iron or soft steel, the characteristic impedance is

$$Z_\infty \approx 100b^2 \text{ kg s}^{-1}, \tag{6}$$

where b is the thickness of the plate in millimeters. Because the elastic parameters of bell bronze and cast iron are very similar, we expect this result to be a good approximation for the bell material. [For other materials, Z_∞ is proportional to $(E\rho)^{1/2}$, where E is the Young's modulus and ρ the material density.] Since, however, the clapper strikes the bell close to a free edge, the effective impedance Z is expected to be about $0.5Z_\infty$. The thickness of the bell wall is not uniform, but is in the range 20–30 mm for bell 29. Adopting this approximation as a guide, therefore, we expect that Z is in the range $2-5\times10^4$ kg s^{-1} for this bell. In some of the calculations to follow, a value of 3×10^4 kg s^{-1} will be assumed.

Referring now to Fig. 2, if d_0 is the diameter of the impact damage flat, then this can be related to the parameter h_0 by

$$h_0 \approx \frac{d_0^2}{8R}, \tag{7}$$

where R is the radius of the clapper ball, or more precisely its radius of curvature, and the approximation is good provided $h_0 \ll R$. Similarly, the parameter h_1, which is the extent of the compression, is related to the diameter d of the contact flat during the impact and to the initial geometric parameter h_0 by

$$h_1 \approx \frac{d^2}{8R}-h_0, \tag{8}$$

provided $h_1+h_0 \ll R$, as is the case in practice. Calculation of the elastic strain in the clapper is very complicated, but it is adequate for our present purposes to assume that it is distributed over a volume of diameter d and depth d/K_1, where K_1 is rather greater than unity. The total force F on the impact surface then approximately satisfies

$$\frac{dF}{dh_1}=\frac{K_1}{d}\left(\frac{\pi d^2}{4}\right)\frac{E_1}{1-\mu_1^2}=\frac{K_1 d}{4\delta_1}, \tag{9}$$

where E_1 and μ_1 are the Young's modulus and Poisson's ratio, respectively, of the clapper material and δ_1 is defined by Eq. (2). The uncertainty implicit in this result is encapsulated in the factor K_1. Its magnitude will be derived later by comparison with the more accurate Hertz theory for a case where they overlap. Substituting Eq. (8) into Eq. (9) and integrating gives

$$F=K_1\frac{(2R)^{1/2}}{3\delta_1}[(h_1+h_0)^{3/2}-h_0^{3/2}]. \tag{10}$$

The elastic deformation of the bell surface presents a similarly difficult problem but, to the same degree of approximation, and by the same arguments used to treat the clapper, we can write

$$F=K_2\frac{(2R)^{1/2}}{3\delta_2}[(h_2+h_0)^{3/2}-h_0^{3/2}], \tag{11}$$

where K_2 is another constant of order unity. It is to be expected that $K_1 \approx K_2$. From Eqs. (10) and (11),

$$(h_2+h_0)^{3/2}=\frac{K_1\delta_2}{K_2\delta_1}(h_1+h_0)^{3/2}+\left(1-\frac{K_1\delta_2}{K_2\delta_1}\right)h_0^{3/2}. \tag{12}$$

There is also a further geometrical connection between the coordinates, namely

$$h_0+h_1+h_2=y_2-y_1+R. \tag{13}$$

Note that, in all these equations, h_0, h_1, and h_2 have been defined so as to be always positive, and F must also be positive since there is no adhesion at the contact.

Finally, the motion of the center of mass of the clapper obeys

$$\frac{d^2y_1}{dt^2}=\frac{F}{m}, \tag{14}$$

where the mass m of the clapper is rather greater than the spherical mass because of its pear shape and attached mechanism. Equations (5)–(14) now define the motion of the whole system.

C. Numerical solution

For a numerical solution, Eqs. (14) and (5) are written in the first-order form

$$\frac{dy_1}{dt}=z, \tag{15}$$

$$\frac{dz}{dt}=\frac{F}{m}, \tag{16}$$

$$\frac{dy_2}{dt}=-\frac{F}{Z} \tag{17}$$

with boundary conditions $y_2=y_0$, $y_1=y_0+R-h_0$, $dy_1/dt=-V$, $dy_2/dt=0$, and $h_1=h_2=0$, where y_0 is an arbitrary constant and h_0 is defined by the diameter d of the initial flat by Eq. (7). It is necessary to write the force F exclusively in terms of the coordinates y_1 and y_2, and to do this h_2 must be eliminated from Eqs. (12) and (13) and the resulting equation solved numerically for h_1. The three equations (15)–(17) then serve as the basis for the computation.

In the carillon under study, the bells are made of cast bronze for which $E_2=1.05\times10^{11}$ Pa and $\mu_2=0.36$, while the clappers are cast iron with $E_1=1.2\times10^{11}$ and $\mu_1=0.27$.[7] These values are sufficiently similar that it is reasonable to assume equality, so that $\delta_1=\delta_2$. In addition, since exact values of the constants K_1 and K_2 in Eqs. (10) and (11) are unknown, it makes sense to assume that $K_1=K_2=K$. These simplifying assumptions then lead to the result $h_1=h_2$ so that, from Eq. (13),

$$h_1+h_0=h_2+h_0=\frac{R+y_2-y_1}{2} \tag{18}$$

and

$$\delta_1=\delta_2=2.4\times10^{-12}\,\mathrm{Pa}^{-1}. \tag{19}$$

These assumptions lead to a great simplification in the subsequent calculations.

The one major formal uncertainty in the extended theory is encapsulated in the value of the elastic parameter K, when all that is known is that it is of order unity. The accompanying uncertainty can, however, be resolved by comparing the computed result for contact time in the case $d_0=0$ and $Z=\infty$ with the known accurate results of the Hertzian treatment for this case as given by Eq. (1). This comparison can be made by assuming that $Z=\infty$ in the simple model and adjusting K to obtain agreement, and leads to the result $K\approx5$. This value is then used in subsequent evaluations.

The theory can now be used to investigate the effect of impact flat diameter upon impact time. Because of uncertainty in the effective value of Z for the bell, and to illustrate the importance of this parameter, a range of values is studied. Once again, the calculations all refer to bell 29.

Figure 3 shows the calculated behavior of the force between the clapper and the bell for an impact velocity of 1 $\mathrm{m\,s^{-1}}$ and three assumed values of the impedance parameter Z. For the case $Z=\infty$ the force curve is symmetrical about its midpoint, in agreement with the Hertzian result (3), while for a finite value of Z, as in reality, the force curve has a longer tail. This asymmetry increases as the impedance Z assumes lower values. It is encouraging to note that these same features are seen in the shapes of curves calculated for the impact of a sphere on a plate of infinite size, as shown in Fig. 71 of Goldsmith.[5] Because of the asymmetry of the curves it is helpful to define the impact time τ^* to be the time from initial contact to the maximum in the force curve. For a

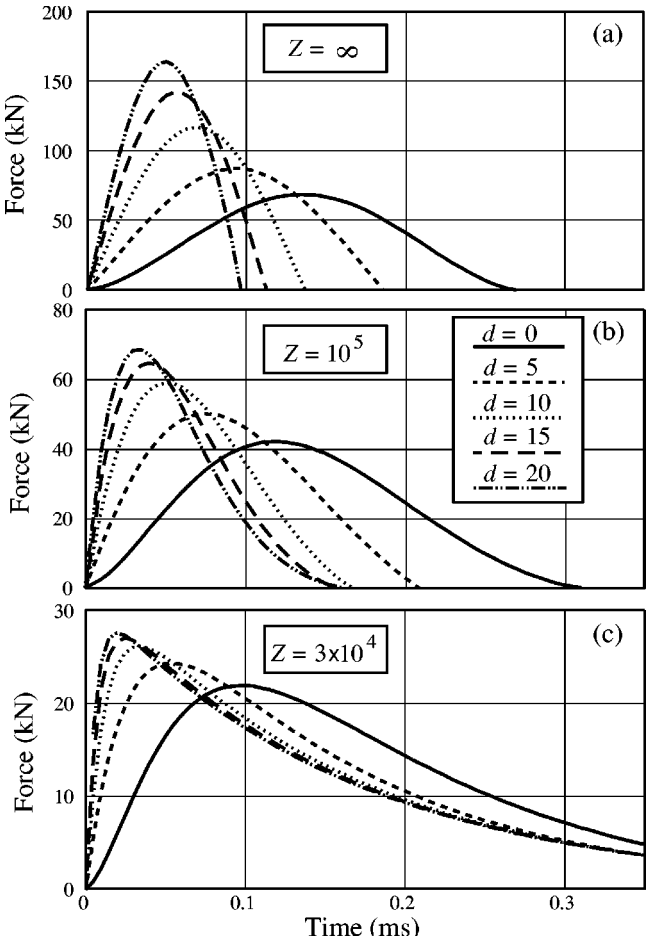


FIG. 3. Calculated time evolution of force F between the clapper and the bell for an impact velocity of 1 $\mathrm{m\,s^{-1}}$ and three different impedance values Z. The diameter of the impact damage flat is shown in millimeters as a parameter. In the case $Z=3\times10^4$ $\mathrm{kg\,s^{-1}}$, the clapper remains in contact with the bell surface.

Hertzian impact ($Z=\infty$) clearly $\tau^*=\tau_H/2$, while τ^* is less than half the contact time for finite values of Z.

Figure 4 shows the calculated impact time τ^*, as derived from the curves of Fig. 3, as a function of impact flat

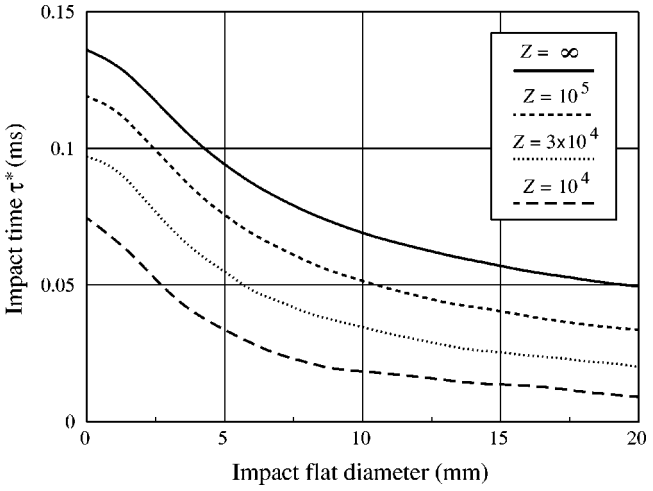


FIG. 4. Calculated impact time τ^* for a spherical clapper of mass 5 kg on a bell surface with effective impedance Z, as a function of the diameter of the damage flat on the clapper surface.

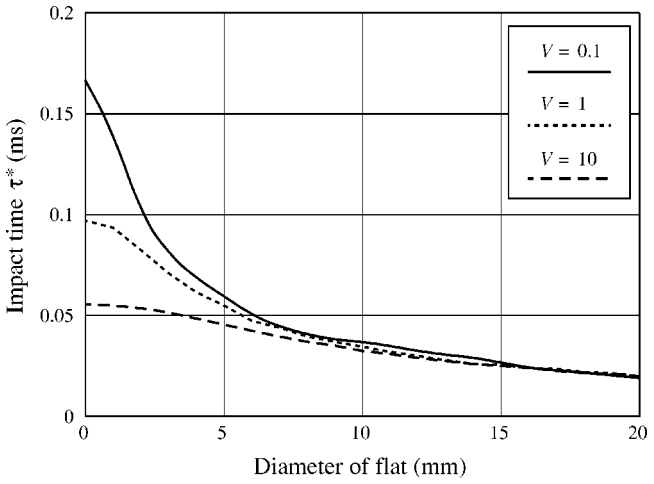


FIG. 5. Calculated variation of the impact time τ^* as a function of impact damage flat diameter for a range of impact speeds, shown in meters per second as a parameter. It is assumed that $Z=3\times10^4$ kg s^{-1}, at which impedance value the clapper comes to rest against the bell after impact.

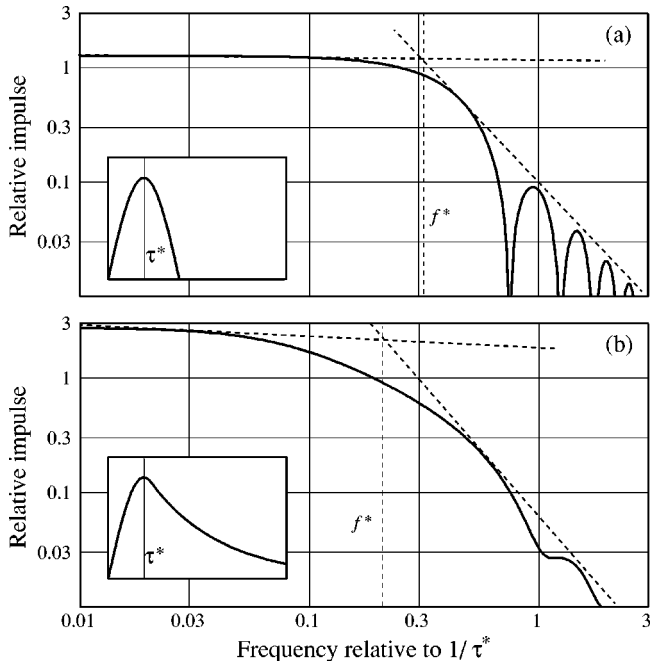


FIG. 6. Fourier transform of the impact force $F(t)$ for (a) a very high bell impedance Z, giving a Hertzian impact as in Fig. 3(a); (b) a low bell impedance giving an asymmetric impact and infinite contact time as in Fig. 3(c). The time variation of the force is indicated in the inset in each case.

diameter for three values of the impedance Z, one of which is the infinite impedance implicitly assumed in the Hertz model. The behavior for $Z=10^5$ kg s^{-1} is not very different from that for $Z=\infty$, though impact times are a little shorter and total contact times a little longer. For Z values smaller than about 5×10^4 kg s^{-1}, however, and specifically for the value 3×10^4 kg s^{-1} considered appropriate for this bell, the clapper actually comes to rest against the bell surface after the impact. The importance of this prediction will be examined in Sec. III D.

Finally, Fig. 5 shows the effect of clapper impact velocity on impact time τ^* for a range of values of impact flat diameter. As already discussed, it is assumed that $Z=3\times10^4$ kg s^{-1} for this bell, so that the clapper remains in contact with the bell after impact and the contact time τ is infinite. For the undamaged clapper, there is a significant shortening of impact time as the impact velocity is increased, but once a flat of moderate diameter develops τ^* becomes independent of impact velocity. This feature is another reason for the clapper voicing program.

D. Other bells

As mentioned earlier, similar measurements and revoicing procedures were carried out on two other bells, one much larger and one much smaller than bell 29. The measurements on all three bells are shown in Table I. Here τ is the total contact time from electrical measurements and f^* is the spectral turnover frequency as shown in Fig. 6. The results for the larger bell, No. 9, are broadly consistent with those analyzed in detail, bell 29. Those for bell 47 are less decisive, since the flat was smaller and the spectral recordings were obscured by wind noise and other problems. It would be of interest to undertake a complete study of spectral balance across the whole carillon, but this was not attempted.

IV. CONTACT TIME DISCREPANCY

There is one aspect in which the theory appears to be in substantial disagreement with experiment, and that is the actual magnitude of the contact time during impact. The experimental contact time for bell 29 ranges from 0.6 ms for the damaged clapper to 1.0–1.2 ms when it has been revoiced. If the bell impedance Z is 10^5 kg s^{-1} or higher, so that the clapper rebounds after contact, then the calculated contact times are about 0.15 and 0.3 ms, respectively—about one-quarter of the measured times. If the bell impedance is substantially lower, as analysis indicates, then the calculated contact times become infinite. While it might be possible to find a precise value of Z giving approximate agreement with experiment for contact times, this balance would be unrealistically sensitive.

One might be tempted to question the reliability of the electrical contact measurements. Other investigators, however, have used almost identical techniques to measure contact times and have obtained results that agree in general magnitude with our own. Jones[8] long ago made measurements on a large bell in place in a chime and found a contact time of about 0.8 ms. More recently Grützmacher *et al.*[9] measured contact times in the range 0.6–1.5 ms in a laboratory setup, in approximate agreement with our results, and also found a decrease in contact time with increasing impact velocity.

An explanation of this discrepancy is actually simple, and is suggested by the observed instability of the contact measurements when the bell is already vibrating. Suppose that the impedance Z that the bell presents to the clapper is less than about 5×10^4 kg s^{-1}, as is appropriate from the plate analogy discussed previously. Then after impact the clapper will remain in contact with the bell surface until it is

dislodged by some mechanical impulse from the bell. From Fig. 3, the clapper impact has a sharp rise time τ^* followed by a slow decay, and thus launches a transverse pulse of this shape in both directions around the sound-bow of the bell. These pulses return to the clapper position after making a complete circuit of the bell, during which they are distorted by dispersion effects, and dislodge the clapper from contact with the surface as they return to the impact site. The clapper then moves away from the bell surface and is further accelerated by the return spring of the action. If the clapper has an impact flat, then the impact time τ^* is shorter, and the generated pulse wave-packet is higher in frequency. The phase velocity of flexural waves, however, is proportional to the square root of the frequency, and the group velocity is twice the phase velocity, so that this higher-frequency pulse travels around the bell more quickly and dislodges the clapper after a smaller delay.

These statements about phase and group velocities follow from the equation for bending wave propagation on a plate, which is

$$\frac{\partial^2 z}{\partial t^2} = -S\nabla^4 z, \tag{20}$$

where z is the normal displacement and S is the plate bending stiffness divided by its mass per unit area. If the wave has the form $z = A\sin(\omega t - kx)$, then Eq. (20) shows that $\omega = S^{1/2}k^2$ and, since the phase velocity is $c = \omega/k$, this gives $c = S^{1/4}\omega^{1/2}$. The group velocity c', however, is given by $c' = \partial\omega/\partial k$ which is just $2c$.

To confirm this explanation of long contact times, it is necessary to introduce quantitative considerations. From the analysis of Skudrzyk,[6] the phase velocity of flexural waves of frequency f on an infinite plate is

$$c = \left(\frac{\pi Z f}{4M}\right)^{1/2}, \tag{21}$$

where M is the plate mass per unit area and Z is the characteristic impedance given by Eq. (6). For bell 29, the wall thickness is about 20 mm so that $Z = 3\times10^4$ kg s^{-1}. At 2 kHz, which is the frequency of maximum excitation as judged from the sound spectrum for the voiced clapper, Eq. (21) gives a phase velocity of about 540 m s^{-1} and a group velocity of 1080 m s^{-1}. This gives a pulse transit time of about 1.2 ms, in fairly good agreement with the experimental contact time of 1.0–1.2 ms for the voiced clapper. For the unvoiced clapper, the excitation maximum is at about 4 kHz, giving a pulse return time of about 0.8 ms, which is comparable with, though a little longer than, the experimental value of 0.6 ms.

It is possible, however, to estimate these propagation velocities quite independently of the validity of the infinite-plate impedance approximation. In the fundamental ''hum'' mode, the bell vibrates with two nodal diameters,[2] and the mode can be considered to be a superposition of two counterpropagating waves, each with two wavelengths around the circumference of the sound-bow. Since the bell circumference is about 140 cm at the strike position, this gives a phase velocity of about 350 m s^{-1} at this frequency, and hence a group velocity of about 1400 m s^{-1} at 2 kHz, in moderately good agreement with the above-given estimate. The contact times predicted from this estimate are about 0.9 ms for the voiced clapper and 0.6 ms for the unvoiced clapper, again in fairly good agreement with experiment.

V. SOUND GENERATION

Sound generation by the bell is a combination of the excitation event and subsequent sound radiation. Since the bell itself is not modified during the voicing process, its mode frequencies, impedances, and radiation efficiencies remain unaltered, and it is necessary only to consider the excitation event. The above-developed theory predicts an impact force $F(t)$ with a time evolution that depends upon the impedance Z of the bell wall as shown in Fig. 3. The spectral shape of this impact event is given by its Fourier transform, which is written most conveniently, to within a constant factor, as

$$F(\omega) = \left[\left(\int_0^\infty F(t)\cos\omega t\,dt\right)^2 + \left(\int_0^\infty F(t)\sin\omega t\,dt\right)^2\right]^{1/2}. \tag{22}$$

The shape of this function, plotted in the same way as the experimental results of Fig. 1, is shown in Fig. 6 for two different cases: one with large Z, giving a nearly Hertzian impact, and one with small Z, giving a lingering contact. In both cases the envelope of the spectrum consists of two straight lines intersecting at a frequency close to $0.2/\tau^*$ where τ^* is the impact time as previously defined and indicated in the insets. When allowance is made for a radiation efficiency that increases smoothly with increasing frequency, there is close qualitative similarity between the spectral envelope of Fig. 6 and that of the bell sound in Fig. 1, for both unvoiced and voiced clappers.

Comparison with experiment. In making comparisons between theoretical predictions and experimental measurements, it must be borne in mind that this was a field experiment rather than a laboratory experiment. This meant that several important parameters had to be estimated rather than measured in a controlled way. With this proviso, the predictions of the theory are in acceptably good agreement with experimental results.

The turnover frequencies f^* predicted for the bell 29 on the basis of the theory set out above and the impact times given in Fig. 4, assuming that $Z = 3\times10^4$ kg s^{-1}, are about 8 kHz for the unvoiced clapper and 2 kHz for the voiced clapper, compared with the measured values of 4 and 1.8 kHz, respectively. While far from exact, this agreement should be considered as satisfactory in a field experiment rather than a laboratory experiment, particularly since the bell impedance parameter Z, the clapper mass m, and the impact velocity V have all been estimated rather than measured. All of these parameters affect the impact time τ^* significantly, as indicated in Figs. 4 and 5. Choice of a rather larger value for Z would improve the agreement.

The measured behavior of the other bells investigated scales in just about the expected manner, as indicated in Table I, though the experimental data for bell 47 is inconclu-

sive. For bell 9, the linear size scaling is by a factor 2.8, so that, by Eq. (6), the scaled value of Z should be about 2.4×10^5 kg s^{-1}. The scaling of the clapper diameter indicates a mass m of about 46 kg. Inserting these values in the theory, together with the impact flat diameter from Table I, gives τ^* values for the damaged and undamaged clapper of 0.1 and 0.23 ms, respectively. The resulting predicted turnover frequencies are 2000 and 870 Hz, respectively, compared with the measured values of 900 and 600 Hz. This agreement is similar to that for bell 29 and would again be improved by choice of a larger value for Z.

It is not appropriate in such a field experiment to seek to refine the agreement between theory and measurement by varying the parameters. Bearing in mind the many approximations involved, the agreement is fairly satisfactory.

VI. CONCLUSIONS

This study of the dynamics of bell clapper impact has elucidated many features of the problem that were presumably solved by the designers and builders of historic bells and carillons on the basis of experience founded upon trial and error. It is somewhat surprising that the bell clapper actually comes to rest against the bell surface until it is pushed away by a returning vibration pulse. On reflection, however, such a design results in the transfer of maximum energy from the clapper to the bell, and thus the achievement of the loudest possible sound for a given impact velocity. The parameter that principally determines this behavior, assuming the bell design to be fixed, is the clapper mass—too light a clapper will bounce, as well as having rather little energy to transfer.

Bell founders have long recognized that the sound of a carillon changes rather rapidly over the first year or so of its use, and then becomes more stable. This change is associated with the development of impact flats on the clappers, and initially proceeds quite rapidly until the clappers, and the inside surface of the bell, become more resistant to plastic deformation through work hardening. A voicing process that restores the shape of the clappers brings the sound closer to its original state, and the work-hardened clappers are now more resistant to plastic deformation. The work hardening itself, which has almost no effect on the elastic moduli, has no direct influence on impact dynamics or sound. Generally nothing is done about the deformation of the bell surface, though some carillons do allow a 180° rotation of the bells to the other equivalent impact position, a routine that ultimately halves the rate of bell deformation.

It is certain that a carillon sounds different to the musical ear and to the carillonneur after it has been re-voiced. It is interesting to determine what are the qualities of a re-voiced carillon that make it an improvement on its state before the operation.

The effects of re-voicing are two: (1) the impact time is lengthened so that the turnover frequency is reduced, giving a more ''mellow'' sound; and (2) the impact time becomes more dependent upon the impact velocity, so that ''soft'' notes have relatively little harmonic development and ''loud'' notes are both louder and ''brighter,'' as in most other musical instruments. These considerations are those that principally motivate the carillon curator to undertake the operation. In this, a carillon is rather different from an isolated bell or from a peal of bells, since the music it is called upon to produce involves harmonies, rather than simply melodic patterns.

Experience with analysis of the re-voicing of this particular carillon shows that details of the clapper impact are important. The analysis also points up the importance of matching each clapper to its bell in order to give a well-balanced tone quality and loudness across the compass. It is these things, as well as the design of the bells themselves, that distinguish a really fine carillon.

[1] X. Boutillon and B. David, ''Assessing tuning and damping of historical carillon bells and their changes through restoration,'' Appl. Acoust. (in press).

[2] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*, 2nd ed. (Springer, New York, 1998), Chap. 21.

[3] *Acoustics of Bells*, edited by T. D. Rossing (Van Nostrand Reinhold, New York, 1984).

[4] A. L. Bigelow, *The Acoustically Balanced Carillon* (School of Engineering, Princeton University, Princeton, NJ, 1961). (Largely reprinted in Ref. 3.)

[5] W. Goldsmith, *Impact* (Arnold, London, 1960).

[6] E. Skudrzyk, *Simple and Complex Vibratory Systems* (Pennsylvania State University Press, University Park, PA, 1968), Chap. 9.

[7] *A Physicist's Desk Reference*, 2nd ed., edited by H. L. Anderson (American Institute of Physics, New York, 1989), p. 37.

[8] A. T. Jones, ''The strike note of bells,'' J. Acoust. Soc. Am. **1**, 373–381 (1930). (Reprinted in Ref. 3).

[9] M. Grützmacher, W. Kallenbach, and E. Nellessen, ''Akustische Untersuchungen an Kirchenglocken,'' Acustica **16**, 34–45 (1965/66). (Reprinted in translation as ''Acoustical investigations on church bells'' in Ref. 3.)

Ultrasonic wave speed measurement using the time-delay profile of rf-backscattered signals: Simulation and experimental results

Fernando R. Pereira
Instituto Politecnico, IPRJ, Universidade do Estado do Rio de Janeiro, UERJ, Caixa Postal 97282, 28601-970, Nova Friburgo, RJ, Brazil

João C. Machado and Wagner C. A. Pereira
Biomedical Engineering Program-COPPE/Federal University of Rio de Janeiro, P.O. Box 68510, Rio de Janeiro, RJ 21945-970, Brazil

(Received 20 April 2001; accepted for publication 25 November 2001)

Conventional methods determine the ultrasonic wave speed measuring the medium path length propagated by a pulsed wave and the corresponding time-of-flight. In this work, the wave speed is determined without the need of the path length. A transmit transducer sends a pulsed wave into the medium (wave speed constant along the beam axis) and the backscattered signal is collected by a hydrophone placed at two distinct positions near the transmitted beam. The time-delay profile, between gated windows of the two rf-signals received by the hydrophone, is determined using a cross-correlation method. Also, a theoretical time-delay profile is determined considering the wave speed as a parameter. The estimated wave speed is obtained upon minimization of the rms error between theoretical and experimental time-delay profiles. A PZT conically focused transmitting transducer with center frequency of 3.3 MHz, focal depth of 30 mm, and beam full width (−3 dB) of 2 mm at the focus was used together with a PZT hydrophone (0.8 mm of aperture). The method was applied to three phantoms (wave speed of 1220, 1540, and 1720 m/s) and, *in vitro*, to fresh bovine liver sample, immersed in a temperature-controlled water bath. The results present a relative speed error less than 3% when compared with the sound speed obtained by a conventional method.
© *2002 Acoustical Society of America.* [DOI: 10.1121/1.1445787]

PACS numbers: 43.80.Ev, 43.80.Vj, 43.58.Dj [FD]

I. INTRODUCTION

As reported in the literature, the speed of sound varies among different types of human soft tissues. Lower values (1440–1490 m/s) were measured in fat-rich tissues,[1] while upper values (1650 m/s) are reported for muscle tissues along the muscular fibers.[2] Variations on the speed of sound can also be found in the same type of tissue, due to the tissue inhomogeneity. Nevertheless, a constant speed of 1540 m/s is typically assumed for ultrasound imaging.

Since phased array ultrasound scanners rely on the speed of sound to position the transmit and receive beam focus as well as to steer the beam, image distortion is introduced due to differences on speed of sound as the beam propagates through the tissues. Besides deterioration in the image quality, detection tasks of clinical relevance may be jeopardized. For instance, erroneous detection of low contrast lesions and of point target within a speckle field may occur.[3]

Tissue condition is also a source for differences on speed of sound in the same type of tissue. *In vitro* measurements in liver[4] and brain[5] have provided different sound speed values between healthy and diseased tissues. During the past couple of years, an effort has been addressed to *in vivo* measurement of speed of sound in the bone and to its clinical use as a tool to diagnose osteoporosis.[6–11]

Therefore, measurement of sound speed plays an important role when ultrasound applied to medicine is considered. Consequently, the development of measurement methods with adequate precision and capable to be used *in vivo* is highly desirable.

When an ultrasound pulse propagates in a medium, three parameters are coupled together: the wave speed, the distance traveled by the pulse, and the time-of-flight. If two of them are known, the third can be calculated. However, some methods have been proposed to determine the wave speed without the simultaneous knowledge of the distance traveled by the pulse and the time-of-flight. They rely on either transmission or pulse-echo ultrasonic methods and on a calibrated medium with known wave speed.[12–14] Other methods, as reviewed by Robinson *et al.*,[15] are based on the analysis of aberrations on the images obtained from different directions of the beam, on the transit-time measurements using intersecting beams, or on the analysis of the medium transaxial deformation imposed by a compression technique.

The transmission methods described by Kuo *et al.*,[12] as well by Hsu and Hughes,[13] have limited success to be applied in medicine and *in vivo*. They need good access from many angles around the medium and suffer with strong refracting structures. On the other hand, the pulse-echo methods[12–14] are limited to *in vitro* or invasive situations, since they are based on reference pulses generated by reflecting surfaces located behind the sample of the tissue under analysis. In general, these methods are very precise.

Other works, based on the inverse-scattering problem applied to backscattered ultrasound signals, have been de-

scribed to determine the wave speed. They rely on deconvolution methods.[16,17] However, the random nature of biological tissues and the attenuation of the incident pulse with respect to frequency preclude the convolution techniques to be used in real biological tissues.[18]

In order to overcome the limitations of the transmission methods and the invasive nature of the reflection methods mentioned previously, and to avoid the complications presented by the deconvolution methods, other approaches have been presented to measure the wave speed.[14,19] In both of them, the method is based on the measurement of time-delay between scattered signals collected with the receive transducer positioned on different locations. Anderson and Trahey,[19] for instance, used an array with 128 piezoelectric elements (aperture=28.2 mm), and measured the interelement time-delays of the echo signal from a target-wire immersed in a liquid, or of the signal scattered by a speckle-generating phantom. They obtained precise sound speed predictions of both liquid and phantom medium. However, their method presents larger errors when a transducer with smaller aperture is used. This is evidenced considering the echo signal, from a point target at the transducer focus, with phase aberration modeled as a random phase error normally distributed with root-mean-squared (rms) phase error of 50 ns and correlation length of 3 mm, defined by the full width at half-maximum of the aberrator's lateral spatial autocovariance function. In this case, simulation results indicate the relative percent error for the sound speed changing from 1.28% to 11.46% as the transducer aperture diminishes from 28 to 12 mm, respectively.

The present work describes a new pulse-echo technique to estimate the sound speed in a fluid containing scatterers, without the need of the propagation distance. The method presented by Anderson and Trahey[19] is based on the measurement of the geometric time-delay pattern, transverse to the transmitted beam axis, between the scattered signal received by each element of an array transducer. On the other hand, the method presented in this work is based on the measurement of the time-delay profile, along the beam axis, between the scattered signal collected by a receive transducer positioned at two locations. Simulation and experimental results are presented for phantoms, made of glass spherical particles immersed in gel-type medium. Also, experimental results are presented for three samples of fresh bovine liver. The whole transducer aperture (including transmission and reception) corresponds to 12 mm and the relative speed error was kept less than 3%.

The methods presented in this work and proposed by Anderson and Trahey,[19] although equivalent, may have different applications while concerning the transducer aperture and keeping low the relative percent error.

II. THEORY

The method used in this work is based on the arrangement illustrated in Fig. 1. The medium, with the scatterers randomly distributed, is immersed in a water bath and irradiated by an ultrasound beam. The scattered signal is received by a hydrophone located near the front surface of the medium and distant d_1 from the transmitted beam axis, as

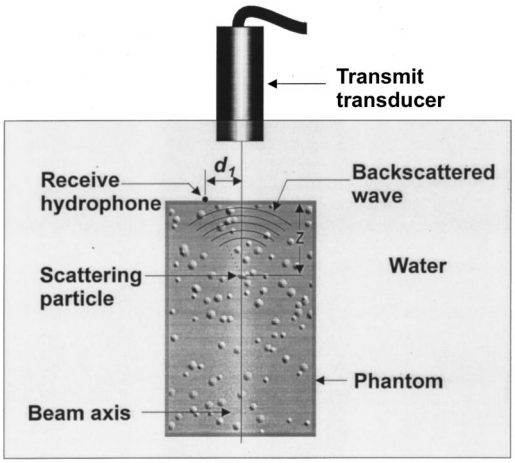


FIG. 1. Schematic setup including the hydrophone distant d_1 from the transmitted beam axis and a point scatterer on the beam axis and at depth z.

depicted in Fig. 1. The time-of-flight, t_1, for the transmitted wave to travel from the phantom face, bounce a point scatterer (at depth z and over the transmitted beam axis), and arrive at the hydrophone is given by

$$t_1=\frac{z}{c}+\frac{\sqrt{z^2+d_1^2}}{c}, \tag{1}$$

with c=wave speed in the medium.

With the hydrophone moved to a new position, d_2, the corresponding time-of-flight, t_2, becomes

$$t_2=\frac{z}{c}+\frac{\sqrt{z^2+d_2^2}}{c}. \tag{2}$$

Combining Eqs. (1) and (2) to eliminate z, and defining Δt ($\Delta t=t_2-t_1$) as the time-delay between the arrival times t_2 and t_1, it results in

$$\Delta t=\frac{-(c^2t_1^2+d_1^2)+\sqrt{(c^2t_1^2-d_1^2)^2+(2ct_1d_2)^2}}{2c^2t_1}. \tag{3}$$

Equation (3) is independent of the scatterer depth location. Therefore, it can be generalized with $t_1=t$ (the time-of-flight of the signal scattered along the beam axis). In this case, Δt becomes the time-delay profile in the axial direction of the beam and is given by

$$\Delta t=\frac{-(c^2t^2+d_1^2)+\sqrt{(c^2t^2-d_1^2)^2+(2ctd_2)^2}}{2c^2t}. \tag{4}$$

The estimated sound speed corresponds to the value of c which minimizes the rms error between the time-delay profiles obtained experimentally and with Eq. (4), after substitution of d_1 and d_2 with the corresponding distances used during the experiments.

III. EXPERIMENTAL METHODS

A. Measurement system

The experimental setup, Fig. 2, consisted of a digital oscilloscope (HP 54500; Hewlett Packard, USA), connected via GP-IB interface to a PC-AT computer. The PZT transmit transducer (12 mm in diameter, center frequency of 3.3

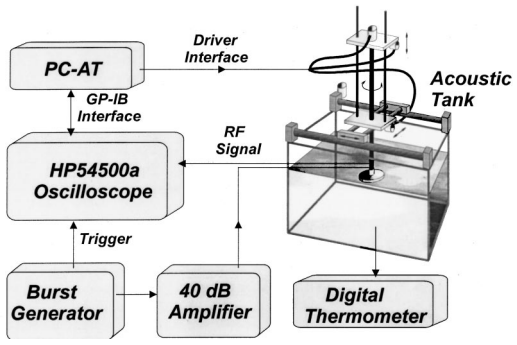


FIG. 2. Experimental setup.

MHz, −3 dB bandwidth of 2 MHz) was focused by an epoxy conical lens providing a focal distance of 25 mm, focal depth of 30 mm, and a −3 dB beam full width of 2 mm. It was excited with a one-cycle sinusoidal burst wave with 16 V of peak amplitude. The hydrophone (PZT disk with 0.8 mm of aperture and coated with epoxy) output signal was sent to the digital oscilloscope after amplification by a rf (radio frequency) amplifier with 40 dB of gain. Each rf signal, digitized with a sample rate of 50 MHz and resolution of 8 bits, contained 1024 samples.

Both transmit transducer and hydrophone were mounted in a positioning head attached to a mechanical moving system controlled by a computer. The positioning head, detailed in Fig. 3, allowed fine adjustments on the direction of the transmitted beam. It moved in a raster mode and the received scattered signal was collected from 121 positions in the sample and over an 11×11 matrix with a 1-mm separation. At every position occupied by the head, one pair of rf signals received by the hydrophone was collected; one signal with d_1 as 5 mm and the other with d_2 as 6 mm. Before moving to the next point in the matrix, the positioning head rotated by 90 degrees and another pair of signals was collected. This procedure was repeated with the positioning head rotated by 180 and 270 degrees. The data collected at four opposed angles contributes to smooth the average delay between the pairs of received signals, since it tends to compensate the delays estimated from scatterers positioned out of the transmitted beam axis. On each experiment, 968 (11×11×4×2) scattered signals were collected, resulting in 484 time-delay estimates.

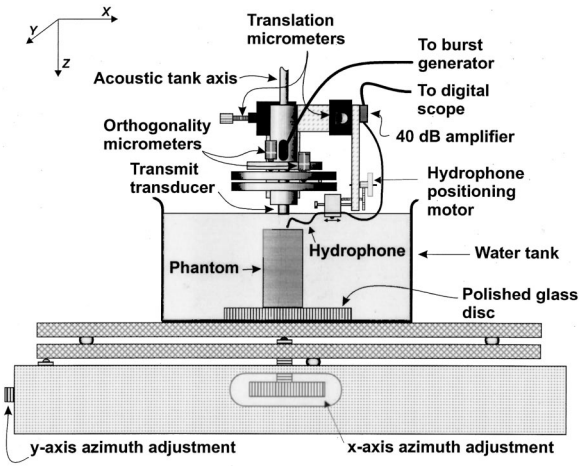


FIG. 3. Positioning head.

The hydrophone was placed in direct contact with the medium front face to avoid the effects of refraction on the ray path of the wave propagated between the medium front face and the hydrophone.

At the beginning of each experiment the reference sound speed was determined with a reference pulse-echo method, based on the method (±2 m/s of precision) described by Ye *et al.*[20]

Both transducers and medium were immersed in a temperature-controlled water bath.

B. Time-delay estimate

In the present work, the time-delay estimate was implemented using a cross-correlation technique, which is extensively used in the detection of tissue displacements[21] and in noninvasive estimates of temperature.[22] An excellent compilation of cross-correlation techniques is given by Hein *et al.*[23]

A 2 μs ($N=100$ samples) time-gated window, shifted by r samples, of the rf signal E_1 collected with the hydrophone set at d_1, was cross-correlated with an N-point time-gated window that spanned over the rf signal E_2, collected with the hydrophone set a d_2. The correlation coefficient, $R(r,m)$, was calculated according to

$$R(r,m)=\frac{\Sigma_{i=0}^{N-1}[E_1(r+i)]\cdot[E_2(r+i+m)]}{\sqrt{\Sigma_{j=0}^{N-1}[E_1(r+j)]^2\cdot\Sigma_{k=0}^{N-1}[E_2(r+k+m)]^2}}. \quad (5)$$

The time-gated window spans over E_2 by varying m. The time-delay, at a depth proportional to r, between gated E_1 and E_2 is given by $m_{\max}$ times the time-sample interval (20 ns), with the maximum of $R(r,m)=R(r,m_{\max})$.

With the typical values used in this work for the focal depth (30 mm), d_1 (5 mm), d_2 (6 mm), and the speed of sound (1200–1800 m/s), the time-delay between gated E_1 and E_2 should be in the range of 100–300 ns. This range of delays restrains m to $5\leqslant m\leqslant 15$.

The sample frequency of 50 MHz provides a resolution of 20 ns for the time-delay measurement. Increasing the sample frequency would improve this resolution. As a tradeoff, more computational time, to determine $m_{\max}$, and memory length, at the oscilloscope, would be necessary. To circumvent this problem, an interpolation technique[24] was used to calculate $m_{\max}$. Basically, this technique interpolates a parabola through three values of the correlation coefficient: the largest one and two adjacent neighbors. Then $m_{\max}$ corresponds to the point where the parabola has its maximum.

The experimental time-delay profile was obtained moving (varying r) the time-gates along E_1 and E_2 and considering only those depths r for which $R(r,m_{\max})$ exceeded 0.9. An average time-delay profile was obtained from the 484 experimental results.

C. Tissue sample

Three samples of fresh bovine liver were cut and placed inside of an acrylic cylindrical container (internal diameter

$=37$ mm, wall thickness$=4$ mm, height$=52$ mm). The container was sealed with a PVC membrane (thickness $=15\ \mu$m) to better define the water/liver interface. The reference sound speed for the bovine liver samples was measured using the reference method[20] and found to be 1592, 1605, and 1630 m/s.

D. Ultrasonic phantom

The phantoms consisted of a uniform distribution of glass spheres (average diameter$=0.8$ mm) in three different mixtures, with different sound speeds. The spherical particles, with a volume concentration of 0.1%–0.5%, were randomly distributed in the phantoms and acted as scatterers. The chosen concentration characterizes single scattering of the ultrasound waves.

The phantom with the lowest reference speed (1220 m/s @17 °C), Phantom LS, was prepared with 8 g of calcium acetate for each 92 g of ethanol. This mixture solidifies under 50 °C and remains stable if well sealed.

For the intermediate reference speed (1501 m/s @17 °C), Phantom IS consisted of a mixture with a concentration of 12 g of natural gelatin, made of porcine skin, to 100 g of water. After solidification, the gelatin phantom was immersed in a formaldehyde solution to avoid deterioration.

The third phantom (reference speed of 1715 m/s @17 °C), Phantom HS, contained a mixture with a concentration of 15 g of natural porcine skin gelatin to of 100 g of ethylene glycol.

All phantom samples were kept inside of cylindrical acrylic containers (internal diameter$=37$ mm, height $=60$ mm). A PVC membrane (thickness$=15\ \mu$m) sealed both faces.

IV. RESULTS

The method was tested theoretically, through computer simulation, and experimentally, using ultrasound phantoms and biological tissue samples.

A. Simulation results

The procedure presented in the Appendix was used to simulate the wave scattered by glass spherical particles (longitudinal wave speed$=5968$ m/s and Poisson coefficient $=0.17$) uniformly distributed (19 spheres/cm^3) inside of a cylindrical region (diameter$=37$ mm, length$=30$ mm) immersed in a fluid with wave speed of 1540 m/s. The dimensions for the cylindrical region were chosen to follow those of the experimentally transmitted ultrasound beam at its focal region. The diameter of the spherical particles was assumed to have a normal distribution with mean$=0.8$ mm and variance$=0.1$ mm^2. The choice of the simulation model used here was based on the work by Insana *et al.*,[25] which proposes the scattering theory of Faran[26] as the most accurate model for an acoustic description of tissue, considering the hypothesis of modeling tissue structures with spherical or cylindrical particles and taking into effect the presence of shear waves due to the elastic properties of collagen.

The choice of spherical particles with 0.8 mm in diameter takes into account the possibility to mimic the liver, where the lobules, displaced as periodic structure with diameter between 0.8 and 2 mm, behave as the scatterers.[27]

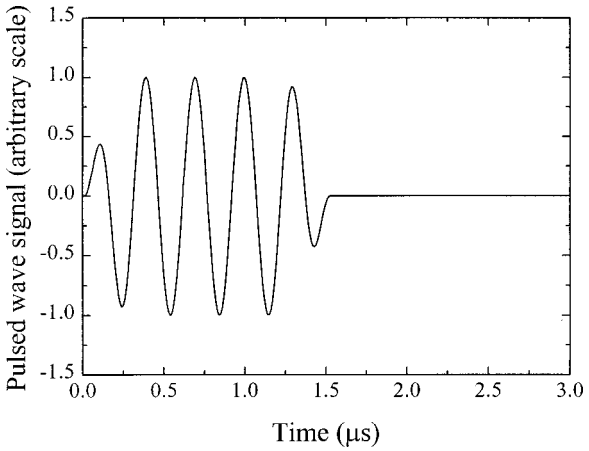


FIG. 4. Transmitted pulsed wave signal, at 3.3 MHz, used to determine the simulation results.

The incident wave was assumed plane and the particles confined with the beam of the transmit transducer.

The signal of the wave scattered by the particles, due to a transmitted pulsed wave signal (depicted in Fig. 4), was calculated at two positions at the face of the cylindrical region containing the particles, considering $d_1=5$ mm and $d_2=6$ mm.

Figure 5 presents the plot for the mean ($\pm$ one standard deviation) time-delay profile, considering 1000 independent simulations of the wave signal scattered by the particles. The simulation signal was corrupted by adding a Gaussian distributed noise with an amplitude 40 dB below the amplitude of the simulation signals and unit variance. Also included in this figure is the plot of the time-delay profile determined from Eq. (4) with $c=1540$ m/s.

B. Experimental results

Figure 6 presents the experimental mean time-delay profile plots for the three phantom types with the corresponding plots of Δt determined from Eq. (4) with $c=1220$, 1501, and 1715 m/s for phantoms LS, IS, and HS, respectively. These

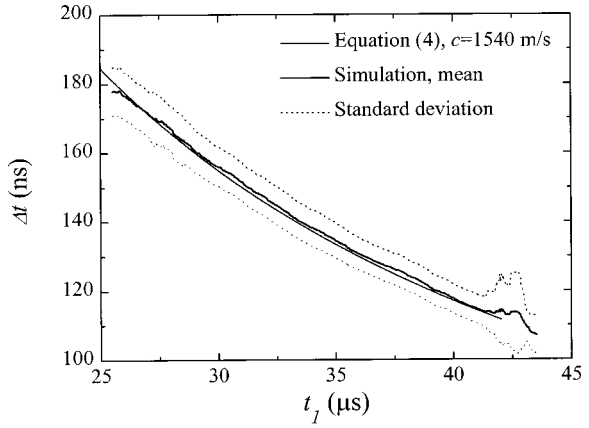


FIG. 5. Plots of the time-delay profile, Δt, calculated from Eq. (4) with $c=1540$ m/s and averaged from 1000 independent simulations of the scattered signal.

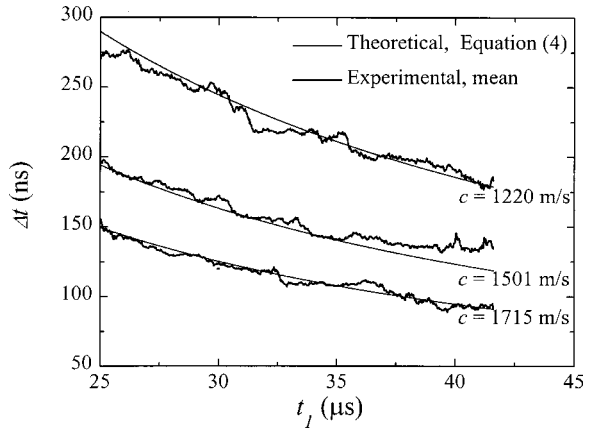


FIG. 6. Plots of the time-delay profile, Δt, obtained experimentally and from Eq. (4), with $c=1220$, 1501, and 1715 m/s, for phantoms LS, IS, and HS, respectively.

results refer to a particle concentration of 0.3% by volume. Similar results were obtained with phantom IS containing particle concentrations of 0.1% to 0.5%.

Figure 7 presents the plot of the experimental time-delay profile (mean $\pm$ one standard deviation) for a sample of bovine liver. Also included is the plot for Δt determined from Eq. (4) with $c=1630$ m/s. The experimental plot of Δt, for the tissue sample, has more fluctuations than the same type of plot for a phantom, as in Fig. 6. The difference among the fluctuations of these plots occurs because the phantom medium is more homogeneous than the tissue medium. As a matter of fact, visual inspection of the bovine liver sample, posterior to the experiments, confirmed the presence of small vessels near the depth corresponding to the largest fluctuations of Δt. The theoretical plots of Δt, in Figs. 6 and 7, were computed using Eq. (4), substituting c with the measurements obtained with the reference method.[20]

The experimental result for the estimated sound speed, c_{est}, was determined upon minimization of the rms error between the experimental mean profile of Δt and the profile determined with Eq. (4), using c as a parameter. A similar procedure, conducted with the experimental standard error of the mean profile of Δt, was used to determine one deviation of the estimated sound speed, Δc_{est}.

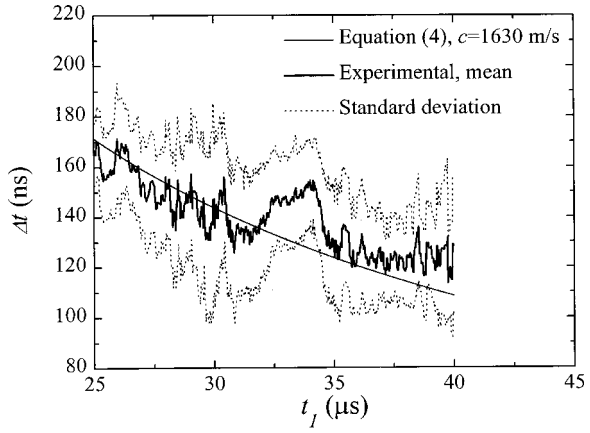


FIG. 7. Plots of the time-delay profile, Δt, obtained experimentally and from Eq. (4), with $c=1630$ m/s, for a sample of fresh bovine liver.

Table I summarizes the experimental results for c_{est}, considering phantom samples, with different particle concentrations and subjected to different temperatures, and three samples of fresh bovine liver. Also included are the corresponding results for Δc_{est}, the reference speed (c_{ref}), measured with the reference method,[20] and the relative speed error$=100.\ (c_{\text{est}}-c_{\text{ref}})/c_{\text{ref}}$.

V. SOURCES OF ERROR FOR THE ESTIMATED SOUND SPEED

The estimated sound speed, determined with the present method, has the influence of two sources of error. One is related to the error on the measurement of Δt and the other refers to the positioning of the hydrophone with respect to the axis of the transmitted beam.

The first case can be analyzed differentiating c with respect to Δt. This is obtained solving Eq. (4) for c as a function of Δt. Thereafter, substitution of t_2 by $t_1+\Delta t$ provides

$$\frac{\partial c}{\partial \Delta t}=\frac{1}{2c}\cdot\left(\frac{d_1^2}{t_1\Delta t^2}-\frac{d_2^2(2\Delta t+t_1)}{(\Delta t^2+t_1\Delta t)^2}\right). \tag{6}$$

Neglecting $2\Delta t$ and Δt^2 in comparison to t_1 and $t_1\Delta t$, then Eq. (6) can be approximated to

$$\frac{\partial c}{\partial \Delta t}=\frac{1}{2ct_1\Delta t^2}\cdot(d_1^2-d_2^2). \tag{7}$$

With typical experimental values for $d_1=5.0$ mm, $d_2=6.0$ mm, $t_1=30$ μs, and $c=1540$ m/s, Eq. (4) yields $\Delta t=152$ ns. With these values in Eq. (7), the change of the estimated sound speed per change in the time-delay estimate is determined as -5.1 m/s/ns.

The second source of error, related to the positioning of the hydrophone with respect to the axis of the transmitted beam, can be analyzed differentiating Δt with respect to d_1 and to Δd, for $\Delta d=d_2-d_1$.

From Eq. (4), after substitution of d_2 by $d_1+\Delta d$, it comes

$$\frac{\partial \Delta t}{\partial d_1}=\frac{1}{c^2t_1}\left[\frac{2c^2t_1^2(d_1+\Delta d)-(c^2t_1^2-d_1^2)d_1}{\sqrt{(c^2t_1^2-d_1^2)^2+[2ct_1(d_1+\Delta d)]^2}}-d_1\right] \tag{8}$$

and

$$\frac{\partial \Delta t}{\partial \Delta d}=\frac{2t_1(d_1+\Delta d)}{\sqrt{(c^2t_1^2-d_1^2)^2+[2ct_1(d_1+\Delta d)]^2}}. \tag{9}$$

Considering again $d_1=5.0$ mm, $d_2=6.0$ mm, $t_1=30$ μs, and $c=1540$ m/s, Eqs. (8) and (9) yield $\partial\Delta t/\partial d_1=27$ ns/mm and $\partial\Delta t/\partial\Delta d=165$ ns/mm, respectively. Therefore, Δt is more sensitive to Δd than to Δd_1.

An error of 50 μm in Δd results in a deviation of 8.25 ns of Δt, which in turn causes an error of -42 m/s for c, according to Eq. (7).

VI. DISCUSSION AND CONCLUSION

This work presents a method to estimate the sound speed using a pulse-echo technique, which precludes the need of

TABLE I. Experimental estimated sound speed (c_{est}) and corresponding deviation (Δc_{est}) for different samples of phantoms and fresh bovine liver, *in vitro*. Also included are the reference speed (c_{ref}), measured with the reference method (Ref. 20), and the relative speed error $= 100 \cdot (c_{est} - c_{ref})/c_{ref}$.

Sample	Particle concentration (% of volume)	Temperature (°C)	c_{ref} (m/s)	c_{est} (m/s)	Δc_{est} (m/s)	$100 \cdot (c_{est} - c_{ref})/c_{ref}$ (%)
Phantom						
LS	0.3	17	1220	1226	5	0.49
IS	0.1	25	1550	1524	6	−1.68
	0.2	25	1550	1540	6	−0.65
	0.3	25	1550	1544	6	−0.39
	0.4	25	1551	1546	6	−0.32
	0.5	25	1550	1532	6	−1.16
	0.5	17	1501	1473	7	−1.87
HS	0.3	17	1715	1712	10	−0.17
Liver		20	1592	1607	26	0.94
		20	1605	1567	32	−2.36
		20	1630	1584	23	−2.82

knowing the distance traveled by the pulsed wave and between transmit and receive transducers. The sound speed was estimated measuring the profile of the time-delay between the rf signals of the wave scattered by the medium and received by a hydrophone at two different positions located near the axis of the transmitted beam. The method was tested theoretically, with simulation rf signals of ultrasound waves scattered by a medium containing a random distribution of glass spherical particles, and, experimentally, with the medium consisting of ultrasonic phantoms or fresh samples of bovine liver.

Most of the results (with two exceptions) presented in Table I for the relative speed error are negatively biased. Part of this bias is systematic to the method and due to the approximate shape of the pulse scattered by a particle and received by the hydrophone placed at two locations close to each other. The approximate shape of these pulses causes the time-delay, between them, to remain constant during a time interval equivalent to their width. It results in a stair-stepped profile of the time-delay, always above the theoretical profile. Therefore, the average of the stair-stepped time-delay profiles becomes biased above the theoretical one. Similar artifactual stair-stepped appearance of time-delay profile has been reported.[15] The simulation results, in Fig. 5, also confirm a time-delay profile, positively biased with respect to the theoretical profile calculated with Eq. (4). An average bias of 1.07 ns exists between the plots of Δt presented in Fig. 5, considering the time-of-flight interval between 27.5 and 40 μs. According to Eq. (7), a positive bias in the time-delay profile results in a bias for the estimated sound speed at a rate of -5.1 m/s/ns. Therefore, the average bias of $\Delta t = 1.07$ ns corresponds to an estimated sound speed biased by -5.46 m/s. This bias gives a relative speed error of -0.35%, considering a reference sound speed of 1540 m/s.

The simulation results allow us to estimate how much of the bias for the relative speed error, in Table I, was due to the method used to estimate Δt and also allow us to conclude that this is not the most important contribution to the error, since the relative speed error varies between -2.82% and 0.94%.

Another possible cause of errors for the estimated sound speed refers to the accuracy of positioning the hydrophone with respect to the transmitted beam axis. As a matter of fact, from Sec. V it is concluded that 5% deviation (50 μm over 1 mm) on the step motion of the hydrophone will cause a deviation of 2.5% on the estimated sound speed, considering $c = 1540$ m/s and the experimental setup for the hydrophone positions used in this work.

However, it is believed that part of the bias is also due to some differences among the reference and the proposed methods used in this work to measure the sound speed. But this is, of course, a subject that needs further attention to be clarified.

The estimated speed deviation in Table I ranges from 5 to 32 m/s. In particular, for the liver tissues it ranges from 26 to 32 m/s. This fluctuation on the measured data is due, in part, to the precision of the method and in part to the random nature of the medium, which imposes a phase aberration on the scattered wave and, consequently, a distortion among the scattered signals received by the hydrophone. Future work on the present method should consider phase aberration correction[28–31] as a tool to improve the precision of the measurements.

The method was repeated three times with samples of fresh bovine liver *in vitro* and the results for the estimated sound speed (with corresponding speed deviation), in m/s, are 1607 (26), 1567 (32), and 1584 (23). These results are in agreement with the data published in the literature[32] and ranging between 1570 and 1600 m/s for human liver. Other results, similar to those obtained in the present work, have also been reported[33] for sound speed ($\pm$ one std dev) of 1538 ($\pm$29) m/s in normal liver, 1610 ($\pm$30) m/s in cirrhotic liver, and 1423 ($\pm$34) m/s in fatty liver.

The recent work of Anderson and Trahey[19] also presents a method for direct prediction of sound speed using pulse-echo ultrasound. They obtained mean relative errors of the sound speed less than 0.4%, which is a better figure if compared with the relative speed errors for the work described here and presented in Table I. They also presented simulation results for predicted sound speed, considering the phase ab-

TABLE II. Simulation results for predicted mean sound speed, $\mu_{\hat{c}}$, its standard deviation, $\sigma_{\hat{c}}$, and relative percent error, $100\sigma_{\hat{c}}/c$, for 10 000 trials of phase aberration modeled as a random phase error normally distributed with rms phase error of 20 and 50 ns and FWHM correlation length of 3 mm. The echo signal is generated by a wire target at the transducer focus (30 mm). The results refer to transducer apertures of 28.16 (presented by Anderson and Trahey Ref. 19) and 12.00 mm.

	Aperture					
	28.16 mm			12.00 mm		
Aberrator	$\mu_{\hat{c}}$ (m/s)	$\sigma_{\hat{c}}$ (m/s)	$100\sigma_{\hat{c}}/c$ (%)	$\mu_{\hat{c}}$ (m/s)	$\sigma_{\hat{c}}$ (m/s)	$100\sigma_{\hat{c}}/c$ (%)
20 ns rms	1539.96	7.77	0.50	1543.99	67.45	4.37
50 ns rms	1539.59	19.71	1.28	1567.73	179.70	11.46

erration introduced in the received signal similar to that observed from measurements in the breast.[34–36] In this case, their results are also excellent, with relative percent error less than 1.60%. However, following the procedure used by them in the simulations, it can be demonstrated that the relative percent error increases when a transducer with smaller aperture is used. Table II reproduces the simulation results for the mean sound speed, its standard deviation, and the relative percent error obtained by them with a 28.16-mm aperture array transducer. It is considered the echo signal from a point target at the transducer focus (30 mm), with phase aberration modeled as a random phase error normally distributed with rms phase error of 20 and 50 ns and correlation length, defined by the full width at half-maximum (FWHM) of the aberrator's lateral spatial autocovariance function, of 3 mm. Also included in Table II are the simulation results obtained for an array transducer with 12 mm in aperture. As it can be observed from this table, the relative percent error for the sound speed increases, up to 11.46%, as the transducer aperture diminishes to 12 mm. In the present work, an aperture of 12 mm, corresponding to the distance between the hydrophone positions opposed by 180 degrees, was used. Even with samples of biological tissue, the relative speed error remained smaller than 3.0%. Therefore, it is concluded that the method proposed in this work presents better accuracy if compared with the method of Anderson and Trahey,[19] when smaller apertures are employed. Nevertheless, considering that sound speed in soft tissues has a typical range of $\pm 5\%$ over 1540 m/s (1450 m/s for fat and 1600 m/s for muscle),[37,38] the error in the present work has to be improved for practical and reliable applications of the method.

The technique described in this work was tested with homogeneous media, with the sound speed constant along the propagation path. However, the technique must be adapted for practical purposes of tissue characterization, since biological medium is nonhomogeneous and the sound speed varies with space. In this case, the medium could be modeled by piecewise continuous layers and the sound speed estimated for each layer. The method would be applied in a recursive way, beginning with measurements of sound speed of the layers proximal to the transducer. In this way, mapping the sound speed as a function of depth would become possible.

Before testing the method with nonhomogeneous media, it is necessary to improve accuracy, to obtain better depth and lateral beam resolution, and to speed up the capture of the rf signals with the receive transducer at the different positions. In this sense, future work should consider an array of receivers arranged along the ring of an annular transducer. This array of receivers would capture the signals simultaneously, eliminating the mechanical rotation of the transducer.

The results obtained so far indicate that the method has the potential to be used in ultrasonic tissue characterization.

ACKNOWLEDGMENTS

The authors appreciate the support received from CNPq and PRONEX/MCT.

APPENDIX: PULSED ACOUSTIC WAVE SCATTERED BY A SPHERICAL PARTICLE

This section describes the procedure used to generate the wave scattered by a spherical particle due to the incidence of a pulsed plane acoustic wave. The result was used in the simulation of the time-delay profile, as presented in Fig. 5, from a random distribution of spherical particles in a fluid medium.

The formulation of sound scattering by spherical particles was well studied, about 50 years ago, by Faran.[26] Based on his work, if an acoustic plane wave, with an unit amplitude, is incident on a solid sphere of radius a immersed in a fluid, the scattered pressure wave, p_s, at a distance R_2, far from the center of the sphere, is given by

$$p_s(R_2) = e^{-ik_3R_1} \cdot \frac{e^{-ik_3R_2}}{k_3R_2} \sum_{m=0}^{\infty} (2m+1) \cdot \sin \eta_m \cdot e^{i\eta_m} \cdot P_m(\cos\theta), \tag{A1}$$

where $e^{-ik_3R_1}$ accounts for the phase added to p_s as the incident wave propagates a distance R_1 between a reference point and the sphere; $k_3 = \omega/c_3$ and c_3 are the wave number and the longitudinal wave speed for the fluid, respectively, and $\omega = 2\pi f$ is the angular frequency of the incident wave; $P_m(\cos\theta)$=Legendre polynomial of order n and θ is the angle between the directions of the incident and scattered waves;

$$\tan \eta_m = \frac{x_3 \cdot j_m'(x_3) - F \cdot j_m(x_3)}{x_3 \cdot n_m'(x_3) - F \cdot n_m(x_3)}, \tag{A2}$$

with $F=-(\rho_3/\rho_1)\cdot\tan[\zeta_m(x_1,\sigma)]$ and $\tan[\zeta_m(x_1,\sigma)]$ given by

$$\tan[\zeta_m(x_1,\sigma)]=\frac{x_2^2}{2}\left[\frac{\dfrac{x_1 j_m'(x_1)}{x_1 j_m'(x_1)-j_m(x_1)}-\dfrac{2(m^2+m)j_m(x_2)}{(m^2+m-2)j_m(x_2)+x_2^2 j_m''(x_2)}}{\dfrac{x_1^2[((\sigma/(1-2\sigma))j_m(x_1)-j_m''(x_1)]}{x_1 j_m'(x_1)-j_m(x_1)}-\dfrac{2(m^2+m)[j_m(x_2)-x_2 j_m'(x_2)]}{(m^2+m-2)j_m(x_2)+x_2^2 j_m''(x_2)}}\right], \tag{A3}$$

considering $j_m(x)$ = spherical Bessel function of order m, $n_m(x)$ = spherical Neumann function of order m, $j_m'(x)$ and $j_m''(x)$ = first and second derivatives of j_m with respect to the argument, respectively, n_m' n_m'' = first and second derivatives of n_m with respect to the argument, respectively, ρ_1 and ρ_3 = density of the sphere and the fluid, respectively, σ = the Poisson's ratio for the sphere material, $x_1=k_1\cdot a$, $x_2=k_2\cdot a$, $x_3=k_3\cdot a$, $k_1=\omega/c_1$, $k_2=\omega/c_2$, with c_1, c_2 representing the longitudinal and the shear wave speeds inside the sphere.

In the original expression derived by Faran[26] for $\tan[\zeta_m(x_1,\sigma)]$, the term $\sigma/(1-2\sigma)$ is misplaced. This was noticed by Hickling[39] and Eq. (A3) is presented on its correct form. On the other hand, the expression presented by Hickling[39] for $\tan\eta_m$ does not have the variable x_3 multiplying $j_m'(x_3)$ and $n_m'(x_3)$. Equation (A2) is the correct form for $\tan(x_3)$.

The computer program used to calculate the wave scattered by a spherical particle, in the time domain, was divided in two parts. The first one, written in C language, provided the scattering amplitude function, $f(\hat{o},\hat{i})$, over the entire spectrum on the incident pulse. Function $f(\hat{o},\hat{i})$ is defined as

$$f(\hat{o},\hat{i})=\sum_{m=0}^{\infty}(2m+1)\cdot\sin\eta_m\cdot e^{i\eta_m}\cdot P_m(\cos\theta), \tag{A4}$$

with $\hat{i}$ and $\hat{o}$ representing the unit vectors in the direction of the incident and scattered waves.

The algorithms used to calculate the spherical Bessel and Neumann functions were derived from Antosiewicz.[40]

The second part of the computer program, written in Matlab© (Math Works, Inc., USA), determined $p_s(R_2)$ in the time domain. The main steps are as follows.

(1) The spectrum of $f(\hat{o},\hat{i})$ is multiplied by that of the wave incident on the particle.
(2) The result from step 1 is multiplied by $e^{-ik_3R_1}\cdot e^{-ik_3R_2}/k_3R_2$.
(3) $p_s(t,R_2)$ is determined through the inverse Fourier transform applied to the result from step 2.

The approach of calling a C function from a Matlab routine reduced the time for the calculation done within the simulation program.

[1] G. Kossoff, E. K. Fry, and J. Jellins, "Average velocity of ultrasound in the human female breast," J. Acoust. Soc. Am. **57**, 1730–1736 (1973).
[2] P. N. T. Wells, *Biomedical Ultrasonics* (Academic, London, 1977).
[3] M. E. Anderson, M. S. McKeag, and G. E. Trahey, "The impact of sound speed errors on medical ultrasound imaging," J. Acoust. Soc. Am. **107**, 3540–3548 (2000).
[4] C. M. Sehgal, G. M. Brown, R. C. Bahn, and J. F. Greenleaf, "Measurement and use of acoustic nonlinearity and sound speed to estimate composition of excised livers," Ultrasound Med. Biol. **7**, 856–874 (1986).
[5] F. W. Kremkau, C. P. McGraw, and R. W. Barnes, "Acoustic properties of normal and abnormal human brain," in *Ultrasound Tissue Characterization II*, edited by M. Linzer (US Government Printing Office, Washington, 1979), National Bureau of Standards Spec. Publ. 525, pp. 81–84.
[6] E. Diessel, T. Fuerst, C. F. Njeh, D. Hans, S. Cheng, and H. K. Genant, "Comparison of an imaging heel quantitative ultrasound device (DTU-one) with densitometric and ultrasonic measurements," Br. J. Radiol. **73**, 23–30 (2000).
[7] M. Weiss, A. B. Ben-Shlomo, P. Hagag, and M. Rapoport, "Reference database for bone speed of sound measurement by a novel quantitative multi-site ultrasound device," Osteoporosis Int. **11**, 688–696 (2000).
[8] K. A. Wear, A. P. Stuber, and J. C. Reynolds, "Relationships of ultrasonic backscatter with ultrasonic attenuation, sound speed and bone mineral density in human calcaneus," Ultrasound Med. Biol. **26**, 1311–1316 (2000).
[9] K. A. Wear and D. W. Armstrong, "Relationships among calcaneal backscatter, attenuation, sound speed, hip bone mineral density, and age in normal adult women," J. Acoust. Soc. Am. **110**, 573–578 (2001).
[10] C. Wu, D. Hans, T. He, B. Fan, C. F. Njeh, P. Augat, J. Richards, and H. K. Genant, "Prediction of bone strength of distal forearm using radius bone mineral density and phalangeal speed of sound," Bone (N.Y.) **26**, 529–533 (2000).
[11] K. I. Kim, I. K. Han, H. Kim, and N. H. Cho, "How reliable is the ultrasound densitometer for community screening to diagnose osteoporosis in spine, femur, and forearm?" J. Clin. Densitom. **4**, 159–165 (2001).
[12] I. Y. Kuo, B. Hete, and K. K. Shung, "A novel method for the measurement of acoustic speed," J. Acoust. Soc. Am. **88**, 1679–1682 (1990).
[13] D. K. Hsu and M. S. Hughes, "Simultaneous ultrasonic velocity and sample thickness measurement and application in composites," J. Acoust. Soc. Am. **92**, 669–675 (1992).
[14] G. Kossoff, "Reflection techniques for measurement of attenuation and velocity," Proc. of the Seminar on Ultrasonic Tissue Characterization (1976), pp. 135–139.
[15] D. E. Robinson, J. Ophir, L. S. Wilson, and C. F. Chen, "Pulse-echo ultrasound speed measurements: progress and prospects," Ultrasound Med. Biol. **17**, 633–646 (1991).
[16] M. Razavy, "Determination of the wave velocity in an inhomogeneous medium from the reflection coefficient," J. Acoust. Soc. Am. **58**, 956–963 (1975).
[17] D. H. Dameron, "Determination of the acoustic velocity in tissues using an inhomogeneous media model," IEEE Trans. Sonics Ultrason. **31**, 69–74 (1979).
[18] M. R. Sedki, "The Deconvolution Problem: An Overview," IEEE Proc. **74**, 82–85 (1986).
[19] M. E. Anderson and G. E. Trahey, "The direct estimation of sound speed using pulse-echo ultrasound," J. Acoust. Soc. Am. **104**, 3099–3105 (1998).
[20] S. G. Ye, K. A. Harasiewicz, C. J. Pavlin, and F. S. Foster, "Ultrasound characterization of normal ocular tissue in the frequency range from 50 MHz to 100 MHz," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 8–14 (1995).
[21] M. O'Donnell, A. R. Skovoroda, B. M. Shapo, and S. Y. Emelianov, "Internal displacement and strain imaging using ultrasonic speckle tracking," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 314–325 (1994).
[22] R. Seip, P. VanBaren, C. A. Cain, and E. S. Ebbini, "Noninvasive real-

time multipoint temperature control for ultrasound phased array treatments,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **43**, 1063–1073 (1996).
[23] I. A. Hein and W. D. O'Brien, ''Current time-domain methods for assessing tissue motion by analysis from reflected ultrasound echoes—a review,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 84–102 (1993).
[24] R. E. Boucher and J. C. Hassab, ''Analysis of discrete implementation of generalized cross correlator,'' IEEE Trans. Acoust., Speech, Signal Process. **29**, 609–611 (1981).
[25] M. F. Insana, R. F. Wagner, D. G. Brown, and T. J. Hall, ''Describing small-scale structure in random media using pulse-echo ultrasound,'' J. Acoust. Soc. Am. **87**, 179–192 (1990).
[26] J. J. Faran, ''Sound scattering by solid cylinders and spheres,'' J. Acoust. Soc. Am. **23**, 405–417 (1951).
[27] T. Varghese and K. D. Donohue, ''Mean-scatterer spacing estimates with spectral correlation,'' J. Acoust. Soc. Am. **96**, 3504–3515 (1994).
[28] S. W. Flax and M. O'Donnell, ''Phase-aberration correction using signals from point reflectors and diffuse scatterers—basic principles,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 758–767 (1988).
[29] M. O'Donnell and S. W. Flax, ''Phase-aberration correction using signals from point reflectors and diffuse scatterers—measurements,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **35**, 768–774 (1988).
[30] L. Nock, G. E. Trahey, and S. W. Smith, ''Phase aberration correction in medical ultrasound using speckle brightness as a quality factor,'' J. Acoust. Soc. Am. **85**, 1819–1833 (1989).
[31] G. C. Ng, P. D. Freiburger, W. F. Walker, and G. E. Trahey, ''A speckle target adaptive imaging technique in the presence of distributed aberrations,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **44**, 140–151 (1997).
[32] J. C. Bamber and C. R. Hill, ''Acoustic properties of normal and cancerous human liver,'' Ultrasound Med. Biol. **7**, 121–133 (1981).
[33] N. Hayashi, N. Tamaki, K. Yamamoto, M. Senda, K. Yonekura, K. Torizuga, T. Ogawa, K. Katakura, and S. Umemura, ''*In vivo* measurement of sound speed in normal and abnormal livers using a high-resolution ultrasonic scanner,'' in *WFUMB '85*, edited by R. W. Gill and M. J. Dadd (Pergamon, Elmsford, NY, 1985), p. 520.
[34] P. D. Freiburger, D. C. Sullivan, B. H. LeBlanc, S. W. Smith, and G. E. Trahey, ''Two-dimensional ultrasonic beam distortion in the breast: *in vivo* measurements and effects,'' Ultrason. Imaging **14**, 398–414 (1992).
[35] L. M. Hinkelman, D-L. Liu, R. C. Waag, Q. Zhu, and B. D. Steinberg, ''Measurement and correction of ultrasonic pulse distortion produced by the human breast,'' J. Acoust. Soc. Am. **97**, 1958–1969 (1995).
[36] M. E. Anderson, M. S. C. Soo, and G. E. Trahey, ''Microcalcifications as elastic scatterers under ultrasound,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **45**, 925–934 (1998).
[37] S. A. Goss, R. L. Johnston, and F. Dunn, ''Comprehensive compilation of empirical ultrasonic properties of mammalian tissues,'' J. Acoust. Soc. Am. **64**, 423–457 (1978).
[38] S. A. Goss, R. L. Johnston, and F. Dunn, ''Compilation of empirical properties of mammalian tissues. II,'' J. Acoust. Soc. Am. **68**, 93–108 (1980).
[39] R. Hickling, ''Analysis of echoes from a solid elastic sphere in water,'' J. Acoust. Soc. Am. **34**, 1582–1592 (1962).
[40] H. A. Antosiewicz, ''Bessel functions of fractional orders,'' in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed., edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1970), Chap. 10, pp. 435–478.

A blood vessel exposed to ultrasound: A mathematical simulation of the temperature field

Boris Krasovitski[a] and Eitan Kimmel
The Department of Agricultural Engineering, Technion, Haifa 32000, Israel

(Received 12 June 2001; accepted for publication 23 December 2001)

In this article we present a mathematical simulation of the temperature field in and around a blood vessel when it is sonicated by a focused ultrasound beam. A simplified geometry is considered: a cylindrical blood vessel is embedded in tissue parallel to a flat skin surface. The ultrasound transducer is placed on the skin above the blood vessel, perpendicular to the skin surface. The 3D geometry of the problem is simplified by transformation, which maps the domain into a parallelepiped. A computational algorithm and computer program were developed. The simulation provides the conditions for successful occlusion of a blood vessel and demonstrates the significant role of the blood flow rate on the temperature difference between the vessel wall and the surrounding tissue. Comparing the predictions with published experimental data tested the validity of the method. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1452741]

PACS numbers: 43.80.Gx [FD]

NOMENCLATURE

$\overline{x}$, $\overline{y}$, $\overline{z}$, $\overline{r}$	space coordinates (m)
τ	time (s)
$\overline{T}$	temperature (K)
$\overline{z_0}$	focal distance of the transducer (m)
$\overline{H_0}$	distance between the vessel axis and the skin surface (m)
$\overline{S}$	area of the transducer surface (m^2)
N_0	power of the transducer (W)
$\overline{I}$	ultrasound intensity (W m^{-2})
$\tilde{\lambda}$	ultrasound wavelength (m)
R_0	radius of the blood vessel (m)
Q	blood flow rate in the vessel (m^3 s^{-1})
a	tissue attenuation coefficient (Np/m)
α	tissue absorption coefficient (Np/m)
a_s	tissue scattering coefficient (Np/m)
w_b	blood perfusion rate (s^{-1})
λ	heat conductivity (W m^{-1} K^{-1})
c	specific heat (J kg^{-1} K^{-1})
κ	heat diffusivity (m^2 s^{-1})
ρ	density (kg m^{-3})
$\overline{\alpha}$	heat transfer coefficient (W m^{-2} K^{-1})
Re	Reynolds number for blood flow
Pr	Prandtl parameter for blood
Subscripts	
t	tissue
b	blood
v	vessel
A	arterial
a	ambient
0	initial or axis
s	surface
w	wall
e	medium without attenuation
f	focal

I. INTRODUCTION

Therapeutic ultrasound (TUS) with frequency in the range 1–3 MHz and wavelength of the order of 10^{-3} m penetrates tissues well. Part of its acoustic energy is absorbed and converted to thermal energy. An ultrasound beam can be focused deep into the body onto a focal zone with a typical dimension of 10^{-3} m. The greatest temperature elevation is induced at the focal zone because the beam intensity there is much higher than in the surrounding tissues. This allows tissue at the focal spot to be selectively destroyed while temperature elevation in the surrounding tissues is relatively small and excessive damage is prevented.

When a blood vessel is located near the focal spot, a substantial temperature increase of the blood and the vessel wall may occur in spite of the cooling effect of the blood flow, which carries away part of the absorbed energy. Vessel spasm and permanent blocking of the vessel may develop under such conditions (Marinot *et al.*, 1994). Occlusion of small blood vessels near the skin surface has therapeutic (e.g., treatment of skin cancer) and cosmetic (e.g., treatment of port-wine stains) applications (Fry *et al.*, 1954; Hill *et al.*, 1994). Deeper blood vessel occlusion is done using focused ultrasonic beam guided by MRI (Hynynen *et al.*, 1996). According to this technique, an ultrasonic transducer is placed on the skin and the focal zone is made to coincide with the blood vessel.

The objective of this study is to evaluate the temperature elevation in and around a blood vessel exposed to a focused ultrasonic beam. The acquired data might provide means to reduce unnecessary damage by minimize overheating of the

[a]Electronic mail: agboris@tx.technion.ac.il

 0001-4966/2002/111(3)/1454/6/$19.00 © 2002 Acoustical Society of America

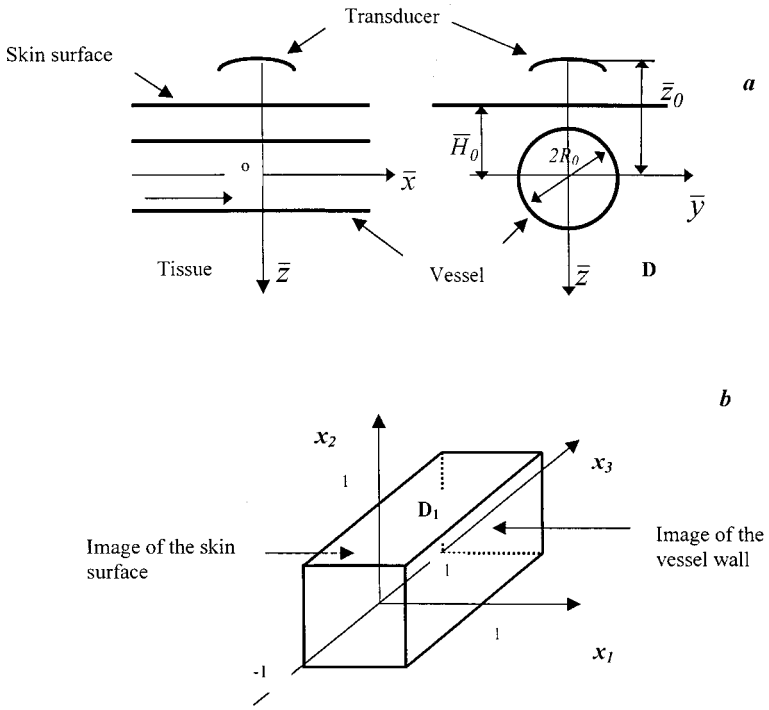


FIG. 1. Scheme of the mathematical statement. (a) Physical domain; (b) transformed domain.

surrounding tissue. Previous theoretical studies of a blood vessel in a focused ultrasonic beam solved a 2D heat transfer problem employing Pennes bioheat equation for the tissue domain and scalar effective thermal conductivity equation for the blood domain (Kolios *et al.*, 1996; Curra *et al.*, 2000). In both studies a 2D problem has been solved assuming a coaxial setup of beam and vessel, and tissue boundaries far away from the vessel. As for heat absorption, Kolios *et al.* (1996) assumed linear absorption function, while Curra *et al.* (2000) took into account nonlinear heat absorption and dispersion. In the problem we solve here the beam axis is perpendicular to the vessel axis. The skin surface is parallel to the blood vessel and even close to it as in the case of subcutaneous blood vessels. This setup requires the solution of a 3D heat transfer problem. Heat absorption is assumed linear based on Curra *et al.* (2000), who showed that when the focus coincides with a blood vessel, the significant influence that nonlinear effects have on the temperature field is limited to a very short initial period of about 0.3 s.

II. INTENSITY DISTRIBUTION OF A FOCUSED ULTRASOUND BEAM IN A TISSUE

Consider a spherically concaved ultrasound transducer that is placed above the skin surface. A nonattenuating medium fills the gap between the transducer and the skin (Fig. 1). Over the transducer surface the amplitude and phase of the normal velocity are assumed constant and intensity is small enough to exclude cavitation and other nonlinear effects.

A recent MRI guided technique for blood vessel occlusion by focused ultrasound allows us to place, with great accuracy, the focal spot into the vessel (Hynynen *et al.*, 1996). As a result, one may expect that steep temperature gradients in the tissue take place in the close proximity of the intersection of the focal spot and the vessel center line. Thereafter, while considering the heat transfer around a heated vessel, an approximate axisymmetric distribution of the ultrasound intensity may be derived in the vicinity of the focal spot,

$$\bar{I}=\bar{N}(\tilde{z})I_e(\tilde{r},\tilde{z}), \tag{2.1}$$

where the nondimensional coordinates are $\tilde{z}=\bar{z}/\bar{z}_0$ and $\tilde{r}=\bar{r}/\bar{z}_0$; the origin of the coordinate system coincides with the focal point (see Fig. 1); and $\bar{N}(\tilde{z})\equiv N(\tilde{z})/N_0$ denotes the nondimensional power, integrated over the cross section of the beam perpendicular to $\tilde{z}$. Here, I_e denotes the intensity distribution for a focused beam emitted by a transducer with power N_0 in a nonattenuating media. The value of I_e at the focal point (I_{ef}) is expressed as follows (O'Neil, 1949) by

$$I_{ef}=I_{00}\exp(\beta), \tag{2.2}$$

where I_{00} is the intensity at the transducer surface, and $\beta=2\ln(\bar{S}/\bar{z}_0\tilde{\lambda})$.

Here, $\bar{S}$ is the emitting surface area of the transducer and $\tilde{\lambda}$ is the wavelength. Following Robinson and Lele (1969), one may claim that in the proximity of the focal zone,

$$I_e=I_0(\tilde{z})\exp[-\xi(\tilde{z})\tilde{r}^{2m}], \tag{2.3}$$

where I_0 is the radiation intensity at the beam axis, described by

$$I_0=I_{ef}\exp(-\beta|\tilde{z}|^n). \tag{2.4}$$

The parameters m and n in Eqs. (2.3) and (2.4) characterize the specific transducer and may be determined during the calibration procedure.

For nonattenuating media, the power is conserved and therefore $\int_0^\infty 2\pi r I_e\,dr=N_0$. In order to meet this condition, the function $\xi(\tilde{z})$ in Eq. (2.3) should be determined as follows:

$$\xi^{1/m}=\frac{1}{S}\exp[\beta(1-|\tilde{z}|^n)], \tag{2.5}$$

where $S=\overline{S}/\pi\overline{z}_0^2$.

Substituting Eqs. (2.4) and (2.5) into Eq. (2.3) yields

$$I_e=I_{00}\exp\left[\beta(1-|\tilde{z}|^n)-\left\{\frac{1}{S}\exp[\beta(1-|\tilde{z}|^n)]\tilde{r}^2\right\}^m\right]. \tag{2.6}$$

For media with nonzero attenuation, part of the acoustic power converts to heat or scatters. Assuming that the attenuation takes place only inside the tissue, one obtains

$$\overline{N}=\exp[-2a\overline{z}_0(\tilde{z}+\overline{H}_0/\overline{z}_0)].$$

The parameter a, the amplitude attenuation coefficient, consists of scattering and absorption contributions, $a=a_s+\alpha$, where a_S is the scattering coefficient and α is the absorption coefficient.

Finally, Eq. (2.1) becomes

$$\overline{I}=I_{00}\exp\left\{-2a\overline{z}_0(\tilde{z}+\overline{H}_0/\overline{z}_0)+\beta(1-|\tilde{z}|^n)-\left[\frac{1}{S}\exp[\beta(1-|\tilde{z}|^n)]\tilde{r}^2\right]^m\right\}. \tag{2.7}$$

III. THE HEAT TRANSFER PROBLEM FOR A BLOOD VESSEL IN TISSUE

A. A mathematical statement

The heat balance for a blood vessel embedded within a tissue includes a convective heat flux term, dissipation of ultrasound energy term [see Eq. (2.7)], and a term accounting for the heat transfer between the vessel wall and the surrounding tissue. Longitudinal conductive heat flux along the vessel axis is usually much smaller than the convective heat flux and here is neglected. In this study (see data below) the convective heat flux is about 500 times greater than the conductive term. We considered also that the blood temperature is uniform in a given cross section of the vessel. In such a statement the heat flux from the blood to the vessel wall is proportional to difference between the averaged blood temperature and the wall temperature and the corresponding heat transfer coefficient $\overline{\alpha}_\nu$ (Rohsenow, 1998). The following analysis involves few dimensionless coordinates (see Fig. 1):

$$x=\frac{\overline{x}}{R_0};\quad y=\frac{\overline{y}}{R_0};\quad z=\frac{\overline{z}}{R_0};\quad r=\frac{\overline{r}}{R_0};$$

dimensionless time $t=\kappa_t(\tau-\tau_0)/R_0^2$; dimensionless focal distance $z_0=\overline{z}_0/R_0$; dimensionless distance between the vessel axis and the skin surface $H_0=\overline{H}_0/R_0$; dimensionless temperatures

$$T_b=\frac{\overline{T}_b-\overline{T}_0}{\overline{T}_0-273};\quad T_t=\frac{\overline{T}_t-\overline{T}_0}{\overline{T}_0-273};$$

$$T_a=\frac{\overline{T}_a-\overline{T}_0}{\overline{T}_0-273};\quad T_A=\frac{\overline{T}_A-\overline{T}_0}{\overline{T}_0-273};$$

and various dimensionless parameters,

$$U=\frac{2\alpha_t R_0^2 N_0}{\lambda_t\overline{S}(\overline{T}_0-273)};\quad W=\frac{\rho_b c_b w_b R_0^2}{\lambda_t};$$

$$N_V=\frac{2\alpha_b\pi R_0^3 N_0}{\rho_b c_b Q\overline{S}(\overline{T}_0-273)};$$

$$P=\frac{\lambda_t R_0}{\rho_b c_b Q};\quad \alpha_s=\frac{\overline{\alpha}_s R_0}{\lambda_t};\quad \alpha_v=\frac{\overline{\alpha}_v R_0}{\lambda_t};$$

$$\gamma=\frac{\pi R_0^{2m}\overline{z}_0^{2(1-m)}}{\overline{S}}.$$

The heat balance for a blood flow in the vessel is described in Lagrangian coordinates by the one-dimensional equation,

$$\frac{\partial T_b}{\partial x}=-q_w+N_v I_{rt};\quad -\infty\leqslant x\leqslant\infty. \tag{3.1}$$

It contains, from left to right, the following three terms: a convective term where T_b is assumed uniform over a given cross section; heat flux across the vessel wall accounting for a vessel length unit $q_w=-P\int_0^{2\pi}(\partial T_t/\partial r)|_{r=1}d\varphi$; and an ultrasound heating term where I_{rt} is the average ultrasound intensity over the vessel cross section and the parameters N_v and P are described above.

The boundary condition for Eq. (3.1) is

$$T_b=0\quad \text{at}\ \ x=-\infty. \tag{3.2}$$

The heat transfer equation for the surrounding tissue, which is subjected to an ultrasound irradiation,

$$\frac{\partial T_t}{\partial t}=\frac{\partial^2 T_t}{\partial x^2}+\frac{\partial^2 T_t}{\partial y^2}+\frac{\partial^2 T_t}{\partial z^2}+W(T_A-T_t)+U\cdot I_t,$$
$$-\infty<x<\infty,\quad z^2+y^2\geqslant 1,\quad z\geqslant -H_0, \tag{3.3}$$

takes into account heat conduction, blood perfusion (W) and arterial temperature (T_A), and ultrasound energy dissipation ($U\cdot I_t$). Here, I_t is the dimensionless ultrasound intensity defined in a 3D system of coordinates similar to Eq. (2.7); U is a dimensionless parameter defined above.

The initial conditions at $t=0$ are

$$T_t=0;\quad T_b=0. \tag{3.4}$$

The boundary conditions at the tissue surface, $z=-H_0$, are

$$\frac{\partial T_t}{\partial z}=\alpha_s(T_t-T_a), \tag{3.5}$$

where T_a is the ambient temperature and α_s is the skin/surrounding heat transfer coefficient.

The boundary conditions at the vessel surface $\{z^2+y^2=1\}$,

$$\frac{\partial T_t}{\partial r}=\alpha_v(T_t-T_b), \tag{3.6}$$

is based on known expression for the heat transfer coefficient $\overline{\alpha}_v$ expressed as a function of Re and Pr of the flowing blood (Rohsenow, 1998).

The boundary conditions at infinite distance from the ultrasound source ($x=\pm\infty \quad y=\pm\infty \quad z=\infty$) are

$$T_t=0; \quad \text{grad}\, T_t=0. \tag{3.7}$$

B. Algorithm of the numerical solution of the problem

Domain D $\{-\infty<x<\infty;\ z^2+y^2\geqslant 1;\ z\geqslant -H_0;\ y\geqslant 0\}$ is bounded by a complex set of boundary surfaces, namely an infinite flat boundary ($z=-H_0$) and a cylindrical boundary ($z^2+y^2=1; y\geqslant 0$) as shown in Fig. 1(a). The numerical solution of the problem [Eqs. (3.1)–(3.7)] in domain D necessitates a complex numerical scheme that has to overcome problems in matching few grid systems. In order to circumvent this difficulty, domain D is mapped by the following transformation,

$$\begin{aligned} y&=\frac{c_0\sin(\pi x_2)}{\cosh(\alpha_0 x_1)-\cos(\pi x_2)},\\ z&=\frac{c_0\sinh(\alpha_0 x_1)}{\cosh(\alpha_0 x_1)-\cos(\pi x_2)}-H_0, \quad x=\tan\left(\frac{\pi}{2}x_3\right), \end{aligned} \tag{3.8}$$

where $c_0=\sqrt{H_0^2-1}$ and $\alpha_0=\ln(H_0+c_0)$, into a parallelepiped shape domain D_1 $\{0\leqslant x_1\leqslant 1; 0\leqslant x_2\leqslant 1; -1\leqslant x_3\leqslant 1\}$ bounded by simple set of finite, parallel and perpendicular flat surfaces $\{x_1=0; x_1=1; x_2=0; x_2=1; x_3=-1; x_3=1\}$ [see Fig. 1(b)]. This transformation [Eq. (3.8)], which was first introduced for a 2D domain by Ufliand (1950), greatly simplifies the computational algorithm.

In D_1 domain, the heat transfer problem for the blood vessel [Eqs. (3.1) and (3.2)] is transformed into

$$\frac{2}{\pi}\cos^2\left(\frac{\pi}{2}x_3\right)\frac{\partial T_b}{\partial x_3}=-q_w+N_v I_{rt}, \quad -1\leqslant x_3\leqslant_1, \tag{3.9}$$

$$T_b=0 \quad \text{at } x_3=-1, \tag{3.10}$$

whereas the heat transfer problem for the surrounding tissue [Eqs. (3.3)–(3.7)] becomes

$$\begin{aligned} \frac{\partial T_t}{\partial t}&=\frac{b(x_1,x_2)}{c_0^2\alpha_0^2}\frac{\partial^2 T_t}{\partial x_1^2}+\frac{b(x_1,x_2)}{c_0^2\pi^2}\frac{\partial^2 T_t}{\partial x_2^2}\\ &+\frac{4}{\pi^2}\cos^2\left(\frac{\pi x_3}{2}\right)\frac{\partial}{\partial x_3}\left(\cos^2\left(\frac{\pi x_3}{2}\right)\frac{\partial T_t}{\partial x_3}\right)\\ &+W(T_A-T_t)+U\cdot I_t, \end{aligned} \tag{3.11}$$

where $b(x_1,x_2)\equiv[\cosh(\alpha_0 x_1)-\cos(\pi x_2)]^2$.

The boundary conditions are, at the tissue surface, $x_1=0$,

$$\frac{\partial T_t}{\partial x_1}=\frac{\alpha_s\alpha_0 c_0}{1-\cos(\pi x_2)}(T_t-T_a); \tag{3.12}$$

and at the vessel surface, $x_1=1$,

$$\frac{\partial T_t}{\partial x_1}=\frac{\alpha_v\alpha_0 c_0}{H_0-\cos(\pi x_2)}(T_b-T_t). \tag{3.13}$$

Symmetry conditions on the plane $y=0$ ($x_2=0$; $x_2=1$) are

$$\frac{\partial T_t}{\partial x_2}=0, \tag{3.14}$$

and, at infinite distance from the ultrasound beam in the direction of the vessel axis, $x_3=\pm 1$:

$$T_t=0, \quad \frac{\partial T_t}{\partial x_3}=0. \tag{3.15}$$

Equation (3.11) with the boundary conditions [Eqs. (3.12)–(3.15)] are solved numerically using a locally one-dimensional finite-difference scheme (Samarski, 1963; Mostinskaya, 1967)

At each time interval $[t_j,t_{j+1}]$ the following subproblems are introduced,

$$\frac{\partial v_1}{\partial t}=\frac{b(x_1,x_2)}{c_0^2\alpha_0^2}\frac{\partial^2 v_1}{\partial x_1^2}+W(T_A-v_1)+U\cdot I_t; \tag{3.16}$$

$$\frac{\partial v_2}{\partial t}=\frac{b(x_1,x_2)}{c_0^2\pi^2}\frac{\partial^2 v_2}{\partial x_2^2}; \tag{3.17}$$

$$\frac{\partial v_3}{\partial t}=\frac{4}{\pi^2}\cos^2\left(\frac{\pi x_3}{2}\right)\frac{\partial}{\partial x_3}\left(\cos^2\left(\frac{\pi x_3}{2}\right)\frac{\partial v_3}{\partial x_3}\right) \tag{3.18}$$

for v_1, v_2 and v_3. The boundary conditions are, at the skin surface $x_1=0$,

$$\frac{\partial v_1}{\partial x_1}=\frac{\alpha_s\alpha_0 c_0}{1-\cos(\pi x_2)}(v_1-T_a), \tag{3.19}$$

and at the vessel surface, $x_1=1$,

$$\frac{\partial v_1}{\partial x_1}=\frac{\alpha_v\alpha_0 c_0}{H_0-\cos(\pi x_2)}(T_b-v_1). \tag{3.20}$$

The symmetry condition on the plane $y=0$ ($x_2=0$; $x_2=1$) is

$$\frac{\partial v_2}{\partial x_2}=0. \tag{3.21}$$

The boundary conditions at infinite distance from the ultrasound beam along the vessel axis $x_3=-1$ are

$$\frac{\partial v_3}{\partial x_3}=0. \tag{3.22}$$

The equation system [Eqs. (3.16)–(3.22)] is then integrated for each time step $[t_j,t_{j+1}]$ using the following initial conditions:

$$v_1(x_1,x_2,x_3,t_j)=T_t(x_1,x_2,x_3,t_j), \tag{3.23}$$

$$v_2(x_1,x_2,x_3,t_j)=v_1(x_1,x_2,x_3,t_{j+1}), \tag{3.24}$$

$$v_3(x_1,x_2,x_3,t_j)=v_2(x_1,x_2,x_3,t_{j+1}), \tag{3.25}$$

$$T_t(x_1,x_2,x_3,t_{j+1})=v_3(x_1,x_2,x_3,t_{j+1}). \tag{3.26}$$

For integration of the boundary value problems the method of forward elimination and back substitution (Godunov and Ryaben'kii, 1987) was used.

Using calculated distribution of the tissue temperature q_w may be determined over the range $-1\leqslant x_3\leqslant 1$ and then Eq. (3.9) is integrated to yield T_b at time t_{j+1}. This procedure is then repeated.

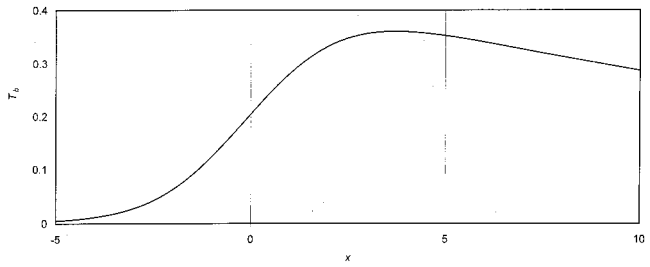


FIG. 2. Blood temperature along the vessel after 1 s sonication. Focal intensity 4.4×10^7 W m^{-2}.

IV. RESULTS AND DISCUSSION

Our method is tested with input data from occlusion experiments of a branch of the renal artery of a rabbit with a focused ultrasound beam (Hynynen *et al.*, 1996). The diameter of the vessel is about 0.6×10^{-3} m and the distance between the vessel axis and the skin surface is about 2×10^{-2} m. We estimate $m=1.065$ in Eq. (2.3) and $n=0.93$ in Eq. (2.4) as the parameters that correspond to the values reported by Hynynen *et al.* (1996) of the beam's half-intensity diameter (10^{-3} m), and length (4.8×10^{-3} m). To further resemble the experimental conditions, we set the transducer power at the simulation to fit the intensities measured by Hynynen *et al.* (1996) at the focal point. Additional properties of kidney were obtained from Duck (1990) and Goss *et al.* (1978) for tissue: heat conductivity $\lambda_t=1.2$ W m^{-1} K^{-1}; specific heat $c_t=4.2\times10^3$ J kg^{-1} K^{-1}; density $\rho_t=10^3$ kg m^{-3}; absorption coefficient $\alpha_t=7.5$ Np/m; blood: arterial flow rate $Q=0.73\times10^{-8}$ m^3 s^{-1}, specific heat $c_b=4.2\times10^3$ J kg^{-1} K^{-1}; density $\rho_t=10^3$ kg m^{-3}; and absorption coefficient $\alpha_b=1$ Np/m. Perfusion rate of $w_b=0.1$ s^{-1} was estimated for the kidney (Guyton, 1986).

Blood temperature (Fig. 2) and tissue temperature (Fig. 3) profiles are shown for a 1 s exposure to ultrasound with focal intensity of 4.4×10^7 Wm^{-2}. As shown in Fig. 2, maximum blood temperature reaches about 50 °C ($T_b=0.36$). At the same time tissue temperatures on the focal axis reach as high as 102 °C and 100 °C ($T_b=1.77$ and 1.7, respectively), while maximal temperature in the horizontal plane of the vessel is about 68 °C ($T_b=0.84$) (see Fig. 3). Note that tissue temperature in between the blood vessel and the skin are greater than deep temperature in spite of the cooling effect of the skin, due probably to greater ultrasound intensity, closer to the transducer. Note also that such high temperatures can damage the tissue.

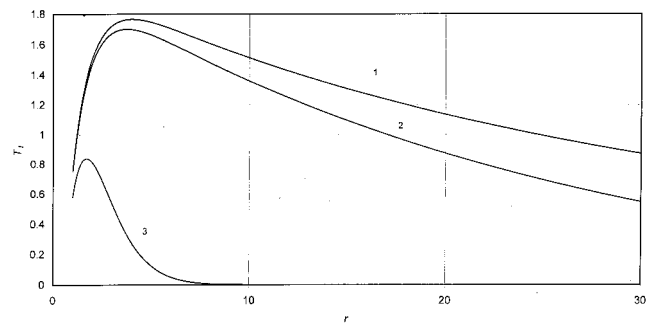


FIG. 3. Temperature profiles in tissue after 1 s sonication. Focal intensity 4.4×10^7 W m^{-2}. (1) z direction towards the skin; (2) z direction away from the skin; and (3) y direction parallel to the skin.

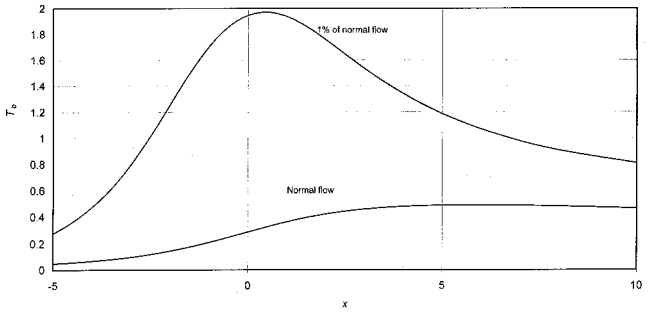


FIG. 4. Blood temperature along the vessel after 10 s sonication of 2.8×10^7 W m^{-2} focal intensity.

Average wall temperature of about 62 °C, which is predicted in this simulation, is in accord with the following two experimental findings: (*i*) Hynynen *et al.* (1996) define focal intensity of 4.4×10^7 W m^{-2} as the lower limit for wall contraction, and (*ii*) Marinot *et al.* (1994) report that tissue temperature higher than 54 °C causes the vessel wall to contract.

The next stage in the protocol of the vessel occlusion included 10 s of sonication with a lower intensity of 2.8×10^7 W m^{-2} (Hynynen *et al.*, 1996). We simulate this stage assuming that the blood flow in the vessel may have been substantially reduced as a result of vessel contraction that followed the first short (1 s) pulse of high-intensity ultrasound. For demonstration, we compared normal blood flow rate with flow rate as low as 1% of the normal. The reduced flow rate results by maximal blood temperature of about 110 °C ($T_b=1.97$) at a location a little downstream to the beam axis (Fig. 4). When similar sonication conditions are applied to a vessel with normal blood flow, the blood temperature is much smaller (about 55 °C, $T_b=0.49$, see Fig. 4). The lower blood temperatures that accompany greater arterial flow are attributed to the convective cooling effect of the blood flow. It is the same cooling effect which is accountable for the large differences in tissue temperature near the artery between the case of very small blood flow (Fig. 5) and much greater normal blood flow (Fig. 6). Far from the blood vessel

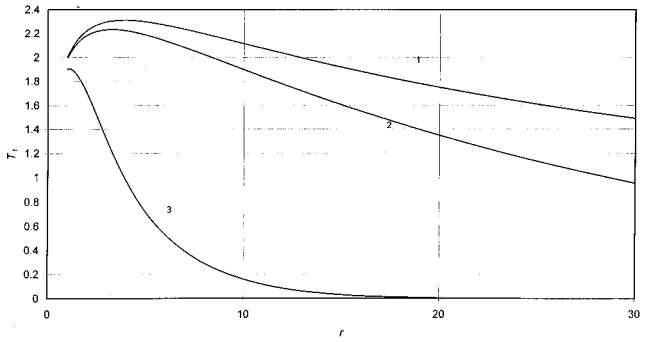


FIG. 5. Temperature profiles in tissue after 10 s sonication of 2.8×10^7 W m^{-2} focal intensity. 1% of normal flow. (1) z direction towards the skin; (2) z direction away from the skin; and (3) y direction parallel to the skin.

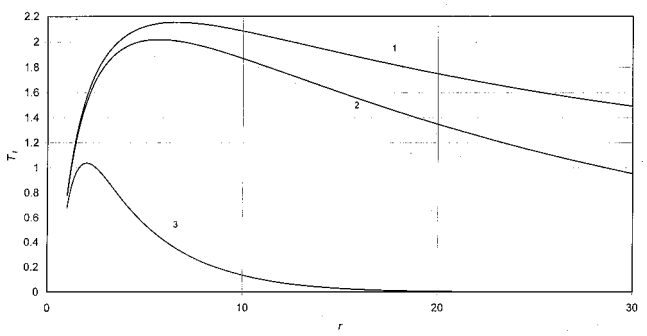


FIG. 6. Temperature profiles in tissue after 10 s sonication of 2.8×10^7 W m^{-2} focal intensity. Normal flow. (1) z direction towards the skin; (2) z direction away from the skin; and (3) y direction parallel to the skin.

(say $r>5$) the temperature profiles in the tissue are about the same in both cases. Yet, substantial temperature differences between the two cases are shown in Figs. 5 and 6 for the close proximity of the vessel wall. In the case of decreased blood flow (Fig. 5), average vessel wall temperature reaches about 109 °C. Such high temperatures are expected to cause welding of the vessel wall and consequently to induce permanent blocking of the vessel (Marinot *et al.*, 1994). In difference, in the case of normal blood flow (Fig. 6) the average wall temperature is much lower—about 62 °C. Such low temperature cannot induce welding of the vessel wall (Marinot *et al.*, 1994).

V. CONCLUDING REMARKS

The above bio-thermal simulations clarify the heat transfer mechanisms involved in the occlusion of a blood vessel using a focused ultrasonic beam and may be used for planning of a more effective procedure. When carrying normal blood flow, the artery acts as a cooling pipe and as such its wall temperature as well as the blood temperature are much lower than the surrounding tissue. For permanent occlusion the artery has to be heated beyond some threshold temperature. In the case of normal blood flow, permanent occlusion therefore is accompanied by excessive damage to the surrounding tissue. Our simulation shows how a much-reduced flow rate in the artery allows us to achieve permanent occlusion without damaging too much the surrounding tissue. The need to induce vessel contraction by a short high-intensity pulse at the beginning of the occlusion procedure is thus justified.

The advantage of our method is that it is applicable to a complex geometrical arrangement of a straight, cylindrical vessel that runs parallel to the skin surface in a homogeneous tissue. In addition, it is applicable, under some adjustment, to any distribution of ultrasound energy. This allows us, for instance, to study occlusion of a blood vessel very close to the skin or to search for the optimal angle between the ultrasonic beam and the vessel. The disadvantage of the method is that it is limited to a specific, simplified configuration, and that no nonlinear effects were taken into account. Further improvement of the numerical scheme to include nonlinearities is planned for future work.

ACKNOWLEDGMENTS

This research was supported by a grant from the Ministry of Science and Arts and by the Center for Absorption in Science, Ministry of Immigrant Absorption.

Curra, F. P., Mourad, P. D., Khohlova, V. A., Cleveland, R. O., and Crum, L. A. (**2000**). ''Numerical simulation of heating patterns and tissue temperature response due to high-intensity focused ultrasound,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **47**, 1077–1088.

Duck, F. A. (**1990**). *Physical Properties of Tissue. A Comprehensive Reference Book* (Academic, San Diego).

Fry, W. J., Mosberg, W. H., Barnard, J. W., and Fry, F. J. (**1954**). ''Production of focal destructive lesions in the central nervous system with ultrasound,'' J. Neurosurg. **11**, 471–478.

Godunov, S. K., and Ryaben'kij, V. S. (**1987**). *Difference Schemes. An Introduction to the Underlying Theory, Studies in Mathematics and Its Applications*, Vol. 19 (North-Holland, Amsterdam).

Goss, S. A., Johnson, R. L., and Dunn, F. (**1978**). ''Comprehensive compilation of empirical ultrasonic properties of mammalian tissues,'' J. Acoust. Soc. Am. **64**, 423–457.

Guyton, A. C. (**1986**). *Textbook of Medical Physiology* (Saunders, Philadelphia).

Hill, C. R., Rivens, I., Vaughan, M. G., and ter Haar, G. R. (**1994**). ''Lesion development in focused ultrasound surgery: a general model,'' Ultrasound Med. Biol. **20**, 259–269.

Hynynen, K., Colluci, V., Chung, A., and Jolesz, F. (**1996**). ''Noninvasive arterial occlusion using MRI-guided focused ultrasound,'' Ultrasound Med. Biol. **22**, 1–7.

Kolios, M. C., Sherar, M. D., and Hunt, J. W. (**1996**). ''Blood flow cooling and ultrasonic lesion formation,'' Med. Phys. **23**, 1287–1298.

Martinot, V. L., Mordon, S. R., Mitchell, V. A., Pellerin, P. N., and Brunetaud, J. M. (**1994**). ''Determination of efficient parameters for argon laser-assisted anastomoses in rats: Macroscopic, thermal and histological evaluation,'' Laser Surgeon Med. **15**, 168–175.

Mostinskaya, S. B. (**1967**). ''The program of solution of two-dimensional parabolic equation with variable coefficients,'' Vychislitel'nye metody i programmirovanie (Computation methods program.) **8**, 165–172 (in Russian).

O'Neil, H. T. (**1949**). ''Theory of focusing radiators,'' J. Acoust. Soc. Am. **21**, 516–525.

Robinson, T. C., and Lele, P. P. (**1969**). ''An analysis of lesion development in the brain and in plastics by high-intensity focused ultrasound at low-megahertz frequencies,'' J. Acoust. Soc. Am. **51**, 1333–1351.

Rohsenow, W. M. (**1998**). *Handbook of Heat Transfer* (McGraw–Hill, New York).

Samarski, A. A. (**1963**). ''Locally one-dimensional difference schemes on nonuniform meshes,'' Zhurnal Vychislitel'noj Matematiki I Matematicheskoj Fiziki (J. Computational Math. Math. Phys.) **3**, 431–466 (in Russian).

Ufliand, Ya. S. (**1950**). *Bipolarnye koordinaty v teorii uprugosti (Bipolar Coordinates in Theory of Elasticity*) (GTTL, Moscow) (in Russian).

Reparable sonoporation generated by microstreaming

Junru Wu[a)]
Department of Physics, University of Vermont, Burlington, Vermont 05405

Joel P. Ross
Department of Chemistry and Physics, Saint Michael's College, Colchester, Vermont 05439

Jen-Fu Chiu
Department of Biochemistry, University of Vermont, Burlington, Vermont 05405

(Received 7 March 2001; accepted for publication 26 September 2001)

Reparable sonoporation was observed in Jurkat lymphocytes in suspension exposed to a vibrating Mason horn tuned to 21.4 KHz. The diameter of the horn tip was 400 μm and its transverse displacement amplitude was 7.8 μm. It was found that the shear stress associated with microstreaming surrounding the Mason-horn tip was the primary reason for the cell reparable sonoporation. The threshold shear stress was determined to be 12 ± 4 Pa for exposure time up to 7 min. It was also found that the shorter the exposure time, the greater the threshold. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1420389]

PACS numbers: 43.80.Gx, 43.35.Ei [FD]

I. INTRODUCTION

Recent *in vitro* studies have shown[1–3] that cell membranes can become temporarily ''open'' to small and large molecules and then ''resealed'' when relatively low-amplitude ultrasound is applied to a cell suspension that contains an ultrasound contrast agent. Consequently, foreign molecules may be trapped in the cells. This phenomenon is called reparable sonoporation (RS).[3] The spatial peak acoustic pressure threshold for generating the reparable sonoporation was found to be 0.12 MPa[4] and 0.1 MPa[1] for 1 and 2.25 MHz ultrasound, respectively. In contrast, the phenomenon in which a cell suffers lethal damage upon exposure to a sound field, and its membrane becomes permanently ''open'' is called lethal sonoporation (LS). In addition, the sonoporation technique was successfully used in gene delivery *in vitro*.[4] The physical mechanism of sonoporation, however, is not well understood.

Our previous study[3] has shown that the bubble–cell spacing played an important role in accurately predicting reparable and lethal sonoporation in the suspension. The percentage of sonoporated cells was found to be proportional to the inverse cubic power of the bubble–cell spacing ($1/r^3$). This result suggests that the possible mechanism involved may be primarily due to a short-range volume interaction between the cell and the nearest bubbles; only the cells that move very close to pulsating bubbles experience this interaction. The probability of this short-range interaction decays quickly as the spacing increases. In other words, the $1/r^3$ dependence also implies that the percentage of sonoporated cells is proportional to the bubble number concentration when the cell concentration is constant. This study also suggests that the mechanism of reparable and lethal sonoporations may be similar, but the severity of the damage incurred by cell membranes is different. When the damage is minor, cells can repair the membrane. When the damage is beyond their self-repairing capacity, cells suffer lethal sonoporation.

One well-known mechanism for short-range interaction involves microstreaming. Acoustic streaming is a steady direct current flow generated in a liquid by an acoustic field.[5,6] The acoustic streaming can be classified into two categories: (1), ''quartz wind''-type streaming in bulk liquids resulting from the attenuation of acoustic energy; and (2), a small-scale vortex-type acoustic streaming (microstreaming) usually occurring near pulsating bubbles at a solid surface or small solid vibrators.[5] The type 2 streaming is usually relevant to bioeffects. It was reported that shear stress associated with microstreaming (details will be discussed later) generated by an oscillating bubble, an oscillating tungsten wire, or the tip of a Mason horn could induce biological effects.[7–10] In these studies, experiments were performed using an erythrocyte suspension subjected to a pulsating gas bubble or an oscillating probe at 20 kHz. It was found that erythrocyte cells experienced lethal sonoporation resulting in hemoglobin release; it was demonstrated that the physical mechanism associated with microstreaming around oscillating bubbles and vibrating wires/tips were similar.[8–10] The shear stress resulting from microstreaming was determined to be the mechanism of hemolysis. The threshold shear stress for hemoglobin release was found to be in the range 300–450 Pa with 34% uncertainty for both cases.

Microstreaming established in the vicinity of a pulsating bubble or a small vibrating probe has a sharp velocity drop across a thin boundary layer whose thickness δ is described by

$$\delta=\sqrt{\eta/(\pi f\rho)}, \tag{1}$$

where η and ρ are, respectively, viscosity and density of the liquid and f is frequency. According to Rooney,[7,8] the velocity gradient G associated with the drop is given by

$$G=2\pi f\epsilon_0^2/(R_0\delta), \tag{2}$$

a)Electronic mail: jwu@zoo.uvm.edu

 0001-4966/2002/111(3)/1460/5/$19.00 © 2002 Acoustical Society of America

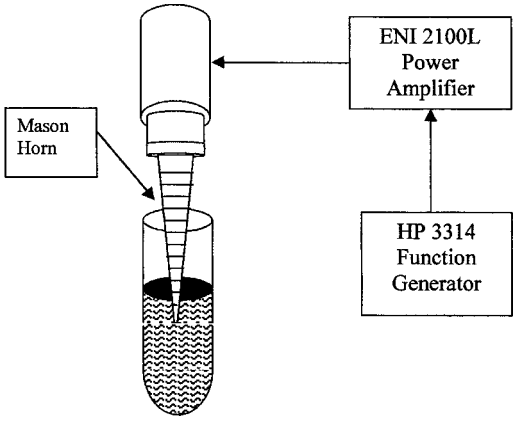


FIG. 1. An illustration of the experimental setup.

where ϵ_0 is the displacement amplitude of a vibrating bubble or a probe of radius R_0. Thus the shear stress S incurred from the velocity gradient is given by

$$S = \eta G = 2\pi^{3/2}\epsilon_0^2(\rho f^3 \eta)^{1/2}/R_0. \tag{3}$$

For the study reported here, it was hypothesized that the reparable sonoporation can also be generated by shear stress related to microstreaming. The threshold shear stress generating reparable sonoporation should be lower than that for hemoglobin release quoted above.

II. EXPERIMENTAL METHODS

The detailed description of the sample preparation may be found in our previous publications.[2,3] Here a brief version is given as follows.

Jurkat lymphocytes were incubated in a 37 °C, 5% CO_2 atmosphere in 25 cm^2 growth area tissue culture flasks (Becton-Dickson, Oxnard, CA) in a solution of RPMI Medium 1640 (GIBCOBRL, Grand Island, NY) supplemented with 10% fetal bovine serum (FBS). For exposures, the cells were diluted to a concentration of 4×10^5 cells/ml and supplemented with 10% (by volume) fluorescein isothyocyanate-dextran (FITC) solution prepared by mixing 1 g FITC (FD-500 S, Sigma Chemical Co., St. Louis, MO) with 10 mL phosphate buffered saline. The FITC has a molecular weight of about 500 kDa, and is normally unable to penetrate the cell membrane. For each trial, 0.5 ml of the cell suspension was placed in a plastic culture tube of 12 mm diam and 75 mm length (VWR Scientific, West Chester, PA) using an aseptic technique. Six samples were prepared for each case, for example, six samples were used for control (sham-exposed) and six samples were used for a specific time exposure to a vibrator (described later) excited at a specific voltage, and so on.

The vibrator used to produce microstreaming was a Mason horn made in this laboratory consisting of a stainless steel stepped (exponentially tapered) bar glued to a cylindrical PZT tube transducer (3 mm thickness, 10.5 cm long, and 2.5 cm diam), as shown in Fig. 1. This vibrator was tuned to 21.4 kHz. The total length of the cylindrical metallic probe was 17 cm and the tip diameter of the smallest section of the probe was 400 μm. The Mason horn was driven by a Hewlett Packard 3314A function generator (Agilent Technologies, Palo Alto, CA) and an ENI 2100L rf power amplifier (ENI, Rochester, NY), as shown in Fig. 1. The oscillation of the tip was along the transverse direction and the transverse displacement was measured using a calibrated microscope. The shear stress produced near the tip of the Mason horn was then calculated using Eq. (3) after the displacement was measured.

It was not possible to measure the acoustic pressure amplitude produced by the oscillating Mason horn inside of the culture tube due to the space limitation. The acoustic pressure produced by the Mason horn was measured in a large tank filled with distilled water by a calibrated 8 mm diam hydrophone (model 8103), a charge amplifier (model 2635, Brüel and Kjær, DK-2850 Nærum, Denmark), and a digital oscilloscope (Infinium 54810, Agilent Technology, Palo Alto, CA). It was found that the acoustic pressure was azimuth dependent but not sensitive to the radial distance of the hydrophone from the horn tip as long as it was rather near the tip of the horn. The hydrophone, placed about 1 mm away from the tip of the Mason horn, was moved around the horn until the maximum reading was registered. In order to find out the transmission coefficient of the culture tube and estimate magnitude of the acoustic pressure inside of the culture tube when the Mason horn vibrated, the acoustic pressure measurement was repeated in the tank when the tip of the Mason horn was surrounded, and not surrounded. Under the former situation, the hydrophone was placed outside of the culture tube.

A specific voltage at a frequency 21.4 kHz was applied to the transducer, with the Mason horn immersed in the culture tube containing the suspension described earlier and the sample was exposed to the sound field produced by the Mason horn for a specific time. Then, the sample was centrifuged twice and the cells were washed with cold RPMI (serum-free) to remove FITC-dextran from the surrounding solution. The cells were resuspended in about 0.4 mL RPMI and 0.1 mL Trypan Blue dye, gently mixed and counted using a hemacytometer. Each count involved pipeting about 100 μL of the solution and placing a small drop in the hemacytometer chamber. A Nikon Optiphot-2 microscope was used for counting. It has both ''normal'' and ''epifluorescent'' modes. For the former, white light was used and for the latter the excitation wavelength was 440 nm generated by a mercury lamp and the wavelength of the fluorescence produced by FITC-dextran molecules trapped within the cells was 530 nm. The number of cells that were stained with Trypan Blue (B) and the total number of cells (T) within the hemacytometer were counted using the ''normal'' mode. The fluorescent cells (F) were counted using the ''epifluorescent'' mode. Thus, the percentage of cells exhibiting fluorescence can be calculated by the quotient F/T. A stained cell indicated that the cell membrane was lethally damaged and the cell was not able to exclude the Trypan Blue (molecular weight is 960 Da). A fluorescent cell showed that the plasma membrane was previously ''open,'' but had since been repaired, which prevented the FITC-dextran molecules from leaving the cell after they entered the cell. Each data point is an average of six independent trials; data are presented as the average, plus and minus one standard deviation.

TABLE I. Mason horn calibration.

Voltage (mV)	50	100	150	200	250	300	350	400	500	600
Peak pressure (Pa)	400	950	1780	2100	2250	2590	2770	3000	3700	4100
Displacement (μm)	2.5	5.1	7.8	10.3	12.9	15.4	18.1	20.6	25.7	31.0
Velocity gradient $\times 10^3$ (1/s)	0.97	4.0	9.4	16.4	25.7	36.7	50.7	65.6	102	149
Shear stress (Pa)	1.2	5.2	12	21	33	47	65	84	131	191

The viscosity of the suspension was measured by using a calibrated capillary-type viscometer tube (Fisher Scientific, Pittsburgh, PA). The density of the suspension was determined by its mass (measured via a digital balance), divided by volume (measured via a volumetric flask).

III. RESULTS

The viscosity and density of the suspension were determined to be 1.28×10^{-3} (Pa s) and 1.01×10^{3} kg/m^3, respectively.

The determined displacement amplitude of the tip of the vibrating Mason horn, the velocity gradient near the tip, the acoustic pressure amplitude, and the shear stress produced by the oscillating Mason horn as functions of voltage applied to the ENI power amplifier input are listed in Table I. The error of shear stress estimated from this procedure was close to that mentioned in Rooney's experiment;[7] it was primarily due to the nonuniform nature of the acoustic streaming and was estimated to be on the order of 34%. Since acoustic pressure was measured in distilled water in a large dimension tank, its value might be different from that inside the culture tube filled with the suspension as reflection did occur at the wall of the culture tube. However, the acoustic pressure amplitudes measured from inside (P_i) and outside (P_t) of the culture tube indicated the acoustic pressure transmission coefficient, that is defined as P_t/P_i, to be 0.77. Considering this result and the fact that the viscosity of the suspension was 1.28 times higher than that of water, it suggested that the standing wave effects were small and the acoustic pressure amplitude inside the culture tube should not be too much higher than the values measured from the open distilled water condition.

Figure 2 shows the percentage of cells exhibiting reparable sonoporation as a function of the displacement amplitude of the vibrating tip after 1 min exposure. When the displacement amplitude increased from 0 (control), the percentage of cells that experienced reparable sonoporation did not change significantly until the displacement reached 10 μ. One-sided *t*-test results indicated the percentage of cells that exhibited reparable sonoporation were statistically identical ($p>0.1$) when the displacement amplitude was below or equal to 7.8 μ. The results corresponding to the tip amplitudes greater than 7.8 μm showed significantly higher than those of the control ($p<0.0006$).

Further study was focused on the reparable sonoporation effect produced by the Mason horn as a function of time. Experiments were performed using the voltages corresponding to 5.1 and 7.8 μm tip oscillation amplitude as the applied voltages. For the case of the 5.1 μm tip oscillation amplitude, the percentage of cells experiencing RS was found to be statistically indistinguishable (data not shown) between 1 and 7 min exposures ($p>0.01$). For the case of the 7.8 μm tip oscillation amplitude, the results were quite different; they varied with time. As shown in Fig. 3(a), for exposure time up to 4 min, the data can be fitted to a linear function of time, resulting in an *R* value of 0.985 (an *R* value of 1 indicates a perfect fit). The percentage of cells experiencing RS leveled off for exposure time longer than 5 min. Figure 3(b) is a plot of the percentage of cells stained with Trypan Blue versus time when the tip amplitude was 7.8 μm. For exposure time up to 4 min, the data can be fitted to a linear function, resulting in an *R* value of 0.996. For exposure time greater than 4 min, the slope increased significantly.

A similar plot that contains average percentage cells showed fluorescence and were stained with Trypan Blue versus time respectively, was presented for the case of the tip amplitude being 25.7 μm, as shown in Fig. 4. The data indicated that that the average percentage cells that showed fluorescence and were stained with Trypan Blue were signifi-

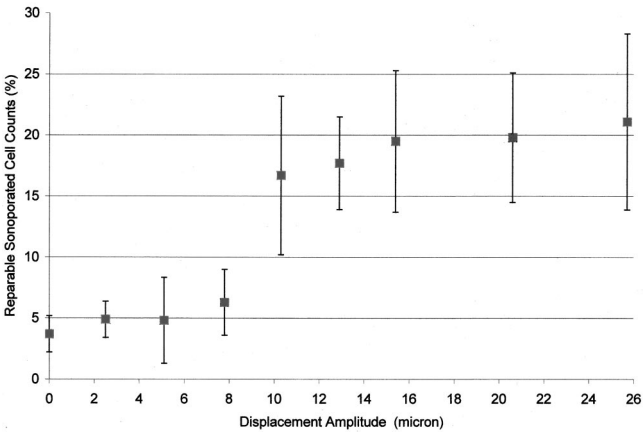
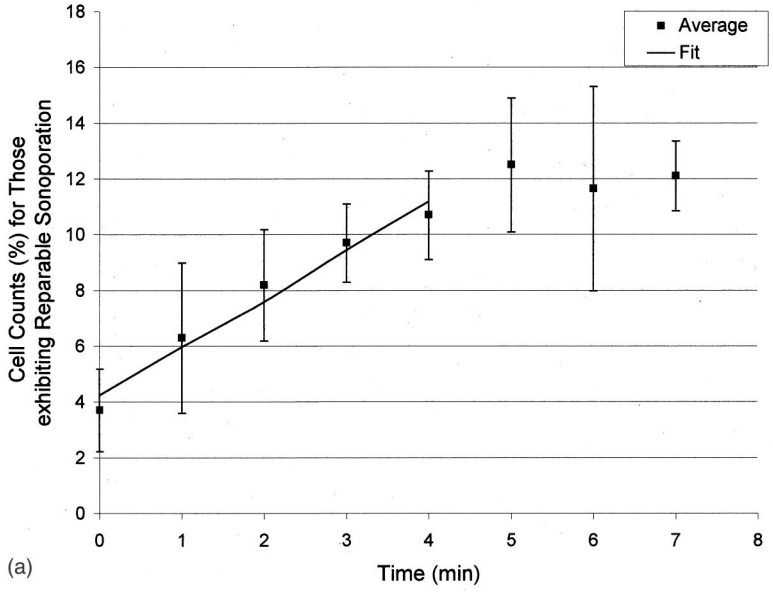


FIG. 2. The percentage of cells exhibiting reparable sonoporation as indicated by fluorescence versus the displacement amplitude of the Mason horn tip. Each data point represents the average and standard deviation of six independent trials.



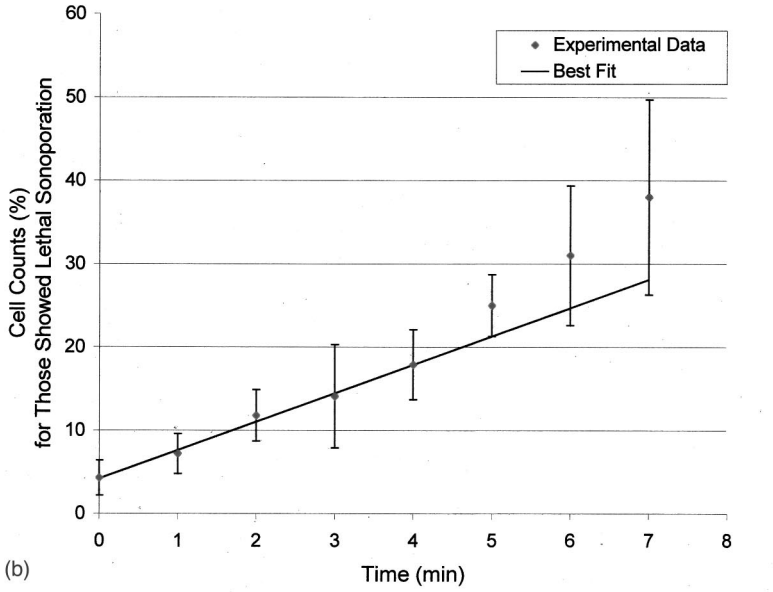


FIG. 3. (a) The percentage of cells exhibiting reparable sonoporation versus time when the displacement amplitude of the tip was 7.8 μm. Each data point represents the average and standard deviation of six independent trials. A linear function was fit to all of the data for exposure time less than 4 min, with an R value of 0.985. (b) The percentage of cells stained with Trypan Blue versus time when the tip amplitude was 7.8 μm. Each data point represents the average and standard deviation of six independent trials. For exposure time up to 4 min, the data can be fit to a linear function resulting in an R value of 0.996. For exposure time greater than 4 min, the slope increased significantly.

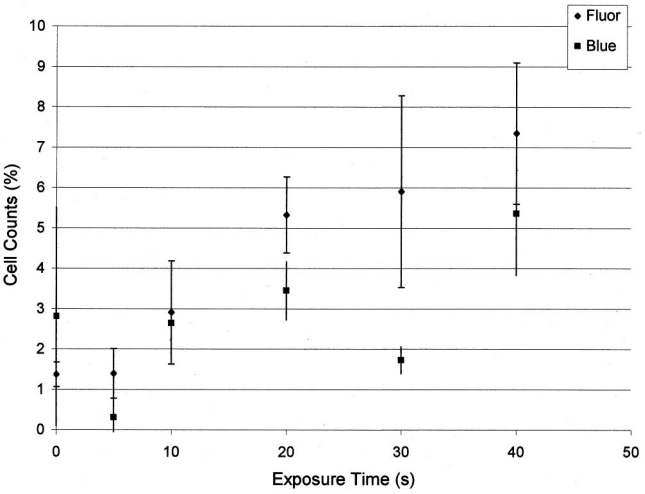


FIG. 4. Percentage of cells exhibiting fluorescence and stained with Trypan Blue versus time when the tip amplitude was 25.7 μm.

cantly different from the control ($t=0$) when the exposure time $\geqslant$10 s ($p<0.03$) and 40 s ($p<0.05$), respectively.

IV. SUMMARY AND DISCUSSION

The measured acoustic pressure amplitude generated by the vibrating Mason horn ($\leqslant$4100 Pa), as shown in Table I, was too low to produce inertial cavitation.[11] Furthermore, no bubble or bubble activity was observed during insonification via usual observation or by the hydrophone as a passive cavitation detector.[12] The microstreaming pattern similar to that shown in Fig. 5 was observed by using small latex particles mixed with the cell suspension as tracers. Therefore, like the earlier experiments[7–10] it is reasonable to conclude that the shear stress associated with microstreaming produced by the Mason horn was the primary reason for the cells showing RS. It has also been demonstrated that the amplitude of the vibrating tip of the Mason horn was the determinating factor for the percentage of cells exhibiting RS. When the ampli-

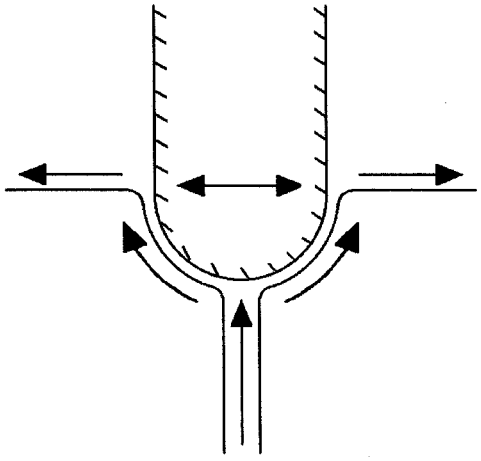


FIG. 5. Microstreaming pattern observed around the Mason horn tip.

tude was less than 7.8 μm, the percentage of cells exhibiting RS was statistically indistinguishable from that of the control ($p>0.05$) and also independent of time up to 7 min. The value of RS became proportional to time for $t\leqslant 4$ min and finally leveled off for $t>4$ min when the amplitude was equal or greater than 7.8 μm. This seems to suggest that the shear stress associated with the microstreaming of the vibrating Mason horn tip when the vibration amplitude is 7.8 μm is the threshold value for RS for exposure time up to 7 min. According to Table I and considering the error discussed earlier, the threshold shear stress of RS should be 12 ± 4 Pa.

An effort was made to count the cells stained with Trypan Blue. As shown in Fig. 3(b), this portion of cells increased linearly with time initially ($\leqslant 4$ min) when the amplitude of the vibrating tip reached 7.8 μm and the slope of this curve increased significantly after 4 min exposure. This result is very similar to what was observed at the stage prior to cell hemolysis; much higher shear stress was needed to cause cell significant hemolysis.[8]

As the data indicate, there were always some cells showing RS or LS, even for freshly prepared cells. This fact suggests that the cells used in the experiment were not uniformly healthy before they were exposed to the sound field. The less healthy cells were apparently much more vulnerable to the effects induced by microstreaming. For the same magnitude of shear stress, they might show LS while other healthy cells exhibited RS. After the cells experienced RS, they might become more vulnerable to the sonic effects and finally became lethally sonoporated.

Since the boundary layer thickness for our case is only 4.34 μm, as suggested by Rooney[8] for his study, cells might pass through the effective treatment region at the boundary layer several times under our situation. They experienced some slight reversible damage during each passage until the damage became large enough to allow FITC entering the cell, yet they could still repair themselves. In this sense, the threshold of reparable sonoporation determined in this study is for treatment time up to 7 min. It has not determined how low the threshold will become when the treatment time is increased to longer than 7 min. Furthermore, the observed range of acoustic microstreaming field as shown in Fig. 4 increases with the amplitude of the vibrating tip. Thus, not only the shear stress increases but the probability of more cells moving to the effective treatment region also increases with the amplitude of the vibrating tip. Eventually, significant cells will suffer irreversible lethal damage at high amplitudes. The data shown in Fig. 4 support the above view point. For the tip displacement amplitude increased to 25.7 μm, it took only 10 and 40 s to observe the reparable and lethal sonoporations, respectively. In general, short exposure needs large tip displacement; this observation agrees well with the results in the literature.[13,14]

There were some fragmented cells after insonification at higher amplitudes of tip oscillation. Although their debris could be observed during a cell counting process, they could not be easily counted. Both the cells stained with Trypan Blue but otherwise still keeping their integrity and those fragmented after exposure to a sound field were lethal damaged cells; the latter were even more severely damaged than the former.

The results of this study may shed light on the possible mechanism of RS demonstrated by using relatively low ultrasound and a contrast agent.[1–3] The microstreaming surrounding a pulsating encapsulated microbubble may generate enough shear stress to generate RS. A further study of RS associated with pulsating encapsulated microbubbles is in process.

ACKNOWLEDGMENT

The authors want to thank Dr. Wesley L. Nyborg for his insightful suggestions.

[1] S. Bao, B. D. Thrall, and D. Miller, Ultrasound Med. Biol. **23**, 953–959 (1997).
[2] M. Ward, J. Wu, and J. F. Chiu, J. Acoust. Soc. Am. **105**, 2951–2957 (1999).
[3] M. Ward, J. Wu, and J. F. Chiu, Ultrasound Med. Biol. **26**, 1169–1175 (2000).
[4] W. J. Greenleaf, M. E. Bolander, G. Sarkar, M. B. Goldring, and J. F. Greenleaf, Ultrasound Med. Biol. **24**, 587–595 (1998).
[5] W. L. Nyborg, "Physical principles of ultrasound," in *Ultrasound: Its Applications in Medicine and Biology*, edited by F. J. Fry (Elsevier, Amsterdam, 1978), Part I.
[6] J. Wu and G. Du, Ultrasound Med. Biol. **19**, 167–176 (1993).
[7] J. A. Rooney, Science **169**, 869–871 (1970).
[8] J. A. Rooney, J. Acoust. Soc. Am. **6**, 1718–1724 (1972).
[9] A. R. Williams, D. E. Hughes, and W. L. Nyborg, Science **169**, 871–873 (1970).
[10] A. R. Williams, *Ultrasound: Biological Effects and Potential Hazards* (Academic, London, 1983).
[11] E. A. Neppiras, Phys. Rep. **61**, 159–251 (1980).
[12] A. R. Roy, S. I. Madanshetty, and R. E. Apfel, J. Acoust. Soc. Am. **87**, 2451–2458 (1990).
[13] R. P. Rand, Biophys. J. **4**, 303–316 (1964).
[14] L. B. Leverett, J. D. Hellums, C. P. Alfrey, and E. C. Lynch, Biophys. J. **12**, 257–273 (1972).

Acoustic features of male baboon loud calls: Influences of context, age, and individuality

Julia Fischer[a)]
Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19104

Kurt Hammerschmidt
German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany

Dorothy L. Cheney
Department of Biology, University of Pennsylvania, Leidy Building, 415 South University Avenue, Philadelphia, Pennsylvania 19104

Robert M. Seyfarth
Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19104

(Received 29 May 2001; revised 28 October 2001; accepted 2 November 2001)

The acoustic structure of loud calls (''wahoos'') recorded from free-ranging male baboons (*Papio cynocephalus ursinus*) in the Moremi Game Reserve, Botswana, was examined for differences between and within contexts, using calls given in response to predators (alarm wahoos), during male contests (contest wahoos), and when a male had become separated from the group (contact wahoos). Calls were recorded from adolescent, subadult, and adult males. In addition, male alarm calls were compared with those recorded from females. Despite their superficial acoustic similarity, the analysis revealed a number of significant differences between alarm, contest, and contact wahoos. Contest wahoos are given at a much higher rate, exhibit lower frequency characteristics, have a longer ''hoo'' duration, and a relatively louder ''hoo'' portion than alarm wahoos. Contact wahoos are acoustically similar to contest wahoos, but are given at a much lower rate. Both alarm and contest wahoos also exhibit significant differences among individuals. Some of the acoustic features that vary in relation to age and sex presumably reflect differences in body size, whereas others are possibly related to male stamina and endurance. The finding that calls serving markedly different functions constitute variants of the same general call type suggests that the vocal production in nonhuman primates is evolutionarily constrained. © *2002 Acoustical Society of America.*
[DOI: 10.1121/1.1433807]

PACS numbers: 43.80.Ka, 43.80.Jz, 43.64.Tk [WA]

I. INTRODUCTION

The loud, two-syllable barks given by adult male baboons (*Papio cynocephalus*) are familiar to anyone who has traveled through the wildlife reserves of Eastern or Southern Africa. These calls, or ''wahoos'' (Hall and DeVore, 1965), are louder than any other calls in the baboons' repertoire and appear to be adapted for long-range communication (Waser and Brown, 1984). Depending on wind conditions they are audible from more than one kilometer (Hall and DeVore, 1965). Wahoos typically consist of a loud bark—the ''wa'' syllable—and a second, lower amplitude, ''hoo'' syllable. Wahoos are given in three broadly different contexts: during predator encounters (alarm wahoos), during aggressive interactions with conspecifics (contest wahoos), and, less commonly, when a male has become separated from the group (contact wahoos). Alarm wahoos are given in response to a variety of different predators, including lions (*Panthera leo*), leopards (*P. pardus*), and crocodiles (*Crocodilus niloticus*). In some cases, the hoo syllable is hardly audible or drops out entirely so that single wa syllables (barks) can also be heard. Contest wahoos appear to play a role in competitive interactions between males. They are frequently heard at dawn, when males commonly participate in a chorus of calling in apparent displays of dominance. They are also given during aggressive interactions, as males either chase each other or herd females (Buskirk *et al.*, 1974; Cheney and Seyfarth, 1977; Saayman, 1971). Males involved in such chases often engage in conspicuous arboreal displays. These chases appear to function at least in part as displays of stamina and fighting ability, and they usually involve higher-ranking males of adjacent ranks who appear to be challenging each other's dominance status. Finally, wahoos may also be given by males who appeared to have lost contact with the group. These calls appear to have the same function as the contact barks given by females, infants, and juveniles (Cheney *et al.*, 1996; Fischer *et al.*, 2001a; Rendall *et al.*, 2000).

In this paper, we present an analysis of the acoustic structure of wahoos given by male chacma baboons (*P.c. ursinus*). We first examine possible differences between contest and alarm wahoos. We then investigate the extent to which contest wahoos given in different contexts (e.g., male–male aggression versus herding of females) differ from one another. Next, we examine age differences within and

[a)]Current address: Julia Fischer, Max-Planck Institute for Evolutionary Anthropology, Inselstr. 22, 04103 Leipzig, Germany. Electronic mail: fischer@eva.mpg.de

between contexts. Finally, we focus on individual variation. Because the first syllable of the wahoo call, the wa, is similar to the single-syllable bark given by females (Fischer *et al.*, 2001a), and the two calls apparently develop from the same call type in juveniles (personal observation), we also explore differences in relation to sex. We examine age- and sex-related differences to test predictions about the relationship between body size and acoustic features.

A number of variables should provide cues to the body size of the caller. Specifically, the duration of loud calls such as wahoos should be related to lung capacity which in turn is related to body size (Fitch and Hauser, in press). Thus, we expect adult males to be able to produce longer calls than either younger males or females. Another acoustic parameter that is frequently related to body size is the fundamental frequency (Darwin, 1872; Morton, 1977; for a discussion see Fitch and Hauser, in press). We therefore predict that adult males will produce calls with a significantly lower fundamental frequency than either younger males or females. Cues to body size can also be provided by formant frequencies. The primary determinant of formant frequencies is the length of the vocal tract (Fant, 1960; Fitch, 1997; Liebermann and Blumstein, 1988). As Fitch (1997) demonstrated, a lengthening of the vocal tract tube leads to a decrease in the average spacing between successive formants (''formant dispersion''). If vocal tract length is correlated with body size, formant dispersion provides an honest cue to body size. In rhesus macaques (*Macaca mulatta*) and dogs (*Canis familiaris*), for instance, there is a close correlation between body size, vocal tract length, and formant dispersion (Fitch, 1997; Riede and Fitch, 1999). Accordingly, we expect a lower formant dispersion in adult males than in animals of other age or sex. By examining the relationship between call structure and function, we aim to formulate hypotheses about the potential information available to the receivers of these signals.

II. METHODS

A. Study site and subjects

The study site lies in the Moremi Wildlife Reserve in the Okavango Delta, Botswana, a huge inland delta fed by the Okavango river. Yearly rainfall in Angola causes the Okavango and its tributaries to rise and flood the grasslands. Only slightly elevated patches of woodland, or ''islands,'' which range in size from less than one to over hundreds of hectares, remain uncovered (Hamilton *et al.*, 1976; Ross, 1987). During the flood, the baboons ford the floodplains to travel from one island to the next.

The average size of a baboon group's home range in this area is 450 ha (range 210–650 ha; Hamilton *et al.*, 1976). The study group, group C, has been observed more or less continuously for more than 20 years. Matrilineal relatedness of all natal animals is known. A number of comprehensive studies describe aspects of the social behavior (Cheney *et al.*, 1996; Palombit *et al.*, 1999, 2000; Silk *et al.*, 1996, 1999), as well as the vocal communication of this population (Cheney *et al.*, 1996; Cheney and Seyfarth, 1997; Fischer *et al.*, 2000, 2001b; Rendall *et al.*, 1999, 2000). During the period of this study, group size ranged from 79 to 84 subjects.

B. Vocal recordings and behavioral observations

We recorded vocalizations *ad libitum* with a Sony WM TCD-100 DAT recorder and a Sennheiser directional microphone (K6 power module and ME66 recording head with MZW66 pro windscreen) during an 18-month period between January 1998 and June 1999. Whenever a male within earshot started calling, we approached the subject and attempted to record his calls. Information regarding caller identity, identity of and distance to subjects in the vicinity, predator type, predator behavior, and predator distance were spoken onto the tape. For the acoustic analysis, we used only calls given at a distance of 5–15 m from the microphone because estimates of some acoustic parameters can be influenced by signal transmission over large distances. This is particularly true for variables that characterize higher frequency components (K. Hammerschmidt and J. Fischer, unpublished data, 2001). Only calls whose context could be determined precisely were subjected to acoustic analysis.

C. Acoustic analysis

We visually inspected and sampled calls that were not disturbed by background noise (i.e., bird song, other animals calling) at a sample frequency of 20 000 Hz using RTS (Engineering Design, Belmont, MA; Beeman, 1996) or COOL EDIT 96 (Syntrillium, Phoenix, AZ). To obtain a better frequency resolution, we first reduced the sample frequency to 10 000 Hz. Next, we used the SIGNAL sound analysis system (Engineering Design, Belmont, MA; Beeman 1996) to conduct a fast Fourier transform (1024-pt FFT; time step: 5 ms; frequency range: 4000 Hz; frequency resolution: approximately 10 Hz). We sampled a mean of 2.9 calls per bout (median: 2, range: 1–13) from a total of 83 call bouts. In the contexts from which we sampled the calls, no high-amplitude vocalizations other than calls categorized as ''wahoos'' occurred. In some of the recording sessions, animals also uttered low-amplitude grunts which were not considered in the present analysis.

We submitted the resulting frequency time spectra to a custom software program (LMA 8.4) that extracts different sets of call parameters from acoustic signals (Hammerschmidt, 1990). Below, we briefly describe the underlying principle for the different groups of measurements. Figure 1 illustrates some of the variables considered in the analyses. First, we calculated an autocorrelation function for every time segment in a given call. Depending on the number of peaks and the periodicity of the autocorrelation function, each time segment was classified as noisy (no peaks could be detected), complex (some peaks could be detected but they were not periodic), or tonal (peaks were periodic). We then determined the percentage of time segments in a given call that was noisy. Second, we measured the statistical distribution of the frequency amplitudes in the spectrum. For each time segment, the overall amplitude is determined. Subsequently, we calculated the frequency at which the distribution of the amplitude in the frequency spectrum (hereafter ''distribution of frequency amplitudes'') reaches the first and second quartile of the total distribution, respectively. Third, we calculated a set of parameters describing the first three dominant fre-

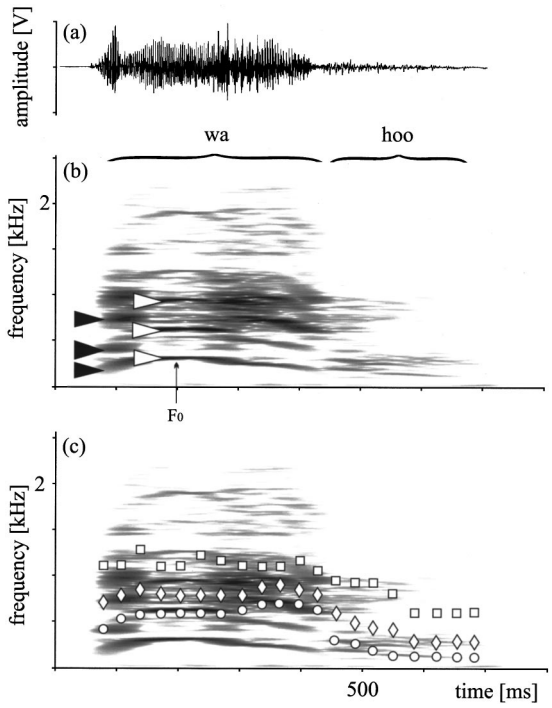


FIG. 1. (a) Amplitude and (b) spectrogram of a wahoo call indicating some of the variables determined in the acoustic analysis. Black triangles denote the first, second, and third dominant frequency at the beginning of the call; white triangles are in that given time segment. Depending on the call structure, the first dominant frequency may or may not correspond to the fundamental frequency. For this call segment, the fundamental corresponds to the first dominant frequency band. The two syllables are typically separated by a short interval of little energy, and hoo syllables consistently exhibit a much lower amplitude than wa syllables. (c) Distribution of frequency amplitudes in the spectrum. Circles mark the first dfa, diamonds the second dfa, and squares the third dfa for each time segment. Subsequently, the start and mean values for the first and second dfa were calculated.

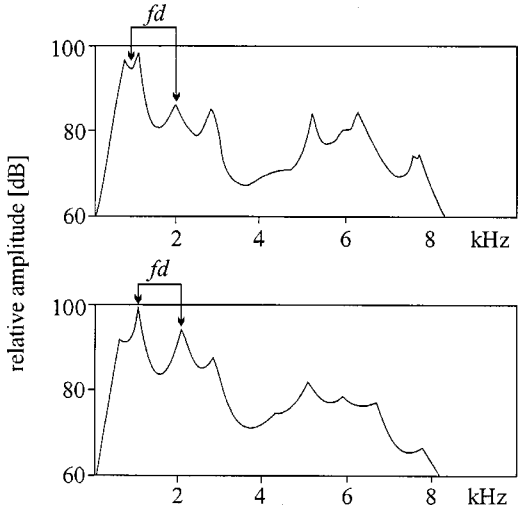


FIG. 2. Filter functions derived from an LPC analysis fitted with 23 coefficients. Both functions are derived from two contest wahoos given by the same individual. The upper graph has a lowest formant with a double peak, the lower graph with a single peak. Formant dispersion (*fd*) was calculated as the difference between the midpoint of the lowest two or the lowest peak, respectively, and the following peak.

quency bands (dfb). The dominant frequency bands are characterized by amplitudes that exceed a given threshold in a consecutive number of frequency bins. Note that the numbers of the dominant frequency bands count from the lowest frequency up; the first dfb is not necessarily the dfb with the highest amplitude. Fourth, we determined the global energy distribution in the calls (''formant-like structures''). Fifth, we specified the location and the modulation of the peak frequency, the frequency with the highest amplitude in a certain time segment. Sixth, we determined the mean and maximum frequency range. We also calculated the duration of both the ''wa'' and the ''hoo'' syllable of the call and the ratio of the maximum amplitude of the wa versus the hoo syllable. Last, we used a cursor to measure the fundamental frequency in the middle third of the wa syllable (see the Appendix for a list of the calculated parameters). A description of the various algorithms is given in Schrader and Hammerschmidt (1997).

Because some calls exhibited diagnostics of period-doubling bifurcations (insertions of subharmonic episodes with approximately $F_0/2$, $3F_0/2$ etc.; see Wilden *et al.*, 1998 for details), it appeared that nonlinear effects also play a role in the production of wahoos. We therefore inspected spectrograms visually and noted whether or not they exhibited signs of period doubling.

Additionally, we conducted a linear prediction coefficient (LPC) analysis using SOUNDSCOPE (GW Instruments, Somerville, MA) or SPEECHSTATION2 (Sensimetrics, Somerville, MA). LPC analysis is an alternative to Fourier analysis for describing the spectrum of a signal segment. We used two approaches. First, to obtain a more detailed analysis of the lower frequency range, we calculated 12 LPC coefficients from a 4096-pt FFT spectrum at a frequency range of 4000 Hz with SPEECHSTATION2. In this analysis, we determined the frequency of the peak of the first three formants (filtering frequency). Second, to obtain a higher resolution, we calculated 23 LPC coefficients from a 4096-pt FFT spectrum at a frequency range of 10 000 Hz with SOUNDSCOPE. For this analysis, we selected a 100-ms time segment of a tonal part of the wa syllable with a nearly constant fundamental frequency. To obtain a better characterization of the higher frequency parts, we used a pre-emphasis filter of 6 dB/octave, increasing the higher frequency components. From this analysis, we determined the location of the first eight peaks, the amplitude ratio between the peak i and peak i+1, and the difference between the frequency of peak i and peak i+1. It was not possible to analyze all of the hoo syllables in this way, as many either had a weak amplitude or were absent. This was particularly true for hoo syllables by adolescent and subadult males.

An inspection of the frequency distribution of the second formant revealed a bimodal distribution with a first peak around 1100 Hz and a second peak around 2000 Hz. This occurred because, in some cases, the peak detection algorithm identified two peaks in the function describing the lower vocal-tract resonances. For those cases that showed such a ''double peak,'' we determined the formant dispersion as the difference between the midpoint of the two lowest and the following peak (Fig. 2). Otherwise, we calculated the difference between the first and the second formant peak.

Finally, we examined call rate. First, we randomly se-

lected one bout per individual and measured the interval from the beginning of a call to the beginning of the subsequent call using COOLEDIT 96. A total of 22 bouts was chosen for analysis, recorded from a total of 16 subjects. On average, we measured 8.9±6.2 ($\bar{x}\pm$s.d.) intervals per bout. We then calculated the number of calls/min. To test for differences in call rate between the alarm and contest contexts, we applied the ''repeated samples with missing values test'' (Mundry, 1999), as not all subjects were represented in both contexts. This is a permutation test that delivers a variant of a Friedman one-way analysis of variance, providing a way to deal with missing values. We used 10 000 permutations. Due to the small sample size, we did not consider the call rate in the contact context for statistical analysis, and simply report the results below.

D. Statistical analysis

We used a discriminant function analysis to identify differences in vocalizations between contexts and among individuals. Discriminant function analysis identifies a linear combination of quantitative predictor variables that best characterizes the differences among groups (Bortz, 1993). Variables are combined into one or more discriminant functions. Variables that fail a tolerance test, i.e., are an almost linear combination of other variables, do not enter the analysis. The discriminant function analysis establishes *N*-1 discriminant functions, where *N* is the number of groups (e.g., contexts or individuals) in the analysis. Discriminant function analysis has been successfully applied to determine acoustic differences between individuals or contexts in a number of studies (e.g., Fischer *et al.*, 2001a; Gouzoules and Gouzoules, 1989; Smith *et al.*, 1982).

Discriminant function analysis provides a classification procedure that assigns each call to its appropriate group (correct assignment) or to another group (incorrect assignment). For external validation, we used a tenfold cross validation in which we randomly selected roughly 90% of the data to calculate the discriminant functions and used the remaining 10% of cases for classification. We iterated this procedure ten times.

The percentage of calls classified correctly is not affected by repeated measures. However, it may be influenced by the number of variables in the analysis and the ratio of the number of variables to the number of cases (Bortz, 1993). This is particularly true for the ''direct method,'' in which all variables are entered simultaneously. We preferred this method over the ''stepwise procedure'' because it preserves all available information. As a result, for those analyses with a smaller sample size we first explored the data set to identify variables that contributed to a discrimination of contexts or individuals and excluded those variables that did not vary among any of them. The amount of correct classification is also unaffected by nonindependent data points, in contrast to the significance test of the discriminant functions. We therefore only considered the classification results. For further statistical evaluation, we calculated the mean values per individual and context to avoid problems associated with pseudoreplication. In case where we used univariate statistics, we applied a sequential Bonferroni correction ($\alpha' = \alpha/(k-n+1)$, where k=number of tests, n=number of significant tests). Statistical analyses were calculated using the statistical package SPSS 9.0. All tests are two-tailed. Unless otherwise stated, the significance level is set at $\alpha = 0.05$.

TABLE I. Number of calls per individual male in the analyses of differences between alarm and contest contexts, and differences within the contest context. Callers marked with an asterisk were used in the analysis of individual differences.

	Between contexts		Within contest context		
Caller	Contest	alarm	Chase	Herd	Contest
AP*	20	6	4	20	6
AU*	18	6	13	11	2
DG		11			
EN*	19		2	14	4
KI	13			13	
KK*	17	9	8	10	20
RB	7	5		7	7
RY*	20		13		11
TH*	13	11	20	11	9
VE*	20		5	6	18
WA*	19	13	3	10	20
ZK	2			2	
Σ	168	61	68	104	90

E. Data sets

The rationale for the selection of calls was to sample an even number of calls per individual and context from as many independent recordings as possible. We first entered all calls that we recorded into a database. Next, we pseudorandomly chose a given number of calls per recording session. In cases where an individual provided less than the required number of calls, all calls available were taken.

For acoustic analysis, we created different data sets. The first set was used to examine differences between contest and alarm wahoos and contained up to 20 calls per individual and context. Some of the acoustic parameters listed in the Appendix (F_0 mean and all parameters describing the structure of the hoo) could not be determined for every call; there were only 193 of the 229 calls for which there were no missing values. Since we aimed to retain as many calls by as many different males as possible in the analysis, we first performed a test run with these 193 calls. We found that F_0 and the variable describing the spectral characteristics of the hoo did not play a role in the discrimination of contexts. We then reverted to the entire data set of 229 calls and excluded these parameters from the analysis. To examine the differences within the contest condition, we raised the number of calls in the analysis so that the maximum number of calls for each male and subcontext was 20. The total number of calls in this analysis was 262 (see Table I for the distribution of calls across individuals and subcontexts).

For analyses in relation to age and context (Table II), we calculated the individual means of ten variables: noise, formant dispersion, mean fundamental frequency, wa duration, hoo duration, amplitude ratio wa/hoo, start and mean first quartile of the distribution of frequency amplitudes, mean first dominant frequency band, mean peak frequency. Finally,

TABLE II. Number of individual males represented in the different age classes and contexts in the analysis of the combined effects of age and context.

Age class	Contest	Alarm	Contact
Adult	11	7[a]	2[b]
Subadult	4	3[c]	3
Adolescent	1	1	1

[a]Six individuals also represented in the contest category.
[b]One individual also represented in contest and alarm, one individual in the contest category.
[c]One individual also represented in the contest, one individual in the contact, and one individual in the contest and contact category.

to compare the calls of males and females, we used the individual means of 20 acoustic variables that could be determined for both sexes (see the Appendix).

III. RESULTS

A. Acoustic characteristics of alarm and contest wahoos

To test whether contest and alarm wahoos are acoustically different, we first conducted a discriminant function analysis with ''context'' as the grouping variable on 229 calls recorded from 12 males (see Table I). Of the 168 contest calls, 37 were given during male–male chases, 74 were recorded during the morning chorus or when a new male had entered the group, and 57 while herding females. We compared these calls to 61 alarm wahoos recorded from seven males. Forty-six of the alarm calls were given to lions, five to cheetahs (*Acinonyx jubatus*), and ten to wild dogs (*Lycaon pictus*). The average correct assignment was 87.3%; a tenfold cross validation yielded an average correct assignment of 82.7%. Figure 3 shows spectrograms of calls that we selected according to the outcome of the acoustic analysis. The spectrograms depict typical exemplars (i.e., calls with a high assignment probability to their respective context category) of alarm and contest calls recorded from three different males, as well as acoustically more ambiguous exemplars (i.e., calls with a low assignment probability).

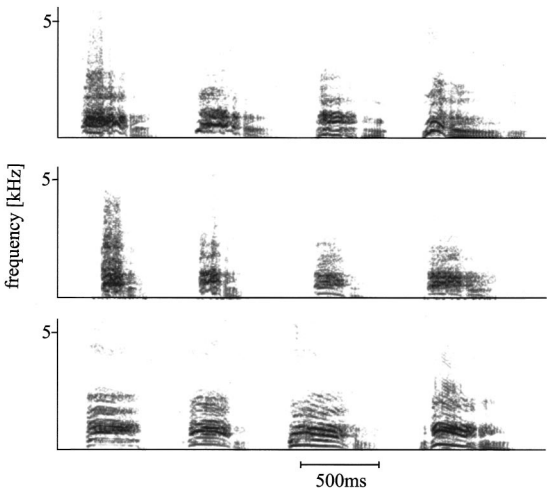


FIG. 3. Spectrograms of wahoo calls from three different males (rows 1–3). For each male, four calls are presented. From left to right: a typical exemplar (according to the outcome of the discriminant function analysis) of an alarm wahoo, an ambiguous exemplar of an alarm wahoo, an ambiguous exemplar of a contest wahoo, and a typical exemplar of a contest wahoo. Note that despite appreciable inter-individual variation, males show structurally similar differences between contexts.

Those variables that—according to the discriminant function analysis—contributed most to a discrimination of contexts were the start of the first quartile of the distribution of frequency amplitudes (univariate analysis of variance with context as fixed factor and subject as random factor; only results for test of differences between contexts are given, since in no case did ''subject'' have a significant effect; $F_{1,17}=17.9$, $P<0.01$), hoo duration ($F_{1,17}=13.3$, $P<0.05$), mean ($F_{1,17}=4.7$, $P<0.1$) and minimum first dominant frequency band ($F_{1,17}=18.4$, $P<0.01$), and the amplitude ratio wa/hoo ($F_{1,17}=8.5$, $P<0.05$; see Table III for more detailed statistics. Note that the inclusion of subject as a factor yields slightly different P levels than in the analysis reported in Table III). In general, contest wahoos had a wa syllable with lower frequency characteristics, a longer hoo duration, and a relatively louder hoo syllable than alarm wahoos. Males also gave alarm and contest wahoos at significantly different rates. While alarm wahoos were given at an $\bar{x}\pm$s.d. rate of 4.3 ± 1.1 calls/min, contest wahoos occurred at a rate of 30.5 ± 4.0 calls/min (repeated samples with missing values test: $M=19.6, N=13, P<0.01$).

An inspection of spectrograms suggested that features diagnostic of nonlinear phenomena did not map systematically onto the classification of calls. Across individuals, 25% $\pm$28% ($\bar{x}\pm$s.d.) of all contest wahoos and 16%$\pm$15% of all alarm wahoos exhibited signs of nonlinear phenomena, mostly in the middle third of the wa syllable (e.g., Fig. 3, row 3, second and fourth call from left). This difference was not significant (repeated samples with missing values test: $M=1.05, N=12$, n.s.).

B. Differences within the contest category

We found only slight differences among contest wahoos given in the three subcontexts: the average correct assignment to the three subcontexts was 65.6% (tenfold cross validation 48.5%). The parameter contributing most to the discrimination of groups was the mean first quartile of the distribution of frequency amplitudes and the start of the first dominant frequency band. However, subsequent univariate analyses of variance yielded no significant differences in any of the parameters.

C. Effects of age and context

Because we also wished to explore how contact calls compared to alarm and contest wahoos, we conducted a renewed analysis on the individual means of ten acoustic variables (see Sec. II E). However, since only few subjects contributed contact calls, we decided to include contact calls from all age classes (see Table III). Therefore, we ran an analysis in which we explored the influence of age and context simultaneously, and in which we also aimed to identify possible interactions between the factors. Figure 4 shows spectrograms of calls given by males of the three age classes in the three contexts under consideration. Since the variable amplitude ratio wa/hoo had some missing values due to the

TABLE III. Mean±s.d. and results (P-values) of univariate analyses of variance for the variables in the analysis regarding the effects of age and context, including alarm, contest, and contact wahoos from males of all three age classes n.s.: >0.1. The amplitude ratio wa/hoo had to be excluded due to missing values; the mean of the first dominant frequency band was excluded due to a significant univariate interaction.

		Context			Statistical effects	
Acoustic variable	Age class	Contest	Contact	Alarm	Age	Context
Formant disp. [Hz]	Adult	916±58	917±0	875±51	0.003	n.s.
	Subadault	906±83	913±64	984±133		
	Adolescent	996	1036	1197		
Fundamental [Hz]	Adult	312±36	297±9	329±56	0.001	n.s.
	Subadult	341±48	392±91	408±80		
	Adolescent	397	490	430		
Noise [%]	Adult	56±20	24±2	63±15	n.s.	n.s.
	Subadult	53±12	44±47	68±15		
	Adolescent	42	52	31		
'Wa' duration [ms]	Adult	351±47	472±66	322±89	0.001	0.09
	Subadult	285±78	297±81	213±16		
	Adolescent	277	323	278		
'Hoo' duration [ms]	Adult	210±64	214±77	116±62	0.001	0.084
	Subadult	98±48	113±61	34±17		
	Adolescent	0	70	0		
dfal start [Hz]	Adult	432±74	397±12	544±62	0.001	0.001
	Subadult	485±37	549±129	743±103		
	Adolescent	597	526	910		
dfal mean [Hz]	Adult	682±54	668±47	691±40	0.002	0.074
	Subadult	689±35	730±121	816±100		
	Adolescent	775	870	893		
pf mean [Hz]	Adult	755±70	742±48	735±44	0.005	n.s.
	Subadult	753±50	867±140	836±170		
	Adolescent	888	988	914		

absence of the hoo syllable in two of the adolescent males, we had to remove this variable from the analysis. Using the remaining nine variables, we observed significant multivariate differences among contest, alarm, and contact wahoos (Wilk's $\lambda=0.138$; $F_{18,30}=3.00, P<0.01$) and in relation to age (Wilk's $\lambda=0.152$; $F_{20,30}=2.78$, $P<0.01$), but no significant multivariate interaction (Wilk's $\lambda=0.108$; $F_{36,61.69}=1.39$, n.s.). However, subsequent inspection of single variables revealed a significant interaction for the mean first dominant frequency band. This variable was therefore excluded from further consideration (Zar, 1999). All other variables in the analysis except for "noise" varied significantly in relation to age (Table III). Both the mean frequency and the formant dispersion decreased with age, whereas wa and hoo duration increased (Fig. 5). Wa duration, hoo duration, and mean and start of the first quartile of the distribution of frequency amplitudes also varied significantly or marginally significantly in relation to context (Table III). *Post hoc* tests

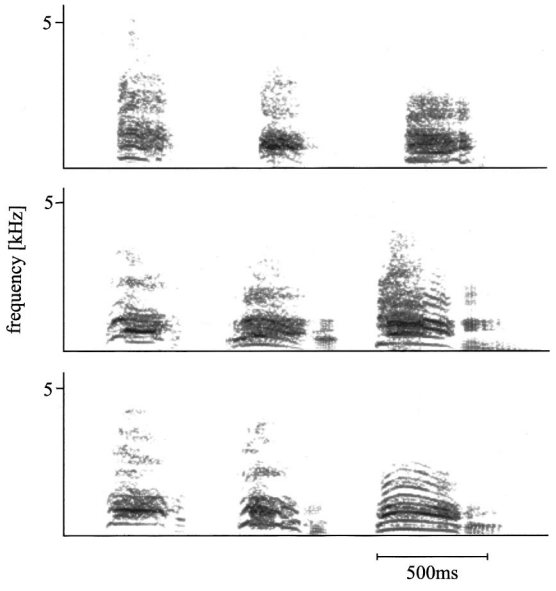


FIG. 4. Spectrograms of wahoo calls from males of three different age classes, given in three different contexts. From left to right: adolescent, subadult, adult. Top row: alarm calls, middle row: contest calls; lower row: contact calls.

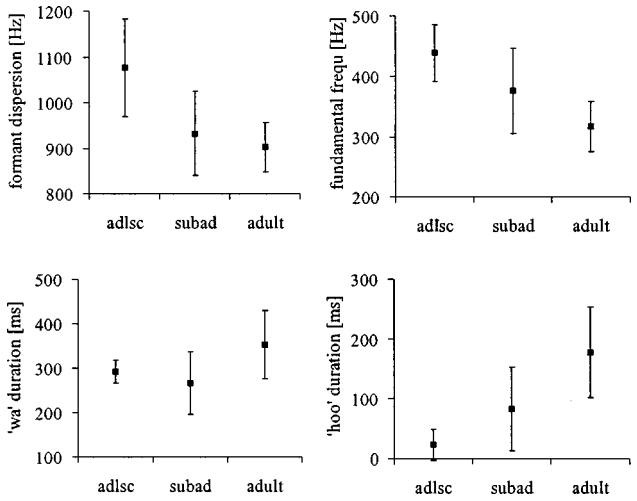


FIG. 5. $\bar{x}\pm$s.d. of four variables that vary in relation to age: formant dispersion, fundamental frequency, wa duration, and hoo duration.

revealed that the hoo duration and the mean first dominant frequency band varied significantly between contest and alarm wahoos, while the start value of the first quartile of the distribution of frequency amplitudes varied significantly between contest and alarm, as well as between contact and alarm wahoos.

For some variables, these effects were limited to specific age groups. For instance, the hoo duration in adolescent contest calls was equivalent to the one in adolescent alarm calls. In sum, we found clear differences in relation to age. Also, across age classes we found consistent differences between alarm and contest wahoos (albeit weaker effects than when the calls of adult males were analyzed separately), and slight differences between alarm and contact wahoos. We failed to find significant differences between contest and contact wahoos. However, due to the small sample size and, as a consequence, the lack of power of this analysis, we cannot rule out the possibility that differences between contest and contact wahoos do exist. Interestingly, we observed a marked difference in the rate at which calls from these two categories were given. Contact calls were given at a much lower rate: 6 of the 14 calls in our sample were given singly, and the remaining calls were given at a rate of 1.5 calls/min (range 0.2–2.4 calls/min).

To make sure that these results also held for a more balanced sample, we replicated this analysis for the calls of adult and subadult males, and included only alarm and contest calls. In this analysis, we also included the amplitude ratio wa/hoo. We observed similar significant multivariate differences in relation to age (Wilk's $\lambda=0.208$; $F_{10,15}=5.72, P<0.001$) and context (Wilk's $\lambda=0.098$; $F_{20,30}=3.28, P<0.01$), and no significant multivariate interaction (Wilk's $\lambda=0.452$; $F_{20,30}=0.73$, n.s.). All variables (see Sec. II E) except formant dispersion varied significantly in relation to age, while wa duration, hoo duration, start and mean first dominant frequency band varied in relation to context. These results support the findings of our previous analysis of all three contexts and age classes.

D. Individual differences

In the analysis of individual differences, we only considered males that contributed at least 16 calls. The data set consisted of 170 calls from eight adult males (16–24 calls per male). Calls were recorded from all major context categories (alarm, contest, and contact). We then repeated this analysis for contest wahoos only (124 calls). We did not have enough alarm or contact wahoos to permit a meaningful separate analysis. The assignment procedure of the discriminant function yielded an average correct assignment of 71.2% (cross validated 59.2%; prior probability 12.5%). The parameters that contributed most to a discrimination of individuals were start first quartile of the distribution of frequency amplitudes (one-way ANOVA $F_{7,169}=17.2, P<0.001$), wa duration ($F_{7,169}=14.3, P<0.001$), and hoo duration ($F_{7,169}=12.6, P<0.001$). However, using discriminant function analysis, all parameters were necessary to distinguish among individuals. If we only considered contest wahoos, the results were essentially replicated: the correct average assignment was 71.8% (cross validated 58.3%) with 124 calls. The same parameters contributed most to distinguishing among males.

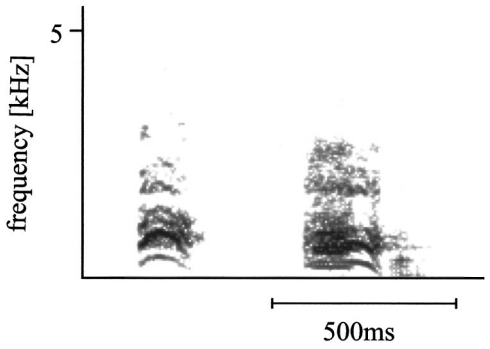


FIG. 6. Spectrograms of a female (left) and a male (right) alarm call.

E. Differences between males and females

First, we compared the acoustic structure of the wa syllable of the male alarm wahoo to its equivalent call in adult females, the single-syllable alarm bark. We used only calls that were given in response to lions. For each individual, we calculated the mean value for those acoustic parameters that were common for both sexes (for details on female alarm calls, see Fischer *et al.*, 2001a). There were nine females and seven males represented in this analysis. The acoustic structure of males' and females' alarm calls varied significantly ($F_{1,14}=376.6, P<0.05$). Figure 6 presents spectrograms of a male and a female alarm call. Males exhibited lower frequency characteristics and a longer call duration than females. A variety of parameters that influence perceived pitch varied significantly in relation to sex, for instance the difference between the first and second dominant frequency band, start and mean values of the first dominant frequency band, and the mean first quartile of the distribution of frequency amplitudes (Table IV).

Next, we investigated whether male and female alarm and contact calls varied in similar ways. We found some similarities in male and female calls with regard to the variables that contribute most to a discrimination of alarm and contact calls. For instance, in both sexes, the start of the first quartile of the distribution of frequency amplitudes revealed higher values for alarm calls than for contact calls (mean $\pm$s.d. for female alarm calls: 774 ± 92 Hz; female contact calls: 569 ± 70 Hz; Wilcoxon matched-pairs signed rank test, $n=16$, $z=-3.46, P<0.001$: for male alarm calls 544 ± 62 Hz; male contact calls 397 ± 12 Hz, for statistics see Table III). Also, in terms of the call duration, alarm calls

TABLE IV. Statistics ($\bar{x}\pm$s.d.) and results of univariate analyses of variance for those variables that vary in relation to sex.

Variable	males	Females	$F_{1,14}$	$P<$
dfb mean difference [Hz]	274 ± 14	357 ± 19	94.8	0.001
dfbl start [Hz]	350 ± 57	473 ± 41	25.8	0.001
dfal mean [Hz]	691 ± 41	824 ± 69	20.5	0.001
dfal start [Hz]	544 ± 63	754 ± 131	15.1	0.005
Duration [ms]	297 ± 77	193 ± 46	14.1	0.005

were shorter than contact calls for both females and males (female alarm calls 194±46 ms; contact calls 291±42 ms, $z=-3.51, P<0.001$; male alarm calls: 322±89 ms, contact calls 472±66 ms). Female alarm calls were significantly noisier than contact calls (female alarm calls 62±21%; contact calls 38±15%, $z=-2.94, P<0.05$; male alarm calls: 63±15%, contact calls 24±12%).

IV. DISCUSSION

A. Acoustic variation in relation to context

Despite their superficial acoustic similarity, alarm and contest wahoos given by adult male baboons exhibit a number of significant acoustic differences. Contest wahoos are given at a much higher rate, exhibit lower frequency characteristics, have a longer hoo duration, and a relatively louder hoo portion than alarm wahoos. Within the different contest contexts, there are only slight differences that do not reach statistical significance. We failed to identify significant differences between contest and contact wahoos in terms of their acoustical structure. However, due to the small sample size, this result should be treated with caution. Contact and contest wahoos clearly differ in terms of the rate at which they occur.

Wahoos also exhibit significant differences among individuals that potentially allow conspecific listeners to identify which male is calling. Similar individual differences have been documented for a variety of other baboon vocalizations, including female grunts (Owren *et al.*, 1997) and alarm barks (Fischer *et al.*, 2001a). Playback studies have shown that these differences are perceptually salient to conspecifics (Fischer *et al.*, 2000; Palombit *et al.*, 1997; Rendall *et al.*, 2000).

We were puzzled by the finding that alarm and contest wahoos represent variations of the same general call type despite the fact that they serve markedly different functions. The same is true for female baboon barks, where ''clear'' and ''harsh'' bark variants function as contact and alarm calls, respectively. Overall, baboons appear to produce only a limited number of different call types, the most common of which are tonal, low-frequency grunts (Owren *et al.*, 1997), noisy screams, and pant barks given in submissive interactions or in response to aversive stimuli such as snakes (personal observation). Additionally, females utter mating calls which consist of a drawn-out tonal component and a series of rhythmic pants, while infants may emit tonal ''begging'' calls.

Why would calls that serve different functions be so similar acoustically? Alarm and contest wahoos may differ only subtly in their acoustic features because they are manifestations of a similar underlying internal state. Predator encounters and male competitive interactions certainly seem to be highly charged contexts that are associated with high states of arousal, and these states may be sufficiently similar that they result in similar-sounding calls. If this affect-based hypothesis is correct, however, it must also account for the subtle but consistent acoustic difference between alarm and contest wahoos—presumably by suggesting that a caller's internal state in these two contexts, while broadly similar, is nonetheless subtly different as well.

Contest and alarm wahoos might also be similar because of limitations in vocal production. The acoustic structure of nonhuman primate calls is determined by oscillation of the vocal folds (and sometimes the vocal lip, see Mergell *et al.*, 1999 for details), articulatory gestures that influence the filtering characteristics of the vocal tract (Brown and Cannito, 1995; Hauser *et al.*, 1993; Hauser and Schön Ybarra, 1994), and respiration (Häusler, 2000). With the present data, we are unable to estimate the influence of the different components. Current evidence suggests that, in nonhuman primates, the anterior cingulate cortex serves to control the initiation of vocalizations, facilitating voluntary control over call emission and onset (Jürgens, 1995). However, the motor coordination of the vocalization appears to take place in the reticular formation of the lower brainstem (Jürgens, 2001), suggesting little voluntary control over the precise structure of the vocal pattern. Thus, contest and alarm wahoos could be linked as well to quite different internal states that converge on similar motor programs (Düsterhoft *et al.*, 2000). Due to methodological constraints, we are currently unable to resolve this question.

B. Acoustic variation in relation to body size

Our findings support the view that animals of bigger size are able to produce longer calls, as males' wa syllables are indeed significantly longer than females' barks. Likewise, we found support for the assumption that large differences in body size lead to differences in fundamental frequency (Darwin, 1872; Morton, 1977): in male wahoos, the fundamental frequency is significantly lower in adults than in younger males or in females. Because we were unable to obtain measurements of adult male body size, it was impossible to determine whether the minimum fundamental is correlated with body size among adults. Studies in adult humans (Fitch and Giedd, 1999) and red deer (McComb, 1991), however, suggest that such a clear correlation is not necessarily the case. Formant frequencies may also serve as reliable indicators of body size, at least across sex and age classes: we observed that formant dispersion decreases with size, but it remains unresolved whether this relation also holds within a given sex and age class. In sum, the loud calls of baboons provide multiple cues to gross size differences. Whether these cues are also reliable within adult males remains an empirical issue. Current theory predicts that the formant frequencies should accurately reflect body size among adult males, whereas fundamental frequency may be less reliable (Fitch and Hauser, in press).

In addition to conveying information about the signaler's body's size, wahoos may also function to advertise a male's endurance and stamina. It seems likely that the hoo syllable represents an exaggerated acoustic trait. This view is supported by the fact that this portion of the call emerges in adolescent males and becomes longer and louder with age, and disproportionately so in contest (and possibly contact) contexts. Further analyses on the relationship between hoo duration and indicators of male stamina within the group of adult males are clearly needed. Further support for the hypothesis that wahoos advertise male quality comes from the fact that, on average, contest wahoos are given at more than five times the rate of alarm wahoos.

C. Listeners' assessment of alarm and contest wahoos

It may seem puzzling at first that calls serving markedly different functions are acoustically so similar, even though they are broadcast over large distances where there may be a cost to ambiguity (Marler, 1965). Given that vocal production in nonhuman primates is heavily constrained (Jürgens, 1995), listeners seem to be under strong selective pressure to differentiate between calls that signal the presence of a predator and those that signal potential aggression. Although call rate varies clearly between alarm and contest wahoos, experiments currently underway suggest that listeners are able to discriminate between these two call types regardless of the rate at which they are presented (D. Kitchen, D. L. Cheney, and R. M. Seyfarth, unpublished data, 2001).

It seems probable that listeners learn to identify those acoustic features that are correlated with each call type, or individual. Infant baboons, for example, appear to learn the distinction between females' contact and alarm barks during development, and their responses to each of these call types becomes more distinctive over time (Fischer *et al.*, 2000). Similarly, through processes of association, and perhaps also by observing the responses of others, listeners attending to wahoos may learn when to escape into trees, when to adopt cryptic positions in order to avoid male aggression, and when it is safe simply to ignore a call.

ACKNOWLEDGMENTS

We thank the Office of the President and the Department of Wildlife and National Parks of the Republic of Botswana for permission to conduct research in the Moremi Wildlife Reserve. Mokupi Mokupi assisted with the data collection, and Markus Metz proved to be a wonderful field companion. Ryne Palombit and Dawn Kitchen kindly made some of their recordings available for analysis. We thank Gustav Peters and two anonymous reviewers for helpful comments on the manuscript. This research was supported by the Deutsche Forschungsgemeinschaft (Fi 707/2-1) and the Kommission zur Förderung des wissenschaftlichen Nachwuchses (J.F.), and by National Science Foundation Grant No. IBN 9514001, National Institutes of Health Grant No. HD-29483, the National Geographic Society, the Research Foundation and the Institute for Research in Cognitive Science of the University of Pennsylvania (D.L.C. and R.M.S.).

APPENDIX: ACOUSTIC PARAMETERS CALCULATED FOR THE "WA" SYLLABLE

Variables used for comparison between males and females are indicated by an asterisk.

Parameter	Description
noise [%]*	percent of time segments in which no harmonic structure could be detected
Dfa1 start [Hz]*	frequency at which the distribution of frequency amplitudes reaches the first quartile at beginning of the call
Dfa1 mean [Hz]*	frequency at which the distribution of frequency amplitudes reaches the first quartile, mean across time segments
Dfa2 mean [Hz]*	frequency at which the amplitude distribution reaches the second quartile, mean across time segments
Dfb1 start [Hz]*	first dominant frequency band, at beginning of the call
Dfb1 max [Hz]*	maximum value of dfb1 in all time segments
Dfb1 mean [Hz]*	mean dfb1 across all time segments
Dfb1 minimum [Hz]	minimum value of dfb1 across all time segments
Dfb1 max location	relative position of maximum in the call; ranges between 0 and 1
Db1 slope*	slope of the trend of the dfb (i.e., the linear regression determined by the least squares method)
dfb trend difference [Hz]	mean difference between regression function and actual value of the dfb
dfb3 %*	percentage of time segments with a third dominant frequency band
dfb max difference [Hz]	max difference between first and second dominant frequency band
dfb mean difference [Hz]*	mean difference between first and second dominant frequency band across all time segments
dfb number*	mean number of dominant frequency bands detected in call
dfb ratio*	amplitude ratio between dfb1 and dfb2 (dfb1/dfb2)
fls1 [Hz]*	mean frequency range of first formant-like structure
fls2 %*	percentage of time segments in which a second fls could be detected
range max [Hz]*	maximum difference between highest and lowest frequency
range mean [Hz]*	mean difference across all time segments
pf mean [Hz]	mean of the frequencies with the highest amplitude across all time segments
pf max [Hz]*	frequency of the maximum frequency of the peak frequency across time segments
pf amplitude max [Hz]	frequency of the maximum amplitude of the peak frequency across time segments
pf max location [Hz]	relative position of maximum in the call; ranges between 0 and 1
pf jump [Hz]*	max difference of the pf in two consecutive time segments
pf slope*	slope of the trend of the pf
pf trend difference [Hz]	mean difference between trend line and pf
pf max jump location	relative location of the max. difference in peak frequency between two adjacent time segments

Beeman, K. (**1996**). *SIGNAL Operation Manual* (Engineering Design, Belmont, MA).
Bortz, J. (**1993**). *Statistik für Sozialwissenschaftler* (Springer, Berlin).
Brown, C. H., and Cannito, M. P. (**1995**). ''Modes of vocal variation in Sykes monkeys (*Cercopithecus albogularis)* squeals,'' J. Comp. Psychol. **109**, 398–415.
Buskirk, W. H., Buskirk, R. E., and Hamilton, W. J. (**1974**). ''Troop-mobilizing behavior of adult male chacma baboons,'' Folia Primatol. (Basel) **22**, 9–18.
Cheney, D. L., and Seyfarth, R. M. (**1977**). ''Behavior of adult and immature male baboons during inter-group encounters,'' Nature (London) **269**, 404–406.
Cheney, D. L., and Seyfarth, R. M. (**1997**). ''Reconciliatory grunts by dominant female baboons influence victims' behavior,'' Anim. Behav. **54**, 409–418.
Cheney, D. L., Seyfarth, R. M., and Palombit, R. (**1996**). ''The function and mechanisms underlying baboon 'contact' barks,'' Anim. Behav. **52**, 507–518.
Darwin, C. (**1872**). *The Expression of the Emotions in Man and Animals* (Murray, London).
Düsterhoft, F., Häusler, U., and Jürgens, U. (**2000**). ''On the search for the vocal pattern generator. A single-unit recording study,'' NeuroReport **11**, 2031–2034.
Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton, The Hague).
Fischer, J., Cheney, D. L., and Seyfarth, R. M. (**2000**). ''Development of infant baboons' responses to graded bark variants,'' Proc. R. Soc. London, Ser. B **267**, 2317–2321.
Fischer, J., Hammerschmidt, K., Cheney, D. L., and Seyfarth, R. M. (**2001a**). ''Acoustic features of female chacma baboon barks,'' Ethology **107**, 33–54.
Fischer, J., Metz, M., Cheney, D. L., and Seyfarth, R. M. (**2001b**). ''Baboon responses to graded bark variants,'' Anim. Behav. **61**, 925–931.
Fitch, W. T., and Hauser, M. D. ''Unpacking ''honesty:'' Vertebrate vocal production and the evolution of acoustic signals,'' in *Acoustic Communication*, edited by A. M. Simmons, R. Fay, and A. N. Popper (Springer, New York) (in press).
Fitch, W. T. (**1997**). ''Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques,'' J. Acoust. Soc. Am. **102**, 1213–1222.
Fitch, W. T., and Giedd, J. (**1999**). ''Morphology and development of the human vocal tract. A study using magnetic resonance imaging,'' J. Acoust. Soc. Am. **106**, 1511–1522.
Gouzoules, H., and Gouzoules, S. (**1989**). ''Design features and developmental modification of pigtail macaque, *Macaca nemestrina*, agonistic screams,'' Anim. Behav. **37**, 383–401.
Hall, K. R. L., and DeVore, I. (**1965**). ''Baboon social behavior,'' in *Primate Behavior: Field Studies of Monkeys and Apes*, edited by I. DeVore (Holt, Rinehart and Winston, New York), pp. 53–110.
Hamilton, W. J., Buskirk, R. E., and Buskirk, W. H. (**1976**). ''Defense of space and resources by chacma (*Papio ursinus*) baboon troops in an African desert and swamp,'' Ecology **57**, 1264–1272.
Hammerschmidt, K. (**1990**). ''Individuelle Lautmuster bei Berberaffen (*Macaca sylvanus*): Ein Ansatz zum Verständnis ihrer vokalen Kommunikation,'' Dissertation, FU Berlin, Berlin.
Hauser, M. D., Evans, C. S., and Marler, P. (**1993**). ''The role of articulation in the production of rhesus monkey, *Macaca mulatta*, vocalizations,'' Anim. Behav. **45**, 423–433.
Hauser, M. D., and Schön Ybarra, M. (**1994**). ''The role of lip configuration in monkey vocalizations: Experiments using Xylocaine as a nerve block,'' Brain Lang **46**, 232–244.
Häusler, U. (**2000**). ''Vocalization-correlation respiratory movements in the squirrel monkey,'' J. Acoust. Soc. Am. **108**, 1443–1450.
Jürgens, U. (**1995**). ''Neuronal control of nonhuman and human primates,'' in *Current Topics in Primate Vocal Communication*, edited by E. Zimmermann, J. D. Newman, and U. Jürgens (Plenum, New York and London), pp. 199–207.
Jürgens, U. (**2001**). ''Neurobiology of primate vocal communication,'' Adv. Ethology **36**, 17.
Liebermann, Ph., and Blumstein, S. E. (**1988**). *Speech Physiology, Speech Perception, and Acoustic Phonetics* (Cambridge University Press, Cambridge).
Marler, P. (**1965**). ''Communication in monkeys and apes,'' in *Primate Behavior: Field Studies of Monkeys and Apes*, edited by I. DeVore (Holt, Rinehart and Winston, New York), pp. 544–584.
McComb, K. (**1991**). ''Female choice for high roaring rates in red deer, *Cervus elaphus*,'' Anim. Behav. **41**, 79–88.
Mergell, P., Fitch, W. T., and Herzel, H. (**1999**). ''Modeling the role of nonhuman vocal membranes in phonation,'' J. Acoust. Soc. Am. **105**, 2020–2028.
Morton, E. S. (**1977**). ''On the occurrence and significance of motivation-structural rules in some bird and mammal sounds,'' Am. Nat. **111**, 855–869.
Mundry, R. (**1999**). ''Testing related samples with missing values: A permutation approach,'' Anim. Behav. **58**, 1143–1153.
Owren, M. J., Seyfarth, R. M., and Cheney, D. L. (**1997**). ''The acoustic features of vowel-like grunt calls in chacma baboons (*Papio cyncephalus ursinus*): Implications for production processes and functions,'' J. Acoust. Soc. Am. **101**, 2951–2963.
Palombit, R., Cheney, D. L., Fischer, J., Johnson, S., Rendall, D., Seyfarth, R. M., and Silk, J. B. (**2000**). ''Male infanticide and defense of infants in Chacma baboons,'' in *Male Infanticide and its Implications*, edited by C. P. van Schaik and C. H. Janson (Cambridge University Press, Cambridge).
Palombit, R. A., Cheney, D. L., and Seyfarth, R. M. (**1999**). ''Male grunts as mediators of social interaction with females in wild chacma baboons (*Papio cynocephalus ursinus*),'' Behaviour **136**, 221–242.
Palombit, R. A., Seyfarth, R. M., and Cheney, D. L. (**1997**). ''The adaptive value of 'friendships' to female baboons: Experimental and observational evidence,'' Anim. Behav. **54**, 599–614.
Rendall, D., Cheney, D. L., and Seyfarth, R. M. (**2000**). ''Proximate factors mediating 'contact' calls in adult female baboons and their infants,'' J. Comp. Psychol. **114**, 36–46.
Rendall, D., Seyfarth, R. M., Cheney, D. L., and Owren, M. J. (**1999**). ''The meaning and function of grunt variants in baboons,'' Anim. Behav. **57**, 583–592.
Riede, T., and Fitch, W. T. (**1999**). ''Vocal tract length and acoustics of vocalization in the domestic dog *Canis familiaris,*'' J. Exp. Biol. **202**, 2859–2867.
Ross, K. (**1987**). *Okavango: Jewel of the Kalahari* (Macmillan, New York).
Saayman, G. S. (**1971**). ''Behavior of adult males in a troop of free-ranging chacma baboons,'' Folia Primatol (Basel) **15**, 36–57.
Schrader, L., and Hammerschmidt, K. (**1997**). ''Computer-aided analysis of acoustic parameters in animal vocalizations: A multiparametric approach,'' Bioacoustics **7**, 274–265.
Silk, J. B., Cheney, D. L., and Seyfarth, R. M. (**1996**). ''The form and function of post-conflict interactinos between female baboons,'' Anim. Behav. **52**, 259–268.
Silk, J. B., Seyfarth, R. M., and Cheney, D. L. (**1999**). ''The structure of social relationshps among female savanna baboons in Moremi Reserve, Botswana,'' Behaviour **136**, 679–703.
Smith, H. J., Newman, J. D., Hoffman, J. J., and Fetterly, K. (**1992**). ''Statistical discrimination among vocalizations of individual squirrel monkeys,'' Folia Primatol (Basel) **37**, 267–279.
Waser, P. M., and Brown, C. H. (**1984**). ''Is there a 'sound window' for primate communication?,'' Behav. Ecol. Sociobiol. **15**, 73–76.
Wilden, I., Herzel, H., Peters, G., and Tembrock, G. (**1998**). ''Subharmonics, biphonation, and deterministic chaos in mammal vocalizations,'' Bioacoustics **9**, 171–197.
Zar, J. H. (**1999**). *Biostatistical Analysis* (Prentice Hall, Upper Saddle River, NJ).

Two-tone suppression in the cricket, *Eunemobius carolinus* (Gryllidae, Nemobiinae)

Hamilton E. Farris[a] and Ronald R. Hoy
Department of Neurobiology and Behavior, Mudd Hall, Cornell University, Ithaca, New York 14853

(Received 12 March 2001; revised 19 November 2001; accepted 14 December 2001)

Sounds with frequencies >15 kHz elicit an acoustic startle response (ASR) in flying crickets (*Eunemobius carolinus*). Although frequencies <15 kHz do not elicit the ASR when presented alone, when presented with ultrasound (40 kHz), low-frequency stimuli suppress the ultrasound-induced startle. Thus, using methods similar to those in masking experiments, we used two-tone suppression to assay sensitivity to frequencies in the audio band. Startle suppression was tuned to frequencies near 5 kHz, the frequency range of male calling songs. Similar to equal loudness contours measured in humans, however, equal suppression contours were not parallel, as the equivalent rectangular bandwidth of suppression tuning changed with increases in ultrasound intensity. Temporal integration of suppressor stimuli was measured using nonsimultaneous presentations of 5-ms pulses of 6 and 40 kHz. We found that no suppression occurs when the suppressing tone is >2 ms after and >5 ms before the ultrasound stimulus, suggesting that stimulus overlap is a requirement for suppression. When considered together with our finding that the intensity of low-frequency stimuli required for suppression is greater than that produced by singing males, the overlap requirement suggests that two-tone suppression functions to limit the ASR to sounds containing only ultrasound and not to broadband sounds that span the audio and ultrasound range. © *2002 Acoustical Society of America.* [DOI: 10.1121/1.1451069]

PACS numbers: 43.80.Lb, 43.66.Gf [WA]

I. INTRODUCTION

Masking refers to the phenomenon in which the level required to detect one stimulus, the signal, is raised by the presence of another stimulus, the masker. Although the underlying mechanisms mediating the various forms of auditory masking may vary, three prominent hypotheses are [see Delgutte (1996) for review] (1) the ''line-busy'' mechanism, in which an excitatory masker causes a reduction in the signal-to-noise ratio in the neural code; (2) adaptive masking, in which adaptation caused by the excitatory masker reduces the excitatory response to the signal; or (3) the masker suppresses or inhibits the excitatory response elicited by the signal. The last has been studied at several different levels in auditory systems using two-tone presentations and includes examples in which one of the two tones may either reduce the excitation produced by the other tone [e.g., as demonstrated by the reduced firing rate or rate suppression in the primary auditory fibers (Kiang, 1965)] or actually inhibit the excitation produced by the first tone. Whether mediated by mechanical or neural mechanisms, this reduced response to the signal tone has been called either two-tone suppression or two-tone inhibition (Kiang, 1965; Sachs and Kiang, 1968; Arthur *et al.*, 1971; these two terms have been used synonymously whether neural inhibition is observed or not).

Two-tone suppression has been studied in a variety of taxa. For example, earlier studies of the frequency response characteristics of the auditory units in the mid and forebrains of microchiropteran bats revealed excitatory frequency bands flanked by inhibitory side-bands (Suga, 1973), presumably sharpening the frequency resolution of these units. Unlike the two-tone rate suppression observed in the auditory afferents of cats (Kiang, 1965), Suga observed that stimulation with frequencies in the suppression side-bands in bats evokes inhibitory or polarizing responses. This class of two-tone suppression is analogous to that demonstrated in the prothoracic nervous system of crickets. In field crickets (Gryllidae, Gryllinae), ultrasound sensitivity is mediated in the CNS by a bilateral pair of ascending interneurons called Int-1 (*sensu*, Moiseff and Hoy, 1983; cf. AN2, Wohlers and Huber, 1982). Whereas the presentation of ultrasound alone elicits excitation in Int-1, the addition of a low-frequency tone (e.g., 5 kHz) can suppress this excitation. Low-frequency suppression in Int-1 is mediated by postsynaptic inhibition; the presentation of low-frequency stimuli alone evokes IPSPs and is thus capable of reducing the excitation elicited by the ultrasound (Nolen and Hoy, 1986b, 1987). Suppression of responses of auditory units in the cricket brain by low-frequency stimuli presumably reflects this ascending activity in Int-1, however, as no IPSPs are evident when excitation elicited by high-frequency stimuli (e.g., 15 kHz) is reduced by the addition of a second tone between 1 and 11 kHz (Boyan, 1981).

From a behavioral perspective, whereas the presentation of songlike (low-frequency) stimuli to field crickets flying on a tether elicits positive phonotaxis, the presentation of batlike ultrasound elicits an acoustic startle response (ASR) comprised in part of the lateral extension of the metathoracic leg, contralateral to the stimulus [response latency: 30–60 ms (Nolen and Hoy, 1986a; Moiseff *et al.*, 1978; Wyttenbach *et al.*, 1996)]. This response presumably functions in the

[a]Present address: Sec. of Integrative Biology, Patterson Hall, University of Texas-Austin, Austin, TX 78712. Electronic mail: hfarris@mail.utexas.edu

avoidance of echolocating insectivorous bats; the directional steering effect of the leg swing causes the cricket to steer away from the ultrasound source (Nolen and Hoy, 1986a). In experimental paradigms similar to those examining single-tone masking, the simultaneous presentation of a low-frequency stimulus (e.g., 5 kHz) with ultrasound raises the level of ultrasound required to elicit the ASR (Nolen and Hoy, 1986b). Because the experimental suppression of Int-1 using hyperpolarizing current is sufficient in suppressing startle responses to ultrasound stimuli (Nolen and Hoy, 1984, 1987), the mechanism for this type of masking (i.e., two-tone suppression using a behavioral assay) is presumably the inhibitory effect of the low-frequency stimulus on Int-1.

Sounds with frequencies $\geqslant$15 kHz also elicit a startle response in flying *Eunemobius carolinus* (Gryllidae, Nemobiinae) (Farris and Hoy, 2000). In contrast to field crickets, the direct presentation of frequencies $<$15 kHz to tethered *E. carolinus* does not elicit a response. These frequencies, however, can suppress the ultrasound-induced startle. Thus, because a direct assay was unavailable, we use the ultrasound-induced startle response and its suppression as an indirect test of low-frequency sensitivity in *E. carolinus*. The experimental paradigms used in this article are modeled after single-tone masking paradigms (e.g., Vogten, 1974) and are used to determine the frequency and temporal sensitivity for sounds with frequencies $\leqslant$15 kHz. From an ethological perspective, these tests examined whether sounds that best suppressed the ASR were similar to those typical of the *E. carolinus* male calling song [e.g., ~6.7 ms pulses of 5.6–6.3 kHz at ~75 pulses/s, these parameters vary with temperature (Alexander, 1957; Walker, 1962)] and thus might reflect specialization for song detection. Some of the results reported here have been previously reported in abstract form (Farris and Hoy, 1997, 1998).

II. MATERIALS AND METHODS

A. General methods

Except where noted, general methods are the same as those in Farris and Hoy (2000).

1. Subject animals

The colony, started from individuals sound-trapped in Lafayette County, MS (see Farris *et al.*, 1997), was reared under a 14 L/10 D h, light schedule and fed ''cricket chow'' *ad libitem.*

2. Acoustic stimuli

Stimuli were generated using Tucker Davis Technologies (TDT) 16-bit, digital-to-analog converters and custom-written software (8 or 6 μs sample period). Stimuli were amplified using a Harman/Kardon HK6150 integrated amplifier. Ultrasound stimuli (i.e., startle stimuli) were broadcast from either a Radio Shack Super tweeter (cat. no. 40-1310b) or through Panasonic 40 kHz transducers (40 kHz stimuli only). Low-frequency stimuli (i.e., suppressor stimuli, 2–15 kHz) were broadcast from either an ESS AMT-1 tweeter or Motorola piezoelectric speakers. All speakers were located 30 cm from the preparation. Stimulus amplitude was adjusted using TDT PA4 programmable attenuators (1 dB minimum step size). The stimuli were calibrated at the position of the test animal using online comparisons of the rms voltage of the stimuli to that generated by a B&K 4220 pistonphone calibrator (125 ms duration samples). The calibration system included a B&K 4135 $\frac{1}{4}$-in. microphone (0 degree angle of incidence), B&K 2639 preamp, and a B&K 5935 microphone power supply. All sound pressure levels (dB SPL) are referenced to 20 μPa. For all speakers, the maximum output level of the system was 107 dB SPL. All pulse onset and offset ramps are raised cosine. Total harmonic distortion of the system was determined to be $<$1% ($-$40 dB) using a Hewlett-Packard 3562A signal analyzer.

3. Startle response and suppression

Long-winged *E. carolinus* were tethered dorsally at the pronotum to a 14-cm-long piece of piano wire using low-temperature melting point wax. Flight can easily be initiated by waving the tethered cricket in the air or by giving it small puffs of air. Once flying, a tethered cricket was placed in a Faraday cage lined with acoustic foam to minimize echoes. The tethered cricket was positioned 7 cm from the floor of the cage and 30 cm from the speakers placed at 90 degrees normal to the cricket. The behavioral components of the acoustic startle response consist of an abrupt cessation of flapping, folding of the hind wings, closure of the tegmina, anterior extension of the prothoracic and mesothoracic legs, posterior extension of the metathorcic legs, and dorsal movement of the antennae (Farris and Hoy, 2000). Although flight usually resumed with the termination of the startle stimulus, reinitiating flight in some subjects became more difficult with increasing numbers of startle responses. To help ensure that most subjects would complete an experiment, we chose to measure thresholds using the adaptive procedures below.

Startle thresholds to ultrasound alone were measured using a 1/2-down, 0/2-up adaptive procedure (see Farris and Hoy, 2000). In this procedure, the amplitude of a single pulse was decreased in 6-dB steps if a startle response was observed in 1 out of 2 stimulus presentations. Stimulus step sizes were then changed to 3 and 1 dB for each reversal until a threshold was determined. This adaptive procedure converges on the stimulus level that elicits a response in 30% of the presentations (Levitt, 1971). Thresholds for two-tone suppression, however, were measured using the 2/2-suppression-down, 1/2-up adaptive procedure diagrammed in Fig. 1. In this procedure, the amplitude of the startle stimulus (ultrasound) was fixed at a level which consistently elicited a startle response (these levels are noted for each experiment; at least 5 dB above startle threshold), whereas the amplitude of the suppressor stimulus was adjusted until the minimum level for suppression was determined. As before, attenuator step sizes were 6, 3, and 1 dB for the sequence of reversals. Two procedures were used to control for any variance in response threshold over time [e.g., habituation or fatigue causing a change in threshold during some portion of the adaptive procedure, see Robert (1989)]. First, following two consecutive presentations in which a startle response was not elicited by the ultrasound+suppressor combination (i.e., 2/2 suppression), the ultrasound was presented alone to verify

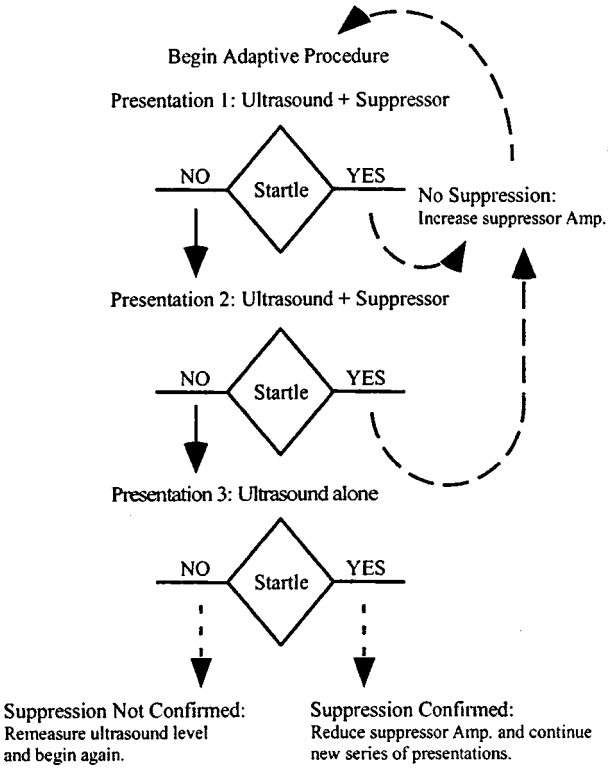


FIG. 1. Diagram of a single trial for the adaptive procedure used to determine suppression thresholds (see methods).

that the subject would still respond to the startle stimulus. If the ultrasound stimulus elicited a startle response, startle suppression for the previous two presentations of two-tone stimuli was thus confirmed, and the amplitude of the suppressor was decreased. If a startle response was not elicited by the ultrasound alone (i.e., suppression was not confirmed), then all tests with that individual were repeated and the startle threshold to ultrasound alone was remeasured to establish a stable reference (see Fig. 1). As part of a second control, the adaptive procedure stepped through the presentations of experimental stimuli in a pseudo-random order so that the same suppressor frequency was not presented in more than two consecutive trials (i.e., two sequences shown in Fig. 1). If a change in startle threshold had occurred during testing (see Sec. III), this control prevented its correlation to any suppressor frequency in particular. This adaptive procedure converges on the stimulus level that elicits suppression in 70.7% of the presentations (Levitt, 1971). The subjects were required to fly for at least 5 s prior to stimulus presentation (i.e., minimum interstimulus interval was 5 s). This interval was also chosen to reduce the probability of any habituation or sensitization. Ambient temperature for all of the experiments was 20–23 °C.

4. Calling song recordings

Singing males ($N=5$) were recorded either in the field (Ithaca, NY) or in the laboratory using a Sony WM D6C Walkman Professional tape recorder (no filter), Memorex CD2 tape, and a Sony ECM 939CT microphone placed at 15 cm dorsal to the singing cricket. Temperature at the position of the male was measured using a Radioshack digital thermometer. Songs were digitized (22.7 or 40 μs sample period) using a TDT AD1 and custom-written software. The mean sound pressure level of the calling songs over all of the recorded males was calculated by comparing the rms value of 15 randomly sampled song pulses (3 pulses from the song of each male) to that generated by a B&K 4220 pistonphone calibrator.

B. Experiments

1. Frequency response of two-tone suppression

Thresholds for startle suppression were measured as a function of the frequency of the suppressor stimuli. Presented simultaneously with the startle stimulus (1 pulse of 40 kHz), the suppressor stimuli had the same temporal structure as the ultrasound (5 ms duration, 1 ms ramps) but varied in frequency from 2 to 15 kHz. For each individual ($N=11$), startle threshold was first determined using the adaptive procedure described above. The startle stimulus was then fixed at 8–10 dB above startle threshold and presented simultaneously with each of the ten suppressor frequencies until the minimum suppressor level required to suppress the startle response was determined. Each individual was tested at all ten suppressor frequencies (2–10, 15 kHz).

2. Equal-suppression contours

In experiments similar to those measuring equal-loudness contours (see Plack and Carlyon, 1995), the relationship between ultrasound intensity and suppressor intensity was tested for seven different suppressor frequencies (3–9 kHz in 1-kHz steps). After determining the startle threshold (at 40 kHz) for each individual, the minimum threshold for startle suppression was measured for four ultrasound intensities above startle threshold (5, 8, 10, 12 dB *re*: startle threshold) at seven suppressor frequencies. Based on evidence from preliminary observations, this 7-dB intensity range for the ultrasound stimulus was chosen to encompass the minimum level (+5 dB *re*: startle threshold) required to consistently elicit a startle for the duration of the experiment and the maximum ultrasound intensity (+12 dB *re*: startle threshold) at which suppression was still possible. In addition to the pseudo-random order of frequency presentation mentioned above, the ultrasound level within frequency presentations was randomly chosen (i.e., 5, 8, 10, or 12 dB *re*: startle threshold). For an individual to be included in the data set, suppression thresholds had to be completed at all four ultrasound levels for at least one suppressor frequency (e.g., startle suppression thresholds at one frequency were determined at 5, 8, 10, and 12 dB above startle threshold). Unlike the frequency response for suppression measured in experiment 1 (above), no individuals flew long enough to measure all four equal suppression contours for all seven frequencies (i.e., 28 suppression thresholds).

Suppression growth (i.e., change in level of suppressor versus change in level at 40 kHz) at each suppressor frequency was calculated using a linear regression of suppression thresholds versus the level of the startle stimulus. Comparison of suppressor growth across the seven suppressor frequencies was accomplished using an analysis of covari-

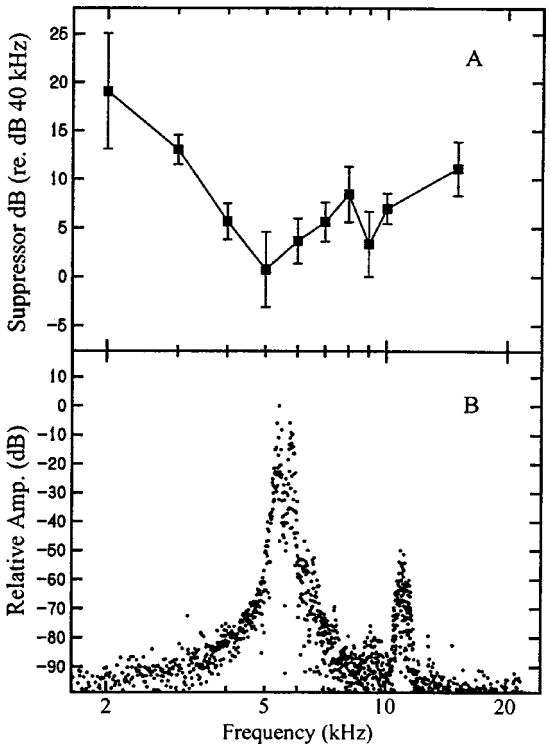


FIG. 2. Suppression tuning. (a) Symbols represent the mean (±SE) threshold required to suppress the ultrasound-induced startle response for ten different suppressor frequencies. Stimuli were 5-ms pulses with 1-ms ramps. The startle stimulus (40 kHz) was set at 8–10 dB above startle threshold for each individual. Individuals ($N=11$) were tested at all frequencies. (b) Amplitude spectrum of an *E. carolinus* calling song recorded at 20 °C.

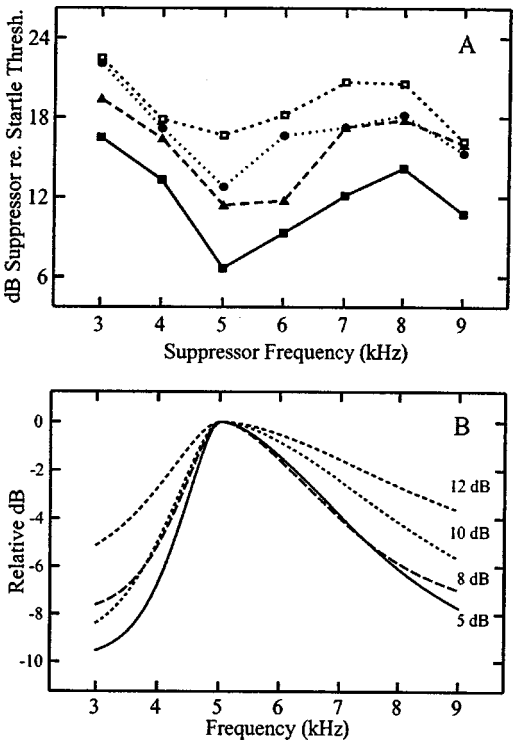


FIG. 3. Equal suppression contours. (a) Threshold to suppress a startle response as a function of frequency and startle stimulus level. Symbols: ■, ▲, ●, □ are the suppression thresholds for startle stimuli at 5, 8, 10, and 12 dB above startle threshold, respectively. Because the level of 40 kHz varied for these measurements, unlike in Fig. 2(a) suppression thresholds are plotted *re*: startle threshold rather than *re*: 40 kHz level. (b) Roex filter functions for each equal suppression contour above (see Table II for the parameters for each filter). Line types are the same in both panels.

ance testing for significant differences between slopes (Zar, 1984, p. 300). Comparison of the suppression slopes and intercepts between males and females at 6 kHz was done using the procedure described in Zar (1984, p. 296).

Shapes of the suppression contours were calculated from the mean suppression thresholds using a method described by Patterson *et al.* (1982; but see also Moore *et al.*, 1990). Briefly, each flank of the tuning curve (i.e., on either side of the center frequency) was modeled using the rounded exponential function roex(p,r):

$$W(g)=(1-r)(1+pg)\exp(-pg)+r,$$

where g is the normalized deviation of frequency from the center frequency, fc, $g=|(f-fc)|/fc$, p is a dimensionless parameter determining the slope of the flanks of the tuning curve, and r is a constant that sets the range limitation of the roex filter. The tuning curves were assumed to be asymmetrical around the center frequency, and separate slopes (p) were calculated for the upper and lower frequency flanks; r was assumed to be the same for each side, however, and equaled the difference between the minimum and maximum thresholds on the tuning curve. After normalizing each equal suppression contour to the threshold at the center frequency (5 kHz), we used a least-squares method to solve for the slope parameter (p) of each flank of the filter. Using the slope parameters, the equivalent rectangular bandwidth (ERB) for each tuning curve was calculated (Patterson *et al.*, 1982), facilitating a comparison of tuning ''sharpness.''

3. Temporal sensitivity: Suppressor duration

Thresholds for startle suppression were measured as a function of suppressor duration. Presented with simultaneous onsets, the startle and suppressor stimuli consisted of single pulses of 40 and 6 kHz, respectively. In part 1 the startle stimulus had a 5-ms duration and suppressor duration varied from 2 to 80 ms, whereas in part 2, the startle stimulus had a 20-ms duration and suppressor duration varied from 2 to 200 ms. The amplitude of the ultrasound stimulus was set 8–10 dB above startle threshold.

4. Temporal sensitivity: Nonsimultaneous suppression

Similar to forward and backward masking, startle suppression was measured while varying the relative onset times of the suppressor (5-ms duration, 1-ms ramps, 6 kHz) and startle (5-ms duration, 1-ms ramps, 40 kHz) stimuli from 3 to −6 ms (*re*: startle onset). The amplitude of the ultrasound stimulus was set 8–10 dB above startle threshold, whereas the amplitude of the suppressor pulse was adjusted until the minimum suppressor level required to suppress the startle was determined for each of eight different relative onset times.

III. RESULTS

A. Frequency response of two-tone suppression

Suppression was tuned to frequencies near 5 kHz, overlapping components in the male calling song (Fig. 2). Whereas a gradual increase in suppression thresholds was measured for frequencies >5 kHz (although a secondary sensitivity peak can be seen at 9 kHz), for frequencies <5 kHz there was a steeper decline in suppression sensitivity [Fig. 2(a)]. Thresholds were measured for at least 10 of the 11 individuals tested for all frequencies except 2 kHz, in which suppression was effective in only three individuals.

B. Equal suppression contours

As in experiment 1, suppression thresholds were lowest for frequencies near 5 kHz with a second sensitivity peak at 9 kHz [Fig. 3(a)]. The equal suppression contours were not parallel, however, as suppressor growth (change in level of suppressor versus change in level at 40 kHz) differed significantly across the seven suppressor frequencies (ANCOVA, $N=236$, $F=3.99$, $P<0.002$; Zar, 1984, p. 300). Because suppressor growth was >1 for 5–7 kHz and <1 for 3–4 and 8–9 kHz (Table I), the equal suppression curves appear to flatten at the higher ultrasound levels. Furthermore, significant positive slopes for suppressor growth (i.e., at five of seven frequencies, Table I) suggests that each individual's startle threshold was stable throughout the procedure, as no relationship would be expected if startle threshold varied.

Although analysis of the equal suppression contours using the roex model did suggest a decrease in the slope parameter for frequencies >5 kHz (and increase in ERB) with increasing ultrasound levels, the model's predictions were not significantly correlated with the data for the highest ultrasound level (i.e., 12 dB *re*: startle threshold) [Fig. 3(b), Table II]. The failure of the roex model for the 12-dB contour is presumably due to the suppression thresholds at 9 kHz. Although the model assumes a center frequency of 5 kHz, at the 12-dB contour, lowest suppression thresholds shifted from 5 to 9 kHz, reducing the amount of variance explained by the model. Indeed, when the thresholds at 9 kHz were removed from the contours, the fit of the roex model was significant at each contour, revealing an increase in ERB with increasing ultrasound levels (Table III).

TABLE I. Relationship between suppression threshold and ultrasound level for each suppressor frequency. Although threshold means are shown in Fig. 3, slopes are calculated from the entire data set using linear regression analysis. Suppressor growth varies significantly across suppressor frequency (ANCOVA, $N=236$, $F=3.99$, $P<0.002$; Zar, 1984, p. 300). Significant positive slopes for five of seven frequencies suggests that reference thresholds (i.e., startle threshold) were stable throughout the experiments.

Suppressor frequency (kHz)	N	Suppressor growth (Δ Suppressor dB/Δ 40 kHz dB)	R^2	P value
3	36	0.903	0.312	0.0004
4	36	0.648	0.070	0.1189
5	32	1.370	0.158	0.0243
6	36	1.354	0.210	0.0049
7	28	1.140	0.192	0.0195
8	32	0.859	0.129	0.0437
9	36	0.728	0.045	0.2145

TABLE II. Filter dimensions and correlations for the least-squares fit of roex(p,r) to the equal suppression contours. Columns are the roex slope parameters (p) for frequencies above and below the center frequency (5 kHz), the equivalent rectangular bandwidth (ERB) of the filter, Q_{3dB} of the filter, and the R^2 and P values of the least-squares fit to the threshold means for the four different startle stimulus amplitudes (*re*: startle threshold).

40 kHz level (dB)	Upper slope p	Lower slope p	ERB (kHz)	Q_{3dB}	R^2	P value
5	4.437	14.802	2.93	2.33	0.737	0.013
8	5.181	13.275	2.683	2.284	0.799	0.007
10	3.245	11.367	3.959	1.676	0.570	0.05
12	2.876	8.746	4.621	1.136	0.218	0.291

C. Temporal sensitivity: Suppressor duration

As suppressor duration increased, suppression threshold did not decrease like that of a simple exponential integrator (Fig. 4). Temporal integration of the low-frequency suppressor was analyzed using a least-squares-fit of the following equation proposed by Plomp and Bouman (1959) for the change in threshold as a function of stimulus duration:

$$\text{Threshold shift } (T) = -10 * \log\left(1 - \exp\left(\frac{-T}{\tau}\right)\right),$$

where τ represents the temporal integration (i.e., summation) time constant that describes the rate at which the threshold reaches an asymptote as a function of T, the duration of the suppressor stimulus. (Note that for the 20-ms ultrasound treatment, because no suppression was measured for suppressor durations <10 ms, these durations were not included in the model.) Correlations between the data and the model were not significant for either the 5 ms ($\tau=12.2$ ms, $r^2=0.25$, $P=0.18$) or 20 ms ($\tau=17.3$ ms, $r^2=0.51$, $P=0.13$) ultrasound treatments and, thus, estimates for the temporal integration time of the suppressor (6 kHz) using a duration versus intensity paradigm were inconclusive. In addition, there was no evidence for tuning to any particular suppressor duration (i.e., similar to pulse durations of the male calling song). For both ultrasound durations (5 and 20 ms), suppression thresholds decreased up to the duration of the startle

TABLE III. Filter dimensions and correlations for the least-squares fit of roex(p,r) to the equal suppression contours not including thresholds at 9 kHz. Columns are the roex slope parameters (p) for frequencies above and below the center frequency (5 kHz), the equivalent rectangular bandwidth (ERB) of the filter, Q_{3dB} of the filter, and the R^2 and P values of the least-squares fit to the threshold means for the four different startle stimulus amplitudes (*re*: startle threshold). Although all upper slopes are affected by the removal of the data at 9 kHz, the lower slope at 12 dB is also changed due to a change in the dynamic range (roex parameter r) of the entire 12-dB filter (see Table II).

40 kHz level (dB)	Upper slope p	Lower slope p	ERB (kHz)	Q_{3dB}	R^2	P value
5	5.963	14.802	2.353	2.902	0.995	0.0001
8	6.068	13.275	2.401	2.552	0.936	0.0016
10	5.066	11.367	2.853	2.325	0.857	0.0081
12	5.706	8.278	2.961	1.678	0.839	0.01

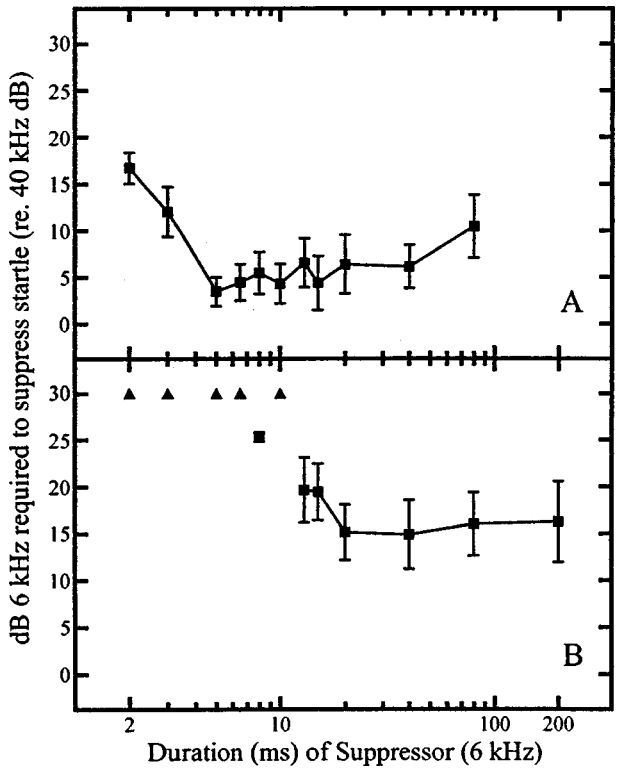


FIG. 4. Suppression threshold (re: startle stimulus dB) as a function of suppressor duration. Suppressor and startle stimuli have simultaneous onsets. Durations of the startle stimulus (40 kHz) were 5 and 20 ms for (a) and (b), respectively. Triangles mark suppressor durations in which <2 individuals exhibited suppression.

stimulus, but varied for greater suppressor durations. Because suppression threshold depended on the duration of the startle stimulus, these results are inconsistent with a hypothetical filter that band-passes durations near 7 ms (i.e., the duration of a typical calling song pulse).

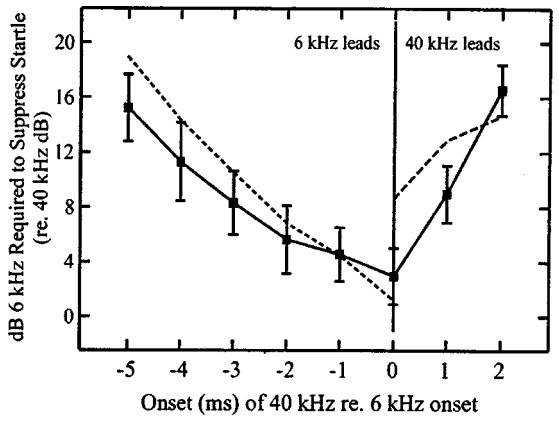


FIG. 5. Suppression threshold (re: startle stimulus dB) as a function of the relative onset times of the suppressor (6 kHz) and startle (40 kHz) stimuli. Both stimuli were 5-ms duration pulses with 1-ms ramps. No suppression occurs when the onset of the suppressing tone is >2 ms after and >5 ms before that the ultrasound stimulus. Dashed curves represent the thresholds predicted by the simple summation model (see results, Fig. 6) when the integration time is 9.395 ms ($r^2=0.987$, $P=0.000\,06$) in the forward suppression condition (6 kHz leads) and 47.07 ms ($r^2=0.914$, $P=0.19$) in the backward suppression condition (40 kHz leads).

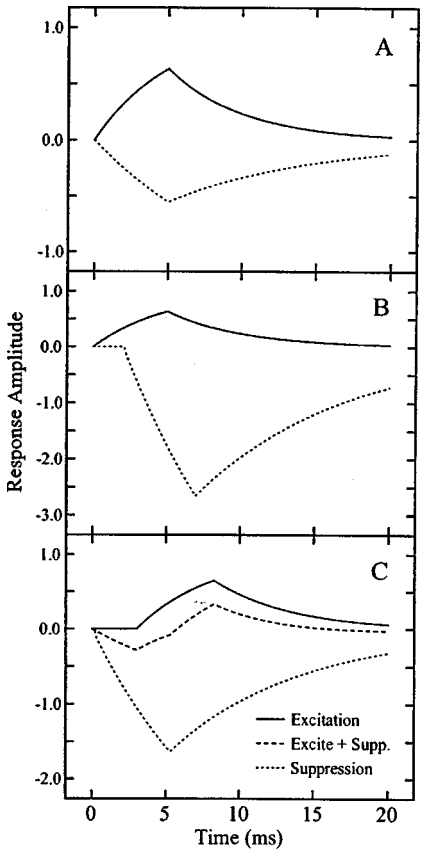


FIG. 6. Diagram of the suppressor and startle stimuli in the nonsimultaneous suppression paradigm when modeled as the interaction of an IPSP (dotted curve) and EPSP (solid curve), respectively. In each condition, the amplitude of the IPSP is multiplied by that (re: EPSP amplitude) required to suppress the startle (Fig. 5). The model assumed that suppression occurred when the sum of the two integrals was −8 dB re: EPSP peak amplitude: (a) simultaneous onset, (b) backward suppression (suppressor trails by 2 ms), and (c) forward suppression (suppressor leads by 3 ms). In (c) the dashed curve represents the sum of the excitation and suppression responses. In this case, because the peak of the summed response is −8 dB (re: the excitatory response), the model deemed that suppression had occurred.

D. Temporal sensitivity: Nonsimultaneous suppression

No suppression could be measured when the onset of the suppressing tone was >2 ms after or >5 ms before that of the ultrasound stimulus (Fig. 5). In both backward and forward suppression, increasing levels of 6 kHz were required to suppress the startle response for greater temporal separations between the two stimuli. Because these data presumably reflect the relative temporal integration times for the two frequencies, we constructed a simple summation model to estimate the integration time of the low-frequency suppressor, which we were unable to measure sufficiently in the previous experiment.

Both the excitatory (40 kHz) and suppression (6 kHz) integrals were modeled as simple exponential integrators that change as

$$1-\exp(-T/\tau)$$

when the stimulus is on, and decay as

$$\exp(-T/\tau)$$

when the stimulus is turned off (Fig. 6). The suppression integral is negative, however, and its amplitude is scaled like that measured at suppression threshold for each of the different relative onset times (e.g., when 40 kHz leads 6 kHz by 2 ms, the amplitude of the 6-kHz pulse is 16.2 dB *re*: 40 kHz at suppression threshold; see Fig. 5) [Fig. 6(b)].

We assumed that the addition of the two integrals was linear [Fig. 6(c)] and that there was no latency difference between the two responses. We also assumed that $\tau = 5.30$ ms for the excitatory integral (Farris and Hoy, 2000) and that suppression occurred when the amplitude of the excitatory integral was reduced by 8 dB (i.e., 40 kHz startle stimulus was set at 8–10 dB above startle threshold) as a result of its addition to the suppression integral. Using a least-squares method, we solved for the integration time (τ) of the suppression stimulus (6 kHz) that produced an 8-dB reduction in excitatory amplitude for each of the relative onset times shown in Fig. 5. Although separate time constants were calculated for both the forward and backward suppression conditions, data for the simultaneous presentation of the two stimuli (0 ms relative onset time) were included in both analyses.

The time constant (τ) for the integration of 6 kHz differed for forward and backward suppression (Fig. 5). In forward suppression (e.g., 6 kHz leads 40 kHz), suppression thresholds were best modeled when the 6-kHz stimuli were integrated over 9.395 ms ($r^2 = 0.987$, $P = 0.000\,06$). In backward suppression, however, integration appeared to occur over a much longer time, as suppression thresholds were best modeled when the integration of 6 kHz lasted 47.07 ms ($r^2 = 0.914$, $P = 0.19$) (Fig. 5). The model for backward suppression was not significantly correlated to the data, however. Potential causes for this difference between the two measurements are addressed below.

IV. DISCUSSION

A. Frequency sensitivity

Because the behavioral context and stimulus structures required to elicit responses to calling songlike stimuli vary (e.g., Loher *et al.*, 1993), it is not clear why tethered *E. carolinus* do not respond to such stimuli when presented alone. Although this lack of a response prevents any direct measure of low-frequency (i.e., <20 kHz) sensitivity (sound trapping notwithstanding, Farris *et al.*, 1997), its measurement is still facilitated by low-frequency suppression of the ultrasound-induced startle. Thus, in experimental paradigms similar to those examining single-tone masking in humans (Vogten, 1974), we used the increase in startle threshold produced by low-frequency sounds to indirectly measure aspects of sensitivity below 20 kHz. Similar to two-tone suppression in field crickets, suppression sensitivity was best at 5 kHz with a secondary peak at 9 kHz (Moiseff and Hoy, 1983). Sensitivity to frequencies across this band overlaps much of the spectrum of male calling songs (Fig. 2) and thus could function in the context of song detection.

Tuning sharpness or the frequency biases of auditory systems are known to change with the amplitude of the stimuli, however. For example, in single-tone masking experiments with humans, the frequency that produces the most masking varies with the intensity of the probe (Vogten, 1974). Furthermore, masking experiments using various broadband maskers show that the shapes of critical bands centered at the same frequency vary with the amount of input or stimulus level (Shailer *et al.*, 1990; Moore *et al.*, 1990). Generally, this change is measured as an increase in critical band ERB as stimulus intensity increases. With respect to tuning across multiple critical bands, a well-known example of such a change is the ''flattening'' of equal loudness contours in humans as a function of the amplitude of the standard [usually 1 kHz; see Plack and Carlyon (1995) for review]. From a physiological perspective, Capranica (1992) pointed out that the frequency tuning of units in the VIIIth nerve of frogs is not maintained at stimulus levels above threshold. Indeed, isointensity measures of spike rate at various frequencies are not necessarily parallel and can reveal either an increased or decreased frequency bias in the unit (Capranica, 1992; Rose *et al.*, 1971). Imaizumi and Pollack (1999) measured the analogous response in the auditory afferents of crickets (*Teleogryllus oceanicus*) and demonstrated that for a 15-dB change in stimulus intensity, the bandwidth of the isointensity response curves for one afferent changed from ~2 to ~12 kHz. With respect to physiological measures of two-tone suppression, the tuning of low-frequency (~4 kHz) suppression of ultrasound-induced excitation in an ascending unit in *T. oceanicus* broadened with increases in suppressor intensity (Hutchings and Lewis, 1984).

Similar to the isointensity measures above, two-tone suppression tuning in *E. carolinus* varied with the sensation level of the ultrasound stimulus. Because the slopes for suppression growth (change in level of suppressor versus change in level at 40 kHz) were not the same for the seven frequencies (Table I), the suppression contours were not parallel. If these equal suppression contours do reflect the frequency response of the calling song channel in *E. carolinus*, the results suggest that comparisons of songs at various frequencies may vary with intensity. Furthermore, the data suggest that there may be greater dynamic range for frequencies near 5 kHz than those near 3 and 9 kHz (i.e., the intensity range between threshold and apparent saturation appears smaller for 3 kHz than for 5 kHz).

The rounded exponential (roex) filter with its rounded peak and exponentially decaying skirts is one of the most widely used methods for describing the shape of auditory filters (see Moore, 1995). Because it describes filter selectivity, we find it surprising that it has been so rarely employed in studies of comparative audition. Indeed, use of the roex model would facilitate the quantitative comparison of tuning curves within and between studies. We used the roex model to compare the shape of suppression tuning at four different intensities of the startle stimulus. When including the data at 9 kHz, the model explained a significant proportion of the variance in the threshold means for the 5-, 8-, and 10-dB suppression curves (Table II; Fig. 3). There was, however, too little selectivity in suppression tuning for the 12-dB startle stimulus to adequately fit the model. For the lower three contours, however, the change in both the upper slope parameter and the ERBs do suggest a decrease in filter tun-

ing (Table II; Fig. 3). Comparison of the 5- and 10-dB parameters suggests a decrease and an increase in the upper slope and ERB, respectively. Although 9 kHz is clearly part of the suppression channel, we explored the filter shape around 5 kHz after removing the 9-kHz thresholds from the data set. Without 9 kHz, the model can be significantly fit to all of the curves (Table III) and reveals a ~600 Hz increase in ERB over the 7-dB change in startle stimulus.

Although our results do suggest that the relative weighting of frequencies by *E. carolinus* varies at intensities above threshold, certain concerns do exist in analyzing the equal suppression contours as exact analogs of equal loudness (or masking) contours. First, because each individual is not represented in every point on the curves (individuals were included only if their thresholds were measured at the four levels for at least one frequency), there is a potential that one individual could bias the results at one frequency. This concern is mitigated, however, by the similarity between the suppression tuning curve in Fig. 2 (in which each individual was tested at all ten frequencies) and the 8- and 10-dB equal suppression contours in Fig. 3. In addition to the tuning being centered at 5 kHz, both experiments revealed a secondary sensitivity peak at 9 kHz. This suggests that the frequency response of the equal suppression contours was not affected by limited replication across frequencies. A second concern arises from the variance around each suppression contour. The mean standard error over all thresholds on the equal suppression contours (Fig. 3) is 2.3 dB (this variance was not presented in Fig. 3 for clarity). Although the variances for some thresholds do overlap, a statistically significant correlation between suppressor level (dB) and startle level (i.e., when regressed against the entire data set to include all of the variance) can still be resolved for all suppressor frequencies except 4 and 9 kHz (Table I).

B. Temporal integration and suppression

Using a duration versus intensity experimental paradigm, Farris and Hoy (2000) found that power integration of single pulses of 40 kHz by *E. carolinus* occurred over ~32 ms. Simple duration versus intensity experiments in which the duration of the suppressor stimulus was varied were insufficient, however, in measuring the integration time of low-frequency stimuli. In trials using a 5-ms ultrasound stimulus, suppression thresholds decreased as the duration of the suppressor increased to 5 ms, but became less effective for longer durations (Fig. 4). From an ethological perspective, these results are interesting as they suggest that there might be some ''duration tuning'' to 6-kHz stimuli. At temperatures near 23 °C, male calling songs consist of 6.7-ms pulses of 6 kHz at 75 pulses/s. Best suppression to 5-ms duration pulses of 6 kHz could thus reflect the response of a temporal bandpass filter applied to frequencies characteristic of male calling songs. Although evidence for such a filter has been found for certain amplitude modulation rates (Schildberger, 1984), there is no prior evidence for a filter that bandpasses certain pulse durations in crickets. In subsequent experiments using a 20-ms ultrasound stimulus, however, we found that suppression thresholds decreased as suppressor duration increased up to 20 ms, with thresholds becoming more variable for longer durations (Fig. 4). Thus, these data suggest that best suppression durations appear to be the result of stimulus overlap rather than the result of a bandpass duration filter. In other words, even though integration of the ultrasound stimulus is known to continue for up to ~32 ms [i.e., 27 ms beyond the end of the 5-ms pulse of 40 kHz (Farris and Hoy, 2000)], we found little evidence for any backward suppression effects produced by increasing the duration of the suppressor stimuli into this time window trailing the startle stimulus. Thus, in order to examine this effect of overlap more closely, we used a nonsimultaneous suppression paradigm to measure the temporal window in which suppression of the ultrasound-induced startle could be elicited.

With the exception of the forward suppression condition in which the startle stimulus started at the offset of the suppressor, stimulus overlap was required for startle suppression. For the 5-ms duration stimuli, no suppression could be measured when the onset of the suppressor was >5 ms before or >2 ms after the ultrasound pulse (Fig. 5). As in human psychoacoustics, nonsimultaneous masking experiments effectively probe the limits of auditory inertia or the time over which the summation of masker and signal occurs. Figure 5 can be thought of as showing the shape of such a temporal window centered at the onset of the 6-kHz pulse. A variety of models have been proposed for describing the characteristics of temporal integration including the temporal analog of the roex filter used above for spectral analysis (Moore *et al.*, 1988). Solutions to this model in humans maximized the signal-to-noise ratio for a signal temporally centered between forward and backward maskers. We chose a different method for analyzing this temporal window, however.

Nolen and Hoy (1986b) found that two-tone suppression in *T. oceanicus* was mediated by low-frequency-induced postsynaptic inhibition in a pair of bilaterally symmetric auditory interneurons called Int-1. Summation of IPSPs effectively reduced excitation below that required to elicit a startle. Central auditory processing of ultrasound in *E. carolinus* appears similar to that in *T. oceanicus*, as extracellular recordings in the neck connectives of *E. carolinus* reveal at least one identifiable auditory unit that responds like Int-1 in *T. oceanicus* (Farris and Hoy, 2000). Thus, the interaction of the suppressor and startle stimulus in the nonsimultaneous paradigm was modeled as the hypothetical summation of inhibitory and excitatory post-synaptic potentials (Fig. 6). Because the time over which nonsimultaneous suppression could be measured was an order of magnitude less than that for temporal integration of a single pulse of ultrasound (32 ms), the model solved for the inhibitory time constant using an excitatory time constant of 5.3 ms, the time constant for paired-pulse integration of ultrasound (Farris and Hoy, 2000). For forward suppression, the data were best modeled when the inhibitory time constant equaled 9.17 ms. In the backward suppression condition the model failed to significantly explain the change in suppression threshold. If the model does represent the underlying neural mechanism of suppression, then at least one hypothesis for the model's failure when the suppressor follows the startle stimulus is related to the production of action potentials. The model only

takes into account graded potentials. If the onset of the suppressor stimulus follows or overlaps with the production of action potentials, the inhibitory effect of the suppressor will be quite different from interactions with only EPSPs.

C. Two-tone suppression and masking in other tympanate orthoptera

Two-tone suppression of the ultrasound-induced startle response is a phenomenon common to both ensiferan and caeliferan orthopterans. For the locust *Locusta migratoria*, Robert (1989) showed that the negative steering behaviors elicited by pulses of 15 kHz are suppressed by the simultaneous presentation of pulses of 5 kHz at 10–15 dB greater amplitude. With respect to crickets, Nolen and Hoy (1986b) found similar suppression tuning in *T. oceanicus*, with best suppression measured between 4 and 5 kHz. Different from *E. carolinus*, however, frequencies >8 kHz were not effective in suppressing the startle response in *T. oceanicus*. Unlike our study, the measurements of suppression threshold in *T. oceanicus* were not normalized to startle threshold, however, as the startle stimulus was presented at 70 dB SPL. The mean startle threshold to 30 kHz in *T. oceanicus* is ~55 dB SPL (Nolen and Hoy, 1986a). Given the limits of their broadcast system, it is likely that the 70 dB SPL startle stimulus used by Nolen and Hoy (1986b; ~15 dB above startle threshold) was too great for any suppression by stimuli with frequencies >8 kHz. Thus, it is not clear if the suppression tuning curves for the higher frequencies (>8 kHz) in the two studies are comparable. Like *E. carolinus*, *T. oceanicus* showed linear suppression growth when tested in a two-tone paradigm. Estimates from Nolen and Hoy's (1986b) study, however, suggest that the slope of suppression growth at 5 kHz for *T. oceanicus* is <1. For *E. carolinus*, when suppression thresholds are normalized to the startle threshold, such low growth rates were measured only for frequencies outside the most sensitive region (Table I). Moreover, we found the greatest suppression growth at the best suppression frequency, 5 kHz (Table I). When the 5-kHz data for *E. carolinus* are *not* normalized to startle threshold, however, and are plotted as sound pressure levels (as in Nolen and Hoy, 1986b), suppression growth is <1 (slope=0.95; $P=0.007$) and thus comparable to that in *T. oceanicus*.

As in *E. carolinus*, suppression of the ultrasound-induced startle in *T. oceanicus* is most easily elicited when the low-frequency suppressor precedes or overlaps the startle stimulus (Nolen and Hoy, 1986b). Although Nolen and Hoy (1986b) used 30-ms duration pulses to measure suppression in the nonsimultaneous presentation paradigm, their measurements are consistent with ours when normalized to pulse duration. In the backward suppression condition, we were unable to measure suppression in *E. carolinus* when the suppressor followed the startle stimulus by >0.4 pulse durations (2 ms). For the similar presentation sequence to *T. oceanicus*, suppression could not be elicited when the suppressor followed the startle stimulus by >0.66 pulse durations (20 ms). In the forward suppression condition (suppressor leads), we were unable to measure any suppression when the startle stimulus followed the suppressor by >1 pulse duration (5 ms). Although the relationship is similar in *T. oceanicus*, the separations in which suppression could be measured are somewhat longer. When the suppressor leads the ultrasound pulse there is a decrease in suppression at a separation of 1 pulse duration (30 ms), but suppression can still be measured when the two stimuli are separated by up to 3 pulse durations (Nolen and Hoy, 1986b).

An alternative method to the two-tone suppression paradigm used here to measure frequency selectivity at different intensities is the classic masking paradigm in which thresholds for the detection of single tones at different intensities are measured as a function of the bandwidth of simultaneously presented maskers [Fletcher, 1940; but see Moore (1995) for review]. While quite common in human psychoacoustics, Ehret *et al.* (1982) uniquely employed this experimental method to measure the critical bands centered at frequencies in the calling song and ultrasound channels of *T. oceanicus*. By measuring the spectrum level of a noise band required to mask either positive or negative phonotaxis to an amplitude modulated tone, critical bands were calculated using the critical ratio formula. Similar to our results using the two-tone suppression paradigm, Ehret *et al.* (1982) found that the critical bandwidth centered around 4.5 kHz (near the calling song carrier frequency) increased with increases in the probe stimulus intensity (model calling song). Although the conclusion that auditory filter bandwidth increases at higher intensities is similar between the two studies, the experiments presumably probe two different underlying mechanisms. Whereas the two-tone suppression paradigm affects threshold using stimuli outside the excitatory band, the masking paradigm used by Ehret *et al.* (1982) presumably changes the signal-to-noise ratio using stimuli within the excitatory band.

D. Function of two-tone suppression in *E. carolinus*

Although two-tone suppression is common across disparate taxa and is mediated by a variety of different mechanisms, most hypotheses for its function are based on the benefits generated from increases in frequency resolution. Indeed, feature extraction in the spectral domain that is important to foraging bats and sexual signaling frogs, for example, is mediated at least in part by nonlinear inhibitory/suppression mechanisms (Capranica, 1965; Suga, 1973; Fuzessery, 1988). With respect to orthopteran insects, the resulting nonlinear response to the coupling of a variety of auditory stimuli has also been shown to modulate the auditory system in a variety of behavioral contexts. For example, the acoustic components of wind modulate the responses of certain auditory units in *Locusta migratoria* to stimulation with other sounds (Boyan, 1986). This modulation effectively shifts the frequency response of the auditory system to higher frequencies when the animal is in flight. The adaptive significance proposed for this shift is one of a contextual filter where noise produced by the wingbeat is prevented from eliciting a startle response (Boyan, 1986; Robert, 1989). In crickets, however, the most likely source of the characteristic sounds that best suppress the ultrasound-induced startle is not the wingbeat, but presumably singing conspecifics. Indeed, our finding that there is some evidence for sexual dimorphism in two-tone suppression sensitivity

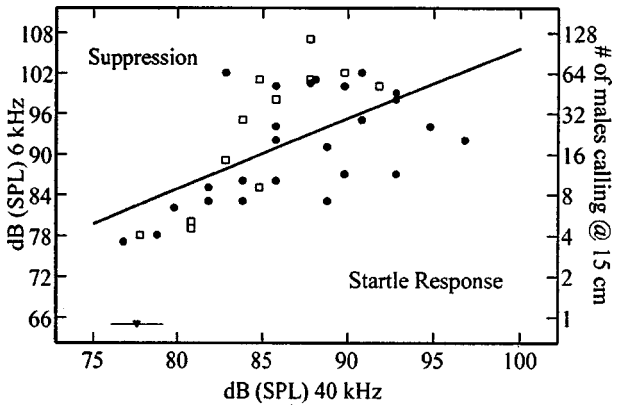


FIG. 7. Plot of suppression thresholds for a 6-kHz suppressor when measured in dB SPL (re: 20 μPa) rather than re: threshold (Fig. 3). Data are the same as those presented in Fig. 3. Based on the mean power output measured for five males calling in the field (@ 15 cm), the right axis is the number of males required to produce an aggregate song with the corresponding sound pressure level. The right axis scale assumes ideal summation of the songs. Circles and squares are the suppression thresholds for females and males, respectively. The triangle represents the mean startle threshold ($\pm$S.E.) to 40 kHz alone for these individuals. Thus, these data suggest that in the field $\sim$8 singing males (@ 15 cm) would be necessary to suppress the startle response, presumably making two-tone suppression unlikely. Note that when measured in SPL the slope for the suppression growth of 6 kHz is 1.04 ($r^2=0.352$, $P<0.001$), whereas in Fig. 3, where suppression thresholds are measured re: startle threshold, the slope is 1.35. This change in scale also reveals a significant difference between suppression growth in males and females ($\text{slope}_{\text{males}}=2.15$, $\text{slope}_{\text{females}}=0.85$, $t=4.518$, $P<0.001$) not found when measured re: startle threshold ($t=0.051$, $P=0.609$).

(see Fig. 7) suggests that the trait may play a role in calling song assessment. There are few data, however, supporting the initial hypothesis that two-tone suppression in crickets functions in the relative comparison of conspecific and predatory signals. As in *T. oceanicus* (Nolen and Hoy, 1986a), we found that relatively high levels of low-frequency (songlike) sounds were required to suppress the ultrasound-induced startle. Based on the average power output of males singing in the field, Fig. 7 shows that under ideal conditions an unrealistically high density of singing males (Farris, personal observation) would be necessary to produce suppressor stimuli. An alternative hypothesis for the function of two-tone suppression is that it could prevent negative phonotaxis to the harmonics of a calling song. Indeed, Nolen and Hoy (1986b) showed that there is enough power in the harmonics of the calling song of *T. oceanicus* to elicit negative phonotaxis. The addition of the low-frequency fundamental was sufficient to prevent negative steering and elicit a positive response. In *E. carolinus*, however, there is relatively little power in the harmonics. The second harmonic of the calling song fundamental is already 50 dB down [Fig. 2; harmonics are not required for effective song attraction (Farris *et al.*, 1997)]. Based on the songs recorded in this study, an aversive song would have to be unrealistically powerful and at close proximity. An alternative context from long-range mate attraction in which two-tone suppression could function is courtship in which aversive behavior in response to a courtship song rich in high frequencies is suppressed. In *E. carolinus*, however, not only is the courtship song similar in spectrum to the calling song (Alexander, 1957; Farris, unpublished observation), but suppression is elicited in the context of flight and, thus, an unlikely adaptation to the close-range communication during courtship. Thus, left without an obvious conclusion for its function, we can only speculate that two-tone suppression of the ultrasound-induced startle by low-frequency sounds in *E. carolinus* functions to restrict startle responses to sounds containing *only* ultrasound. Broadband sounds that span the audio and ultrasound bands (potentially produced by tettigoniids) would thus be prevented from eliciting any antipredator behavior. Further research is needed to elucidate the behavioral ecology of this trait.

ACKNOWLEDGMENTS

Thanks are owed to T. Forrest, A. Mason, and M. Oshinsky for technical help and advice on the project. D. McFadden, D. Bodnar, B. Carlson, B. Wyttenbach, and P. Faure offered useful comments to the article and project. Also, the article benefited from the suggestions provided by two anonymous reviewers. Special thanks go to C. Gilbert, C. Clark and P. Sherman for providing instruction in the design analysis of these experiments. H. Bass at the Nation Center for Physical Acoustics kindly provided equipment and T. Walker graciously helped with collecting animals. This project was presented in partial fulfillment of a doctoral degree at Cornell University. HEF was supported by NIMH training Grant No. 2 T32 MH15793 and NIH Grant No. R01 DC00103 to RRH.

Alexander, R. D. (**1957**). "The song relationships of four species of ground crickets (Orthoptera: Gryllidae: *Nemobius*)," Ohio J. Sci. **57**, 153–163.

Arthur, R. M., Pfeiffer, R. R., and Suga, N. (**1971**). "Properties of 'two-tone inhibition' in primary auditory neurones," J. Physiol. London **212**, 593–609.

Boyan, G. (**1981**). "Two-tone suppression of an identified auditory neurone in the brain of the cricket *Gryllus bimaculatus* (De Geer)," J. Comp. Physiol. **144**, 117–125.

Boyan, G. (**1986**). "Modulation of auditory responsiveness in the locust *Locusta Migratoria*," J. Comp. Physiol. **158**, 813–826.

Capranica, R. R. (**1965**). "The evoked vocal responses of the bullfrog: a study of communication by sound," Research Monograph 33 (MIT, Cambridge, MA).

Capranica, R. R. (**1992**). "The untuning of the tuning curve: is it time?," Semin. Neurosci. **4**, 401–408.

Delgutte, B. (**1996**). "Physiological models for basic auditory percepts," in *Auditory Computation*, edited by H. L. Hawkins, T. A. McMullen, A. N. Popper, and R. R. Fay (Springer-Verlag, New York).

Ehret, G., Moffat, A. J. M., and Tautz, J. (**1982**). "Behavioral determination of frequency resolution in the ear of the cricket Teleogryllus oceanicus," J. Comp. Physiol. **148**, 237–244.

Farris, H. E., and Hoy, R. R. (**1997**). "Acoustic startle and two-tone suppression in a Nemobiine cricket, *Eunemobius carolinus*," 27th Annual Meeting of the Society for Neuroscience, New Orleans, 25–30 October 1997, Society for Neuroscience Abstracts, Vol. **23**, p. 1070.

Farris, H. E., and Hoy, R. R. (**1998**). "Two-tone suppression of the ultrasound induced startle response in a cricket," Proceedings of the 16th International Congress on Acoustics and the 135th Meeting of the Acoustical Society of America, Seattle, WA, Vol. **2**, pp. 687–688.

Farris, H. E., and Hoy, R. R. (**2000**). "Ultrasound sensitivity in the cricket, *Eunemobius carolinus* (Gryllidae: Nemobiinae)," J. Acoust. Soc. Am. **107**, 1727–1736.

Farris, H. E., Forrest, T. G., and Hoy, R. R. (**1997**). "The effects of calling song spacing and intensity on the attraction of flying crickets (Orthoptera: Gryllidae: Nemobiinae)," J. Insect Behav. **10**, 639–653.

Fletcher, H. (**1940**). "Auditory patterns," Rev. Mod. Phys. **12**, 47–65.

Fuzessery, Z. M. (**1988**). ''Frequency tuning in the anuran central auditory system,'' in *The Evolution of the Amphibian Auditory System*, edited by B. Fritzsch, M. J. Ryan, W. Wilczynski, T. E. Hetherington, and W. Walkowiak (Wiley, New York).

Hutchings, M., and Lewis, B. (**1984**). ''The role of two-tone suppression in song coding by ventral cord neurones in the cricket *Teleogryllus oceanicus* (Le Guillou),'' J. Comp. Physiol., A **154**, 103–112.

Imaizumi, K., and Pollack, G. S. (**1999**). ''Neural coding of sound frequency by cricket auditory receptors,'' J. Neurosci. **19**, 1508–1516.

Kiang, N. Y-S. (**1965**). ''Discharge patterns of single fibers in the cat's auditory nerve,'' Research Monograph No. 35 (MIT, Cambridge, MA).

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

Loher, W., Weber, T., and Huber, F. (**1993**). ''The effect of mating on phonotactic behaviour in *Gryllus bimaculatus* (De Geer),'' Physiol. Entomol. **18**, 57–66.

Moiseff, A., and Hoy, R. R. (**1983**). ''Sensitivity to ultrasound in an identified auditory interneuron in the cricket: A possible neural link to phonotactic behavior,'' J. Comp. Physiol. [A] **152**, 155–167.

Moiseff, A., Pollack, G. S., and Hoy, R. R. (**1978**). ''Steering responses of flying crickets to sound and ultrasound: mate attraction and predator avoidance,'' Proc. Natl. Acad. Sci. U.S.A. **75**, 4052–4056.

Moore, B. C. J. (**1995**). ''Frequency analysis and masking,'' in *Hearing*, edited by B. C. J. Moore (Academic, London).

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1990**). ''Auditory filter shapes at low center frequencies,'' J. Acoust. Soc. Am. **88**, 132–140.

Moore, B. C. J., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). ''The shape of the ear's temporal window,'' J. Acoust. Soc. Am. **83**, 1102–1116.

Nolen, T. G., and Hoy, R. R. (**1984**). ''Initiation of behavior by single neurons: the role of behavioral context,'' Science **226**, 992–994.

Nolen, T. G., and Hoy, R. R. (**1986a**). ''Phonotaxis in flying crickets. I. Attraction to the calling song and avoidance of bat-like ultrasound are discrete behaviors,'' J. Comp. Physiol., A **159**, 423–439.

Nolen, T. G., and Hoy, R. R. (**1986b**). ''Phonotaxis in flying crickets. II. Physiological mechanisms of two-tone suppression of the high frequency avoidance steering behavior by the calling song,'' J. Comp. Physiol., A **159**, 441–456.

Nolen, T. G., and Hoy, R. R. (**1987**). ''Postsynaptic inhibition mediates high-frequency selectivity in the cricket *Teleogryllus oceanicus*: Implications for flight phonotaxis behavior,'' J. Neurosci. **7**, 2081–2096.

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

Plack, C. J., and Carlyon, R. P. (**1995**). ''Loudness perception and intensity coding,'' in *Hearing*, edited by B. C. J. Moore (Academic, London).

Plomp, R., and Bouman, M. A. (**1959**). ''Relation between hearing threshold and duration for tone pulses,'' J. Acoust. Soc. Am. **31**, 749–758.

Robert, D. (**1989**). ''The auditory behaviour of flying locusts,'' J. Exp. Biol. **147**, 279–301.

Rose, J. E., Gross, N. B., Geisler, C. D., and Hind, J. E. (**1971**). ''Some effects of stimulus intensity on response of auditory nerve fibers in the squirrel monkey,'' J. Neurophysiol. **34**, 685–699.

Sachs, M. B., and Kiang, N. Y.-S. (**1968**). ''Two-tone inhibition in auditory-nerve fibers,'' J. Acoust. Soc. Am. **43**, 1100–1128.

Schildberger, K. (**1984**). ''Temporal selectivity of identified auditory neurons in the cricket brain,'' J. Comp. Physiol., A **155**, 171–185.

Shailer, M. J., Moore, B. C. J., Glasberg, B. R., and Watson, N. (**1990**). ''Auditory filter shapes at 8 and 10 kHz,'' J. Acoust. Soc. Am. **88**, 141–148.

Suga, N. (**1973**). ''Feature extraction in the auditory system of bats,'' in *Basic Mechanisms in Hearing*, edited by A. R. Moller (Academic, New York).

Vogten, L. L. M. (**1974**). ''Pure-tone masking; a new result from a new method,'' in *Facts and Models in Hearing*, edited by E. Zwicker and E. Terhardt (Springer-Verlag, New York).

Walker, T. J. (**1962**). ''Factors responsible for intraspecific variation in the calling songs of crickets,'' Evolution **16**, 407–428.

Wohlers, D. W., and Huber, F. (**1982**). ''Processing of sound signals by six types of neurons in the prothoracic ganglion of the cricket *Gryllus bimaculatus* L.,'' J. Comp. Physiol. **146**, 161–173.

Wyttenbach, R. A., May, M. L., and Hoy, R. R. (**1996**). ''Categorical perception of sound frequency by crickets,'' Science **273**, 1542–1544.

Zar, J. H. (**1984**). *Biostatistical Analysis* (Prentice-Hall, Englewood Cliffs, NJ).