CODEN: JASMAN

The Journal of the Acoustical Society of America

ISSN: 0001-4966

Vol. 116, No. 6, December 2004

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 116, NO. 6, DECEMBER 2004

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 116, NO. 6, DECEMBER 2004

CONTENTS—*Continued from preceding page*

(Continued)

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA VOL. 116, NO. 6, DECEMBER 2004

CONTENTS—*Continued from preceding page*

BIOACOUSTICS [80]

INDEX TO VOLUME 116

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.*

Announcement of the 2005 Election

In accordance with the provisions of the bylaws, the following Nominating Committee was appointed to prepare a slate for the election to take place on 6 May 2002:

Richard Stern, *Chair*
Peggy B. Nelson
Emily A. Tobey
Elizabeth A. McLaughlin
James M. Sabatier
Ewart A. Wetherill

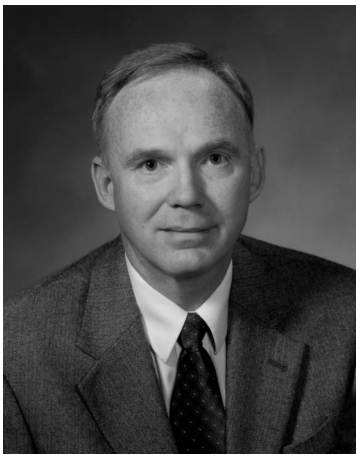
The bylaws of the Society require that the Executive Director publish in the *Journal* at least 90 days prior to the election date an announcement of the election and the Nominating Committee's nominations for the offices to be filled. Additional candidates for these offices may be provided by any Member or Fellow in good standing by letter received by the Executive Director not less than 60 days prior to the election date and the name of any eligible candidate so proposed by 20 Members or Fellows shall be entered on the ballot.

Biographical information about the candidates and statements of objectives of the candidates for President-Elect and Vice President-Elect will be mailed with the ballots.

CHARLES E. SCHMID
Executive Director

The Nominating Committee has submitted the following slate:

FOR PRESIDENT-ELECT



Anthony A. Atchley

Henrik Schmidt

FOR VICE PRESIDENT-ELECT

Whitlow W. L. Au

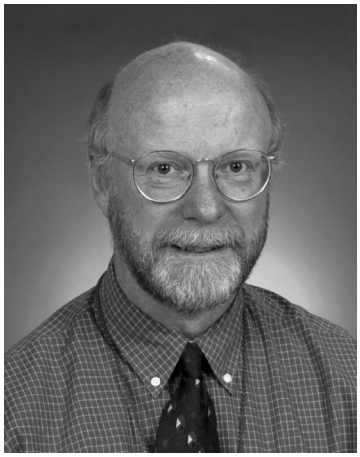
Uwe J. Hansen

 0001-4966/2004/116(6)/3237/3/$20.00

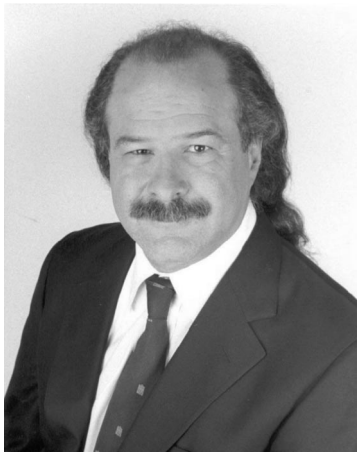
FOR MEMBERS OF THE EXECUTIVE COUNCIL

Sid P. Bacon

Diane Kewley-Port

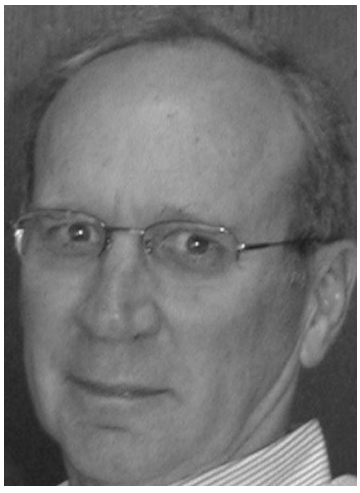


Clark S. Penrod



Neil A. Shaw

Stephen C. Thomson



Neal F. Viemeister

USA Meetings Calendar

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

2005

16–20 May	149th Meeting joint with the Canadian Acoustical Association, Vancouver, Canada [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW:http://asa.aip.org].
16–19 May	Society of Automotive Engineering Noise & Vibration Conference, Traverse City, MI [Patti Kreh, SAE International, 755 W. Big Beaver Rd., Ste. 1600, Troy, MI 48084, Tel.: 248-273-2474; E-mail: pkreh@sae.org].
18–22 July	17th International Symposium on Nonlinear Acoustics, State College, PA [Anthony Atchley, The Pennsylvania State University, 217 Applied Research Lab Building, University Park, PA 16802; Tel.: 814-865-6364; E-mail: ISNA17@outreach.psu.edu; WWW: http://www.outreach.psu.edu/c&i/isnal7/].
17–21 Oct.	150th Meeting joint with Noise-Con, Minneapolis, MN [Acoustical Society of America, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Tel.: 516-576-2360; Fax: 516-576-2377; E-mail: asa@aip.org; WWW:http://asa.aip.org].

Cumulative Indexes to the Journal of the Acoustical Society of America

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

Some indexes are out of print as noted below.

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

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

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

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

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

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

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

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

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

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

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

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

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

ACOUSTICAL NEWS—INTERNATIONAL

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

18th International Congress on Acoustics held in Kyoto, Japan, 4–9 April 2004

International Congress on Acoustics

Over 1000 papers were presented at the 18th International Congress on Acoustics (ICA) which was held in the historical city of Kyoto, Japan. The dates of April 4th through 9th were chosen to correspond with the blooming of the cherry blossoms, which provided added beauty to a city with aesthetic architectural and landscape designs. The Congress met in the Kyoto International Conference Hall, which played a part in recent history as it was the site for the adoption of the 1997 environmental ''Kyoto Accord'' on climate change.

The statistics for attendance appeared in the October issue of the *Journal of the Acoustical Society of America*. Over 1300 acousticians attended from 49 countries. Most of the attendees were from Japan (723), followed by the USA (89), Korea (69), Germany (58), and China (52).

The theme of the Congress was ''Acoustical Science and Technology for Quality of Life.'' A majority of the 94 structured sessions, which were covered by 56 titles, focused on this theme. Over 100 experts from around the world organized these sessions. Topics which comprised the theme of quality of life included the medical uses of sound (ultrasonic imaging, therapeutic ultrasound, sonochemistry, ultrasonic actuators, auditory neuroscience), architectural acoustics (classroom acoustics, computer simulations), noise control (noise barriers, environmental noise policies), speech processing, and of course musical acoustics. Like ICA meetings in the past, the presentations in Kyoto contained a mixture of new research and reviews of past work in acoustics. The *ICA2004 Proceedings*, consisting of papers 2 to 4 pages in length, were published on a CD-ROM, and is still available for purchase. (Please indicate ''ICA2004 Proceedings'' when you contact the Acoustical Society of Japan by e-mail ⟨KYM05145@nifty.ne.jp⟩ or fax: +81-3-5256-1022.)

The opening ceremony on Monday was made memorable with traditional Japanese Noh music, a videotaped welcome to attendees by Prime Minister Koizumi, and speeches by the conference organizers. On Thursday, Hideki Tachibana, the General Chairperson for the Conference, arranged for a special concert at the Kyoto Concert Hall which was open to the public. Ms. Mariko Senji performed on her Stradivarius violin with the Ritsumeikan University Symphony Orchestra. Before the concert she demonstrated the use of concert hall acoustics by showing how she optimizes her position on stage, and by asking various sections of the orchestra to play from nontraditional positions on stage. Another interesting feature of the Congress were posters and demonstrations on education in acoustics which were given in the exhibit area, including a lively demonstration of physical acoustics by the Stray Cats Group consisting of physics teachers from the Nagoya area of Japan.

FIG. 1. Ilene Busch-Vishniac, President of ASA, with Sonoko Kuwano, Secretary General for the 18th International Conference of Acoustics, after the opening ceremony.

International Commission for Acoustics

Over 40 delegates from ICA member societies gathered at the General Assembly of the International Commission for Acoustics (ICA) (The acronym ICA is used for both the International Commission for Acoustics and the International Congress on Acoustics.) on April 7th. They elected officers and new members for open seats on the ICA Board, and also stated their appreciation to retiring ICA officers (see Fig. 2). Outgoing president Gilles Daigle was thanked for his efforts to increase the activities of the Commission in addition to the usual planning for ICA meetings held every 3 years. These activities included support for specialty conferences and symposia, travel grants for young scientists, and Gilles Daigle's work on the proposal that the ICA join the ICSU (International Council of Scientific) as a separate scientific union. In addition, ICA Early Career Awards were presented to Timothy Leighton (UK) and Oleg Sapozhnikov (Russia).

The Commission selected Sydney, Australia as the site of the 2010 ICA meeting. At its previous meeting in Rome in 2001, the Commission had chosen Madrid, Spain to be the site of the 2007 Congress which will be held September 2nd through the 7th (see ⟨www.ica2007madrid.org⟩).

CHARLES SCHMID
ASA Delegate to the International Commission for Acoustics

FIG. 2. Officers of the International Commission for Acoustics (l to r): Sonoko Kuwano (Secretary-General), Hugo Fastl (Treasurer), Philip Nelson (President), Suk Wang Yoon (Vice President), Volker Mellert (Past Treasurer), Gilles Daigle (Past President).

International Meetings Calendar

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

December 2004

8–10	**Tenth Australian International Conference on Speech Science & Technology**, Sydney, Australia (Web: www.assta.org/sst/2004)

March 2005

14–17	**31st Annual Meeting of the German Acoustical Society (DAGA '05)**, Munich, Germany (Web: daga2005.de).
14–17	**Spring Meeting of the Acoustical Society of Japan**, Koganei, Japan (Acoustical Society of Japan, Nakaura 5th-Bldg., 2-18-20 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan; Fax: +81 3 5256 1022; Web: www.asj.gr.jp/index-en.html).
21–22	∗**Sonar Transducers and Numerical Modeling in Un-**

derwater Acoustics, Teddington, UK (Fax: +44 1727 850553; Web: www.npl.co.uk/acoustics/events/ioaconference2005)

April 2005

18–21 **International Conference on Emerging Technologies of Noise and Vibration Analysis and Control**, Saint Raphaël, France (Fax: +33 4 72 43 87 12; e-mail: goran.pavic@insa-lyon.fr).

May 2005

16–20 **149th Meeting of the Acoustical Society of America**, Vancouver, British Columbia, Canada (ASA, Suite 1NO1, 2 Huntington Quadrangle, Melville, NY 11747-4502; Fax: +1 516 576 2377; Web: asa.aip.org)

June 2005

20–23 **IEEE Oceans05 Europe**, Brest, France (ENST Bretagne—Technopôle Brest Iroise, 29238 Brest Cedex, France; Fax: +33 229 00 1098; Web: www.oceans05europe.org).

23–24 **2nd Congress of the Alps-Adria Acoustical Association (AAAA2005)**, Opatija, Croatia (Web: had.zea.fer.hr).

28–1 **International Conference on Underwater Acoustic Measurements: Technologies and Results**, Heraklion, Crete, Greece (Web: UAmeasurements2005.iacm.forth.gr).

July 2005

4–8 *__Turkish International Conference on Acoustics 2005: New Concepts for Harbor Protection, Littoral Security, and Underwater Acoustic Communications__, İstanbul, Turkey (Web: www.tica05.org/tica05)

11–14 **12th International Congress on Sound and Vibration**, Lisbon, Portugal (Web: www.iiav.org).

August 2005

6–10 **Inter-Noise**, Rio de Janeiro, Brazil (Web: www.internoise2005.ufsc.br).

28–2 **EAA Forum Acusticum Budapest 2005**, Budapest, Hungary (I. Bába, OPAKFI, Fö u. 68, Budapest 1027, Hungary; Fax: +36 1 202 0452; Web: www.fa2005.org).

September 2005

4–8 **9th Eurospeech Conference (EUROSPEECH'2005)**, Lisbon, Portugal (Fax: +351 213145843; Web: www.interspeech2005.org).

5–9 **Boundary Influences in High Frequency, Shallow Water Acoustics**, Bath, UK. (Web: acoustics2005.ac.uk).

11–15 **6th World Congress on Ultrasonics**, Beijing, China (Secretariat of WCU 2005, Institute of Acoustics, Chinese Academy of Sciences, P.O. Box 2712 Beijing, 100080 China; Fax: +86 10 62553898; Web: www.ioa.ac.cn/wcu2005).

14–16 **Autumn Meeting of the Acoustical Society of Japan**, Sendai, Japan (Acoustical Society of Japan, Nakaura 5th-Bldg., 2-18-20 Sotokanda, Chiyoda-ku, Tokyo 101-0021, Japan; Fax: +81 3 5256 1022; Web: www.asj.gr.jp/index-en.html).

October 2005

19–21 **36th Spanish Congress on Acoustics Joint with 2005 Iberian Meeting on Acoustics**, Terrassa (Barcelona), Spain (Sociedad Española de Acústica, Serrano 114, 28006 Madrid, Spain; Fax: +34 914 117 651; Web: www.ia.csic.es/sea/index.html).

June 2006

5–6 **6th European Conference on Noise Control (EURONOISE2006)**, Tampere, Finland (Fax: +358 9 7206 4711; Web: www.acoustics.hut.fi/asf)

26–28 **9th Western Pacific Acoustics Conference (WESPAC9)**, Seoul, Korea (Web: www.wespac9.org)

July 2007

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

September 2007

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

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

June 2008

23–27 **Joint Meeting of European Acoustical Association (EAA), Acoustical Society of America (ASA), and Acoustical Society of France (SFA)**, Paris, France (e-mail: phillipe.blanc-benon@ec-lyon.fr)

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:

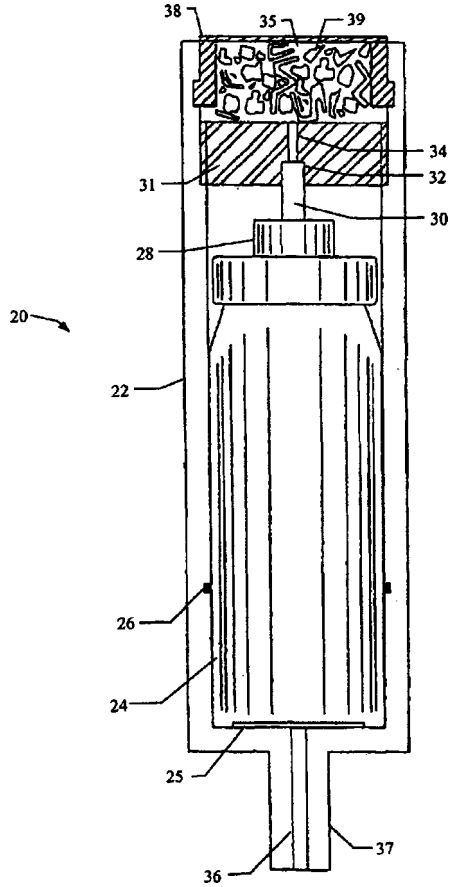
GEORGE L. AUGSPURGER, *Perception, Incorporated, Box 39536, Los Angeles, California 90039*
JOHN M. EARGLE, JME Consulting Corporation, 7034 Macapa Drive, Los Angeles, California 90068
MARK KAHRS, *Department of Electrical Engineering, University of Pittsburgh, Pittsburgh, Pennsylvania 15261*
DAVID PREVES, *Starkey Laboratories, 6600 Washington Ave. S., Eden Prarie, Minnesota 55344*
DANIEL R. RAICHEL, *2727 Moore Lane, Fort Collins, Colorado 80526*
CARL J. ROSENBERG, *Acentech Incorporated, 33 Moulton Street, Cambridge, Massachusetts 02138*
NEIL A. SHAW, Menlo Scientific Acoustics, Inc., Post Office Box 1610, Topanga, California 90290
WILLIAM THOMPSON, JR., *Pennsylvania State University, University Park, Pennsylvania 16802*
ROBERT C. WAAG, *Univ. of Rochester, Department of Electrical and Computer Engineering, Rochester, New York 14627*

6,749,481

43.28.Hr SOUND PRODUCING APPARATUS

Leong Kheng Yap and Tiang See Neo, assignors to TechnoCracker Private Limited
15 June 2004 (Class 446/213); filed in Singapore 31 August 1998

Fireworks and firecrackers (and gunpowder) are an ancient Chinese discovery dating back to the 8th century. Unfortunately, they are explosive and dangerous to life and limb, so then the question becomes how to create a safe and sane equivalent. The inventors propose using compressed air that rapidly ruptures a tightly sealed cap **38** also emitting confetti **39**.—MK

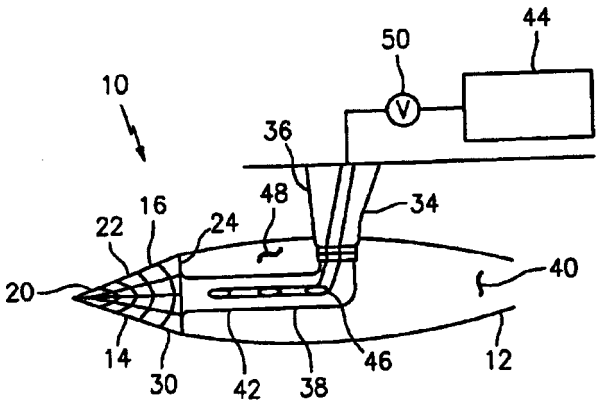


6,744,694

43.30.Es GASEOUS CAVITY FOR FORWARD-LOOKING SONAR QUIETING

Thomas J. Gieseke and Robert Kuklinski, assignors to The United States of America as represented by the Secretary of the Navy
1 June 2004 (Class 367/1); filed 6 October 2003

A system for minimizing the hydrodynamic noise output of a sonar array mounted on a high speed surface vessel consists of a cavitation device **14** for intentionally producing cavitation downstream of the forward-looking array **16**, which is mounted on the cavitator, and hence forming a cavity of gas **40** behind the array. Struts **34** and **38** support the cavitator **14** and the array **16** from the hull of the vessel. The cavitator could have the shape of a flat plate, a disk, a hemisphere or, as illustrated, a cone. The flow, which is laminar over the face of the array, separates aft of the base **24** of the cavitator forming gas bubble **40**, whose outer boundary is suggested by line **12**. The size of the bubble can be augmented by pumping air from on-board source **44** out through holes **46** in strut **38**. The gaseous cavity **40** serves to eliminate direct acoustic paths between the array **16** and sources of noise aft of the array, hopefully without being a source of noise itself.—WT

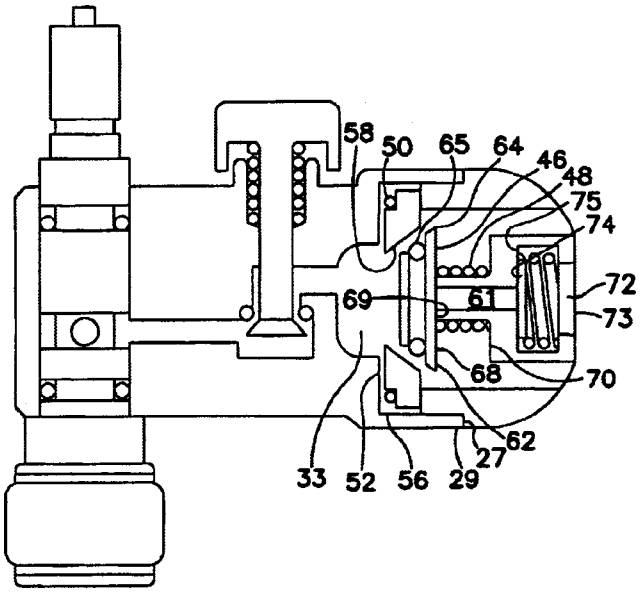


6,755,147

43.30.Nb UNDERWATER SIGNALLING DEVICES

Peter Katz, Manly Vale, NSW, Australia
29 June 2004 (Class 116/27); filed in Australia 8 December 1999

An underwater diver's percussive signaling device, which is energized by the diver's own compressed air supply, is described. Some of the air flowing through the channel at the left of the figure can be diverted to chamber **33** by depressing the spring-loaded, finger-actuated switch near the top center of the figure. The increase in pressure in closed chamber **33** forces both the valve seat **58** and its associated valve **46** to the right. First, the valve seat bottoms against an annular stop associated with the housing, thus creating percussive sound that radiates to the surrounding medium through openings in the housing. The valve itself continues movement to the right.



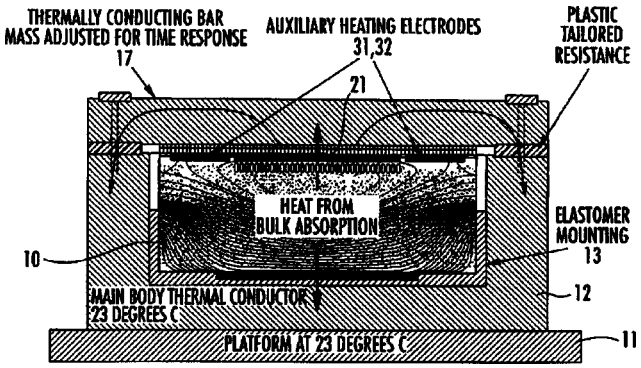
Its attached centering rod **61** ultimately contacts a disk abutting spring **74**, again radiating some sound through opening **73**. Springs **48** and **74** ultimately stop valve **46** and force it, as well as the movable valve seat **58**, back towards the closed position. The disk at the left end of spring **74**, now moving to the left, then bottoms against a stop due to its housing **75**, thus radiating more percussive-type sound. Similarly, the valve, reseated in valve seat **58**, will hit stop **52** of the main housing, radiating more sound. Now that chamber **33** is again sealed, the increase in pressure due to the input air flow starts the cycle again.—WT

6,765,709

43.35.Sx ACTIVE COMPENSATION FOR TRANSIENT THERMAL EFFECTS IN ACOUSTO-OPTIC DEVICES

Robert Morris Montgomery *et al.*, assignors to Harris Corporation
20 July 2004 (Class 359/285); filed 28 April 2003

Acousto-optic systems are generally employed in optical modulation systems to process high-frequency signals. Owing to the small physical volume of such subsystems and the typically random-data-pattern nature of the signals they process, acousto-optical signal processing provisions are subject to thermal transients, which can lead to distortion of the optical beam and degrade system performance. In order to mitigate the thermal



distortion problem, a multichannel acousto-optical modulator contains an integrated structure for providing active compensation for transient thermal effects. Multiple electro-thermal elements in the form of resistive strips are interleaved with acoustic wave-emitting transducers, to which signals are applied for modulating the acoustic waves launched into an acousto-optic medium. The resistive strips receive electrical signals that controllably introduce thermal energy components into the acousto-optic medium in a manner that causes the overall spatial distribution of thermal energy to have a prescribed characteristic. It is said that by establishing a thermal gradient characteristic across all the channels of the rf signal processor and compensating each channel on an individual basis, the device compensates for time-dependent variations in heating, thereby resulting in a substantially constant thermal behavior.—DRR

6,751,161

43.38.Ar FLUIDBORNE SOUND PROJECTOR WITH SWEEP CLEANING FACILITIES

John W. Henry IV *et al.*, assignors to The United States of America as represented by the Secretary of the Navy
15 June 2004 (Class 367/143); filed 26 September 2001

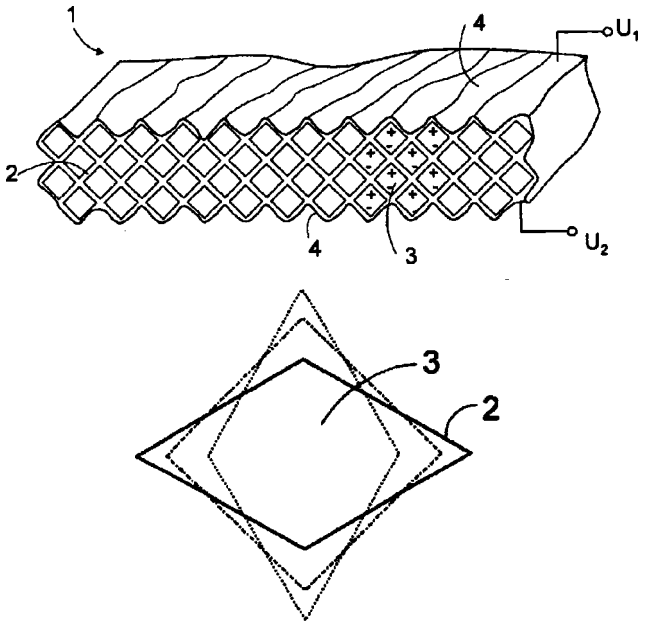
The operation of a high-power, low-frequency piston projector, used to input acoustical energy into a high-static-pressure piping system, is compromised by the formation of air bubbles in the fluid in those pipes. The patent describes a system for dynamically sweeping those bubbles away by injecting a flow of pressurized liquid into a vented chamber adjacent to, but communicating with, the chamber housing the piston projector.—WT

6,759,769

43.38.Ar ELECTROMECHANIC FILM AND ACOUSTIC ELEMENT

Kari Kirjavainen, Tampere, Finland
6 July 2004 (Class 307/400); filed in Finland 25 November 1999

This patent describes a bilateral compound moving structure that acts as something of a cross between an electret foil and a piezoelectric element. As shown in the figure, the structure undergoes deformation under the influence of sound waves. It is described as lightweight, and the net charge



potential is in the 100-volt range. Thus, the output should be substantial and uniform over a wide frequency range. The patent also discusses manufacturing details.—JME

6,757,399

43.38.Bs ANTI-NOISE-ELECTRET PICK-UP WITH AN ELECTRET

Ziyi Cheng, Nankai District, Tianjing, China
29 June 2004 (Class 381/356); filed in China 22 July 1998

A good noise-canceling (antinoise) microphone is a pressure gradient design that is mechanically symmetrical enough to cancel sound arriving at a distance, while at the same time allowing proximity effect to generate an adequate first-order, low-frequency boost for close-in speaking. The front/back gradient distance determines the speech sensitivity of the microphone, and as that distance is increased for the sake of sensitivity, the effectiveness of high-frequency noise canceling will be compromised. The patent deals with several mechano-acoustical adjustments that result in improved overall performance.—JME

6,744,370

43.38.Dv VIBRO-TACTILE ALERT AND MASSAGING SYSTEM HAVING DIRECTIONALLY ORIENTED STIMULI

Charles G. Sleichter III *et al.*, assignors to InSeat Solutions, LLC
1 June 2004 (Class 340/576); filed 13 July 1999

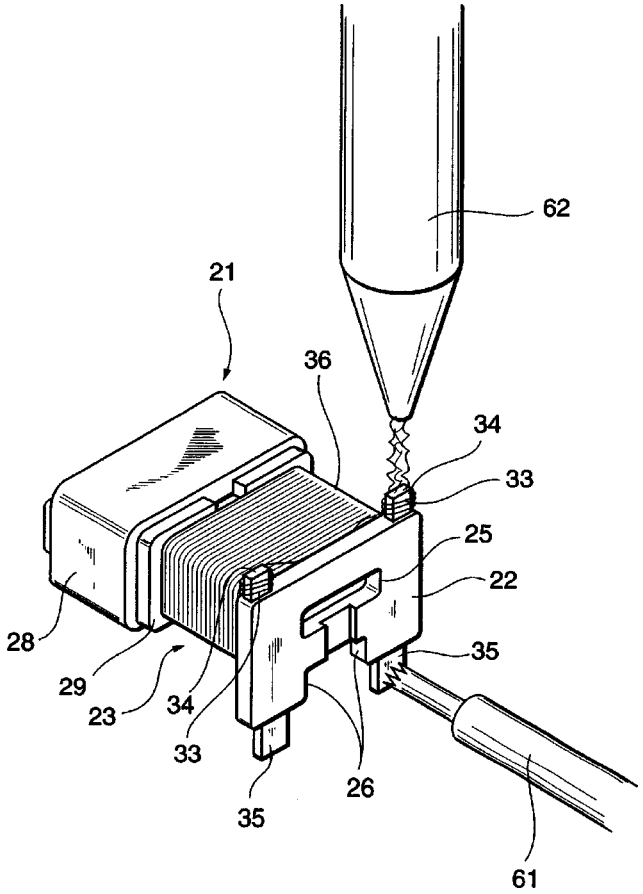
Six or eight electromagnetic vibrators are embedded in the driver's seat of a vehicle. The patent describes an elaborate programming arrangement that allows the system to provide a variety of relaxing massage routines as well as serve as a warning device.—GLA

6,757,403

43.38.Dv ELECTROACOUSTIC TRANSDUCERS

Kiyoshi Urushibata and Tomonari Suzuki, assignors to Star Micronics Company, Limited
29 June 2004 (Class 381/418); filed in Japan 16 March 2000

With the ever-increasing exposure of humans to noise in this modern world, the need for more robust transducers for use in hearing aids will only increase. The patent describes in detail a means of producing more robust



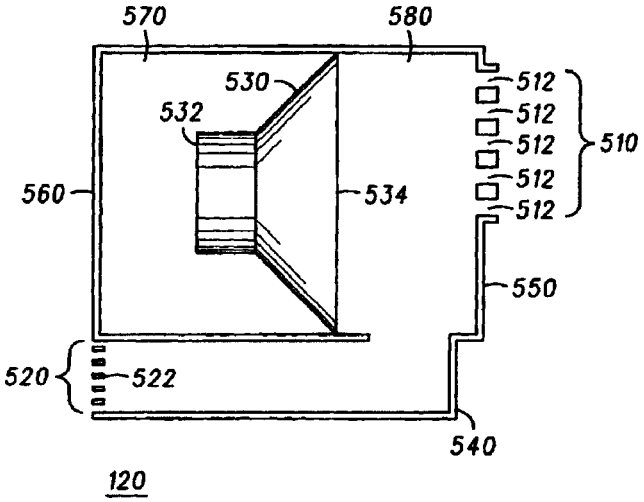
electroacoustic transducers for use in hearing aids, especially the means of easily attaching the small and fragile lead wires from the coil to the signal input terminals.—NAS

6,758,303

43.38.Dv ELECTRONIC DEVICE HAVING A MULTI-MODE ACOUSTIC SYSTEM AND METHOD FOR RADIATING SOUND WAVES

Robert A. Zurek *et al.*, assignors to Motorola, Incorporated
6 July 2004 (Class 181/155); filed 31 October 2002

A cell phone, or other personal communications device, normally has at least two modes of operation. For two-way communication the cell phone ''clamshell'' structure is opened and both transmitter and receiver are engaged at the user's head. For receive-only operation, the device is closed and the backside of the receiver is positioned at the user's ear. In this patent, both opened and closed modes of receiver operation are individually equalized, via acousto-mechanical networks, so that the quality of sound will be



120

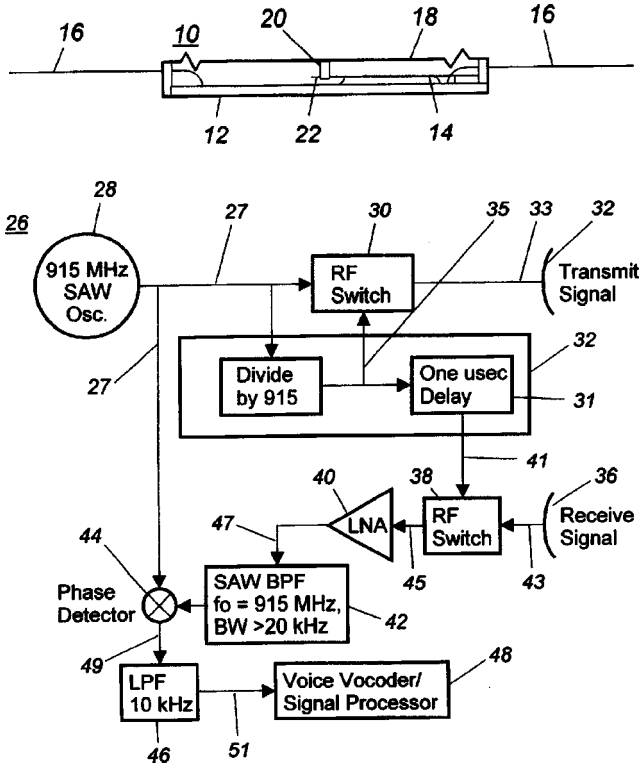
uniform in both modes of operation. The back chamber of the receiver is sealed and both outputs are in acoustical parallel. The main output section **510** is equalized by the volume **580** and the resistance produced by elements **512**. The alternate (clamshell closed) output **520** is similarly fed through the volume **580** and the resistance produced by elements **522**.—JME

6,760,454

43.38.Fx PASSIVE VOICE-ACTIVATED MICROPHONE AND TRANSCEIVER SYSTEM

Gregory A. Shreve *et al.*, assignors to TRW Incorporated
6 July 2004 (Class 381/110); filed 4 August 2000

In advanced automotive control systems there is a need for verbal commands by the driver to directly actuate various functions. Microphones may be positioned in the cab to pick up these commands, but noise often intrudes. Body-worn microphones are better as regards noise, but they may be clumsy to implement. This patent describes an unusual passive microphone, located in the upper portion of the driver's seatbelt, that has the benefit of high signal-to-noise ratio and requires no particular effort on the part of the driver. Figure 1 shows details of a passive microphone based on



an acoustic surface wave (ASW) element. Diaphragm **18** physically actuates the ASW element **14** when the driver talks. An electronic actuator/receiver assembly is located in the dashboard, or in the sun visor, and this assembly (shown in Fig. 2) emits a continuous set of rf pulses in the 1-GHz range. These pulses (emitted by element **32**) are picked up by antennas **16** on the microphone and are voice modulated by the bending motion of element **14**. They are then, during the interval between rf pulses, reradiated back to the receiving element **36** and further processed.—JME

6,757,394

43.38.Hz MICROPHONE ARRAY

Naoshi Matsuo, assignor to Fujitsu Limited
29 June 2004 (Class 381/92); filed in Japan 18 February 1998

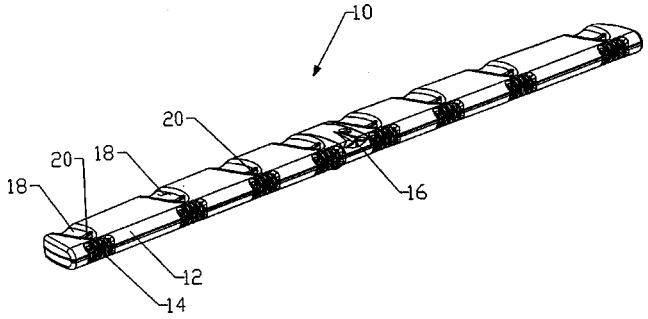
One way of synthesizing higher-order directional microphone pickup is to use a ''delay-and-sum'' technique, which enhances signal pickup along a target axis while allowing noise components to add more or less randomly. The patent explores methods for further processing of individual signals to result in lower noise levels.—JME

6,763,118

43.38.Hz HIGH DIRECTIVITY MICROPHONE ARRAY

Martin Reed Bodley *et al.*, assignors to GN Netcom, Incorporated
13 July 2004 (Class 381/355); filed 28 October 2002

This patent is a divisional part of United States Patent 6,473,514 [reviewed in J. Acoust. Soc. Am. **113**(6), 2957 (2003)], which dealt with electronic performance. The present patent deals with packaging and housing



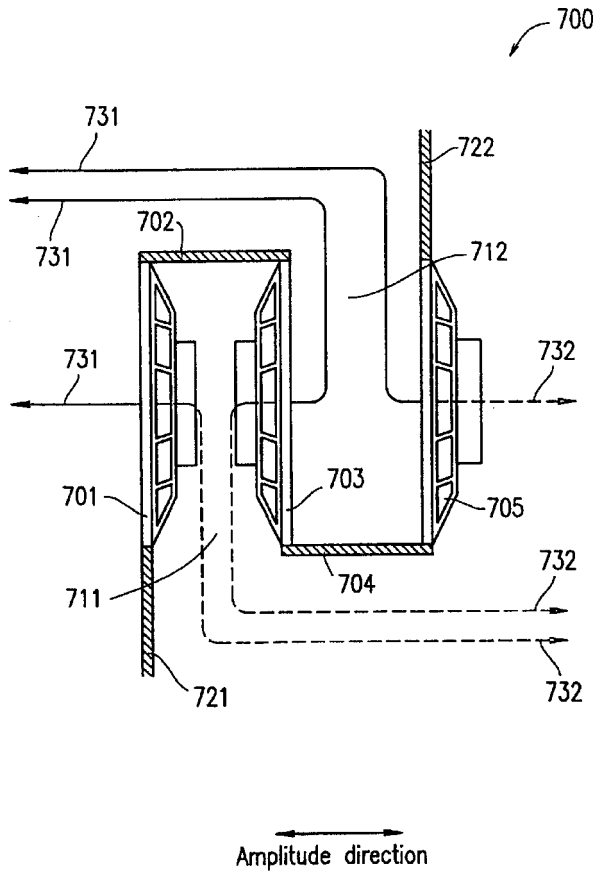
aspects and stresses the various mechanical features of an in-line microphone array intended to be mounted on the top of a video monitor and aimed as required.—JME

6,739,424

43.38.Ja SPEAKER SYSTEM

Takashi Ogura and Kosaku Murata, assignors to Matsushita Electric Industrial Company, Limited
25 May 2004 (Class 181/145); filed in Japan 22 January 2001

This is a curious patent. For the moment, consider only loudspeakers **701** and **703** mounted in a U-shaped baffle. The speakers can be either piezoelectric or electrodynamic devices. If the cones move in opposite directions, then their combined front radiation is indicated by arrows **731** and rear radiation by arrows **732**. This is the basic arrangement from which all



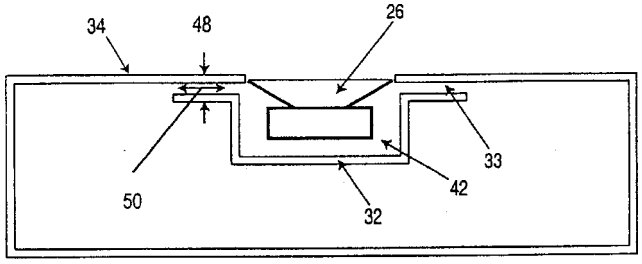
43 patent claims are derived, yet it has been used elsewhere—the Electro-Voice ''Manifold Technology'' design, for example. If additional cells are included as shown, then the arrangement would seem to be a variant of Heil's earlier ''air motion transformer.'' Neither the Heil nor Electro-Voice patents are included in the references cited.—GLA

6,741,717

43.38.Ja DEVICE FOR REDUCING STRUCTURAL-ACOUSTIC COUPLING BETWEEN THE DIAPHRAGM VIBRATION FIELD AND THE ENCLOSURE ACOUSTIC MODES

Stephane Dedieu and Philippe Moquin, assignors to Mitel Knowledge Corporation
25 May 2004 (Class 381/345); filed in the United Kingdom 28 September 2001

A small loudspeaker that might be used in a loudspeaking telephone is characterized by a tiny magnet to reduce cost and a lightweight moving assembly to maximize efficiency. As a result, there is a pronounced peak at



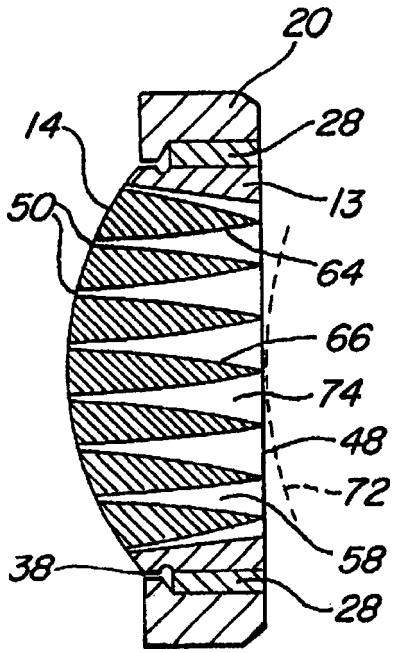
system resonance. In this design, a small, cup-shaped secondary enclosure **32** vents into the main rear cavity. If the acoustic elements are properly selected, the peak can be flattened and low-frequency response can be extended.—GLA

6,744,899

43.38.Ja DIRECT COUPLING OF WAVEGUIDE TO COMPRESSION DRIVER HAVING MATCHING SLOT SHAPED THROATS

Robert M. Grunberg, Edgecliff, NSW, Australia
1 June 2004 (Class 381/343); filed 21 September 2000

To obtain maximum efficiency over a wide bandwidth, the diaphragm of a high-frequency horn driver must be coupled to a much smaller horn throat in such a way that very high audio frequencies are not attenuated by path length differences. Traditionally, sound pickup channels can take the



form of circular slots, radial slots, or tubular conduits. In many cases, an additional coupler is required to connect the circular driver opening to a rectangular horn throat. The clever design shown here is a kind of three-dimensional Venetian blind that picks up sound through parallel linear slots and combines the channels into a single rectangular opening.—GLA

6,744,903

43.38.Ja MULTIPLE DAMPING DEVICE OF SPEAKER SYSTEM FOR VIDEO DISPLAY EQUIPMENT

Chang Wook Jeon, assignor to LG Electronics Incorporated
1 June 2004 (Class 381/388); filed in the Republic of Korea 15 April 1999

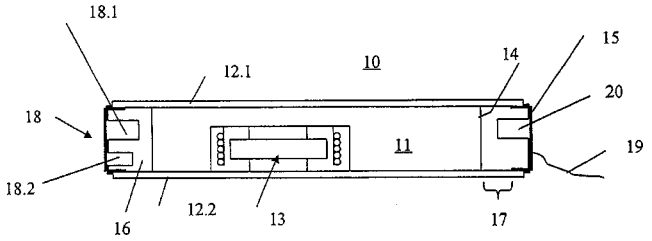
In some tabletop TV sets, self-contained speaker assemblies are attached to the cabinet using resilient fasteners. These are intended to improve sound quality and minimize video jitter induced by external vibrations. This patent describes an improved ''multiple damping device'' that is said to provide better mechanical isolation, especially at low frequencies. In effect, two or three isolation mounts are coupled together with little Z brackets.—GLA

6,748,090

43.38.Ja MULTI-MODE RADIATOR PANELS

Wolfgang Bachmann *et al.*, assignors to Harman Audio Electronic Systems GmbH
8 June 2004 (Class 381/152); filed in Germany 19 September 1998

This panel-type speaker assembly is made up of core layer **11** plus cover layers **12.1** and **12.2**. A conventional moving-coil driver **13** is located in cavity **16** between the two cover layers. Driver electronic circuitry is also



located in cavity **16**, which can be hermetically sealed. This allows the entire assembly to be ''operated without problem in listening areas which pose an explosion hazard or are subjected to a hostile environment and/or climate.''—GLA

6,751,329

43.38.Ja LOUDSPEAKER DRIVER

Martin Colloms and Ian David MacFarlane, assignors to New Transducers Limited
15 June 2004 (Class 381/353); filed in the United Kingdom 21 September 2000

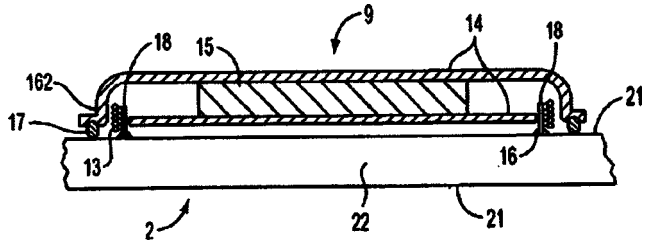
Fifteen or 20 years ago, patents were regularly awarded to inertial audio transducer designs. These devices were intended to be screwed to wall panels or table tops and functioned somewhat like wideband tuning forks. This patent describes a more sophisticated variant that can be affixed to a point-of-purchase display panel. A housing contains a transducer and an associated power amplifier. At least one adhesive pad secures the housing to the panel and vibrations are transmitted to the panel through a separate adhesive pad.—GLA

6,751,333

43.38.Ja INERTIAL VIBRATION TRANSDUCERS

Henry Azima *et al.*, assignors to New Transducers Limited
15 June 2004 (Class 381/396); filed in the United Kingdom 2 September 1995

The design shown here is an inertial driver assembly for distributed-mode, panel-form loudspeakers. Moving coil **18** is glued to panel **22**, while the enclosed magnetic assembly **14**, **15** is supported on a resilient suspen-



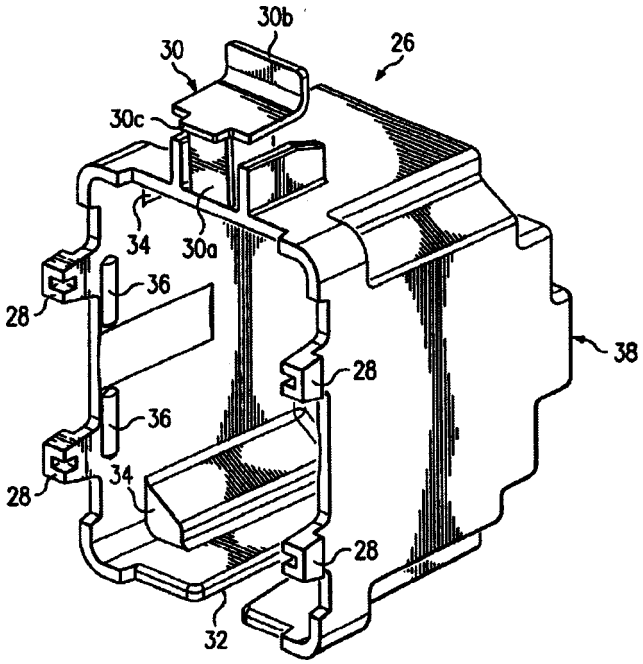
sion **17**. The patent includes a fair amount of general information about such loudspeakers, and the claims cover both the inertial vibration transducer and a loudspeaker incorporating such a transducer.—GLA

6,755,277

43.38.Ja SPEAKER RESONANCE VOICEBOX

Mark L. Rehmann, assignor to Dell Products L.P.
29 June 2004 (Class 181/155); filed 2 August 2001

This patent is one of many that concerns itself with what happens with the sound that emanates from the rear of a loudspeaker mounted in an enclosure, which will radiate from the transducer towards the rear of the enclosure, where it is then reflected back towards the transducer, and then outward through the transducer. The peak in response due to this reflection from the rear of the enclosure is said to be reduced by the invention. A soft material lining the rear, and sometimes the sides, of the enclosure is often



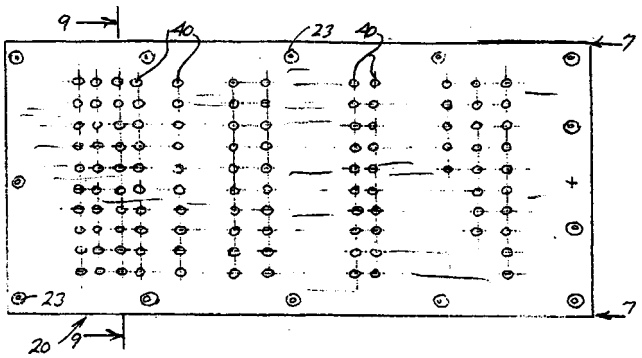
used to reduce this peak in response. Although this simple method is often effective, it can reduce the level of the reflected sound and therefore the total sound from the loudspeaker and enclosure. For laptop computers, the goal in some cases is to maximize the sound output from the tiny loudspeaker and its tiny backbox. ''Therefore, what is needed is a speaker box that varies the natural response and avoids a peak in the perceived speaker volume, but does not absorb sound energy and is less costly to manufacture.''—NAS

6,760,462

43.38.Ja PLANAR DIAPHRAGM LOUDSPEAKERS WITH NON-UNIFORM AIR RESISTIVE LOADING FOR LOW FREQUENCY MODAL CONTROL

F. Bruce Thigpen, assignor to Eminent Technology Incorporated
6 July 2004 (Class 381/431); filed 9 January 2003

Planar film speakers such as ribbons and electrostatics are often tensioned to hold the diaphragm in place within the speaker's structure. Even when tensioned, modal flexing of the film diaphragm can still occur, along with the attendant response anomalies and buzzing. Acoustic resistance in the form of optimum vent holes can create an air load to minimize this effect. This patent presents a means of controlling the low-frequency modal behavior of electrodynamic or electrostatic ''flat panel'' loudspeakers by the



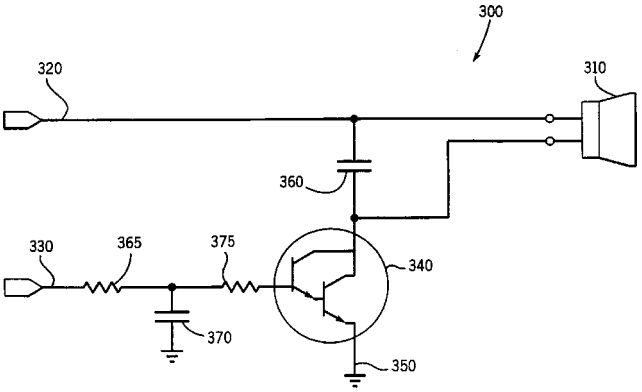
use of nonuniformly spaced openings **40** in the stator plate or plates **7**. The text includes a brief and clear review of the problems that are inherent due to the transducer mechanical design. The author has several other patents concerned with planar loudspeakers which are licensed to some of the major players in this market.—NAS

6,725,109

43.38.Lc SOUND GENERATOR CIRCUIT SYSTEM AND METHOD

Wayne B. Hile and James B. Henrie, assignors to Palm, Incorporated
20 April 2004 (Class 700/94); filed 12 March 2001

This easy-to-understand disclosure covers a simple sound generator for a palm computer. The inventors argue that using an amplifier together with PCM samples will be too expensive in both memory and battery consumption. So, they propose a simple three-component low-pass filter coupled to a Darlington pair driving a piezoelectric transducer. The input is a serial PWM stream. Not exactly hi-fi, but very economical.—MK



6,437,229

43.38.Md EQUIPMENT AND PROCESS FOR MUSIC DIGITALIZATION STORAGE, ACCESS, AND LISTENING

Fernando H. Nobumoto, assignor to Itautec Phico S/A
20 August 2002 (Class 84/615); filed 17 November 2000

An MP3 jukebox? Filed in 2000, this patent is at least 15 years behind the times.—MK

6,439,723

43.38.Md ORNAMENTAL IMAGE DISPLAY AND SOUND DEVICE

Robert S. Tano, Ellicott City, Maryland
27 August 2002 (Class 353/15); filed 10 March 2000

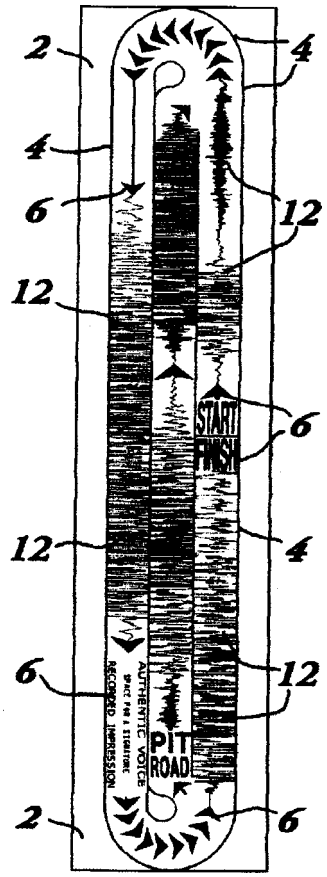
The patent author wants to add audio to Christmas ornaments. Given the extensive prior art on stored greetings, the novelty of this part of the patent is very suspect (The image projector is anyone's guess).—MK

6,440,497

43.38.Md PRINTED AUDIO LAYER FOR PACKAGING, SPORTING EQUIPMENT AND TOYS

Donald Giles Hohenstein, Mooresville, North Carolina
27 August 2002 (Class 427/256); filed 30 May 2000

You thought vinyl was dead? Not so fast. This inventor wants to use raised inks to create a soundtrack. And, you can combine it with ordinary quasiflat inks to have labels, too. The playback transducer is a fingernail guided perpendicularly along the track.—MK



6,699,098

43.38.Md ANIMATED MUSICAL ALLIGATOR

Ben Kau, assignor to Ben Kau
2 March 2004 (Class 446/298); filed 20 August 2002

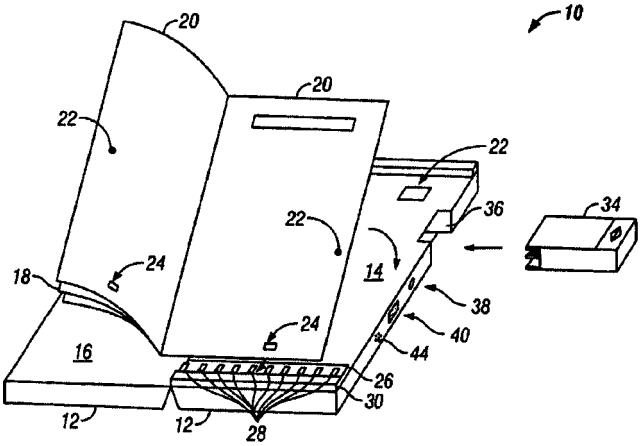
Although the inventor promises to make this animated alligator sing along with the music it emits, there is absolutely no discussion about how this might be accomplished. Florida residents may wish their alligators sang too.—MK

6,763,995

43.38.Ne METHOD AND SYSTEM FOR ILLUSTRATING SOUND AND TEXT

Jin K. Song, assignor to PIL, L.L.C.
20 July 2004 (Class 235/375); filed 25 April 2000

The patent describes a method and system for illustrating sound and text through a book holder **12** adapted to accept a special type of book with pages containing illustrations and/or text, with some of the pages containing magnetic signatures. The magnetic signals may be placed at specific locations **24** of the book and used by the reader to retrieve electronic representations of the text and/or illustrations for audible delivery. The book holder contains a reading surface **14** incorporating a magnetic signature sensor **16**, a cartridge slot **36**, a reading controller, a speaker **44**, and a power supply. A



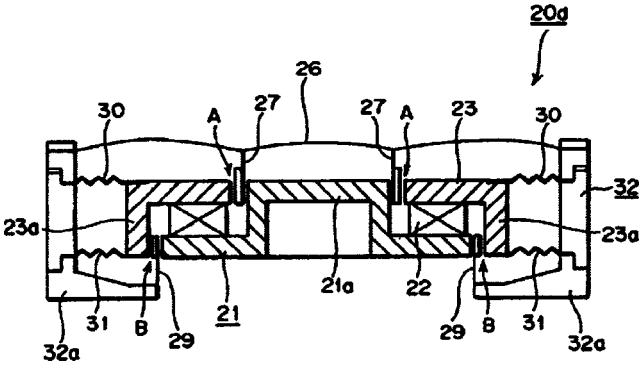
cartridge **34** containing stored audio representations related to the illustrations and/or text of the pages inserts into slot **36**. The magnetic signature sensor **16** detects magnetic signatures **24** on the pages as they are turned by a user viewing the book. The reading controller interacts with the magnetic reading sensor to ascertain what page or pages the user is viewing and to retrieve the corresponding audio representations. The reading controller reproduces the audible sounds via loudspeaker **44**.—DRR

6,741,721

43.38.Si INFORMING APPARATUS FOR MOBILE COMMUNICATION APPARATUS

Toshifumi Fukawatase *et al.*, assignors to Minebea Company, Limited
25 May 2004 (Class 381/401); filed 22 February 2001

Most pagers and cellular telephones allow the user to select between an audible beep or an inaudible vibration as a call signal. The patent argues that using a single magnetic circuit and voice coil to accomplish both functions is unsatisfactory. In this design, coil **A** drives diaphragm **26** to generate an audible signal, whereas coil **B** moves the entire magnetic assembly to produce tactile vibrations.—GLA



6,754,546

43.38.Tj ELECTRONIC AUDIO CONNECTION SYSTEM AND METHODS FOR PROVIDING SAME

Debby Hindus *et al.*, assignors to Interval Research Corporation
22 June 2004 (Class 700/94); filed 9 October 1998

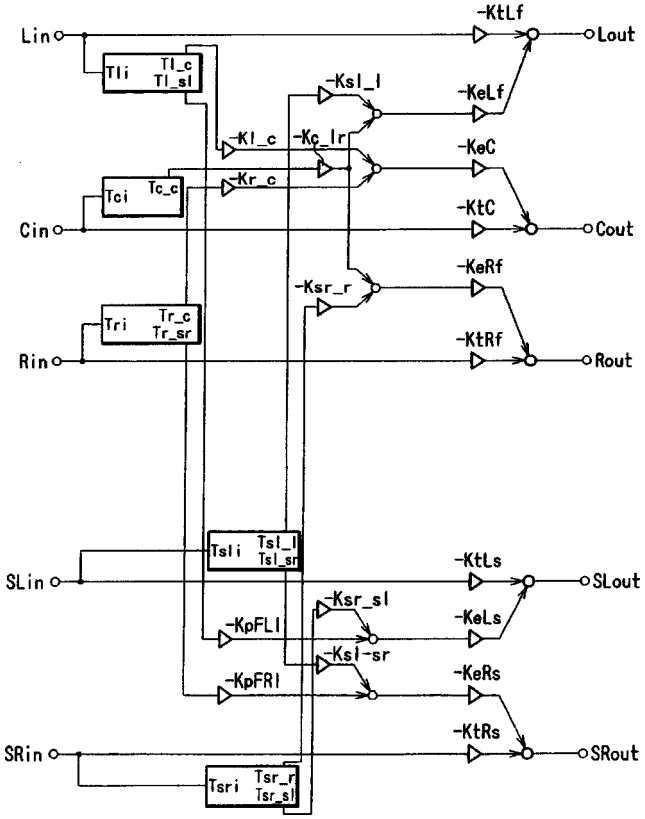
This patent relates to the audio interconnection between two or more spaces which will enable parties in these spaces to be continuously in contact, yet remain ''in the background'' when that is appropriate. There is a human dimension to the approach. For example, the usual telephone link between spaces, whether in the home or in a business context, is something of an intrusion and normally allows only a one-to-one communication path. What is proposed here is a system in which several microphones in one space communicate with one or more other spaces so that there is a clear mutual sensing among them. As the abstract states, ''a central server processes the summed audio signals and in turn transmits playable audio signals to each separate space... . This creates an audio space shared by each of the active separate physical spaces.'' The patent further describes relevant hardware and signal processing for refining these aims.—JME

6,754,352

43.38.Vk SOUND FIELD PRODUCTION APPARATUS

Takashi Kanai *et al.*, assignors to Sony Corporation
22 June 2004 (Class 381/63); filed in Japan 27 December 1999

As if your discrete five-channel home playback system for movies did not already present enough in the way of directional and ambience cues, this patent adds a rich array of additional options, as shown in the figure, to ''recreate'' the additional cues that might be present in an actual motion



picture theater. Never mind the fact that today's cinemas (at least in the United States) tend to be fairly ''dry'' acoustic spaces, and that the five basic channels are intended by the film producers and ''sound designers'' to convey all necessary ambience and delay cues.—JME

6,754,350

43.38.Vk SURROUND REPRODUCING CIRCUIT

Yositugu Sugimoto, assignor to New Japan Radio Company, Ind.
22 June 2004 (Class 381/17); filed in Japan 22 March 2001

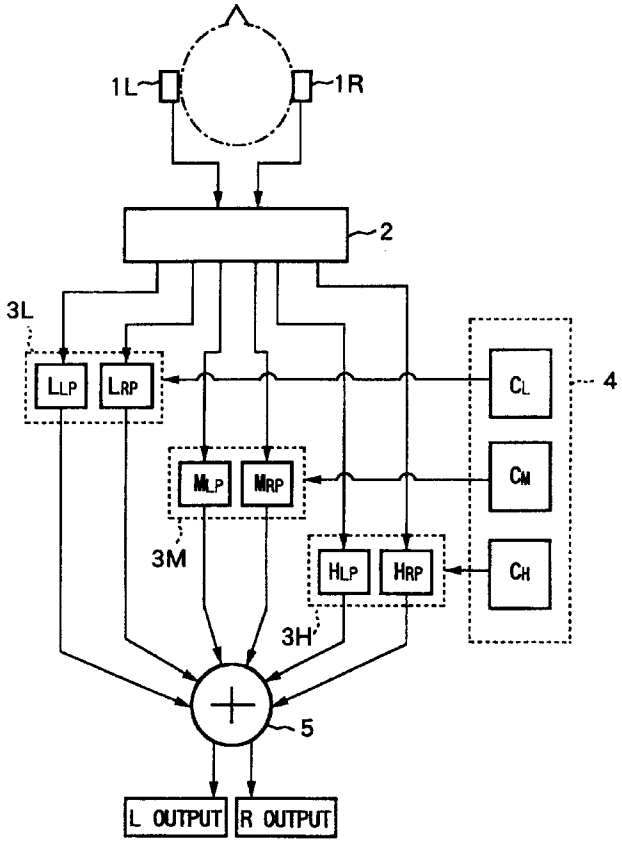
The patent deals with the expansion of two-channel stereo playback into a wide-front presentation for the listener. Using basic circuit elements such as equalizers, signal-adaptive gain, and all-pass networks, some of the effects of cross-talk cancellation over critical portions of the frequency range can be generated without the use of relatively expensive delay elements. The patent is clearly written, and the device takes its place in an already crowded field of audio signal processing for portable radios and television sets.—JME

6,763,115

43.38.Vk PROCESSING METHOD FOR LOCALIZATION OF ACOUSTIC IMAGE FOR AUDIO SIGNALS FOR THE LEFT AND RIGHT EARS

Wataru Kobayashi, assignor to OpenHeart Limited; A Limited Responsibility Company, Research Network
13 July 2004 (Class 381/309); filed in Japan 30 July 1998

The figure pretty much speaks for itself. A binaural recording head feeds a processor **2** that divides the joint program into three frequency bands, each pair of bands being operated on by a control signal. The dynamic behavior of the system is program directed, but does not relate directly to original cues in the initial pickup. As the abstract states, ''Then, the divided audio signal of each band is subjected to a processing for controlling



an element for a feeling of the direction of a sound and an element for feeling of the distance up to that sound source, which are appealed to person's auditory sense and outputting the processed audio signal.'' The reader gets the clear impression that the original English translation, presumably made in Japan, was put into the patent with no copy editing whatever.—JME

6,721,015

43.38.WI SOUND BITE AUGMENTATION

Henry C. Yuen *et al.*, assignors to E Guide, Incorporated
13 April 2004 (Class 348/460); filed 24 July 2002

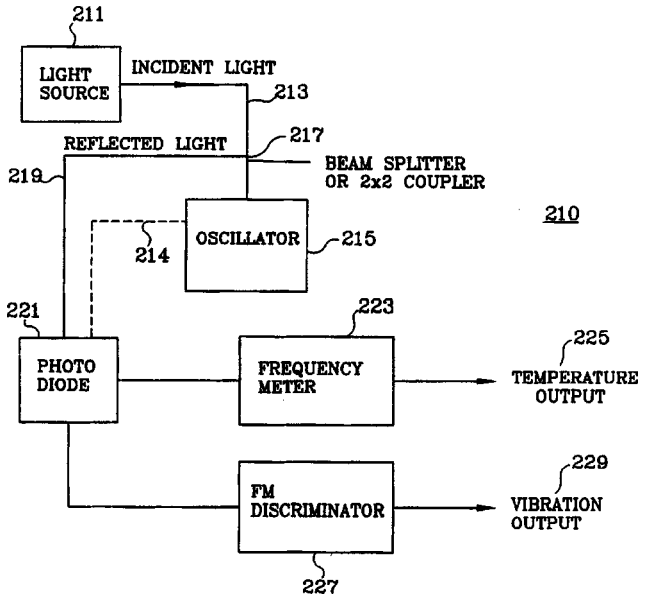
This patent describes the concept of a brief audio condensation of a television program, referred to here as a ''sound bite,'' which would be used to help locate the show while using an onscreen TV guide, or other similar applications. It is only a concept, because no method of producing the condensed clip is presented, nor even a suggestion of how long such a clip should be. The claims cover a variety of applications for the concept.—DLR

6,763,723

43.40.Yq FIBER-OPTIC VIBRATION SENSOR BASED ON FREQUENCY MODULATION OF LIGHT-EXCITED OSCILLATORS

J. David Zook *et al.*, assignors to Honeywell International Incorporated
20 July 2004 (Class 73/657); filed 25 October 2000

The patent relates to a vibration sensor that makes use of FM technology whereby the device uses a strain-sensitive oscillator in which resonant membranes are driven and sensed by a single multimode optical filter. The device includes a laser, an optical fiber for transmitting the laser signal, and an oscillator that receives the laser signal and reflects it as a frequency-



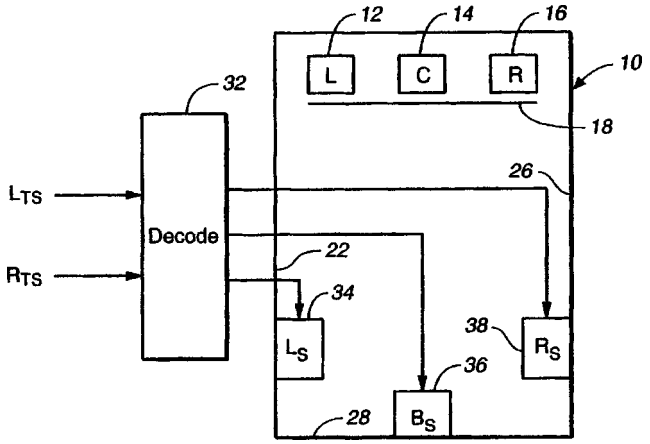
modulated beam. A second optical filter transmits the FM beam and a discriminator receives the FM signal and produces a signal responsive to vibration of the oscillator. In the preferred embodiment, the discriminator also includes a frequency meter to indicate the average number of cycles per unit time, thus providing a second signal indicative of the temperature of the oscillator.—DRR

6,760,448

43.55.Jz COMPATIBLE MATRIX-ENCODED SURROUND-SOUND CHANNELS IN A DISCRETE DIGITAL SOUND FORMAT

Kenneth James Gundry, assignor to Dolby Laboratories Licensing Corporation
6 July 2004 (Class 381/23); filed 5 February 1999

The number of channels in many motion picture soundtracks is now 5.1, where there are five full bandwidth channels and one ''0.1'' channel.



There are now schemes where 7.1 and 10.2 soundtracks are envisioned and even produced. The patent describes a brief and abridged history of channel encoding/decoding schemes, as well as describing a method of encoding three surround channels in a system originally designed for two surround channels.—NAS

6,758,305

43.55.Ti COMBINATION SOUND-DEADENING BOARD

Lawrence J. Gelin *et al.*, assignors to Johns Manville International, Incorporated
6 July 2004 (Class 181/285); filed 16 October 2001

The combination board comprises a structural skin (gypsum board or plywood, for example) and a sound-deadening material (a lightweight fiber board, for example). Bonded together, the combination board provides some resiliency for the structural skin, which in turn improves the transmission loss of a wall or floor/ceiling system. It is as easy as regular wallboard to apply.—CJR

6,751,485

43.58.Jq SOUNDING ALERT FOR RECENT CALLS

Jukka Ranta, assignor to Nokia Corporation
15 June 2004 (Class 455/567); filed 28 June 2001

Cache the last few phone numbers in a memory in your cell phone. If the phone is called by one of those numbers, then generate a different ringing tone. That's it. Nothing more, and certainly nothing less. If Amazon can patent the ''one click'' user interface, then surely Nokia can do this one.—MK

6,688,545

43.58.Wc FISHING-REEL SOUNDING DEVICE

Keigo Kitajima and Hirokazu Hiraoka, assignors to Shimano Incorporated
10 February 2004 (Class 242/306); filed in Japan 22 January 2002

Bicycle component followers know that Shimano makes many different parts. This patent applies their mechanical know-how to fishing reel design, specifically, how to design a gear-pawl mechanism so that the sound generated when the reel is under line pressure is uniform regardless of the force direction. Detailed mechanical drawings are included.—MK

6,741,264

43.58.Wc METHOD OF GENERATING AN AUDIBLE INDICATION OF DATA STORED IN A DATABASE

Michael F. Lesser, assignor to GIFIC Corporation
25 May 2004 (Class 345/727); filed 11 May 1999

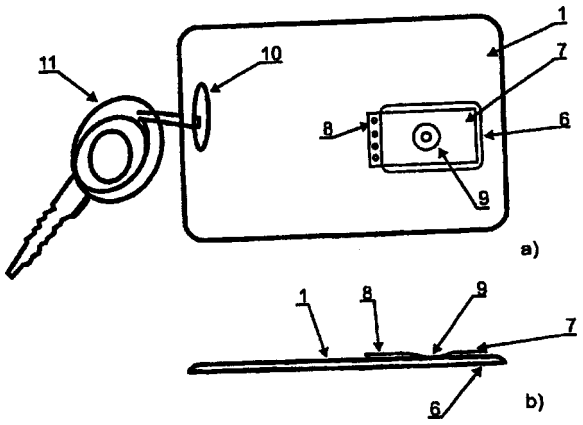
The idea is to generate audible indicators (called an earcon in the auditory display community) for various visual displays. The inventor proposes choosing sounds based on the circle of fifths—without any psychoacoustical justification.—MK

6,742,713

43.58.Wc ACCESS CONTROLLED SYSTEM WITH ACOUSTIC MECHANICAL VERIFICATION TRANSMITTER

Marian Trinkel, assignor to Deutsche Telekom AG
1 June 2004 (Class 235/492); filed in Germany 4 June 1999

Magnetic card readers are susceptible to vandalism and weather, to say the least. This acoustic card uses a variety of techniques including slots,



bars, and holes to create a unique acoustical signature when the plate **9** is deformed and released. The inventor says nothing at all about how well this works in practice, or how to design a card to fit a particular code.—MK

6,750,758

43.58.Wc REMOTELY CONTROLLED BEEPER AND METHOD

Timothy T. Duncan *et al.*, assignors to Tri-Tronics, Incorporated
15 June 2004 (Class 340/384.1); filed 27 December 2001

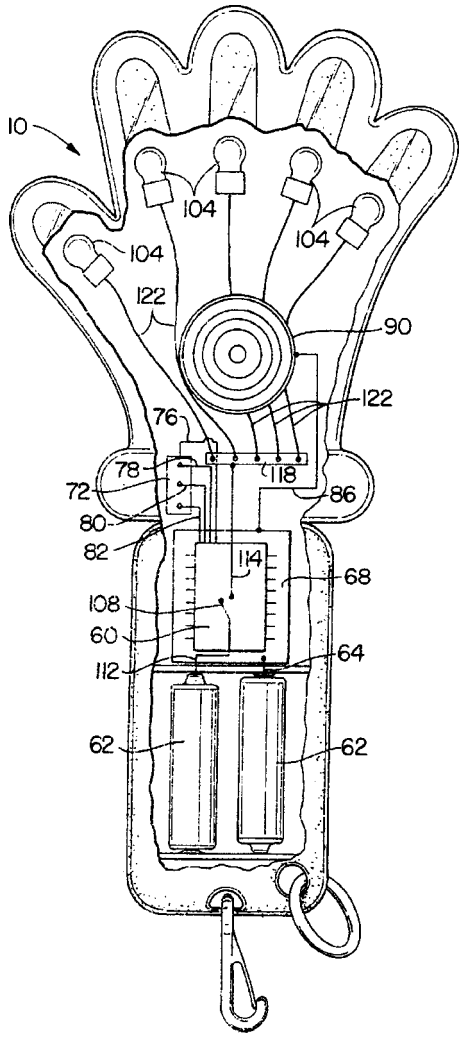
Imagine the following scene: you are accompanying your trusty hunting dog to obtain a fresh wild duck for dinner. The dog goes into the tall grass where s/he cannot be seen and only heard. The collar beeps as long as the dog moves. Finally, the prey are spotted, floating by. Freezing, the dog ''points.'' The collar recognizes the absence of movement and emits a hawk screech via piezoelectic speaker. The ducks take flight, where now it's your turn.—MK

6,538,565

43.58.Wc APPLAUSE DEVICE

Bradley L. Gotfried, Hobe Sound, Florida
25 March 2003 (Class 340/384.3); filed 19 July 2000

Showing approval and disapproval can be tiring—clapping and booing



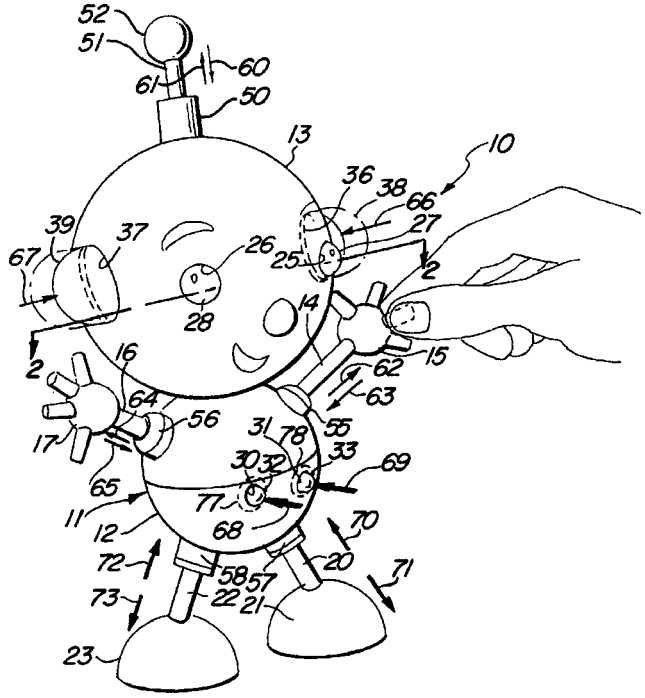
take physical energy. Why not use a battery-controlled device instead? Takes all the fun out of it, doesn't it.—MK

6,461,217

43.58.Wc TALKING DOLL HAVING EXTENDIBLE APPENDAGES

Danutcha Pestonji, assignor to Mattel, Incorporated
8 October 2002 (Class 446/297); filed 4 August 2000

A doll with extendible appendages can also include the now-



ubiquitous sound chip for generating various sound effects.—MK

6,760,447

43.60.Dh SOUND RECORDING AND REPRODUCTION SYSTEMS

Philip Arthur Nelson *et al.*, assignors to Adaptive Audio Limited
6 July 2004 (Class 381/17); filed in the United Kingdom 16 February 1996

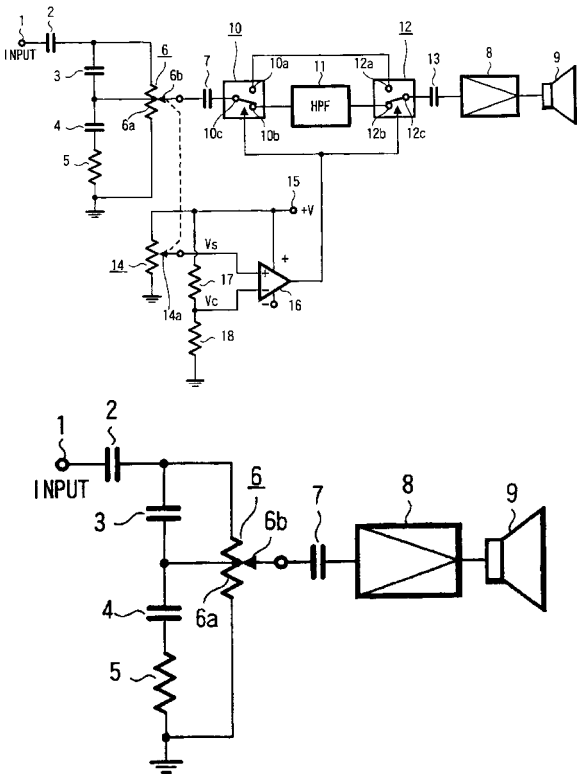
Clocking in at 59 pages, this patent may be the most detailed description yet given of what Cooper and Bauck referred to as ''transaural'' sound a quarter of a century ago (JAES 37,1/2). It discusses virtually all aspects of cross-talk cancellation and HRTFs in creating an accurate binaural listening experience via a pair of two closely spaced loudspeakers, and should be required reading for anyone involved in the field.—JME

6,763,113

43.60.Dh SPEAKER DRIVING CIRCUIT

Yoshimichi Maejima, assignor to Sony Corporation
13 July 2004 (Class 381/109); filed in Japan 13 March 2000

The modification of a signal to compensate for decreased human perception of low- and high-frequency sound at low listening levels has been known for sometime and is called loudness compensation. This type of control is still found in many audio preamplifiers, integrated receivers, and



similar equipment. The patent describes a method of enhancement to some existing methods of changing the low- and high-frequency equalization as a function of level.—NAS

6,760,449

43.60.Fg MICROPHONE ARRAY SYSTEM

Naoshi Matsuo, assignor to Fujitsu Limited
6 July 2004 (Class 381/92); filed in Japan 28 October 1998

Almost any three-dimensional microphone array can be used to identify the directional bearing of an arbitrary signal source in space. The art is in the refining of the array so that it uses the minimum number of elements to provide the greatest acuity with the highest signal-to-noise ratio. This well-written patent examines these problems and their solutions in detail.—JME

6,760,450

43.60.Fg MICROPHONE ARRAY APPARATUS

Naoshi Matsuo, assignor to Fujitsu Limited
6 July 2004 (Class 381/92); filed in Japan 26 June 1997

A divisional part of United States Patent 6,317,501 [reviewed in J. Acoust. Soc. Am. **112**(1), 15 (2002)], this very complex patent describes an array of omnidirectional microphones intended to detect a sound source with both accuracy and high signal-to-noise ratio. The specific aim is to exceed the performance of other array approaches in the areas of noise, speed, and accuracy of localizing a sound source. It is very complex, and the reading makes for very difficult going, due primarily to its ''grammatical density.''—JME

6,755,083

43.60.Tj METHOD FOR DISTINGUISHING MULTIPLE TARGETS USING TIME-REVERSAL ACOUSTICS

James G. Berryman, assignor to The Regents of the University of California
29 June 2004 (Class 73/602); filed 22 April 2002

This heady four-page patent has no figures and is expressed largely in conceptual terms. It deals with, among other things, a method of locating anomalies in an extended, random three-dimensional field. In the example given, ''an acoustic array sends a signal into a medium, and then receives the returned/reflected signal. This ... signal is then time-reversed and sent back to the medium again, and again, until the signal being sent and received is no longer changing. At that point, the array has isolated the largest eigenvalue/eigenvector combination and has effectively determined the location of a single target in the medium (the one that is most strongly reflecting).''—JME

6,766,028

43.66.Qp HEADTRACKED PROCESSING FOR HEADTRACKED PLAYBACK OF AUDIO SIGNALS

Glenn Norman Dickens, assignor to Lake Technology Limited
20 July 2004 (Class 381/310); filed in Australia 31 March 1998

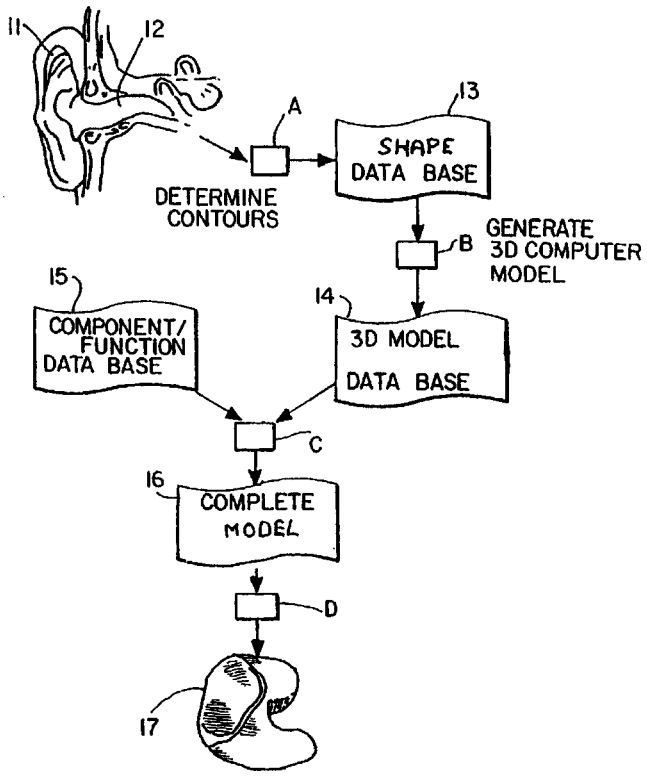
Traditionally, headtracking for use with binaural headphones has required intensive HRTF lookup data, and in some cases operation of the systems has suffered from lag. This well-written, fairly short patent proposes that the spatial reference be modeled on a number of previously determined loudspeaker positions. As the listener's head is moved, the required coefficients are calculated as differences from the pre-existing head position. As such, the lookup process is more efficient. The technique as described is aimed at performance with entertainment systems, specifically Dolby Digital, with its five symmetrically arrayed discrete channels.—JME

6,748,093

43.66.Ts HEARING AID WITH A TIGHTENING RING

Jan Tøpholm, assignor to Tøpholm & Westermann APS
8 June 2004 (Class 381/322); filed in the European Patent Office 26 March 2001

Described is a method of automating the manufacture of custom in-the-ear hearing aids. The shape of the auditory canal of an individual is acquired and converted to a 3D computer model of a custom hearing-aid shell for that ear canal. A faceplate model is formulated automatically to fit



with the shell model, forming a complete hearing-aid housing model that is used for manufacturing the hearing aid under computer control. The outer surface of the shell model may be increased selectively to form a rib for improving the acoustic seal.—DAP

6,754,336

43.66.Ts TELEPHONE APPARATUS WITH HEARING ASSISTANCE

Shinichiro Ito, assignor to Matsushita Electric Industrial Company, Limited
22 June 2004 (Class 379/353); filed in Japan 27 August 1996

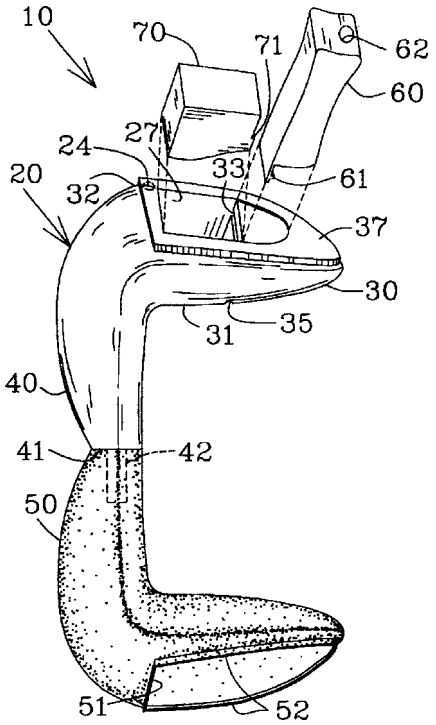
From the rather nebulous description given in the patent document, this device appears to be nothing more than a system whereby the receiving portion receives frequency-converted signals and turns them into voice signals that are preadjusted into a format especially suitable for the needs of the person at the receiving end.—DRR

6,754,357

43.66.Ts EXPANDABLE IN-EAR DEVICE

Ian McIntosh and Roger Saulce, assignors to Sonomax Hearing HealthCare Incorporated
22 June 2004 (Class 381/322); filed 21 January 2003

A method is outlined for using a settable compound to form an expandable in-ear device that is customizable *in situ* to the shape of an individual ear canal. The device contains a removal handle and may function as an ear plug or hearing aid for either left or right ears.—DAP

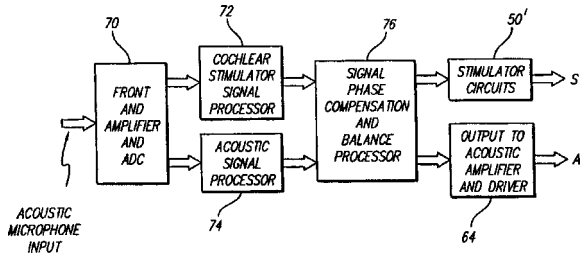


6,754,537

43.66.Ts HYBRID IMPLANTABLE COCHLEAR STIMULATOR HEARING AID SYSTEM

William Vanbrooks Harrison *et al.*, assignors to Advanced Bionics Corporation
22 June 2004 (Class 607/57); filed 13 November 2001

A combination hearing-aid system provides direct stimulation of high-frequency auditory neurons via a cochlear implant with a short electrode



array and low-frequency acoustic amplification. The patent states that the device can also provide tinnitus suppression.—DAP

6,760,457

43.66.Ts AUTOMATIC TELEPHONE SWITCH FOR HEARING AID

Mark A. Bren and Timothy S. Peterson, assignors to Micro Ear Technology, Incorporated
6 July 2004 (Class 381/331); filed 11 September 2000

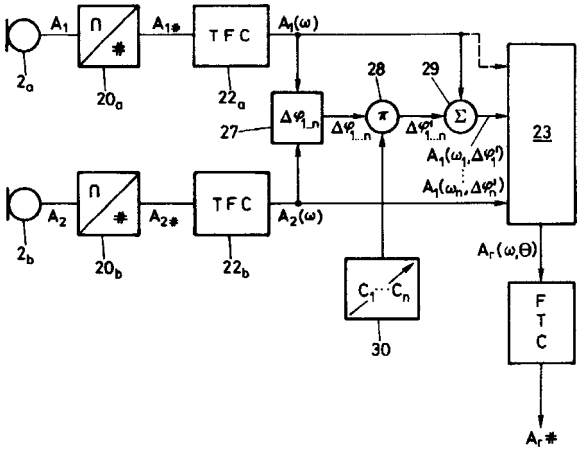
A hearing-aid input is automatically switched from microphone input to a telecoil input by sensing the magnetic field from a telephone handset when it is brought into close proximity to the ear. A switching network is recommended to connect and disconnect the microphone and telecoil.—DAP

6,766,029

43.66.Ts METHOD FOR ELECTRONICALLY SELECTING THE DEPENDENCY OF AN OUTPUT SIGNAL FROM THE SPATIAL ANGLE OF ACOUSTIC SIGNAL IMPINGEMENT AND HEARING AID APPARATUS

Joseph Maisano, assignor to Phonak AG
20 July 2004 (Class 381/313); filed 3 September 1998

An acoustical beamformer has at least two input transducers spaced at



a predetermined distance. The outputs of the transducers have spatial weights so as to selectively amplify sources from desired locations.—DAP

6,760,408

43.66.Ts SYSTEMS AND METHODS FOR PROVIDING A USER-FRIENDLY COMPUTING ENVIRONMENT FOR THE HEARING IMPAIRED

David Crosson and Mike Vermilyea, assignors to Cingular Wireless, LLC
6 July 2004 (Class 379/52); filed 3 October 2002

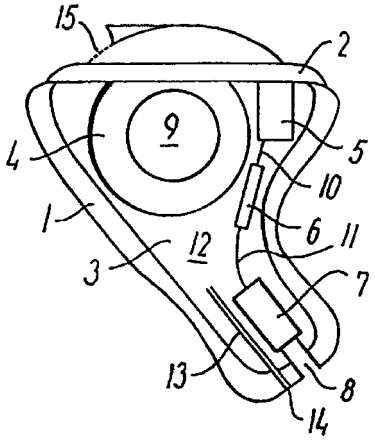
This is a method for signed instructions or animated prompting in lieu of spoken instructions or other means of audio prompting for hearing-challenged users of computer programs. The object of the method is to provide a user-friendly computing environment for the hearing impaired by supplying a program or help menu ''patch'' that provides signed instructions or responses to queries. An existing computer program contains information that is translated into a format usable by the hearing-impaired user. Desired program information is retrieved and translated into animation, digital movie, or still image formatted to be displayed on a monitor. A hyperlink is associated with the translated data and is stored in a memory device.—DRR

6,766,031

43.66.Ts IN-THE-EAR HEARING AID WITH REDUCED OCCLUSION EFFECT AND A METHOD FOR THE PRODUCTION AND USER-FITTING OF SUCH A HEARING AID

Soren Erik Westermann, assignor to Widex A/S
20 July 2004 (Class 381/328); filed in Denmark 15 April 1997

The sensation of occlusion and unnatural sound of the hearing-aid wearer's own voice often results from placing a custom hearing aid in the ear. A ventilation tube is inserted into an in-the-ear hearing aid on the side facing the interior of the wearer's ear canal. The tube forms a Helmholtz



resonator with the residual cavity volume inside the hearing-aid housing, resulting in a notch filter response in the frequency range at which excess energy build-up in the ear canal causes the occlusion effect.—DAP

6,726,636

43.70.Dn BREATHALYZER WITH VOICE RECOGNITION

Viken Der Ghazarian and Ohanes Der Ghazarian, assignors to Loran Technologies, Incorporated
27 April 2004 (Class 600/532); filed 28 March 2001

The prior technology for the detection of human alcohol consumption, as discussed here, includes breath analysis, blood sample analysis, and various methods of assessing motor skills. This device, intended for use in an automobile to prevent operation by an inebriated driver, tests some of these conditions, but also performs a speech recognition test to detect slurring. Given that the system could be thoroughly trained with the user's normal voice, it might just work.—DLR

6,726,486

43.71.Es METHOD AND APPARATUS FOR AUTOMATED TRAINING OF LANGUAGE LEARNING SKILLS

Elizabeth H. Budra *et al.*, assignors to Scientific Learning Corporation
27 April 2004 (Class 434/169); filed 26 September 2001

This patent deals with learning and speech perception, but does not involve the assignee's frequently patented formant modification technology. Here, a number of words are presented on a computer screen, as well as audibly, and the student places them into one of two or more categories based on any of semantic, syntactic, phonological, and morphological criteria. The nearly 200-page patent includes details for programming a large number of such games.—DLR

6,570,991

43.72.Ar MULTI-FEATURE SPEECH/MUSIC DISCRIMINATION SYSTEM

Eric D. Scheirer and Malcolm Slaney, assignors to Interval Research Corporation
27 May 2003 (Class 381/110); filed 18 December 1996

Scheirer and Slaney's music/speech discriminator is well known—and here is the patent. The idea is simple enough: Use 13 features as the input to a classifier. Three different classifiers are discussed in the patent and compared: Gaussian, nearest neighbor, and k-d trees. The patent is easy to read—and it only took 7 years from the filing date.—MK

6,430,529

43.72.Gy SYSTEM AND METHOD FOR EFFICIENT TIME-DOMAIN ALIASING CANCELLATION

Shay-Jan Huang, assignor to Sony Corporation; Sony Electronics Incorporated
6 August 2002 (Class 704/229); filed 26 February 1999

The Dolby AC-3 audio coder uses time domain alias cancellation (TDAC). The TDAC algorithm given by Dolby in the specification is not optimized for real-time performance. This patent describes specific optimizations and, what's more, real C code is included.—MK

6,725,199

43.72.Ja SPEECH SYNTHESIS APPARATUS AND SELECTION METHOD

Paul St John Brittan and Roger Cecil Ferry Tucker, assignors to Hewlett-Packard Development Company, L.P.
20 April 2004 (Class 704/258); filed in the United Kingdom 4 June 2001

This is a speech synthesizer intended for use in a complex environment involving a recognizer with elaborate natural language capabilities, the entire system operating under the control of a dialogue manager. As speech synthesis proceeds, an analysis of the output is performed to monitor the syntactic and semantic coherence and relevance of the output to the ongoing state of the dialogue manager. Depending on the results, any of several speech synthesizer strategies may be selected.—DLR

6,724,866

43.72.Ne DIALOGUE DEVICE FOR CALL SCREENING AND CLASSIFICATION

Roland Kuhn *et al.*, assignors to Matsushita Electric Industrial Company, Limited
20 April 2004 (Class 379/88.21); filed 8 February 2002

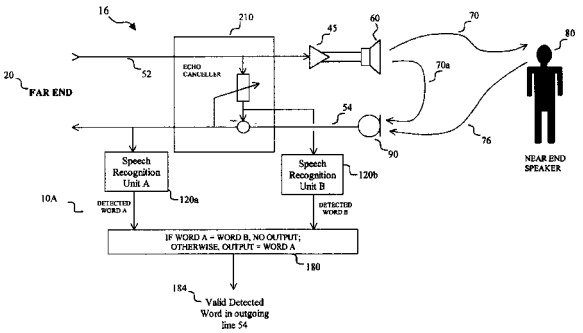
This patent discusses some of the mechanics underlying the use of a speech recognition system for the initial interaction with a telephone caller, particularly in the case of a telemarketing call. Noting the limitations of a caller ID system for this purpose, the ''better way'' devised here is for the system to learn a speaker model for all ''acceptable'' callers. In an automatic mode, the system saves a new speaker model whenever the user talks to that caller longer than a preset time limit. The user may also instruct the system specifically to save, or not save, a certain caller's speaker model. There is no discussion of specific recognition technology.—DLR

6,725,193

43.72.Ne CANCELLATION OF LOUDSPEAKER WORDS IN SPEECH RECOGNITION

Thomas J. Makovicka, assignor to Telefonaktiebolaget LM Ericsson
20 April 2004 (Class 704/233); filed 13 September 2000

On the face of it, this patent seems to be based on a rather preposterous notion, that a speech recognizer could function as an echo canceller. Such a system would be used together with a typical signal-based canceller and is intended to ''clean up'' any residual echo signal which gets through the traditional system. Both the incoming and outgoing streams of speech are processed by the recognizer. If the same word occurs in both streams with the appropriate timing, that word is assumed to have been picked up from a



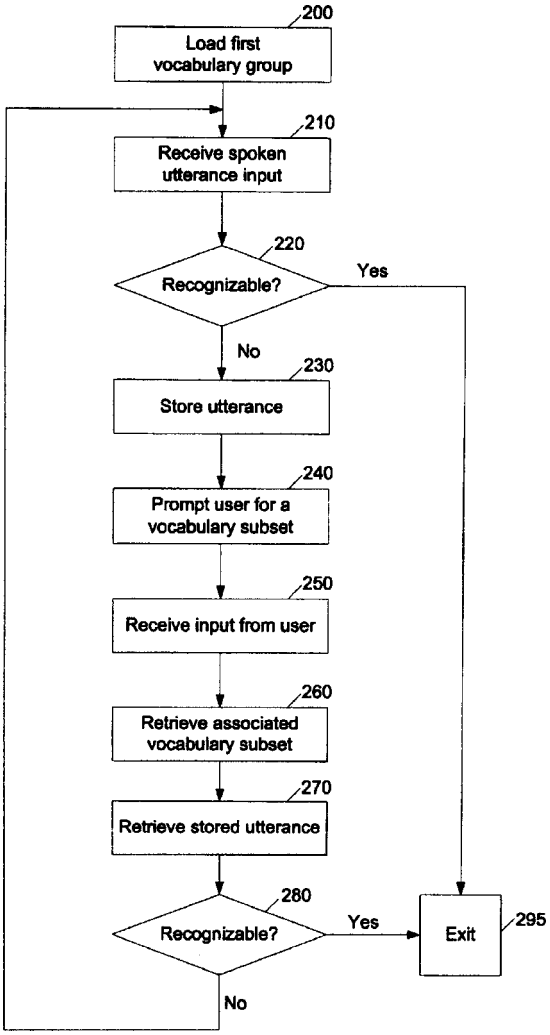
line echo. The system output (but not what the user hears, according to the figure) is the output of a comparator, which is turned off when the same word occurs in both streams. It is not said how this output is to be used.—DLR

6,751,595

43.72.Ne MULTI-STAGE LARGE VOCABULARY SPEECH RECOGNITION SYSTEM AND METHOD

Senis Busayapongchai and Pichet Chintrakulchai, assignors to BellSouth Intellectual Property Corporation
15 June 2004 (Class 704/275); filed 9 May 2001

To reduce processing time and improve accuracy of speech recognition, the total vocabulary is divided up into smaller vocabulary subsets



which correlate with certain keywords. The user provides a spoken input and the utterance is matched to the appropriate vocabulary subset.—DAP

6,757,655

43.72.Ne METHOD OF SPEECH RECOGNITION

Stefan Besling *et al.*, assignors to Koninklijke Philips Electronics N.V.
29 June 2004 (Class 704/270.1); filed in Germany 9 March 1999

Some client-server speech recognition systems utilizing stored speech inputs require excessively long processing times due in part to the limited processing power of the client. A methodology is recommended here to speed up processing time by performing speech recognition on a speech recognizer connected to the client through a communications network rather than by the client itself.—DAP

6,689,946

43.75.Bc AID FOR COMPOSING WORDS OF SONG

Tomoyuki Funaki, assignor to Yamaha Corporation
10 February 2004 (Class 84/478); filed in Japan 25 April 2000

The inventor proposes a user interface design to assist the construction (composition) of songs. Specifically, the music is presented piano-roll style while words are composed below. The editor offers a list of rhyming words and grammatical rules (free-verse poets need not apply).—MK

6,740,802

43.75.Bc INSTANT MUSICIAN, RECORDING ARTIST AND COMPOSER

Bernard H. Browne, Jr., Clearwater, Florida
25 May 2004 (Class 84/609); filed 6 September 2000

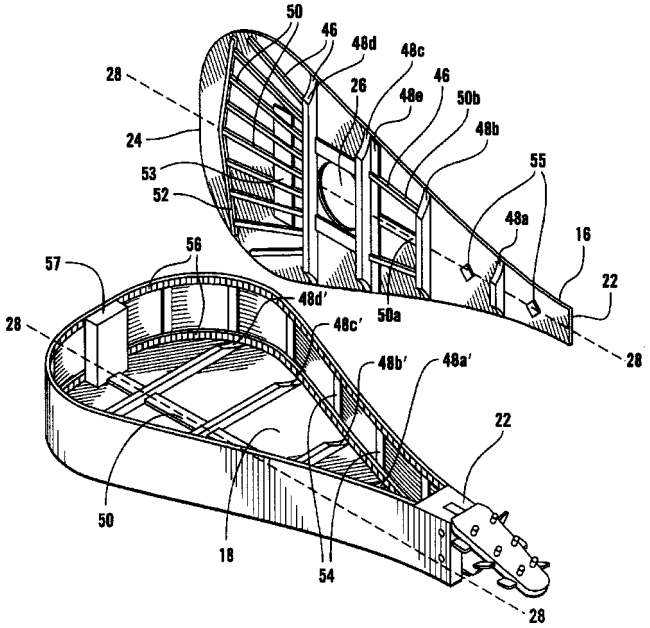
This inventor proposes a general all-in-one, all purpose, ''slices and dices'' music composition program for the naive user. Everything proposed here has been implemented before by others. If anything could possibly be novel, it would be the user interface. But, even that tests the obvious. Algorithmic details are notably absent.—MK

6,693,233

43.75.Gh NECKLESS LAP GUITAR

David L. Sewell, Madison, Wisconsin
17 February 2004 (Class 84/291); filed 3 March 2003

The inventor proposes a novel neckless teardrop-shaped ''lap guitar.''



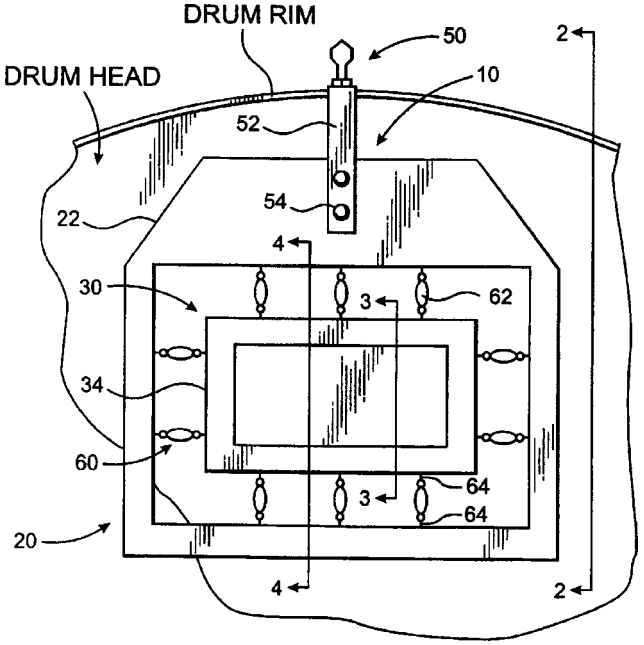
Avoiding the neck–body joint affords a certain sturdiness. The references to past guitar construction are interesting.—MK

6,586,664

43.75.Hi DRUM MUTING SYSTEM

Francis Park Hubbell, Miami, Florida
1 July 2003 (Class 84/411 M); filed 25 June 2001

There are many reasons to mute a vibrating drum head including respect for your neighbors, small recording studio room acoustics, and not frightening small children. This inventor proposes an externally attached muffler that damps the head via an inner frame **30**. The damper resembles a



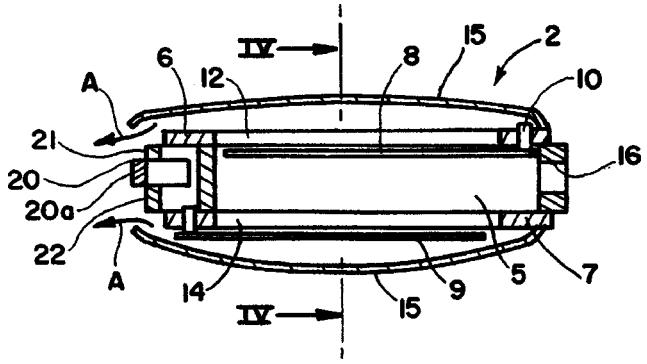
piano damper with the point positioned to soften the drum membrane on large excursions. The frame design is rather arbitrary: What would happen if it was designed to match the well-described vibratory modes of the drum head?—MK

6,462,261

43.75.Pq HARMONICAS

Richard Smith, Portway, Burghill Heresford, the United Kingdom
8 October 2002 (Class 84/377); filed in the United Kingdom 3 November 1998

To amplify a harmonica, the player holds a microphone close to the rear of the instrument. This can amplify breath noise and feedback from the



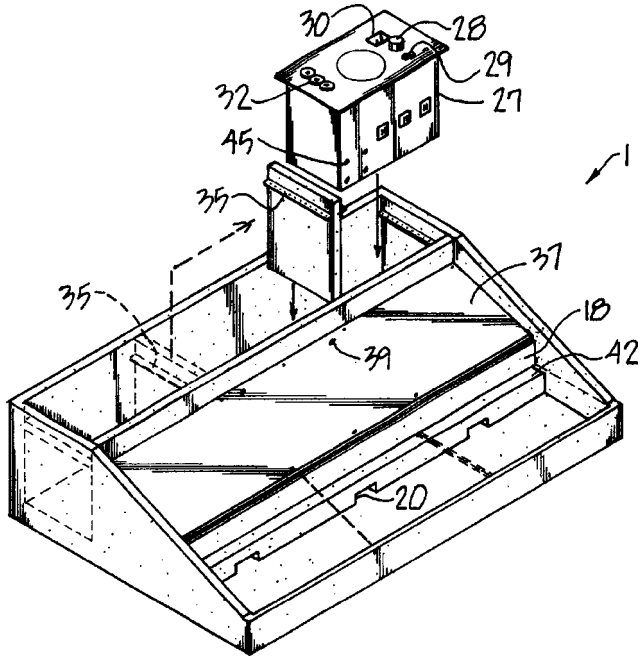
amplification system. The inventor proposes adding a shell **15** and a small (undoubtedly electret) microphone **20**. The acoustic effect of creating this small cavity was not contemplated.—MK

6,538,185

43.75.Tv PEDAL BOARD ASSEMBLY

Michael K. Stratton, assignor to MKS Professional Stage Products, Incorporated
25 March 2003 (Class 84/422.1); filed 10 October 2001

For more than 40 years, electronic test instruments have featured ''plug-ins,'' whereby a common power supply (and often CRT) are shared



by many instruments. Now, think of creating an effects box with a common backplane and you have exactly the same idea.—MK

6,538,189

43.75.Tv WIND CONTROLLER FOR MUSIC SYNTHESIZERS

Russell A. Ethington, Irvington, New York
25 March 2003 (Class 84/615); filed 1 February 2002

Given the MIDI clock rate (31.25 kB/s=32-μs interval), key transitions on a wind controller (like the well-known WX-7) can result in glitches. The inventor proposes using two switches per key to produce a ''trit'' per key. These are interpreted by a microcontroller (and converted to MIDI). As a bonus, the microcontroller can also offer chords to the chord-hungry wind player.—MK

6,434,520

43.75.Wx SYSTEM AND METHOD FOR INDEXING AND QUERYING AUDIO ARCHIVES

Dimitri Kanevsky and Stephane H. Maes, assignors to International Business Machines Corporation
13 August 2002 (Class 704/243); filed 16 April 1999

The inventors sketch the high-level design of an audio retrieval system. Data are assigned features and segmented appropriately. This is further used to index the data in the database. The patent writing is clear but intentionally lacking in details and specifics.—MK

6,570,081

43.75.Wx METHOD AND APPARATUS FOR EDITING PERFORMANCE DATA USING ICONS OF MUSICAL SYMBOLS

Hideo Suzuki and Masao Sakama, assignors to Yamaha Corporation
27 May 2003 (Class 84/622); filed in Japan 21 September 1999

Yamaha describes how to accept raw MIDI data and convert them to musical ''icons.'' The choice of the word icon is poor—''figure'' would be a better choice. Essentially, the inventors use an extremely simplistic form of pattern matching to do the dirty work of converting the MIDI stream into another stream with different figures.—MK

6,757,574

43.80.Nd METHODS AND APPARATUS FOR PRODUCING ANIMAL SOUNDS TO LURE ANIMALS

William James Gardner and Tommy J. Martin, assignors to G&B Patent Holdings, LLC
29 June 2004 (Class 700/94); filed 14 January 2000

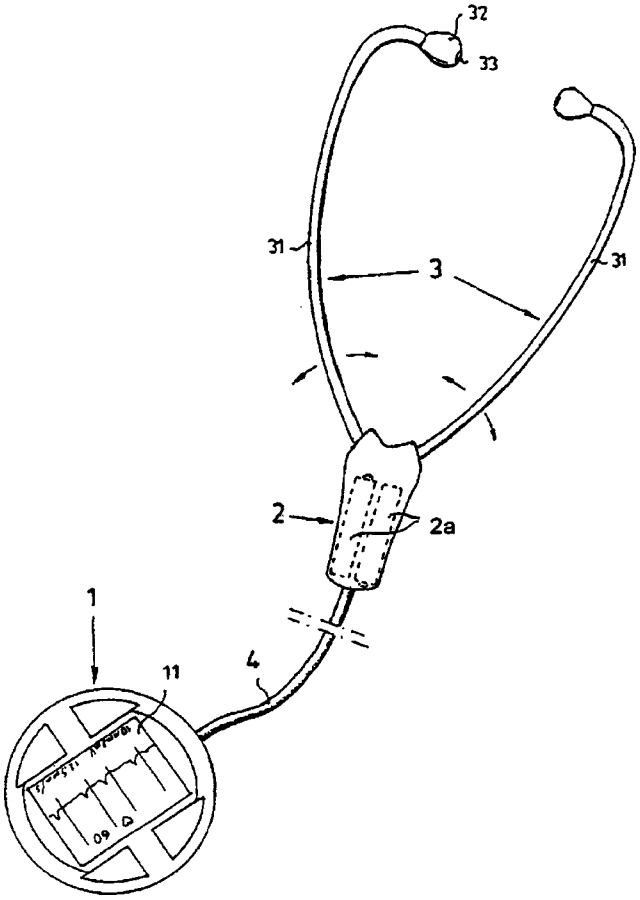
The methodologies and the associated apparatus (which is nothing more than a recorded sound playback system) purport to be the electronic equivalent of the hunter's moose call or means of attracting other animals of specific species by mimicking their sounds. In one embodiment, a prerecorded sound that simulates an environmental contact sound made by the animal species in the environment is played back in order to lure members of the species to the area. In a second embodiment, two separate prerecorded sounds simulating the sounds made by the species are played back simultaneously.—DRR

6,757,392

43.80.Qf ELECTRONIC STETHOSCOPE

Artemio Granzotto, Zürich and Fridolin Voegeli, Rotkreuz, both of Switzerland
29 June 2004 (Class 381/67); filed in Switzerland 6 July 1995

This is an electronic stethoscope that permits simple electrocardiographic investigations as well as the more customary auscultation. The device consists of a headpiece **1** that is placed on the patient, a chestpiece **2**, and two earpieces **3**. Several sensors are arranged around the microphone in the headpiece, at least three of which are electrodes arranged in an Einthoven Triangle format for the purpose of obtaining an electrocardiogram. A



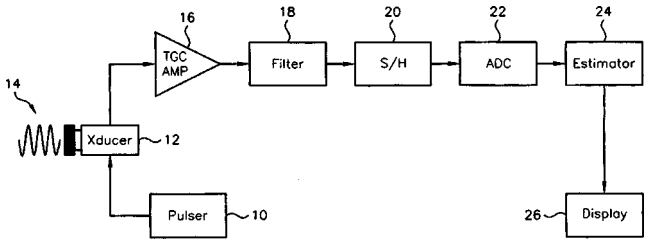
display is also incorporated in the headpiece for visual display of body function values gathered by the sensors, these values being collected by an evaluation unit and electronically processed for the display. The chestpiece contains rechargeable batteries, and the earpieces terminate in ''olives'' **32** containing microphones that generate audio signals constituting the auscultation.—DRR

6,758,815

43.80.Qf APPARATUS AND METHOD FOR INDICATING MECHANICAL STIFFNESS PROPERTIES OF BODY TISSUE

Richard Bruce Bernardi, Wayne, Pennsylvania
6 July 2004 (Class 600/437); filed 17 October 2001

Mechanical stiffness properties of body tissues are measured through the use of Doppler imaging techniques. Ultrasonic signals that are transmitted to the target not only are reflected to yield images of the target but, by



appropriate selection of their intensity, the body tissues under investigation are deformed or moved by the impinging signals. The deformation or movement of the body tissues is imaged and is therefore representative of the stiffness of the targeted body tissue.—DRR

6,761,693

43.80.Qf DEVICE AND METHOD FOR DETECTING OPENING OF PASSAGE IN BODILY CAVITY

Steen B. Rasmussen, assignor to Rhinometrics A/S
13 July 2004 (Class 600/462); filed in Denmark 2 November 1999

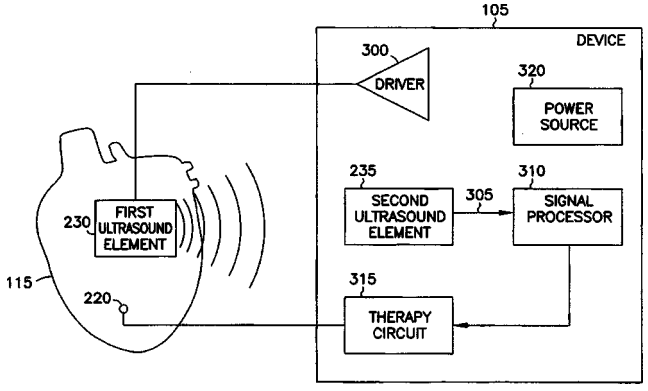
This device, designed to detect the opening of a passage in a body cavity, consists of an electrical signal source, a transducer that converts the electrical signal into an acoustic signal, and a second transducer to receive the acoustic signal. The two transducers are adapted to be placed in the body cavity at opposite sides of an obstruction. The degree to which a passage has opened up is indicated by the quality of the acoustic signal from the emitting transducer received by the receiving transducer.—DRR

6,757,563

43.80.Sh CARDIAC RHYTHM MANAGEMENT SYSTEM WITH ULTRASOUND FOR AUTOCAPTURE OR OTHER APPLICATIONS

Robert J. Sweeney, assignor to Cardiac Pacemakers, Incorporated
29 June 2004 (Class 607/28); filed 17 December 2002

The principal feature of this cardiac rhythm management system is that it provides ultrasound autocapture capability for determining whether a stimulation has evoked a desired response from the heart. It also adjusts the energy of the stimulation on the basis of the observed response from the heart. An ultrasound element is deployed on a lead inside the heart. Another ultrasound element is located elsewhere in the heart or in the implanted device. An autocapture determination circuit determines whether motion of



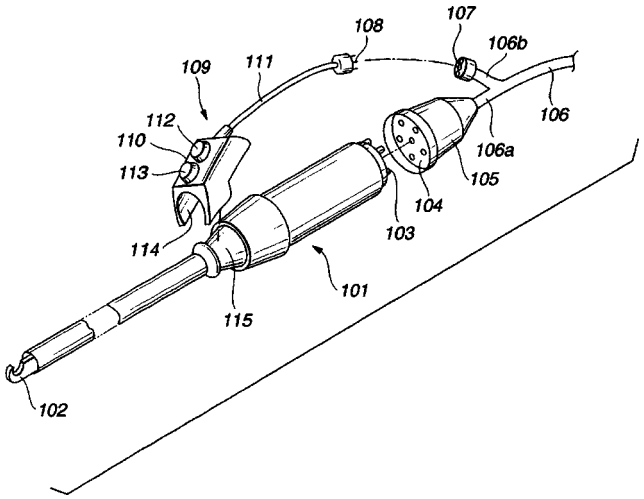
the heart chamber indicates a contraction in response to the stimulation and adjusts the stimulation energy to provide just enough energy needed to obtain capture (this saves energy and prolongs the life of the implanted device). Other applications include using ultrasound for (1) determining the strength of heart contractions; (2) determining dissociation between electrical and mechanical heart activity; (3) measuring the volume of the heart; (4) determining the origin of sensed intrinsic electrical heart activity signals; (5) detecting particular arrhythmias; (6) disrupting cell membranes for lowering stimulation thresholds; (7) controlling steroid delivery; and (8) obtaining blood flow information.—DRR

6,761,698

43.80.Sh ULTRASONIC OPERATION SYSTEM

Norikiyo Shibata *et al.*, assignors to Olympus Corporation
13 July 2004 (Class 601/2); filed in Japan 28 July 2000

The object of this ultrasonic surgical tool is to facilitate surgery through ultrasonic cutting by monitoring feedback from strategically deployed transducers. A main unit includes an ultrasonic vibrator and may be



connected by means of a detachable cable to any of several different transducers. The main unit receives an identification signal in an identification circuit, automatically controls drive signals through a control circuit, and automatically displays the identified instrument on a display.—DRR

6,764,450

43.80.Sh METHOD AND APPARATUS FOR INTRAVASCULAR TWO-DIMENSIONAL ULTRASONOGRAPHY

Paul G. Yock, assignor to SciMed Life Systems, Incorporated
20 July 2004 (Class 600/466); filed 13 May 2003

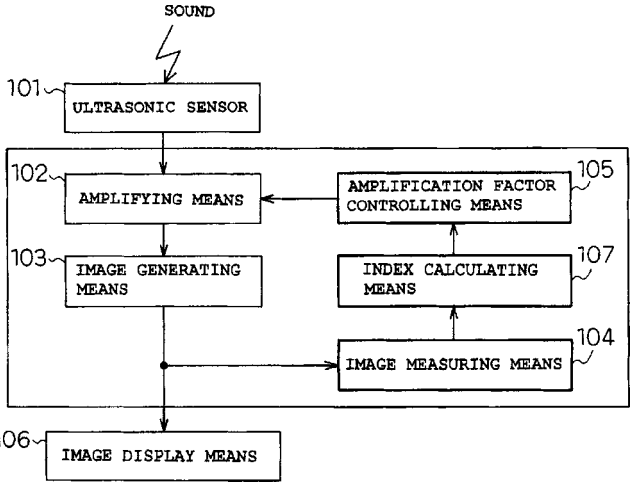
This catheter is designed for insertion into a blood vessel of a patient for ultrasonically imaging the vessel wall. The catheter includes an ultrasound emitter and features a tubular element and an internally housed drive cable for executing circumferential scan about the catheter. Both tubular element and drive cable are of a size and flexibility adapted to allow their introduction into the blood vessel and subsequent advancement through the vessel to the location of the vessel wall where imaging is to be executed.—DRR

6,755,785

43.80.Vj ULTRASONIC IMAGE GENERATING APPARATUS AND ULTRASONIC IMAGE GENERATING METHOD

Jun Omiya, assignor to Matsushita Electric Industrial Company, Limited
29 June 2004 (Class 600/443); filed in Japan 20 November 2001

Analysis of an image produced by this apparatus is used to determine where the ultrasonic beam is blocked by a relatively impenetrable region



and the direction of the beam is controlled to reduce the area over which the beam is blocked.—RCW

6,755,786

43.80.Vj ULTRASONIC DIAGNOSTIC IMAGING WITH ELEVATION BIPLANE IMAGES

Janice Frisa and McKee Dunn Poland, assignors to Koninklijke Philips Electronics N.V.
29 June 2004 (Class 600/443); filed 30 May 2003

Two planes at different elevations in a volume are scanned in real time. The planes may have corresponding depths separated by the same selectable distance in elevation. Alternatively, one plane may have a fixed orientation relative to the imaging probe and the location of the other plane can be selected.—RCW

6,755,787

43.80.Vj MEDICAL DIAGNOSTIC ULTRASOUND SYSTEM AND METHOD FOR VERSATILE PROCESSING

John A. Hossack *et al.*, assignors to Acuson Corporation
29 June 2004 (Class 600/447); filed 19 November 2002

A multirow or a two-dimensional transducer is used to obtain sequential, parallel, or related frames of image data in elevation. Two-dimensional images are obtained by compounding the frames. Three-dimensional images are interpolated using multiple two-dimensional frames spaced in elevation and compounded into a number of two-dimensional frames also spaced in elevation. An anisotropic filter is employed in this process.—RCW

6,752,762

43.80.Vj METHOD AND APPARATUS FOR ULTRASOUND CONTRAST IMAGING

Nico DeJong and Peter J. Frinking, assignors to Acuson Corporation
22 June 2004 (Class 600/458); filed in the United Kingdom 21 January 1999

Subharmonic acoustic signals produced by a fundamental frequency are received and focused to produce an image.—RCW

6,755,788

43.80.Vj IMAGE ORIENTATION DISPLAY FOR A THREE DIMENSIONAL ULTRASONIC IMAGING SYSTEM

Douglas Armand Demers and Mckee Dunn Poland, assignors to Koninklijke Philips Electronics N.V.
29 June 2004 (Class 600/447); filed 12 May 2003

A number of planes are scanned in real time and images of the planes are displayed. An icon on the display depicts the relative inclination of the planes. The icon includes a perspective view of one plane with the aspect ratio of the outline varying to show the instantaneous inclination of the planes.—RCW

6,755,789

43.80.Vj ULTRASONIC VASCULAR IMAGING SYSTEM AND METHOD OF BLOOD VESSEL CANNULATION

Bradley J. Stringer *et al.*, assignors to Inceptio Medical Technologies, LLC
29 June 2004 (Class 600/461); filed 5 February 2002

Two perpendicular linear arrays of transducers are used to obtain images of a blood vessel in corresponding perpendicular planes. Doppler beams at an angle beneath one of the transducer arrays are used to determine blood flow velocity. The assembly is enclosed by a protective sheath and secured to a marked cover that facilitates orientation of the assembly and guidance of a needle during a cannulation procedure.—RCW

6,761,689

43.80.Vj BIPLANE ULTRASONIC IMAGING

Ivan Salgo *et al.*, assignors to Koninklijke Philips Electronics N.V.
13 July 2004 (Class 600/447); filed 31 October 2002

One of two images in a volume has a fixed planar orientation relative to the transducer. The other image can be rotated or tilted relative to the fixed image. An icon is displayed with the images to show their relative orientation.—RCW

LETTERS TO THE EDITOR

This Letters section is for publishing (a) brief acoustical research or applied acoustical reports, (b) comments on articles or letters previously published in this Journal, and (c) a reply by the article author to criticism by the Letter author in (b). Extensive reports should be submitted as articles, not in a letter series. Letters are peer-reviewed on the same basis as articles, but usually require less review time before acceptance. Letters cannot exceed four printed pages (approximately 3000–4000 words) including figures, tables, references, and a required abstract of about 100 words.

An alternate approach to constructing distortion product otoacoustic emission (DPOAE) suppression tuning curves[a)]

Tiffany A. Johnson,[b)] Stephen T. Neely, Darcia M. Dierking, Brenda M. Hoover, and Michael P. Gorga
Boys Town National Research Hospital, Omaha, Nebraska 68131

(Received 22 June 2004; revised 14 September 2004; accepted 15 September 2004)

Distortion product otoacoustic emission (DPOAE) suppression tuning curves (STCs) typically are constructed using suppression criteria of 3 or 6 dB. This paper describes a technique for generating DPOAE STCs using criteria ranging from 3 dB to complete suppression. The criterion effect was examined at various primary levels ($f_2=4$ kHz) in normal-hearing ears. As expected, Q_{ERB} and tip-to-tail differences decreased as probe level (L_2) increased. Q_{ERB} values were not systematically dependent on criterion. For low L_2's, tip-to-tail differences decreased as criterion increased. Given similarities in methodology, DPOAE STCs constructed from completely suppressed responses might be most appropriate for comparison to psychophysical tuning curves. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1815134]

PACS numbers: 43.64.Jb, 43.64.Kc [BLM] Pages: 3263–3266

I. INTRODUCTION

Distortion product otoacoustic emissions (DPOAEs) have been measured using a suppression paradigm. Suppression tuning curves (STCs) constructed from these measures may be compared to tuning curves obtained from other methods, such as psychophysical masking experiments in humans. Because the slope of functions relating amount of suppression to suppressor level varies with suppressor frequency, the shape of the DPOAE STC constructed from these functions might depend on the suppression criterion used to construct it. DPOAE STCs frequently are constructed using suppression criteria of either 3 or 6 dB (e.g., Abdala, 2001; Abdala and Fitzgerald, 2003; Abdala *et al.*, 1996; Brown and Kemp, 1984; Gorga *et al.*, 2002; Gorga *et al.*, 2003; Martin *et al.*, 1998b, 2003). These criteria, however, represent only partial reduction of the DPOAE response and, as such, are not directly analogous to tuning curves generated from psychophysical masking experiments in which the probe is rendered inaudible. In this paper, we propose a method for generating DPOAE STCs, based on the fully suppressed condition, that may be more similar to the conditions from which psychophysical tuning curves are constructed. We also examine the influence of suppression criterion on the shape of the DPOAE STC as described by Q_{ERB} and the tip-to-tail difference.

II. METHODS

A. Subjects

Data from 22 ears of 20 subjects with normal hearing were included in the analyses presented here. Normal hearing was defined as thresholds ≤15 dB HL (re: ANSI, 1996) at 4 kHz. The mean threshold at 4 kHz was 5.7 dB HL.

B. Stimuli

Stimulus generation (and data collection) was accomplished using custom designed software (EMAV, Neely and Liu, 1993) that controlled a sound card (CardDeluxe, Digital Audio Labs) housed in a PC. The stimuli were delivered via loudspeakers housed in an ER-10C probe-microphone system. The probe frequency (f_2) was fixed at 4 kHz with a primary frequency ratio (f_2/f_1) of 1.22. The level of f_2 (L_2) varied in 10-dB steps from 20 to 70 dB SPL. The level of f_1 (L_1) was set according to the equation $L_1=0.4L_2+39$ dB (Kummer *et al.*, 1998). Seventeen suppressor frequencies (f_3) were used, varying from 1 octave below to $\frac{1}{2}$ octave above f_2. The level of the suppressor (L_3) was varied in 5-dB steps from 5 to either 85 or 90 dB SPL, depending on f_3.

[a)]Portions of this work were presented in "DPOAE suppression tuning curves: Effect of suppression criteria," 27th Midwinter Meeting of the Association for Research in Otolaryngology, Daytona Beach, FL, February, 2004.

[b)]Electronic mail: johnsonta@boystown.org

C. Data collection

For each L_2, DPOAE levels were measured first in response to the primaries presented alone (the control condition). Next, the primaries were presented in the presence of a fixed suppressor frequency (f_3) that varied in level. This process was repeated until the f_3-level sequence had been repeated for each of 17 f_3's and each of 6 L_2's.

D. Analysis

Following data collection, DPOAE levels were converted to decrements (amount of suppression) by subtracting the level measured during each suppression condition from the level measured during the control condition. Mean decrement-versus-L_3 functions for 3 f_3's (one below, one approximately equal to, and one above probe frequency, f_2) are shown as solid lines in the upper panel of Fig. 1. In this figure, $f_2=4$ kHz and $L_2=40$ dB SPL. Because we were interested in constructing DPOAE STCs for a range of suppression criteria, including the condition where the response is fully suppressed, the maximum amount of decrement was calculated for each L_2. The signal-to-noise ratio (SNR) in the control condition (28 dB at this L_2) defines the maximum measurable amount of suppression and was used to define the point at which the response was suppressed to the noise floor. In the top panel of Fig. 1, the upper horizontal line is equivalent to the control-condition SNR when $L_2=40$ dB SPL. The lower horizontal line in the same panel indicates the point at which the response had been decremented (suppressed) by 3 dB.

Before generating DPOAE STCs, the decrement-versus-L_3 functions were transformed according to

$$D=10\log(10^{\mathrm{decr}/10}-1). \tag{1}$$

When $D=0$, the decrement is 3 dB. This equation served to linearize the functions and allowed points with high SNRs (i.e., decrements between 1 and 3 dB) to be included in the generation of a linear model for the data. The transformed data were then fit with a linear equation with the constraint that only decrements $\geqslant 1$ dB were included in the fit and that there must be at least 3 L_3's producing decrements greater than 1 dB. Additionally, only functions including at least one condition for which at least 3 dB of suppression occurred were transformed with a linear equation. The r^2 values resulting from these fits ranged from 0.8 to >0.99 with 86% greater than 0.9. Those r^2 values falling below 0.9 were equally divided between low- and high-frequency suppressors. The linear equations were subsequently solved for decrements ranging from 3 dB to the fully suppressed condition (i.e., the point when the decrement equals the SNR in the control condition). DPOAE STCs were generated using these different suppression criteria.

These approaches to data analysis are also illustrated in Fig. 1. In the upper panel, the linear fits to the transformed data are shown as dashed lines. The open symbols represent 3-dB decrements (calculated from these linear fits) for the three suppressor frequencies shown in the figure. The closed symbols represent the corresponding complete-suppression condition. In the lower panel, these points are re-plotted as a function of frequency. The lines shown as tuning curves represent spline fits to these data, in combination with the data from the other 14 suppressor frequencies. The STCs that were generated using decrement criteria of both 3 dB and complete suppression are shown. It is important to note that, for some conditions (especially when $f_3>f_2$), it was not always possible to fully suppress the DPOAE response, even when $L_3=85$ or 90 dB SPL. However, using the above-described approach, it was possible to extrapolate to the L_3 that would be expected to fully suppress the response. In the lower panel, points above the horizontal, dashed line represent extrapolated values of L_3. The greatest difference in the STCs constructed from the 3-dB and fully suppressed criteria

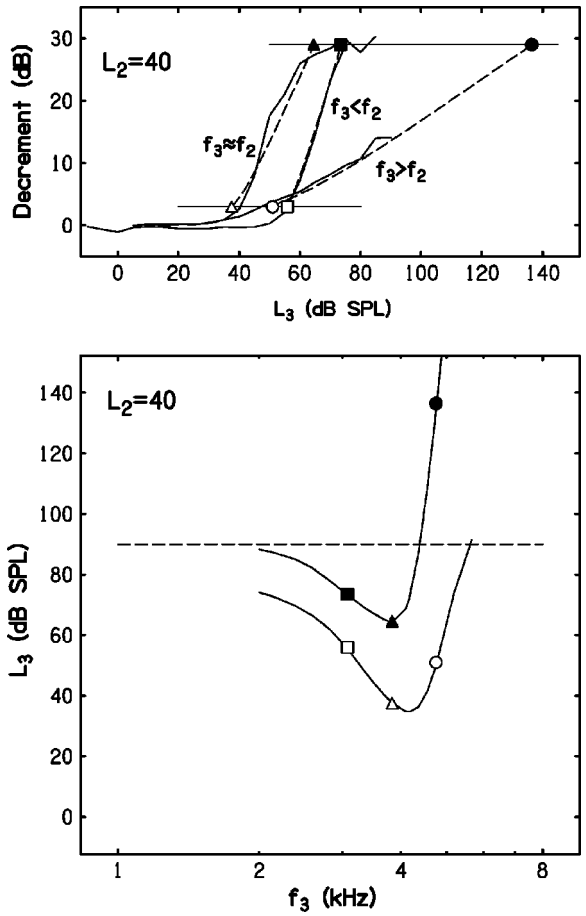


FIG. 1. Illustration of the approach to data analysis. In both panels, the probe frequency (f_2) = 4 kHz and the probe level (L_2) = 40 dB SPL. Upper panel: Mean decrement-vs-L_3 functions for 3 f_3's are shown as solid lines. The upper horizontal line represents the maximum amount of decrement and was determined by calculating the SNR in the control condition (28 dB in this example). This horizontal line, therefore, represents the point at which the response is fully suppressed into the noise floor. The lower horizontal line represents the point at which the response has been suppressed by 3 dB. The dashed lines represent the linear fits to the transformed data (as described in the text). The linear fits were solved for a range of decrements. The 3-dB decrement points are shown as open symbols for three different suppressor frequencies. The corresponding fully suppressed points are shown as closed symbols. Lower panel: Mean DPOAE STCs plotted as suppressor level (L_3) vs suppressor frequency (f_3) for two different suppression criteria, 3 dB (lower curve) and complete suppression (upper curve). The symbols plotted on the STCs correspond to the symbols displayed in the upper panel. The points on the upper curve above the horizontal, dashed line represent extrapolated values of L_3 (see the text for more details).

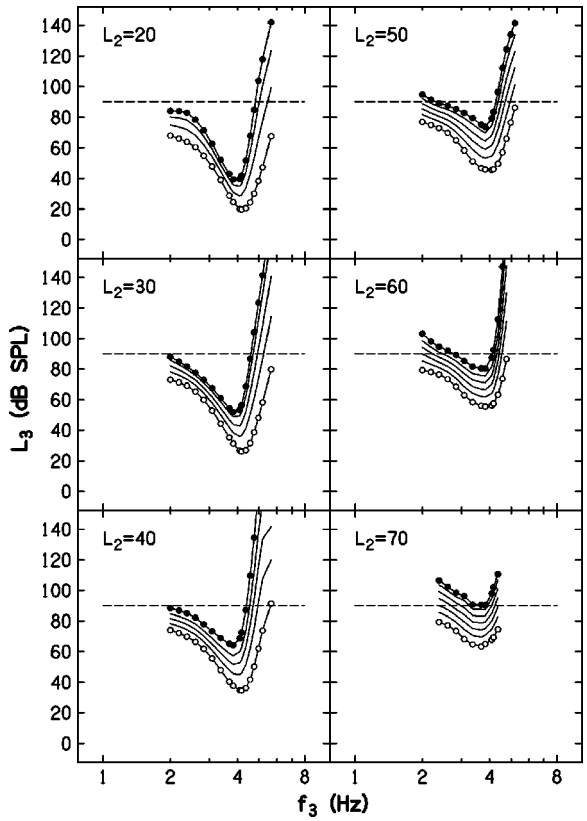


FIG. 2. Mean DPOAE STCs. Each panel represents data for a different L_2. Within each panel, STCs are shown for suppression criteria ranging from 3 dB (open symbols) to complete suppression (closed symbols) in 6-dB increments. As in Fig. 1, points falling above the horizontal, dashed lines do not represent measured suppressor levels, but rather are values extrapolated from the linear fits to the transformed decrement-vs-L_3 functions.

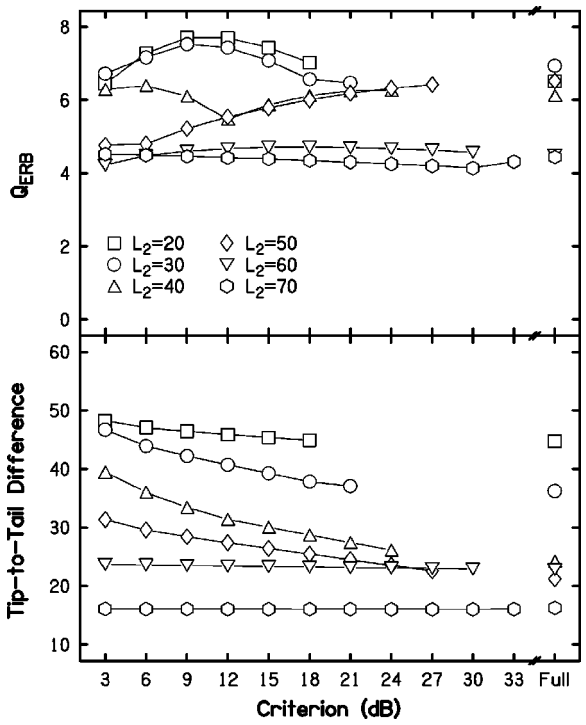


FIG. 3. Q_{ERB} (upper panel) and tip-to-tail difference (lower panel) as a function of suppression criterion (dB). In both panels the parameter is L_2.

was observed when $f_3 > f_2$, less when $f_3 \approx f_2$, and least when $f_3 < f_2$.

III. RESULTS

A. Suppression tuning curves

Mean DPOAE STCs are shown in Fig. 2, with data for a different L_2 shown in each panel. DPOAE STCs are shown for suppression criteria ranging from a decrement of 3 dB to the fully suppressed condition (in 6-dB increments). In each panel, the lowest curve represents the DPOAE STC constructed from the 3-dB criterion and the highest curve represents the fully suppressed condition. In each panel, the circles shown on the upper- and lower-most curves indicate f_3's where the decrement-versus-L_3 function met the linear-fit inclusion criteria. Even for the highest L_2 (70 dB SPL), where it was more difficult to suppress the response, linear-fit inclusion criteria were met for 10 suppressor frequencies. At lower L_2's, the inclusion criteria were met for more suppressor frequencies. Data falling above the dashed, horizontal lines represent extrapolated values of L_3, as described earlier. As the criterion progressed from minimum (3-dB decrement) to maximum suppression, there was a systematic upward shift in the STC. At low L_2's, there tended to be a greater shift at the tip than on the low-frequency tail of the tuning curve. This effect was less evident at higher L_2's. Additionally, the low- and high-frequency slopes of the STCs changed as the suppression criterion was varied. As the suppression criterion increased, the slope of the high-frequency flank became steeper while the slope of the low-frequency flank became shallower.

B. Q_{ERB} and tip-to-tail difference

The upper panel of Fig. 3 plots the Q_{ERB} values for the mean DPOAE STCs as a function of the suppression criterion used to construct the DPOAE STC. The lower panel plots the tip-to-tail difference for the same conditions, where the tip was defined as the lowest L_3 on each curve and the tail was defined as the L_3 when $f_3 = 2$ kHz (an octave below f_2). Within both panels, the parameter is L_2. Q_{ERB} is defined as the best frequency (BF) divided by the equivalent rectangular bandwidth (ERB). For any filter, the corresponding ERB is the bandwidth of the rectangular filter with the same BF response that passes the same total power. Q_{ERB} values for low probe levels were larger (indicating sharper tuning) than Q_{ERB} values for higher probe levels. The criterion used to construct the DPOAE STC had little systematic influence on Q_{ERB}. Across all criteria, the tip-to-tail difference decreased as L_2 increased, going from a maximum of about 45–48 dB for $L_2 = 20$ dB SPL down to approximately 16 dB for $L_2 = 70$ dB SPL. For $L_2 = 20-50$ dB SPL, the tip-to-tail difference decreased as the DPOAE response was more fully suppressed. At higher L_2's, the suppression criterion did not influence the tip-to-tail difference.

IV. DISCUSSION

With the exception of the tip-to-tail differences at $L_2 \leq 50$ dB SPL, the criterion used to construct the DPOAE STC had little influence on measures of cochlear tuning de-

rived from the DPOAE STC (e.g., Q_{ERB}). Given the variation in the slopes of the decrement-versus-L_3 functions across suppressor frequency (see Fig. 1), it was expected that the shape of the DPOAE STCs constructed using different suppression criteria would change. Indeed, there are several reports in the literature that estimates of tuning became sharper as the criterion for constructing the DPOAE STC was varied from minimal suppression to greater amounts of suppression (e.g., Kummer *et al.*, 1995; Martin *et al.*, 1998a; Taschenberger and Manley, 1998). In the present data, the slope of the high-frequency side of the DPOAE STC increased as the response was more fully suppressed; however, this effect had little influence on estimates of tuning (Q_{ERB}) around the tip of the STC.

Tip-to-tail differences on DPOAE STCs previously have been viewed as measures related to ''cochlear-amplifier gain'' (Mills, 1998; Pienkowski and Kunov, 2001), and have been shown to decrease as probe level increases (Gorga *et al.*, 2002, 2003). The decrease in gain, based on these measures, is a consequence of differences in growth of response to on- and low-frequency suppressors. The observations in the present study, in which tip-to-tail differences decreased as the response was more fully suppressed, are a consequence of the same underlying processes. That is, growth of response to suppressors below probe frequency (f_2) was more rapid, compared to suppressors close to f_2. As the criterion increases, this frequency-dependent difference in response growth results in a decrease in the difference between the L_3 required to produce the criterion amount of suppression for on- and low-frequency suppressors.

In our view, the most important result was the demonstration of an approach that enabled us to construct DPOAE STCs from any suppression criteria, including the fully suppressed condition. Because decrement-versus-L_3 functions were well described by linear fits to transformed data, it was possible to extrapolate to suppressor levels that were above the levels that could be produced by the hardware. Assuming this extrapolation was appropriate, this enabled us to extend the dynamic range of our measurements. Furthermore, by using the entire decrement-vs-L_3 function to develop a linear model, reliable estimates of L_3 were possible for all conditions, including those for which the response was close to the noise floor (i.e., those conditions approaching the fully suppressed condition).

While there was little systematic influence of criterion on estimates of Q_{ERB} from DPOAE STCs, criterion did influence estimates related to cochlear-amplifier gain, particularly at the low to moderate L_2's that are similar to the levels used to construct psychophysical tuning curves. Furthermore, constructing DPOAE STCs based on the fully suppressed criterion may be methodologically more similar to conditions under which psychophysical tuning curves are constructed. In psychophysical tuning-curve experiments, the probe frequency and level are fixed and the level of the masker is varied until the probe is ''at threshold'' or just audible. Contrast this condition with the DPOAE condition, in which the response is only partially suppressed. The need to present probe and suppressor simultaneously complicates comparisons to other measurements that describe only the ''excitatory'' pattern of response. However, the completely suppressed response may be a more appropriate criterion for constructing DPOAE STCs when comparisons will be made with tuning curves constructed from threshold measurements such as psychophysical masking experiments.

ACKNOWLEDGMENTS

The authors would like to thank Denis Fitzpatrick for his assistance with applying the spline fits to these data. This work was supported by the NIH (R01 DC02251 and T32 DC00013). Recruitment of subjects was supported by a core grant from the NIH (P30 DC04662).

Abdala, C. (**2001**). ''Maturation of the human cochlear amplifier: Distortion product otoacoustic emission suppression tuning curves recorded at low and high primary levels,'' J. Acoust. Soc. Am. **110**, 1465–1476.

Abdala, C., and Fitzgerald, T. S. (**2003**). ''Ipsilateral distortion product otoacoustic emission ($2f_1-f_2$) suppression in children with sensorineural hearing loss,'' J. Acoust. Soc. Am. **114**, 919–931.

Abdala, C., Sininger, Y. S., Ekelid, M., and Zeng, F.-G. (**1996**). ''Distortion product otoacoustic emission suppression tuning curves in human adults and neonates,'' Hear. Res. **98**, 38–53.

ANSI (**1996**). *ANSI S3.6-1996, Specifications for Audiometers* (American National Standards Institute, New York).

Brown, A. M., and Kemp, D. T. (**1984**). ''Suppressibility of the $2f_1-f_2$ stimulated acoustic emissions in gerbil and man,'' Hear. Res. **13**, 29–37.

Gorga, M. P., Neely, S. T., Dierking, D. M., Dorn, P. A., Hoover, B. M., and Fitzpatrick, D. F. (**2003**). ''Distortion product otoacoustic emission suppression tuning curves in normal-hearing and hearing-impaired human ears,'' J. Acoust. Soc. Am. **114**, 263–278.

Gorga, M. P., Neely, S. T., Dorn, P. A., and Konrad-Martin, D. (**2002**). ''The use of distortion product otoacoustic emission suppression as an estimate of response growth,'' J. Acoust. Soc. Am. **111**, 271–284.

Kummer, P., Janssen, T., and Arnold, W. (**1995**). ''Suppression tuning characteristics of the $2f_1-f_2$ distortion-product otoacoustic emission in humans,'' J. Acoust. Soc. Am. **98**, 197–210.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). ''The level and growth behavior of the $2f_1-f_2$ distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss,'' J. Acoust. Soc. Am. **103**, 3431–3444.

Martin, G. K., Jassir, D., Stagner, B. B., and Lonsbury-Martin, B. L. (**1998a**). ''Effects of loop diuretics on the suppression tuning of distortion-product otoacoustic emissions in rabbits,'' J. Acoust. Soc. Am. **104**, 972–983.

Martin, G. K., Jassir, D., Stagner, B. B., Whitehead, M. L., and Lonsbury-Martin, B. L. (**1998b**). ''Locus of generation for the $2f_1-f_2$ vs $2f_2-f_1$ distortion-product otoacoustic emissions in normal-hearing humans revealed by suppression tuning, onset latencies, and amplitude correlations,'' J. Acoust. Soc. Am. **103**, 1957–1971.

Martin, G. K., Villasuso, E. I., Stagner, B. B., and Lonsbury-Martin, B. L. (**2003**). ''Suppression and enhancement of distortion-product otoacoustic emissions by interference tones above f_2. II. Findings in humans,'' Hear. Res. **177**, 111–122.

Mills, D. M. (**1998**). ''Interpretation of distortion product otoacoustic emission measurements. II. Estimating tuning characteristics using three stimulus tones,'' J. Acoust. Soc. Am. **103**, 507–523.

Neely, S. T., and Liu, Z. (**1993**). ''EMAV: Otoacoustic emission averager,'' Tech. Memo No. 17, Boys Town National Research Hospital, Omaha, NE.

Pienkowski, M., and Kunov, H. (**2001**). ''Suppression of distortion product otoacoustic emissions and hearing thresholds,'' J. Acoust. Soc. Am. **109**, 1496–1502.

Taschenberger, G., and Manley, G. A. (**1998**). ''General characteristics and suppression tuning properties of the distortion-product otoacoustic emission $2f_1-f_2$ in the barn owl,'' Hear. Res. **123**, 183–200.

High-sensitivity piezoelectric microphones based on stacked cellular polymer films (L)

Joachim Hillenbrand[a] and Gerhard M. Sessler[b]
Department of Telecommunications, Darmstadt University of Technology, Merckstrasse 25, 64283 Darmstadt, Germany

(Received 30 July 2004; revised 1 September 2004; accepted 7 September 2004)

Improvements of the sensitivity of piezoelectric microphones based on charged cellular polymer films are reported. The improvements are achieved by (1) an increase of the piezoelectric d_{33}-coefficient of the cellular polypropylene films by pressure expansion and (2) stacking of the films. Microphones consisting of a single film of such material have sensitivities of about 2 mV/Pa at 1 kHz, independent of size, while for a microphone with five stacked films a sensitivity of 10.5 mV/Pa was measured. The equivalent noise level is about 37 dB(A) for the single-film transducer and 26 dB(A) for the stacked version. Advantages of these new piezoelectric transducers include their simple design, low cost, and small weight, as well as a large range of shapes and sizes possible. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1810272]

PACS numbers: 43.38.Fx, 43.38.Ar, 43.38.Kb [AJZ] Pages: 3267–3270

I. INTRODUCTION

Cellular polypropylene (PP), after appropriate electrical charging, is highly piezoelectric.[1–3] In particular, the piezoelectric d_{33}-coefficient of this material reaches values of about 150 pC/N in the audio frequency range and is thus about five times as high as that of polyvinylidenefluoride (PVDF), the best conventional piezoelectric polymer. It was therefore suggested to use charged cellular PP in electro-acoustic and electromechanical transducers. In particular, implementations of loudspeakers,[4] microphones[5,6] and hydrophones[6] have been described in the literature. Although of considerably simpler design than conventional transducers, the previously implemented new devices did not yet reach the electro-acoustic performance of older capacitive and piezoelectric systems.

Recently, the piezoelectric activity of cellular PP has been significantly increased by thickness-expansion of the polymer.[7–11] Microphones with such improved films, showing sensitivities of about 2.2 mV/Pa at 1 kHz, have already been described by the present authors.[12] The use of film stacks in such microphones, not yet implemented experimentally, is a method to further enhance the sensitivity of these transducers. This suggests to build advanced microphones and to examine their electro-acoustic properties.

In the present letter, the implementation and characterization of such microphones are described. In particular, the cellular films are briefly specified in Sec. II, the microphone design and measuring methods are outlined in Sec. III, electro-acoustic measurements on the new systems are reported in Sec. IV, and the properties of these microphones are discussed in Sec. V.

II. CELLULAR PP

The upper part of Fig. 1 shows a SEM photograph of the cross sectional area of a cellular film (HS01 by VTT, Tampere, Finland). The film is usually charged on its surface by a corona discharge. Due to the ensuing electric field in the interior of the film, discharges occur in the voids and charging as shown in the lower part of the figure is achieved. A charge distribution of this kind in a nonhomogeneous material causes the piezoelectric effect.[2]

All piezoelectric films used in the present microphones were made of a commercial cellular PP film (VHD40 by Treophan, Neunkirchen, Germany). Expansion of these films is achieved by a pressure treatment consisting in the application of an increased gas pressure for some time, followed by pressure reduction to atmosphere. This results in an increase of the thickness of the lenslike voids which originally extends up to about 5 μm and approximately up to about 10 μm after expansion. According to the Paschen law, the increased thickness after expansion lowers the electric field required for breakdown. Thus, for a given voltage, more voids experience stronger breakdown, causing an increase of the piezoelectric coefficient. Since metallization after the expansion causes shrinking of the film, a second expansion after metallization softens the material and thus results in another increase of d_{33}.[10]

The frequency response of the d_{33}-coefficient of such samples was measured, utilizing the inverse piezoelectric effect, by sinusoidal electrical excitation and interferometric measurement of the resulting surface deflection. A typical response of an expanded sample is shown in Fig. 2. In the audio frequency range, d_{33}-coefficients of about 420 pC/N were found. These values are larger than those previously reported for nonexpanded samples by a factor of 3. The slow decrease of d_{33} up to about 30 kHz is due to an increase of Young's modulus, while the resonance at approximately 140 kHz is determined by Young's modulus and the mass of the film.[9]

The pressure dependence of d_{33} was determined quasistatically by measuring the generated charge upon pressure application.[9] Typical results show that up to pressures of a few kPa d_{33} increases slightly. This indicates that the stress–

[a]Electronic mail: j.hillenbrand@nt.tu-darmstadt.de
[b]Electronic mail: g.sessler@nt.tu-darmstadt.de

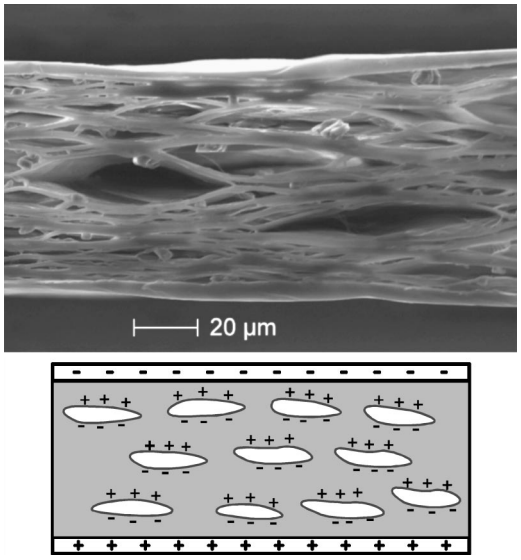


FIG. 1. SEM photograph of cross section of cellular PP film (HS01) of 70 μm thickness (top) and schematic view of charge distribution in this material (bottom).

strain relationship of the cellular films is not exactly linear in this pressure range, as expected for cellular materials.[13]

III. MICROPHONE DESIGN AND MEASURING METHODS

The expanded cellular VHD40 films were used to construct piezoelectric microphones.[6,12] These transducers consist simply of a piece of the cellular material of 0.3 cm^2 size that is metallized on both sides. For a film thickness of 55 μm, the capacitance of the microphone is 8 pF. Shielding requires the mounting of the film in a small housing. The microphone output is fed directly into a preamplifier of unity gain (B&K 2669)[14] and its output into an audio analyzer (R&S UPD).

In the stacked microphones, the single film is substituted by a stack of films metallized on both sides which are glued on top of each other. Since the sound pressure acts on all films and since the films are electrically connected in series, the output voltages of all layers add up and the open circuit

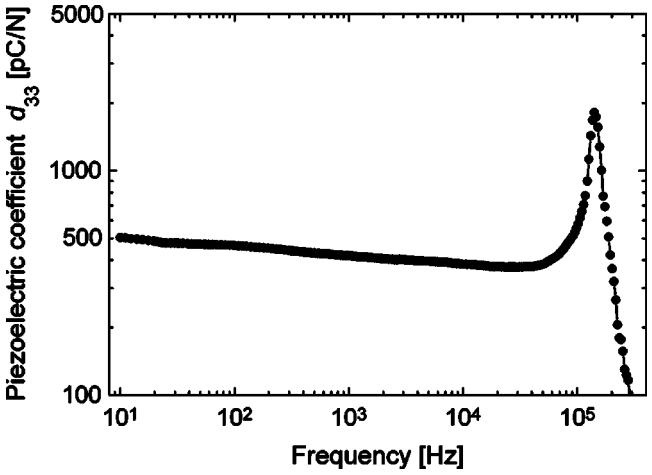


FIG. 2. Interferometrically measured d_{33}-coefficient for expanded cellular PP film (VHD40) of 55 μm thickness.

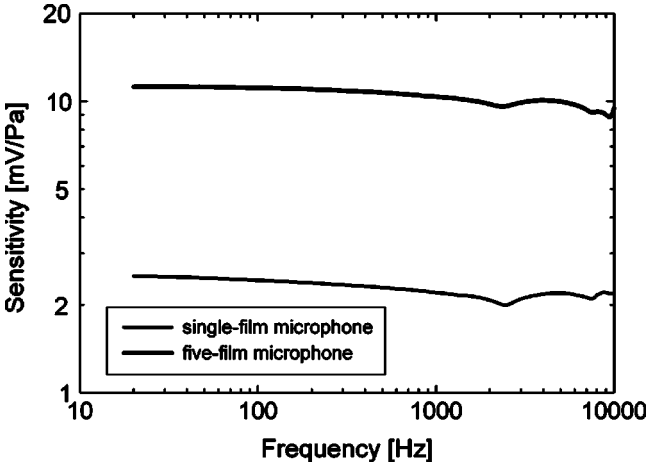


FIG. 3. Frequency response of cellular PP microphones with single film and a stack of five films, determined by a comparison method in an acoustic coupler. Films of about 55 μm thickness (VHD40) were used.

sensitivity of a microphone with n layers should be n times as high as that of a single-film transducer, while its capacitance and resonance frequency (see Sec. V) decrease by a factor of n.

Electroacoustic measurements were carried out by placing the microphone in an acoustic coupler with a volume of 0.4 cm^3. A 1/8-in. condenser microphone (B&K 4138), also extending into the coupler, serves as a reference. The sound pressure in the coupler is generated with a small electrodynamic speaker in a separate cavity coupled by a 4.5-cm-long metallic pipe into the measuring coupler.

Electronic data recording and processing is carried out with the above-mentioned audio analyzer, which allows the evaluation of the frequency response of the sensitivity, its amplitude dependence, the total harmonic distortion, and the noise spectrum of the microphone and its amplifier.

IV. EXPERIMENTAL RESULTS

The measured frequency responses of microphones with one and five films of cellular PP films are shown in Fig. 3. As seen from the figure, the open-circuit sensitivity of the five-film microphone is about 10.5 mV/Pa at 1 kHz and thus, as expected (see above), almost five times larger than the 2.2 mV/Pa of the single-film microphone at this frequency. Both responses decrease by about 1 dB from 20 Hz to 1 kHz, as expected from the frequency response of the d_{33}-coefficient shown in Fig. 2. The ripples seen at 2 kHz and above are due to the fact that the dimensions of the pressure chamber with its connector are comparable to a quarter wavelength at these frequencies.

To test the linearity of the microphones, the dependence of the sensitivity on applied sound pressure was examined. The data (not shown in this letter) indicates that there is a 3% sensitivity increase up to 3.2 kPa (164 dB SPL). This result is in qualitative agreement with the increase of d_{33} with pressure, discussed above, and thus an indication of nonlinearities of the stress–strain relationship.

The total harmonic distortion (THD) of the cellular microphone is also related to this nonlinearity. As measurements show, THD increases approximately proportionally to

sound pressure and is less than 1% at 164 dB SPL. This very small distortion originates probably to some part from the loudspeaker used in these experiments. The part generated by the microphone is again due to the small nonlinearities of the stress–strain relationship.

The A-weighted noise voltages of the single-film transducer and the five-film stack transducer, combined with a preamplifier, are 3.0 and 4.2 μV, respectively. From these values, total equivalent noise levels (ENLs) of 37 and 26 dB(A), respectively, are obtained. The noise corresponds closely to that of the preamplifier, as specified by the manufacturer.[14] The improvement of the ENL for the stack microphone by 11 dB is due to the increase of the sensitivity (14 dB), reduced by the increase of the preamplifier noise which is mainly due to the lowering of the capacitance (-3 dB).

V. DISCUSSION AND CONCLUSIONS

The sensitivity M of the single-film microphone described above is related to the d_{33}-coefficient by[15]

$$M = d_{33}\frac{(s_1+\varepsilon s_2)}{\varepsilon\varepsilon_0}, \tag{1}$$

where s_1 and s_2 are the combined thicknesses of all solid or gas parts of the cellular film, respectively, and ε_o and ε are the absolute and relative permittivities, respectively. From Eq. (1) the observed sensitivity of 2.2 mV/Pa is obtained for $s_1=26$ μm, $s_2=30$ μm (calculated from the density and total thickness of the film), and $\varepsilon=2.35$ by substituting $d_{33}=475$ pC/N, which is close to values actually measured interferometrically (cf. Fig. 2).

The measured sensitivity of 10.5 mV/Pa for a five-film microphone is very high for a piezoelectric microphone and is comparable with sensitivities of electret condenser microphones. Even higher sensitivities may be possible by further increasing the d_{33}-coefficients which can be achieved by increasing the charge density and by decreasing Young's modulus of the cellular films.[10] The sensitivity may also be improved by increasing the number n of piezoelectric films in the stack microphone. Since such an increase lowers the capacitance of the device, stray capacitances and the input capacitance of the preamplifier have an adverse effect on the sensitivity. For this reason, the sensitivity of the present experimental design will not gain very much by increasing n beyond 5. However, systems with reduced stray capacitance and/or with larger transducer area will show improved sensitivities for $n>5$. The eventual limit will be reached when the film stack capacitance becomes smaller than the input capacitance of the preamplifier.

Equally important is the equivalent noise level which is at 37 and 26 dB(A) for the single- and five-film microphones, respectively. Particularly the latter value is again comparable with that for typical electret microphones[16] and is much better than that of previous cellular microphones [52 dB(A)].[6]

A drawback of the present cellular microphones is their decrease of sensitivity at temperatures in excess of 60 °C due to the instability of the electret charges in the cellular PP film. Efforts are underway to produce cellular films of polymers with better charge stability than that of the presently used PP types.

Additional features of the cellular microphones are their low harmonic distortion and their high resonance frequencies. As Fig. 2 shows, the resonance frequency of a single-film microphone is expected to be at about 140 kHz. For a stack of n films, the resonance frequency decreases by a factor of n since the mass of the system increases and the stiffness decreases by this factor. This suggests a resonance frequency of 28 kHz for the five-film transducer.

In addition to these features, the cellular PP microphones have a simple design. The transducers consist essentially only of one or several pieces of charged and metallized cellular films, equipped with suitable backing and shielding. No miniature air gaps, as in electret microphones, have to be maintained. Cellular microphones can therefore be manufactured at very low cost. These features make such microphones very suitable for a wide range of applications.

Because of the ambiguous meaning of the term ''cellular microphones'' and in view of the correlation of electret and piezoelectric properties in cellular films, we suggest calling these transducers ''*piezo*electret microphones.''

ACKNOWLEDGMENTS

The authors are grateful to Dr. Xiaoqing Zhang for preparation of the expanded cellular films, to Treofan for supplying the original films, and to the Deutsche Forschungsgemeinschaft and the Volkswagen Foundation for financial support.

[1] J. Lekkala, R. Poramo, K. Nyholm, and T. Kaikkonen, ''EMF force sensor–a flexible electret film for physiological applications,'' Med. Biol. Eng. Comput. **34**, 67–68 (1996).
[2] G. M. Sessler and J. Hillenbrand, ''Electromechanical response of cellular electret films,'' Appl. Phys. Lett. **75**, 3405–3407 (1999).
[3] S. Bauer, R. Gerhard-Multhaupt, and G. M. Sessler, ''Ferroelectrets: Soft Electroactive Foams for Transducers,'' Phys. Today **57**, 37–43 (February 2004).
[4] M. Antila, T. Muurinen, J. Linjama, and H. Nykänen, ''Measurement methods of flat panel electromechanical film loudspeakers,'' Active **97**, 607–618 (1997).
[5] H. Nykänen, M. Antila, J. Kataja, J. Lekkala, and S. Uosukainen, ''Active control of sound based on utilizing EMFI-technology,'' Active **99**, 1159–1170 (1999).
[6] R. Kressmann, ''New piezoelectric polymer for air-borne and water-borne sound transducers,'' J. Acoust. Soc. Am. **109**, 1412–1416 (2001).
[7] M. Paajanen, M. Wegener, and R. Gerhard-Multhaupt, ''Understanding the role of the gas in the voids during corona charging of cellular electret films—a way to enhance their piezoelectricity,'' J. Phys. D **34**, 2482–2488 (2001).
[8] J. Hillenbrand, X. Zhang, J. Zhang, and G. M. Sessler, ''Pressure-treated cellular polypropylene with large piezoelectric coefficients,'' *2003 Annual Report, Conf. Electric. Insul. and Diel. Phenom.* (2003), pp. 40–43.
[9] X. Zhang, J. Hillenbrand, and G. M. Sessler, ''Piezoelectric d_{33}-coefficient of cellular polypropylene subjected to expansion by pressure treatment,'' Appl. Phys. Lett. **85**, 1226–1228 (2004).
[10] X. Zhang, J. Hillenbrand, and G. M. Sessler, ''Improvement of piezoelectric activity of cellular polymers by a double-expansion process,'' J. Phys. D **37**, 2146–2150 (2004).
[11] M. Wegener, W. Wirges, R. Gerhard-Multhaupt, M. Dansachmüller, R. Schwödiauer, S. Bauer-Gogonea, S. Bauer, M. Paajanen, H. Minkkinen,

and J. Raukola, ''Controlled inflation of voids in cellular polymer ferroelectrets: Optimizing electromechanical transducer properties,'' Appl. Phys. Lett. **84**, 392–394 (2004).
[12] J. Hillenbrand and G. M. Sessler, ''New piezoelectric transducers based on expanded cellular polymer electrets,'' in *Proceed. of ICA 2004, The 18th Internat. Congress on Acoustics*, Kyoto, Japan, 2004, Vol. I, 349-352.
[13] L. J. Gibson and M. F. Ashby, *Cellular Solids*, 2nd ed. (Cambridge U.P., Cambridge, 1997).
[14] See data sheet ''B&K microphone preamplifier type 2669,'' www.bksv.com/pdf/bp1422.pdf
[15] J. Hillenbrand and G. M. Sessler, ''Piezoelectricity in cellular electret films,'' IEEE Trans. Dielectr. Electr. Insul. **7**, 537–542 (2000).
[16] G. M. Sessler, ''Microphone,'' in *Encyclopedia of Science & Technology*, 9th ed., Vol. 11 (McGraw-Hill, New York, 2002), pp. 88–95.

Sound equalization in a large region of a rectangular enclosure (L)

John C. Sarris[a] and Finn Jacobsen[b]
Acoustic Technology, Ørsted DTU, Technical University of Denmark, Ørsteds Plads, Building 352, DK-2800 Kgs. Lyngby, Denmark

George E. Cambourakis[c]
School of Electrical and Computer Engineering, National Technical University of Athens, Heroon Polytechniou 9, 157 73, Athens, Greece

(Received 13 February 2004; revised 14 September 2004; accepted 24 September 2004)

The work presented by Santillán [J. Acoust. Soc. Am. **110**, 1989–1997 (2001)] about equalization at low frequencies in rectangular enclosures is extended, and topics that remained unaddressed in the original study are treated in this paper. A modification is introduced to the original cost function that leads to a least-squares problem formulation that demonstrates increased robustness, and the multiple-error LMS (Least Mean Square) adaptive algorithm is employed to approximate the coefficients of the equalization filters to their optimum values. Other issues studied in this paper include the dependence of the limits of the zone of equalization on factors such as the damping constant of the modes of the enclosure, the number of sources, and the driving frequency. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1819500]

PACS numbers: 43.55.Br, 43.60.Pt, 43.60.Mn [NX] Pages: 3271–3274

I. INTRODUCTION

The purpose of equalization in room acoustics is to compensate for the undesired modification that a particular enclosure introduces to signals as, for example, audio or speech. Traditional multichannel methods introduce digital filters to preprocess the input signal before it is fed to a set of loudspeakers so that the spectral coloration and the reverberation tail associated with the transmission path are removed and the obtainable signals at a set of receiving points are the best approximations to a set of desired signals.[1] Such techniques exhibit a high performance exactly at the receiving points, but the zone where equalization can be performed is limited to small areas in the vicinity of the points.[2]

The problems that traditional equalization methods face were overcome in Santillán's approach.[3] In this work a plane wave moving along a rectangular enclosure was generated in order to equalize the sound field at low frequencies in a continuous three-dimensional region that occupied almost the complete volume of the enclosure. Theoretical results showed that the generation of a traveling plane wave is possible within a rectangular enclosure provided that sets of sources are arranged in the two walls that are perpendicular to the direction of propagation. The traditional least-squares approximation has been employed to estimate the coefficients of FIR filters for the equalization of broadband signals.[3]

However, some issues remained unaddressed in the original study, and thus the purpose of this paper is to extend the previous work and discuss some matters more thoroughly on the basis of simulation and experimental results.

II. EQUALIZATION OF HARMONIC SOUND FIELDS

A. Problem formulation and optimum solution

In the work by Santillán[3] the optimum strengths of the sources for the equalization of a harmonic sound field by simulating a propagating plane wave are obtained according to the multiple point method[4] by minimizing the mean-squared error at a number of receivers distributed in the zone of equalization. However, under certain conditions this least-squares formulation fails to provide a solution or provides one that lacks robustness.

A more robust solution for the optimum strengths of the sources can be obtained by including a term proportional to the effort of the sound sources in the cost function. Minimization of this modified cost function is guaranteed to provide a solution in all cases, although it increases the minimum value of the cost function and thus in general makes the solution slightly less perfect. Furthermore, since in practice the strengths of the sources are approximated to their optimal values by using an adaptive algorithm such as, for example, the multiple-error LMS algorithm,[5] this modification ensures convergence. In the following section simulation results are presented to support this.

B. Simulation results

1. An ill-conditioned system

The purpose of the first simulation is to demonstrate the different performance of the cost function used by Santillán[3] and the modified cost function proposed in this study when the system is ill-conditioned. The enclosure considered in the simulation is rectangular, with dimensions $L_x = 2.7$ m, L_y

[a]Now at: School of Electrical and Computer Engineering, National Technical University of Athens, Heroon Polytechniou 9, 157 73, Athens, Greece. Electronic mail: jsarris@central.ntua.gr
[b]Electronic mail: fja@oersted.dtu.dk
[c]Electronic mail: gcamb@cs.ntua.gr

$=5$ m, and $L_z=2.5$ m. All modes with a natural frequency up to 1 kHz are used to simulate the sound pressure. Thirty-two sources and 36 receivers are considered placed as described by Santillán.[3] The driving frequency is 200 Hz and a damping constant with value equal to 0.01 for all the modes is assumed.

The transfer function matrix that results from the arrangement of sources and receivers described above is rank deficient, and thus the least-squares formulation proposed by Santillán[3] does not have a numerically stable solution. However, the cost function with effort weighting leads to a formulation that provides a stable solution. With the LMS adaptive formulation of the cost function without effort weighting there are some modes with infinite time constants, since the transfer function matrix is rank deficient. The inclusion of the effort weighting in the definition of the cost function compensates for the rank deficiency and ensures that there are no such modes.

2. Limits of the zone of equalization

According to Santillán's study[3] the zone of equalization, defined as the continuous three-dimensional region of the enclosure where the deviations of the sound pressure from the desired value are within ± 3 dB, is constantly limited to be 0.6 m away from the two parallel walls where the sources are placed. This limit is a conservative estimate for most practical cases, and thus the dependence of this restriction on factors like the damping constant of the modes of the enclosure, the number of sources, and the driving frequency calls for further examination.

From simulation results it can be concluded that the limits of the zone of equalization depend on all three factors, i.e., the number of sources, the driving frequency, and the damping constant of the modes of the enclosure. In practical cases where a finite number of sources is used the zone of equalization will be limited with respect to the walls where the sources are placed, and the distance of the limits from these walls will be proportional to the driving frequency and inversely proportional to the number of the sources. Furthermore, the results of simulations show that in general damping tends to limit the zone of equalization further.

III. EQUALIZATION OF BROADBAND SOUND FIELDS

A. Problem formulation and optimum solution

As shown by Santillán, it is possible to equalize a broadband signal in a large region in a rectangular room by generating a traveling plane wave.[3] In his study, equalization of a broadband signal was performed by FIR equalization filters whose coefficients were determined by minimization of the expectation of the sum of squared errors at a number of receivers distributed in the zone of equalization.[6] However, as in the case of harmonic sound-field equalization, this least-squares formulation may fail to provide a solution for certain arrangements of sources and receivers.

The introduction of a term proportional to the sum of the squared equalization coefficients to the cost function results in a least-squares formulation of increased robustness and, as in the case of the adaptive implementation for the equalization of a single-tone sound field, the convergence stability of the LMS formulation of the modified cost function is guaranteed.

B. Experimental results

Experimental results have been obtained in a rectangular enclosure constructed of 22-mm MDF fiberboard with internal dimensions $L_x=1.2$ m, $L_y=2$ m, and $L_z=0.2$ m. For simplicity the enclosure was much smaller in one dimension than in the other two, so that the resulting sound field is two-dimensional in the frequency range considered. Two holes were made on each side perpendicular to the y axis (the direction of the traveling wave), and two 5-in. loudspeakers were placed in a line at $z=0.1$ m with their respective centers positioned at $x=0.3$ m and $x=0.9$ m in each line. Thus, a total of four sources was used. According to the specification provided by the manufacturer (Vifa-Speak) the loudspeakers have a flat frequency response in the frequency range from about 140 Hz up to approximately 2 kHz. Holes were also made in one of the large sides of the enclosure, and five 9.7-mm electret microphones were distributed within the enclosure in three lines parallel to the x axis, with two microphones in each of the two first lines and a single microphone in the last one. The number of microphones was limited by the memory of the available DSP board. The first line of microphones was positioned at $y=0.6$ m, and the distance between adjacent microphones was 0.43 and 0.8 m in the y- and x direction, respectively. Since the sound field in the frequency range of interest should be practically independent of the z direction, all the microphones were placed in a single plane at $z=0$ m. In order to measure the sound pressure at various points in the enclosure, 19 holes were evenly made in one side of the enclosure along the y direction. Thus, by traversing a 9.7-mm electret microphone to 13 evenly spaced positions along the x axis ($x=0$ to 1.2 m) in each of the 19 holes the sound pressure could be measured at 247 positions in the enclosure.

The input signal was random noise generated and bandpass filtered in the frequency range from 100 Hz up to 300 Hz by the Brüel & Kjær PULSE system and additionally low-pass filtered by an analog filter with a cutoff frequency of 310 Hz. The input signal was then amplified, and passed through an A/D converter before it was fed to the DSP board. Each of the signals from the microphones was amplified by amplifiers that were adjusted so that each channel had the same sensitivity, passed through an A/D converter, and then transferred to the DSP board. The signals to drive the loudspeakers were generated in the DSP board, passed through a D/A converter, amplified, and then fed to the loudspeakers. Analog antialising filters with a cutoff frequency of 400 Hz were also used to low-pass filter the amplified signals from the microphones, and analog reconstruction filters with a cutoff frequency of 400 Hz were used to low-pass filter the output signals of the DSP before they were fed to the amplifiers and loudspeakers. The card with the A/D converter, the card with the D/A converter, and the DSP board were housed in a Pentium 233-MHz PC. The DSP board was manufactured by Loughborough Sound Images Ltd. and had implemented a Motorola DSP 96002 processor with a 33.3-MHz

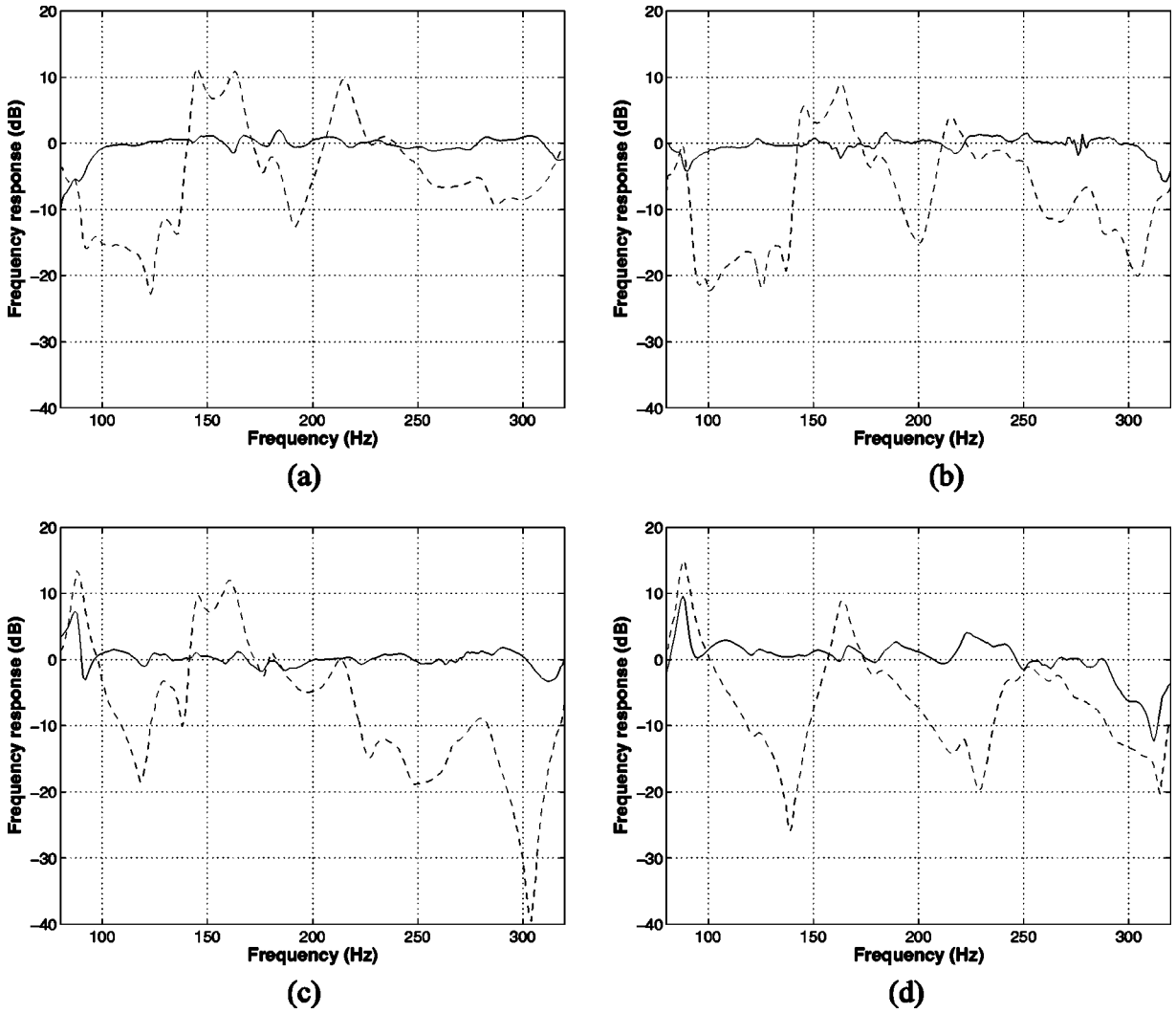


FIG. 1. Frequency responses before (- -) and after (——) equalization at the points (a) (0.2, 1.03, 0.1) m; (b) (0.4, 1.1, 0.1) m; (c) (1.0, 1.6, 0.1) m; and (d) (0.6, 0, 0.1) m corresponding to the location of a microphone used in the equalization, a point within the zone of equalization, a point just outside the zone of equalization, and a point far from the zone of equalization.

clock frequency. The A/D converter was a 32-channel analog acquisition card, and the D/A converter was a 16-channel analog output card, both manufactured by Loughborough Sound Images Ltd.

The impulse responses of the transmission paths between the loudspeakers and the microphones were represented by FIR filters with 150 coefficients, estimated by the multiple-error LMS algorithm implemented as an identification algorithm. Once the identification was completed the coefficients of the identification FIR filters were kept fixed and the equalization process began. The equalization filters were FIR filters with 60 coefficients. The sampling frequency was 1600 Hz, and the propagating plane wave was generated by applying delays of 20 samples to obtain the desired signal for the two receivers in the first line and additional delays of two samples for the receivers in the following lines. These two sample delays correspond to the time that the propagating wave requires to travel from each line of receivers to the next.

The results obtained for the equalization of the broadband sound field by the multiple-error LMS implementation of the modified cost function are illustrated in Fig. 1 at four positions in the enclosure, along with the unequalized frequency responses obtained by exciting the enclosure with the loudspeaker placed at (0.3, 0, 0.1) m. The frequency responses at the four positions correspond to the location of a microphone used in the equalization, a point within the zone of equalization, a point just outside the zone of equalization, and a point far from the zone of equalization. It can be seen that the frequency responses are improved at all positions after the equalization. The peaks and dips of the original frequency responses are removed and practically flat responses are obtained in the frequency range of interest. Even at the point far from the zone of equalization a reasonably flat response is obtained after the equalization almost in the entire frequency range of concern. Similar results have been obtained at other positions in the enclosure. The spatial distribution of the sound pressure in the 200-Hz, one-third-octave band before and after the broadband sound field equalization is shown in Fig. 2. The sound pressure before the equalization was obtained by exciting the enclosure with the source placed at (0.3, 0, 0.1) m. It can be seen that the equalization removed the spatial fluctuations almost in the entire volume of the enclosure, and thus considerably im-

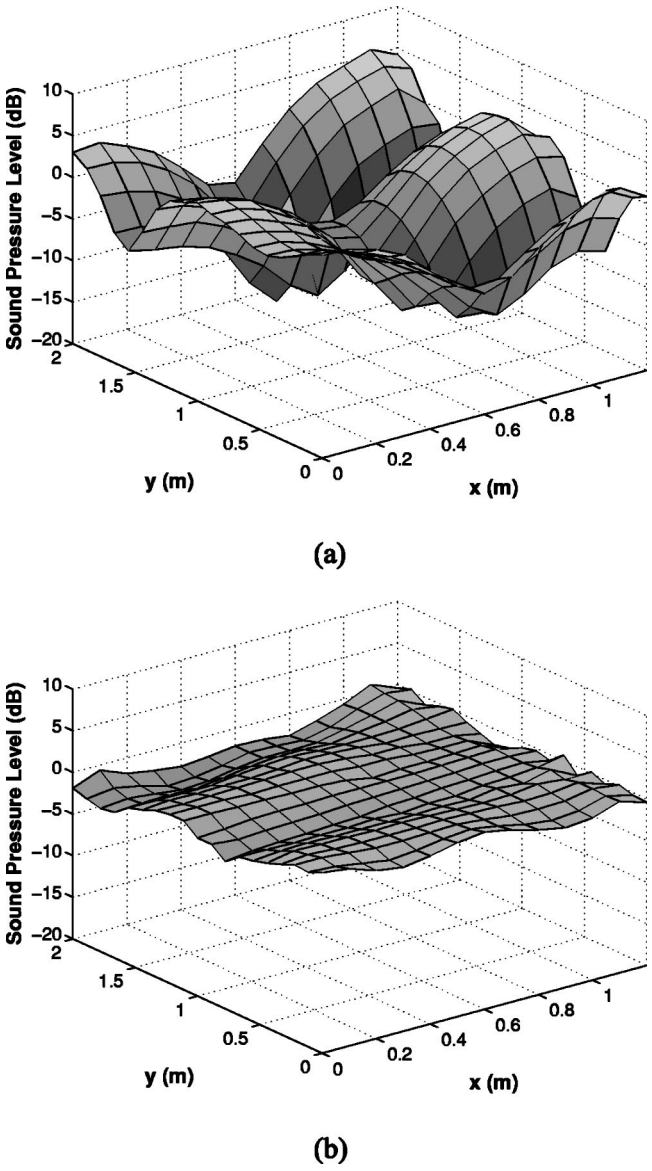


FIG. 2. Sound-pressure level in the 200-Hz one-third-octave band in the experimental enclosure (a) before, and (b) after equalization with random noise as the input signal. The sound pressure before the equalization is obtained by exciting the enclosure with the source placed at (0.3, 0, 0.1) m.

proved the spatial distribution of the sound field.

To examine the robustness of the adaptive implementation of the modified cost function in the presence of modeling errors, three additional holes were made at the points (0.2, 0.63, 0) m, (0.23, 1.46, 0) m, and (1, 1.49, 0) m, and the three microphones originally placed at (0.2, 0.6, 0) m, (0.2, 1.46, 0) m, and (1, 1.46, 0) m were moved at these positions after the identification and before the equalization process. Two of these new positions were shifted from the original positions along the direction of the wave propagation, while the last one was shifted in the x direction. It should be noted that this type of error corresponds to the worst case of model uncertainties.[7] By proper adjustment the multiple-error LMS implementation of the modified cost function was found to be stable, and the results obtained for the equalization of the broadband sound field were similar to the ones presented above.

IV. CONCLUSIONS

As shown theoretically by Santillán,[3] the most efficient way to equalize the sound pressure in a large region in a room is to generate a plane propagating wave. This work supplements Santillán's simulation study; the method has been examined experimentally, and several issues that remained unaddressed in the original study have been treated in this paper. The modification that is introduced to the original cost function leads to a least-squares problem formulation that is guaranteed to provide a solution, and the multiple-error LMS adaptive formulation that is based on the modified cost function has been shown to alleviate problems concerning convergence stability and convergence speed that can occur in the adaptive implementation of the original cost function. Also, it has been shown that in the practical case where a finite number of sources is used the zone of equalization will be limited with respect to the walls where the sources are placed, and the distance of the limits from these walls will be proportional to the driving frequency and inversely proportional to the number of the sources.

ACKNOWLEDGMENTS

The authors would like to thank A. O. Santillán for making part of the simulation code available and for many helpful discussions. The work of the first author was supported by a Marie Curie Fellowship under the program of European Doctorate in Sound and Vibration Studies.

[1] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," IEEE Trans. Acoust., Speech, Signal Process. **36**, 145–152 (1988).
[2] P. A. Nelson, F. O. Bustamante, and H. Hamada, "Inverse filter design and equalization zones in multichannel sound reproduction," IEEE Trans. Speech Audio Process. **3**, 185–192 (1995).
[3] A. O. Santillán, "Spatially extended sound equalization in rectangular rooms," J. Acoust. Soc. Am. **110**, 1989–1997 (2001).
[4] F. Asano and D. C. Swanson, "Sound equalization in enclosures using modal reconstruction," J. Acoust. Soc. Am. **98**, 2062–2069 (1995).
[5] S. J. Elliott, I. M. Stothers, and P. A. Nelson, "A multiple error LMS algorithm and its application to the active control of sound and vibration," IEEE Trans. Acoust., Speech, Signal Process. **35**, 1423–1434 (1987).
[6] S. J. Elliott and P. A. Nelson, "Multiple-point equalization in a room using adaptive digital filters," J. Audio Eng. Soc. **37**, 899–907 (1989).
[7] A. Omoto and S. J. Elliott, "The effect of structured uncertainty in the acoustic plant on multichannel feedforward control systems," IEEE Trans. Speech Audio Process. **7**, 204–212 (1999).

Annoyance of bandpass-filtered noises in relation to the factor extracted from autocorrelation function (L)

Yoshiharu Soeta[a)]
Institute for Human Science and Biomedical Engineering, National Institute of Advanced Industrial Science and Technology (AIST), Midoriga-oka, Ikeda, Osaka, 563-8577 Japan

Takuya Maruo[b)] and Yoichi Ando[c)]
Graduate School of Science and Technology, Kobe University, Rokkodai, Nada, Kobe 657-8501, Japan

(Received 2 September 2003; revised 18 February 2004; accepted 25 June 2004)

This study examines the annoyance of pure-tone and bandpass noises with center frequencies of 1000 and 2000 Hz under the condition of equal sound-pressure level. The bandwidths of the source signal were 0, 40, 80, 160, and 320 Hz, with a 2068-dB/octave sharp filter to control the autocorrelation function (ACF) of the source signal. Scale values of annoyance were obtained using a paired-comparison method. The results show that the annoyance of pure-tone and bandpass noises inside the critical band is not constant. The annoyance of pure tones is greater than that of sharply filtered noises within the critical band. The annoyance of bandpass noises increases with increasing effective duration of the ACF, τ_e, which represents the repetitive feature or tonal component of the source signal within the critical band. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1782931]

PACS numbers: 43.50.Qp, 43.66.Lj [DKW] Pages: 3275–3278

I. INTRODUCTION

Although changes in sound intensity are highly correlated with loudness changes, the relationship is not perfect. That is, changes in frequency, duration, and bandwidth all affect the perceived loudness of a stimulus, even though the level of the stimulus remains fixed (Yost, 2000). Previous studies have concluded that the loudness of a noise remains constant as the bandwidth of the noise increases until the bandwidth reaches the critical band. Loudness then increases with increasing bandwidth under the same sound-pressure level conditions (Zwicker *et al.*, 1957; Greenwood, 1961; Scharf, 1962). However, it was found that the loudness of a sharply (2068 dB/octave) filtered noise increases as the effective duration of the autocorrelation function (ACF), τ_e, increases, even when the bandwidth of the signal is within the critical band (Sato *et al.*, 2002). τ_e represents repetitive features within the signal itself.

Industrial development has substantially increased environmental noise. It has become important to determine the various effects of such noise on human beings. Recently, a theory of primary and spatial sensations to environmental noise has been proposed based on the auditory–brain system (Ando, 2001). The model includes the ACF and the interaural cross-correlation function (IACF) mechanisms. The characteristics of a flushing toilet noise in a bedroom on the downstairs floor, which was perceived to be annoying although the sound-pressure level was only about 35 dBA, were examined based on this model (Kitamura *et al.* 2002). The results showed that the ACF and IACF factors for the flushing toilet noise changed dramatically as a function of time and might influence annoyance. It was also found that the ACF factors τ_e and ϕ_1 had a significant effect on annoyance (Fujii *et al.*, 2002).

Basically, psychoacoustic annoyance depends on the loudness, timbre, and temporal structure of sounds. Loudness and annoyance have been distinguished and defined (Berglund *et al.*, 1975; Hellman, 1982). These authors relate annoyance to an individual's reaction to noise within the context of a given situation and loudness directly to sound intensity. The present study examines the annoyance of bandpass noises in terms of the factors extracted from the ACF. Scale values of annoyance of sharply filtered bandpass noises centered on 1000 and 2000 Hz were obtained using a paired-comparison method.

II. METHOD

A. Subjective annoyance test

Pure-tone and bandpass noises with center frequencies of 1000 and 2000 Hz were used as source signals. The basic stimulus in this experiment was a maximum-length sequence (MLS) signal (order 21; sampling frequency 44 100 Hz) which was bandpass filtered. The MLS signal used during this experiment is called pseudorandom white noise. To control the ACF of the bandpass noise, the filter bandwidth was varied at 0, 40, 80, 160, and 320 Hz by using a cutoff slope of 2068 dB/octave, obtained by a combination of two filters (Sato *et al.*, 2002). In fact, the filter bandwidth of 0 Hz was the only slope component. The critical bandwidth of the center frequency of 1000 and 2000 Hz are approximately 160 and 300 Hz (Zwicker and Terhardt, 1980). Figure 1 shows the power spectra of some of the sound used in this study.

[a)]Electronic mail: y.soeta@aist.go.jp
[b)]Present address: Graduate School of Design, Kyushu University, 4-9-1 Shiobaru, Minami-ku, Fukuoka 815-8540, Japan
[c)]Present address: Makizono, Kagoshima, 899-6603 Japan

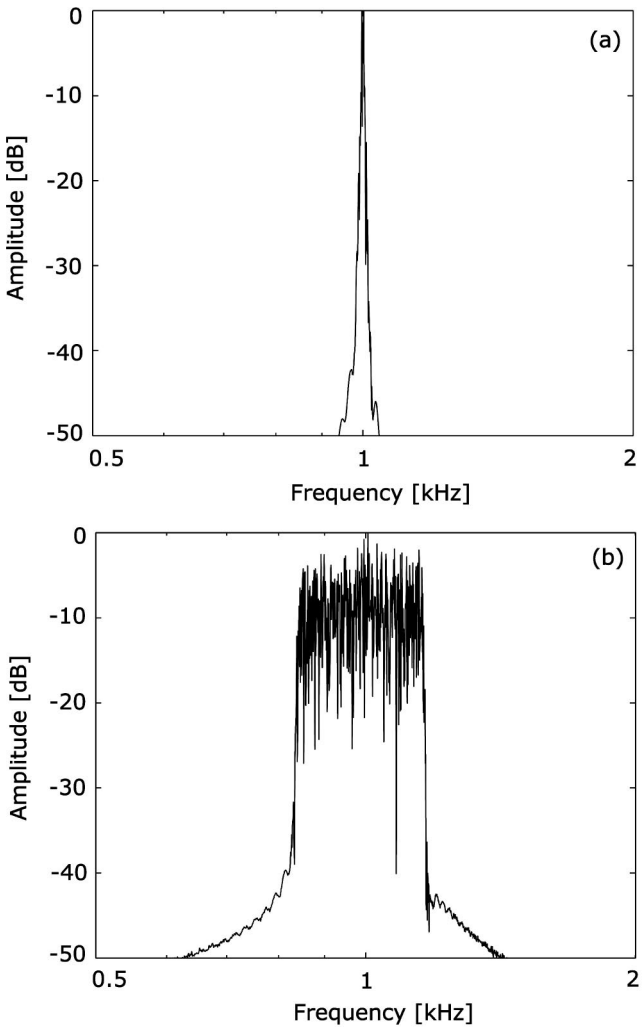


FIG. 1. Power spectra of bandpass noise with (a) 0 Hz bandwidth; and (b) 320-Hz bandwidth of 1000-Hz center frequency.

,The ACF provides the same information as the power spectral density of a signal. A normalized ACF of a signal, $p(t)$, can be expressed by

$$\phi(\tau)=\frac{\Phi(\tau)}{\Phi(0)}, \tag{1}$$

where

$$\Phi(\tau)=\frac{1}{2T}\int_0^{2T} p'(t)p'(t+\tau)dt, \tag{2}$$

where $2T$ is the integral interval, τ is the time delay, and $p'(t)$ is obtained after passing through the A-weighted network, which corresponds approximately to the sensitivity of the ear, $s(t)$, so that $p'(t)=p(t)*s(t)$. From the ACF analysis the following are determined: (1) the energy represented at zero delay, $\Phi(0)$; (2) τ_e, which is defined by the time delay at which the envelope of the normalized ACF becomes 0.1; (3) the time delay of the first maximum peak, τ_1; and (4) its amplitude, ϕ_1, (Ando, 2001). The source signals were characterized by ACF factors, τ_1, ϕ_1, and $\tau_e \cdot \tau_1$ corresponds to the center frequency of the bandpass noise. ϕ_1 and τ_e increase as the filter bandwidth decreases (Fig. 2), and there is a certain degree of coherence between ϕ_1 and τ_e.

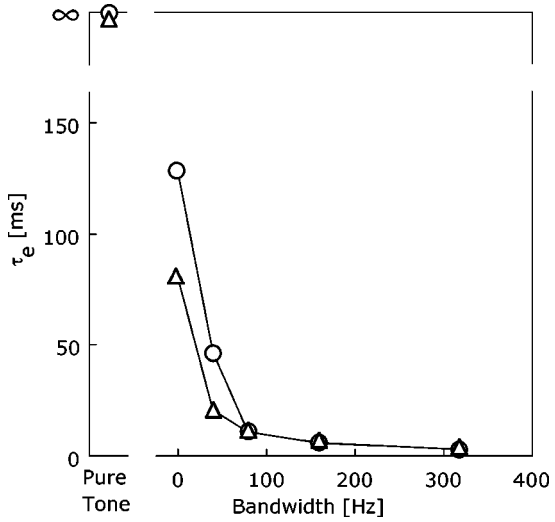


FIG. 2. Measured effective duration of ACF, τ_e of the source signal as a function of the bandwidth. Different symbols indicate different frequencies: (○): 1000 Hz; (Δ): 2000 Hz.

The auditory stimuli were binaurally presented using headphones (Sennheiser HD-340). All stimuli were fixed at the same sound-pressure level [74 dB(A)] by measuring $\Phi(0)$. The sound-pressure level was calibrated by using a dummy head with 1/2-in. condenser-type microphones at both ears. The magnitude of the interaural cross-correlation function (IACC) was kept constant at nearly unity because the signals fed to both ears were identical.

Eight 21-to 23-year-old subjects with no histories of hearing disorders participated in the experiment. They were seated in a dark, soundproof room with a comfortable thermal environment and heard the sound stimuli. The paired-comparison tests were performed for all combinations of the pairs of pure-tone and bandpass noise, i.e., 15 pairs $[N(N-1)/2, N=6]$ of stimuli interchanging the order in each pair per session, with the pairs presented in random order. A total of ten sessions was conducted for each subject. The duration of the stimuli was 2.0 s, the rise and fall times were 50 ms, the silent interval between the stimuli was 1.0 s, and the interval between pairs was 3.0 s, which was the time allowed for the subjects to respond by pushing one of two buttons. They were asked to judge which of two sound signals was more annoying. The scale values of the annoyance to each subject were calculated according to case V of Thurstone's theory (Thurstone, 1927) and the model of case V for all data was reconfirmed by a goodness-of-fit test (Mosteller, 1951).

III. RESULTS

The relationship between the scale values of annoyance and the filter bandwidths is shown in Fig. 3. Statistical significance was determined by a Bonferroni test. For the center frequency of 1000 Hz tested, the averaged scale value of annoyance for pure tones was significantly greater than those of sharply filtered noises under the condition of equal sound-pressure levels ($p<0.01$). For the center frequency of 2000 Hz tested, the scale value of annoyance for the pure tone was significantly greater than that of the bandpass noises with 40-, 80-, and 160-Hz bandwidths ($p<0.05$). There was an

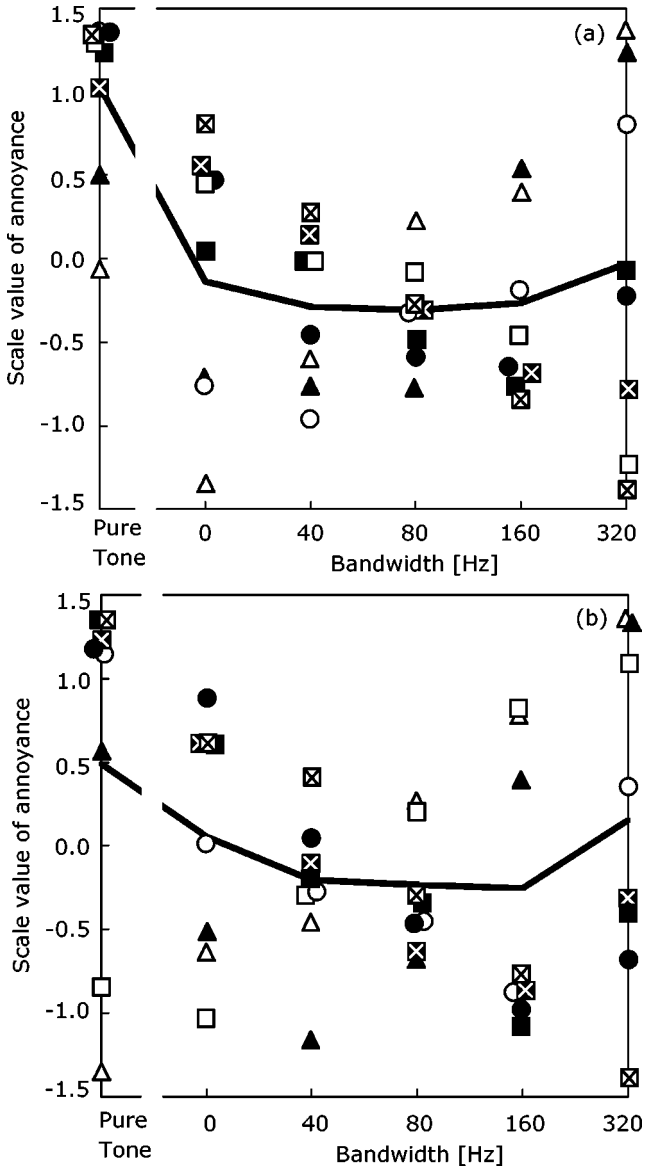


FIG. 3. Scale value of annoyance as a function of the bandwidth of bandpass noise with center frequency of (a) 1000 and (b) 2000 Hz. Each symbol represents one subject. The line represents the mean scale value of eight subjects.

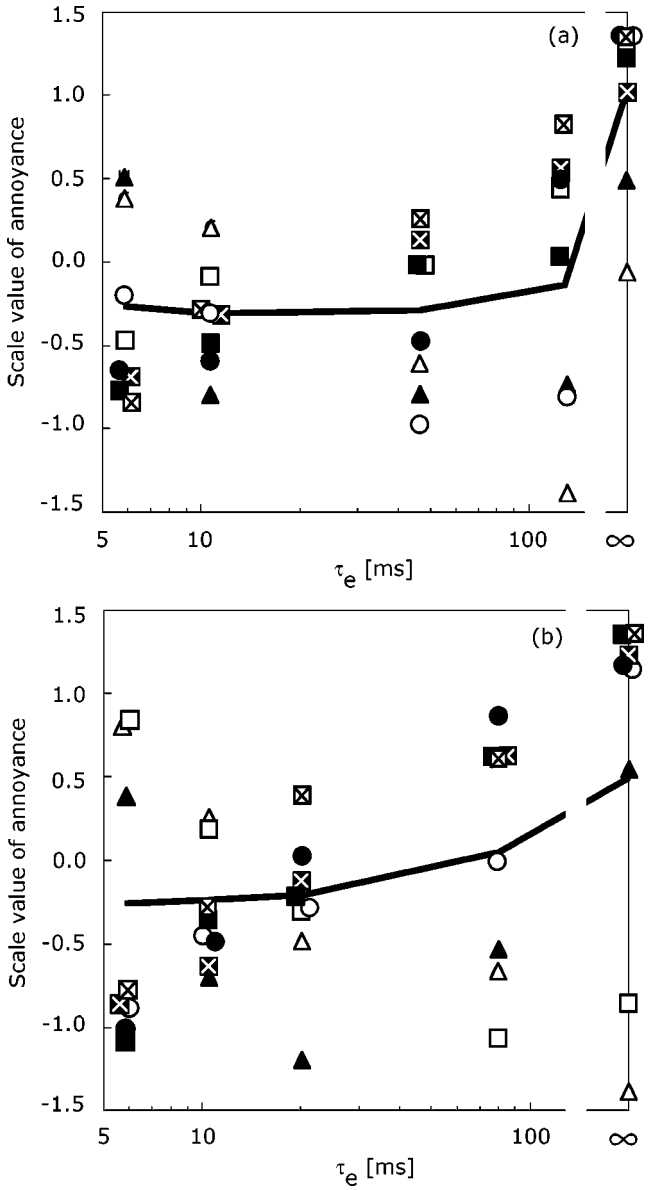


FIG. 4. Scale value of annoyance as a function of the τ_e of bandpass noise with center frequency of (a) 1000 and (b) 2000 Hz. Each symbol represents one subject. The line represents the mean scale value of eight subjects.

agreement among subjects that the most annoying stimulus was a pure-tone or bandpass noise with a 320-Hz bandwidth. Relatively large individual differences were found for bandpass noises with a 320-Hz bandwidth.

IV. DISCUSSION AND CONCLUSIONS

Averaged scale values of annoyance increased as the τ_e of sound source increased within the critical band for both center frequencies of 1000 and 2000 Hz as shown in Fig. 4. The τ_e represents the repetitive feature or tonal component of the source signals. The previous research suggests that tonal components increase the perceived annoyance and noisiness of broadband noise (e.g., Kryter and Pearsons, 1965; Hargest and Pinker, 1967; Fujii *et al.*, 2002). These are consistent with our results. Two of the eight subjects had increasing annoyance from pure tone to band noise with widest bandwidth of a center frequency of 2000 Hz. This could indicate that the effects of τ_e on annoyance depend on the individual.

Loudness increases with increasing bandwidth under the same sound-pressure level conditions beyond the critical band (Zwicker *et al.*, 1957; Greenwood, 1961; Scharf, 1962). A significant increase of annoyance beyond the critical band was not found in our study. Relatively large interindividual variations in annoyance were found for bandpass noises with a 320-Hz bandwidth. The loudness of bandpass noises with center frequencies of 1000 and 2000 Hz was also investigated. The same subjects who had joined in the annoyance tests also participated in the loudness tests. The auditory stimuli were presented in the same way as in the annoyance tests. The subjects were asked to judge which of two sound signals was the louder signal. The relationship between the scale values of loudness and the filter bandwidths is shown

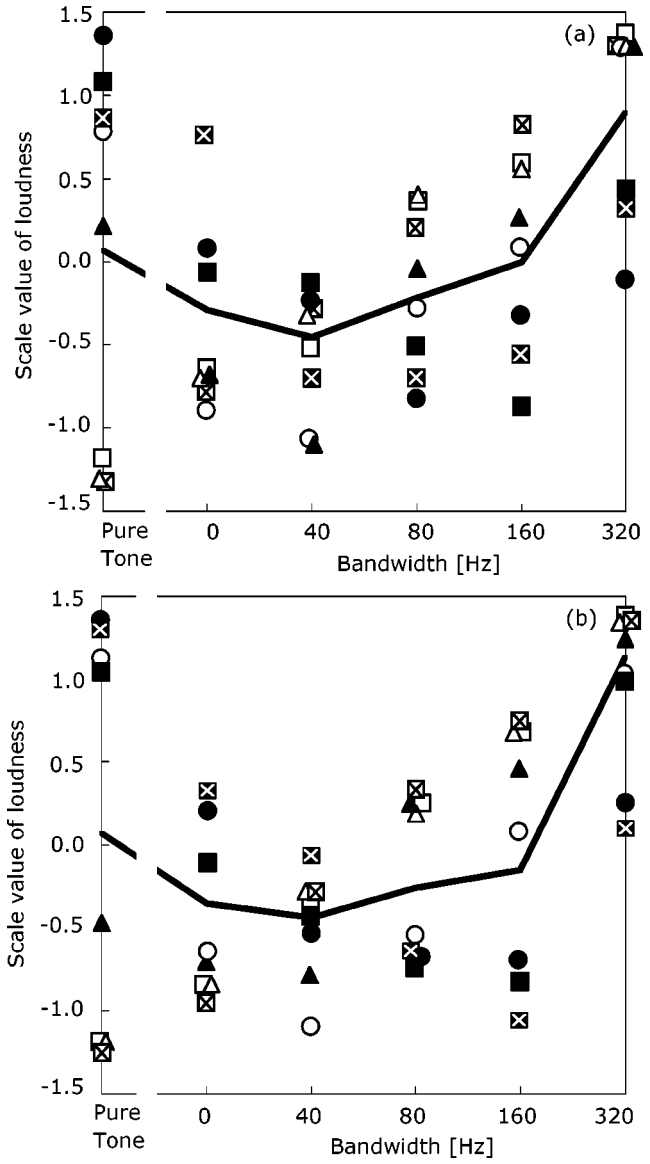


FIG. 5. Scale value of loudness as a function of the bandwidth of bandpass noise with center frequency of (a) 1000 and (b) 2000 Hz. Each symbol represents one subject. The line represents the mean scale value of eight subjects.

in Fig. 5. The averaged scale value of loudness for the bandpass noises with a 320-Hz bandwidth was significantly greater than those for the other bandpass noises for both center frequencies of 1000 and 2000 Hz, which is consistent with previous research by Zwicker *et al.* (1957), Greenwood (1961), Scharf (1962), and Sato *et al.* (2002), one of which showed that the loudness for the pure tone was significantly greater than that of the bandpass noises, and loudness increased with increasing τ_e, although one subject in the study had increasing loudness from pure tone to widest bandwidth (Sato *et al.*, 2002). In this study, three of eight subjects had increasing loudness with increasing bandwidth. The increase of loudness and annoyance by pure tone was not found in two subjects, and this is indicated with an open triangle and square in Figs. 3(b), 4(a), and 4(b). This indicates that there could have been a subject who was hardly affected by the τ_e on loudness or annoyance judgment. That is, the effect of τ_e for loudness or annoyance judgment seems to be partially dependent on the individual.

Almost all the subjects showed an increase of loudness above the critical band. However, the increase of annoyance above the critical band showed individual variations. The increase in annoyance for pure tones was more evident than that of loudness. Loudness and annoyance produced different results. This is consistent with the findings obtained by Berglund *et al.* (1975), Hellman (1982, 1984), and Kuwano *et al.* (1988).

Environmental noise from sources such as airplanes, railways, and so on has been evaluated based on the ACF and IACF theory (Fujii *et al.*, 2001, 2002; Kitamura *et al.*, 2002. The ACF factors τ_e and ϕ_1 had a significant effect on annoyance (Fujii *et al.*, 2002). In this study, annoyance increased as the τ_e of the sound source increased within the critical band. Therefore, factors extracted from the ACF, such as τ_e, could be useful criteria for evaluating environmental noise.

Ando, Y. (**2001**). ''A theory of primary sensations and spatial sensations measuring environmental noise,'' J. Sound Vib. **241**, 3–18.

Berglund, B., Berglund, U., and Lindvall, T. (**1975**). ''Scaling loudness, noisiness, and annoyance of aircraft noise,'' J. Acoust. Soc. Am. **57**, 930–934.

Fujii, K., Soeta, Y., and Ando, Y. (**2001**). ''Acoustical properties of aircraft noise measured by temporal and spatial factors,'' J. Sound Vib. **241**, 69–78.

Fujii, K., Atagi, J., and Ando, Y. (**2002**). ''Temporal and spatial factors of traffic noise and its annoyance,'' J. Temporal Design Arch. Environ. **2**, 33–41; http://www.jtdweb.org.

Greenwood, D. D. (**1961**). ''Auditory masking and critical band,'' J. Acoust. Soc. Am. **33**, 484–502.

Hargest, T. J., and Pinker, R. A. (**1967**). ''The influence of added narrow-band noises and tones on the subjective response to shaped white noise,'' J. R. Aeronaut. Soc. **71**, 428–430.

Hellman, R. P. (**1982**). ''Loudness, annoyance, and noisiness produced by single-tone-noise complexes,'' J. Acoust. Soc. Am. **72**, 62–73.

Hellman, R. P. (**1984**). ''Growth rate of loudness, annoyance, and noisiness as a function of tone location within the noise spectrum,'' J. Acoust. Soc. Am. **75**, 209–218.

Kitamura, T., Shimokura, R., Sato, S., and Ando, Y. (**2002**). ''Measurement of temporal and spatial factors of a flushing toilet noise in a downstairs bedroom,'' J. Temporal Design Arch. Environ. **2**, 13–19; http://www.jtdweb.org.

Kryter, K. D., and Pearsons, K. S. (**1965**). ''Judged noisiness of a band of random noise containing an audible pure tone,'' J. Acoust. Soc. Am. **38**, 106–112.

Kuwano, S., Namba, S., and Fastl, H. (**1988**). ''On the judgment of loudness, noisiness, and annoyance with actual and artificial noise,'' J. Sound Vib. **127**, 457–465.

Mosteller, F. (**1951**). ''Remarks on the method of paired comparisons. III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed,'' Psychometrika **16**, 207–218.

Sato, S., Kitamura, T., and Ando, Y. (**2002**). ''Loudness of sharply (2068 dB/octave) filtered noises in relation to the factors extracted from the autocorrelation function,'' J. Sound Vib. **250**, 47–52.

Scharf, B. (**1962**). ''Loudness summation and spectrum shape,'' J. Acoust. Soc. Am. **34**, 228–233.

Thurstone, L. L. (**1927**). ''A law of comparative judgment,'' Psychol. Rev. **34**, 273–289.

Yost, W. A. (**2000**). *Fundamentals of Hearing; An Introduction* (Academic, San Diego).

Zwicker, E., Flottorp, G., and Stevens, S. S. (**1957**). ''Critical bandwidth in loudness summation,'' J. Acoust. Soc. Am. **29**, 548–557.

Zwicker, E., and Terhardt, E. (**1980**). ''Analytical expressions for critical-band rate and critical bandwidth as a function of frequency,'' J. Acoust. Soc. Am. **68**, 1523–1525.

Eigenmodes of triaxial ellipsoidal acoustical cavities with mixed boundary conditions

M. Willatzen[a)]
Mads Clausen Institute for Product Innovation, University of Southern Denmark, Grundtvigs Alle 150, DK-6400 Sønderborg, Denmark

L. C. Lew Yan Voon[b)]
Department of Physics, Worcester Polytechnic Institute, 100 Institute Road, Worcester, Massachusetts 01609

(Received 19 December 2003; revised 27 September 2004; accepted 27 September 2004)

The linear acoustics problem of resonant vibrational modes in a triaxial ellipsoidal acoustic cavity with walls of arbitrary acoustic impedance has been quasi-analytically solved using the Frobenius power-series expansion method. Eigenmode results are presented for the lowest two eigenmodes in cases with pressure-release, rigid-wall, and lossy-wall boundary conditions. A mode crossing is obtained as a function of the specific acoustic impedance of the wall; the degeneracy is not symmetry related. Furthermore, the damping of the wave is found to be maximal near the crossing. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1819391]

PACS numbers: 43.20.Ks [MO] Pages: 3279–3283

I. INTRODUCTION

The knowledge of analytical acoustic cavity eigenmodes is still important in spite of advances in computational acoustics.[1] Indeed, there remains computational difficulties in obtaining the eigenmodes of irregular three-dimensional cavities using techniques such as the boundary-element method.[2] Nevertheless, the normal modes are known to be analytically obtainable in cavities of various shapes, describable in terms of eleven orthogonal curvilinear coordinate systems.[3] The solutions for rectangular, cylindrical, and spherical cavities are now textbook examples.[4] Other shapes that have been solved exactly are the spheroid,[5] the elliptic cylinder,[6] the parabolic cylinder,[7] and the parabolic rotational lens.[8] Yet, the most general cavity with a one-coordinate surface, the triaxial ellipsoid, has not been solved. Nevertheless, physical models in acoustics using the latter shape abound including, for example, for the study of ocean acoustics,[9] of cavitation in bubbles,[10] and of the human body;[11] these works have mostly dealt with scattering problems. Indeed, the approximation of an irregular body by an ellipsoid (the so-called Brillouin ellipsoid[12]) is common. We also note that previous work on analytic shapes in acoustics were often only done for rigid-wall and pressure-release boundary conditions. One exception is the discussion in Morse and Ingard[13] for rectangular rooms.

The ellipsoidal cavity is fundamentally important since it can be viewed as the limit of a highly distorted spherical cavity and exact results will still hold even when perturbation theory on the sphere is no longer valid. The lowered symmetry compared to a sphere and a spheroid makes it ideal as an analytical model of a nonsymmetric cavity. In addition, the one-parameter nature of the bounding surface also allows a straightforward implementation of a uniform boundary condition. Finally, the cavity might have acoustical characteristics that are more desirable than either the spherical or the spheroidal cavity since the mode degeneracies present in the latter are expected to be completely removed due to the lower shape symmetry.

In this Letter, we present the solution to the triaxial ellipsoidal cavity problem analytically in terms of Lamé wave functions. The construction of the latter functions can be carried out independently of the boundary condition. This allows an efficient computation of the complex eigenfrequencies for walls with arbitrary acoustic impedance.

II. THEORY

The mathematical formulation of the problem is finding the discrete spectrum of the Helmholtz equation,[14]

$$\nabla^2 P(\mathbf{r}) + k^2 P(\mathbf{r}) = 0, \tag{1}$$

subject to the mixed boundary condition on the surface of the ellipsoidal cavity:

$$\eta \frac{\partial P}{\partial n} + ikP = 0, \tag{2}$$

where $P(\mathbf{r})$ is the acoustic pressure field, k the wave number, and η is the specific impedance. The latter is given by $Z/(\rho_0 c_0)$ where Z is the wall acoustic impedance, c_0 is the speed of sound, and ρ_0 the background or equilibrium medium mass density. In order to reduce the number of degrees of freedom in the problem, we will only consider the case where the acoustic impedance is real and frequency independent, and the surface is locally reactive.[13] Nevertheless, the general boundary condition, Eq. (2), is still complex and frequency dependent.

[a)]Electronic mail: willatzen@mci.sdu.dk; phone: +45 65 50 16 82; fax: +45 65 50 16 60.
[b)]Electronic mail: llew@wpi.edu; phone +1 508 831 5249; fax; +1 508 831 5886.

A. Ellipsoidal coordinates and the triaxial ellipsoid

Ellipsoidal coordinates (EC) are defined by three families of orthogonal confocal quadric surfaces. The latter are defined by the equation

$$\frac{x^2}{\xi^2-a^2}+\frac{y^2}{\xi^2-b^2}+\frac{z^2}{\xi^2}=1, \quad a\geqslant b\geqslant 0, \tag{3}$$

where different types of surfaces are obtained for different values of the parameter ξ. Those three types of surfaces are obtained as follows.[3] If $\xi\equiv\xi_1>a$, all three terms on the left hand side of Eq. (3) are positive and the equation describes an ellipsoidal surface. If $a>\xi\equiv\xi_2>b$, the first term on the left hand side of Eq. (3) is negative; the quadric surfaces are then a family of confocal hyperboloids of one sheet. If $b>\xi\equiv\xi_3>0$, the first two terms on the left hand side of Eq. (3) are negative; the quadric surfaces are confocal hyperboloids of two sheets.

The relationship of the EC ξ_1,ξ_2,ξ_3 to the Cartesian ones is[3,15,16]

$$x=\frac{(\xi_1^2-a^2)^{1/2}(\xi_2^2-a^2)^{1/2}(\xi_3^2-a^2)^{1/2}}{a(a^2-b^2)^{1/2}},$$
$$y=\frac{(\xi_1^2-b^2)^{1/2}(\xi_2^2-b^2)^{1/2}(\xi_3^2-b^2)^{1/2}}{b(b^2-a^2)^{1/2}}, \quad z=\frac{\xi_1\xi_2\xi_3}{ab}, \tag{4}$$

with

$$\xi_1>a>\xi_2>b>\xi_3>0. \tag{5}$$

One set of ξ_1,ξ_2,ξ_3 corresponds to eight Cartesian points.

In general, a and b can take on any values subject to the convention given in Eq. (5). It turns out the values are fixed when space is partitioned by an ellipsoid. Thus, let the Cartesian coordinates of the points of intersection of the ellipsoid with the Cartesian axes be $\pm x_0,\pm y_0,\pm z_0$. Then, Eq. (3) for $\xi=\xi_1$ relates x_0,y_0,z_0 to a,b:

$$\xi_{1ref}=z_0, \tag{6}$$

$$a=(\xi_{1ref}^2-x_0^2)^{1/2}, \tag{7}$$

$$b=(\xi_{1ref}^2-y_0^2)^{1/2}, \tag{8}$$

where ξ_{1ref} is the value of ξ_1 on the ellipsoidal surface (recall that ξ_1 alone defines an ellipsoidal surface). We observe that the ordering $a>b>0$ implies $z_0>y_0>x_0$.

B. Separation of variables in ellipsoidal coordinates

The above problem [Eqs. (1)–(2)] is separable in EC. Let

$$\Psi=X_1(\xi_1)X_2(\xi_2)X_3(\xi_3). \tag{9}$$

The insertion of Eq. (9) in Eq. (1) leads to three ordinary differential equations in the separated functions X_i:

$$\sqrt{(\xi_1^2-a^2)(\xi_1^2-b^2)}\frac{d}{d\xi_1}\left[\sqrt{(\xi_1^2-a^2)(\xi_1^2-b^2)}\frac{dX_1}{d\xi_1}\right]$$
$$=(-k^2\xi_1^4+\alpha_2\xi_1^2-\kappa)X_1,$$

$$\sqrt{(a^2-\xi_2^2)(\xi_2^2-b^2)}\frac{d}{d\xi_2}\left[\sqrt{(a^2-\xi_2^2)(\xi_2^2-b^2)}\frac{dX_2}{d\xi_2}\right]$$
$$=-(-k^2\xi_2^4+\alpha_2\xi_2^2-\kappa)X_2, \tag{10}$$

$$\sqrt{(a^2-\xi_3^2)(b^2-\xi_3^2)}\frac{d}{d\xi_3}\left[\sqrt{(a^2-\xi_3^2)(b^2-\xi_3^2)}\frac{dX_3}{d\xi_3}\right]$$
$$=(-k^2\xi_3^4+\alpha_2\xi_3^2-\kappa)X_3,$$

The constants α_2 and κ are separation constants determined by the boundary conditions and the fact that any acceptable solution is finite and differentiable within the ellipsoidal enclosure. Equations (10) can be rewritten as $(i=1,2,3)$

$$(\xi_i^2-a^2)(\xi_i^2-b^2)\frac{d^2X_i}{d\xi_i^2}+\xi_i[2\xi_i^2-(a^2+b^2)]\frac{dX_i}{d\xi_i}$$
$$+[k^2\xi_i^4-\alpha_2\xi_i^2+\kappa]X_i=0, \tag{11}$$

an equation known as the ellipsoidal or Lamé wave equation.[3] By making the transformation,

$$t_i=\frac{\xi_i^2}{b^2}, \tag{12}$$

we obtain the form due to Arscott *et al.*[17]

$$t_i(t_i-1)(t_i-c)\frac{d^2X_i}{dt_i^2}+\frac{1}{2}[3t_i^2-2(1+c)t_i+c]\frac{dX_i}{dt_i}$$
$$+[\lambda+\mu t_i+\gamma t_i^2]X_i=0, \tag{13}$$

where

$$\frac{a^2}{b^2}=c, \quad \kappa=4b^2\lambda, \quad \alpha_2=-4\mu, \quad k^2=\frac{4}{b^2}\gamma. \tag{14}$$

It is always possible to write X_i in the general form[17]

$$X_i(t_i)=t_i^{\rho/2}(t_i-1)^{\sigma/2}(t_i-c)^{\tau/2}F(t_i), \tag{15}$$

where ρ, σ, and τ are either 0 or 1, i.e., eight different types of X_i are possible. The function F must be found using a quasi-analytical approach. Inserting Eq. (15) into Eq. (13) leads to the differential equation[17]

$$t_i(t_i-1)(t_i-c)\frac{d^2F_i}{dt_i^2}+\frac{1}{2}(A_2t_i^2-2A_1t_i+A_0)\frac{dF_i}{dt_i}$$
$$+(\lambda-\lambda_0+(\mu+\mu_0)t_i+\gamma t_i^2)F_i=0, \tag{16}$$

where

$$\lambda_0=\frac{1}{4}[(\rho+\tau)^2+(\rho+\sigma)^2c],$$

$$\mu_0=\frac{1}{4}(\rho+\sigma+\tau)(\rho+\sigma+\tau+1), \quad A_0=(2\rho+1)c, \tag{17}$$

$$A_1=(1+\rho)(1+c)+\tau+\sigma c, \quad A_2=2(\rho+\sigma+\tau)+3.$$

Equation (16) can be solved formally by using the Frobenius power-series expansion method,[8] i.e.,

$$F_i(t_i)=\sum_{r=0}^{\infty}a_r(t_i-t_0)^r, \tag{18}$$

where t_0 is an expansion parameter. The insertion of Eq. (18) into Eq. (16) leads to a five-term recursion formula in the coefficients a_r. First, solutions are found for $\gamma=0$ (the Lamé equation). Solutions exist whenever

$$\mu=-\mu_0-n(n-1)-\frac{1}{2}A_2 n, \tag{19}$$

where n is an integer $(0,1,2,\ldots)$. The possible values of λ are then found by solving an $n+1$-polynomial equation in λ, i.e., $n+1$ real λ solutions exist for each n $(\lambda=\lambda_{nm};m=0,1,2,\ldots,n)$.[17] These solutions give initial values at $\gamma=0$ for the subsequent computation of separation constants $\mu(\gamma)$ and $\lambda(\gamma)$ at finite γ values. Values of $\mu(\gamma)$ and $\lambda(\gamma)$ are obtained next using Newton's method which is known to be locally convergent when employing sufficiently small steps in γ. In the present work, $\Delta\gamma$ is chosen to be 0.01 starting from $\gamma=0$ and t_0 is chosen to be 1 for all the eight cases: ρ,σ,τ equal to 0 or 1.

Once a set of solution parameters γ, μ, and λ has been obtained, the total eigenfunction within one octant of the ellipsoid becomes

$$\Psi(\xi_1,\xi_2,\xi_3)=X(\xi_1)X(\xi_2)X(\xi_3), \tag{20}$$

where X satisfies the ellipsoidal wave equation [Eq. (11)]. A second solution Y to Eq. (11) exists, however, $\vec{\nabla}Y$ is not well-defined everywhere inside the ellipsoid and Y can thus be disregarded as a possible solution. For example, $\vec{\nabla}Y$ diverges as $\xi_1\to a+$, $\xi_2\to a-$, and $\xi_3\to 0+$.

III. NUMERICAL RESULTS AND DISCUSSIONS

Having determined the characteristic values (i.e., μ and λ versus γ), we impose the boundary condition [Eq. (2)] so as to obtain a discrete set of solutions for the separable constants: $(\gamma_i,\mu_i,\lambda_i)$ where $i=1,2,3,4,\ldots$, for each of the eight cases: $\rho=\pm1,\sigma=\pm1,\tau=\pm1$. Once these separation constants are determined, eigenfrequencies and associated eigenfunctions (pressure modes) are easily specified as will be described next. The following relations apply [refer to Eq. (14)]:

$$k_i=\frac{2\sqrt{\gamma_i}}{b},\quad f_i=\frac{\omega_i}{2\pi}=\frac{k_i c_0}{2\pi}=\frac{\sqrt{\gamma_i}c_0}{\pi b}, \tag{21}$$

where k_i, ω_i, and f_i are the wavenumber, angular frequency, and the frequency of pressure mode i, respectively. For finite values of η, the separation constants are complex; then, both the eigenvalues and eigenfunctions are also complex. In the following, we consider an ellipsoidal acoustical enclosure with absolute semiaxes $(x_0,y_0,z_0)=(1.0,1.5,2.0)$ m associated with the parameter values: $\xi_{1ref}=2.0$ m, $a=1.732$ m, $b=1.323$ m, and $c=a^2/b^2=1.714$.

A. Modes in a triaxial ellipsoid with pressure-release walls

In this subsection, it is assumed that the ellipsoidal cavity walls are characterized by a specific impedance equal to zero, i.e., $\eta=0$ corresponding to pressure-release boundary conditions. The general boundary condition then degenerates into a Dirichlet boundary condition:

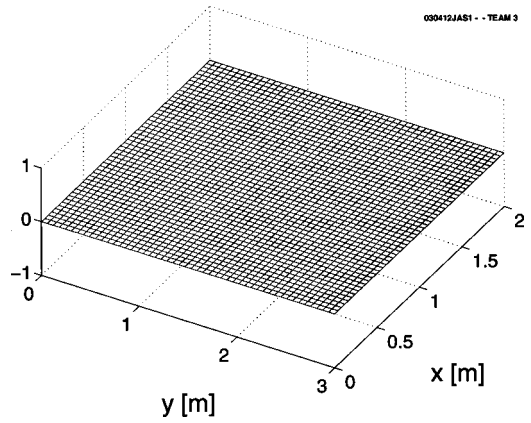


FIG. 1. The fundamental and first-excited modes plotted along three orthogonal planes (from left to right: xy, yz, xz). They correspond pairwise to pressure-release (i.e., first two rows), rigid-wall (middle two rows), and lossy-wall (with specific impedance $\eta=0.6$) boundary conditions. For the latter case, the plots shown are the modulus of the complex eigenmode.

$$\Psi|_{(\xi_1=\xi_{1ref})}=0. \tag{22}$$

The fundamental mode with $\gamma=2.41$, $\rho=\sigma=\tau=0$ is shown as a function of x,y in the $z=0$ plane (left plot), x,z in the $y=0$ plane (middle plot), and y,z in the $x=0$ plane (right plot) in the first row of of Fig. 1. The corresponding eigenfrequency is 128 Hz for a cavity containing air at room temperature (with a sound speed of $c_0=343$ m/s). Note that the fundamental-mode frequency for a spherical enclosure of the same volume is 119 Hz (in agreement with the intuitive rule that the fundamental mode frequency corresponding to Dirichlet boundary conditions must increase with increasing spatial asymmetry at a constant volume). It is evident that this state has no nodes along the three planes and peaks at the center of the ellipsoid similar to what is found for the groundstate of a spherical cavity. It is important to realize that since this mode is obtained as an exact series solution, it is smooth and differentiable. This contrasts to solutions obtained via purely numerical techniques such as finite difference and finite element methods that only give the eigenmodes at the grid or nodal points and, due to the severe memory requirements of a three-dimensional problem, are often not very smooth.

In the second row of Fig. 1, similar plots are shown for the first excited state along the three planes. The associated eigenfrequency is 161 Hz ($\gamma=3.79$) and indices ρ,σ,τ are 1,0,0, respectively. As one observes, the eigenmode is zero in the $z=0$ plane because $\xi_3=0$ when $z=0$, and

$$\Psi(\xi_1,\xi_2,\xi_3)=\frac{\xi_1\xi_2\xi_3}{b^3}F(\xi_1)F(\xi_2)F(\xi_3), \tag{23}$$

where F is the solution to Eq. (16) (note that $t=\xi_i^2/b^2$). However, Ψ is nonzero when plotted in the $y=0$ and $x=0$ planes corresponding to $\xi_2=b>0$ (or $\xi_3=b>0$) and $\xi_1=a>0$ (or $\xi_2=a>0$), respectively. Observe also that the eigenmode is nodeless in both planes. This solution resembles the first excited state for the sphere in having a nodal plane. However, all solutions are found to be nondegenerate in contrast to the spherical case. For the sphere

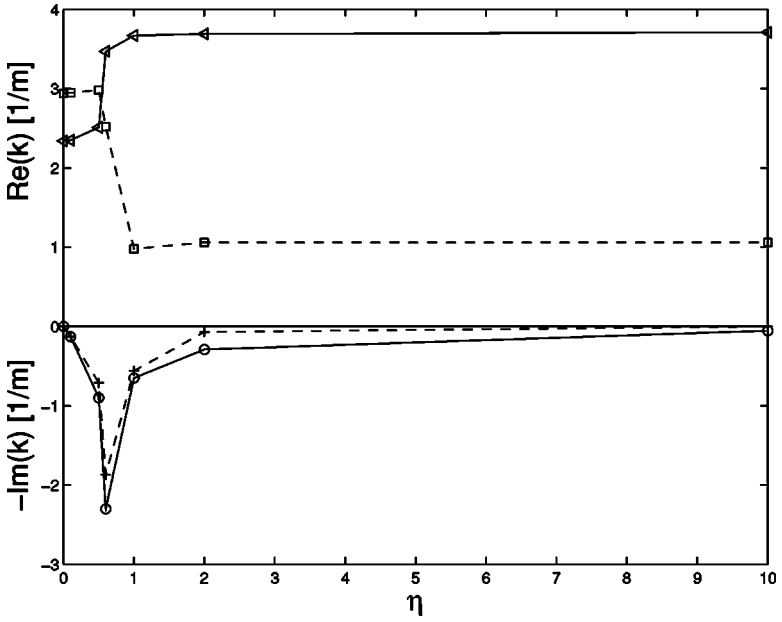


FIG. 2. The real and imaginary part of the wave number as a function of specific impedance for the $(\rho,\sigma,\tau)=(0,0,0)$ (solid line) and $(\rho,\sigma,\tau)=(1,0,0)$ (dashed line) modes.

problem, the first excited state would be three-fold degenerate.

B. Modes in a triaxial ellipsoid with rigid walls

In the third row of Fig. 1, the fundamental mode is plotted along the three planes mentioned above corresponding to the case with rigid walls ($\eta=\infty$), i.e.,

$$\left.\frac{\partial\Psi}{\partial n}\right|_{(\xi_1=\xi_{1ref})}=0, \tag{24}$$

where $\partial\Psi/\partial n$ denotes the partial derivative along the surface normal. This solution corresponds to an eigenfrequency of 58 Hz ($\gamma=0.49$) and the indices n,m, ρ,σ,τ equal to 0,0,1,0,0, respectively (the corresponding frequency for a spherical enclosure having the same volume is 79 Hz). There is another solution with a smaller eigenfrequency, namely $f=0$ ($\gamma=0$). However, this is the trivial solution where $F=$const and $\rho=\sigma=\tau=0$. Since the fundamental mode has the same set of indices in ρ,σ,τ as the first-excited state with pressure-release walls this eigenmode is zero in the $z=0$ plane [refer to Eq. (23)]. The middle and right plots of Fig. 1, third row reveal that the fundamental mode satisfies Neumann boundary conditions on the curved ellipsoid boundary side as it must.

In the fourth row of Fig. 1, the first excited state is shown. This state has an eigenfrequency of 76 Hz ($\gamma=0.84$) and is associated with indices n,m, ρ,σ,τ equal to 0,0,0,1,0. The form of this eigenmode is

$$\Psi(\xi_1,\xi_2,\xi_3)=\left(\frac{\xi_1^2}{b^2}-1\right)^{1/2}\left(\frac{\xi_2^2}{b^2}-1\right)^{1/2}\left(\frac{\xi_3^2}{b^2}-1\right)^{1/2}\times F(\xi_1)F(\xi_2)F(\xi_3), \tag{25}$$

where, again, F is the solution to Eq. (16). Hence, this state is zero in the $y=0$ plane (right plot) where $\xi_2=b$ or $\xi_3=b$. The left and middle plots (fourth row of Fig. 1) show that Neumann boundary conditions are fulfilled along the curved ellipsoid boundary side.

C. Modes in a triaxial ellipsoid with lossy walls

Next, let us consider the general case with a finite, nonzero specific impedance such that the boundary conditions are of mixed type [Eq. (2)]. The variation in the complex wave number is plotted in Fig. 2 for the fundamental mode for $\eta=0$ and the corresponding first excited mode. As is commonly done in plotting complex electronic band structures, we have plotted the real (imaginary) part of the wave number along the positive (negative) y axis. This plot reveals a number of interesting results. It is evident that the calculated eigenfrequencies are real when $\eta=0$ and $\eta=\infty$, however, for a finite, nonzero specific impedance, γ becomes complex due to the mixed boundary condition with an imaginary coefficient. In other words, the pressure mode decays exponentially with time when the specific impedance is finite and acoustic losses take place at the ellipsoidal walls. Furthermore, there is a rapid convergence of the real part of the wave number with η; indeed, the results with $\eta=2$ are less than 2% from the $\eta=\infty$ results. While the real parts for the two modes shown differ significantly, the imaginary parts are quite similar. They both reach a maximum near the crossing of the real parts. The similarity in the damping constants can be understood in terms of earlier results that the damping constant is mostly influenced by the wall impedance. A possible explanation of the near coincidence of the maximum in the damping constant with the mode degeneracy is due to the correspondence of the damping constant with the linewidth of the resonance (in frequency space). Thus, if the resonance width is larger than the resonance separation, the neighboring resonances cannot be resolved and they are quasi-degenerate. We also note that the imaginary part of the wave number is usually much smaller than the real part[13] except near the crossing.

The fundamental mode plotted along the three planes $z=0$, $x=0$, and $y=0$ is shown in the fifth row of Fig. 1. In general, the eigenfunction is complex and we have plotted the modulus. Since this state is characterized by $\rho=\sigma=\tau=0$, this eigenmode has no nodal plane along the three Car-

tesian axes. In the sixth row of Fig. 1, the first-excited state for $\eta=0.6$ is plotted along the three planes. As mentioned earlier, due to $\rho=1$, this state has a nodal plane ($z=0$).

IV. CONCLUSION

A quasi-analytical solution of the vibrational modes in triaxial ellipsoidal cavities was obtained for the case with arbitrary acoustic impedance boundary conditions. Complex wavenumber data and eigenmode plots are given for the cases: ρ,σ,τ equal to 0,0,0 and 1,0,0 being the fundamental-mode indices corresponding to pressure-release and rigid-wall boundary conditions, respectively. A mode crossing of the lowest two modes was found to occur at an intermediate specific acoustic impedance with a corresponding peak in the damping constant.

ACKNOWLEDGMENTS

The Work of LYV was partially supported by a NSF CAREER award (NSF Grant No. 9984059) and a Balslev award (Denmark).

[1] H. Levine, ''Acoustical cavity excitation,'' J. Acoust. Soc. Am. **109**, 2555–2565 (2001).
[2] A. Ali, C. Rajakumar, and S. M. Yunus, ''Advances in acoustic eigenvalue analysis using boundary element method,'' Comput. Struct. **56**, 837–847 (1995).
[3] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953).
[4] H. Lamb, *The Dynamical Theory of Sound* (Dover, New York, 1960).
[5] C. T. M. Chang, ''Natural resonant frequency of a prolate acoustical resonator,'' J. Acoust. Soc. Am. **49**, 611–614 (1971).
[6] K. Hong and J. Kim, ''Natural mode analysis of hollow and annular elliptical cylindrical cavities,'' J. Sound Vib. **183**, 327–351 (1995).
[7] M. Willatzen and L. C. Lew Yan Voon, ''Acoustic eigenmodes in parabolic cylindrical structures,'' in the *Proceedings of the Tenth International Congress on Sound and Vibration*, 7–10 July, Stockholm, Sweden, 2003, pp. 3279–3286.
[8] M. Willatzen and L. C. Lew Yan Voon, ''Acoustic cavity modes in lens-shaped crystals,'' J. Acoust. Soc. Am. **115**, 84–90 (2004).
[9] J. Yan and K. K. Yen, ''A derivation of three-dimensional ray equations in ellipsoidal coordinates,'' J. Acoust. Soc. Am. **97**, 1538–1544 (1995).
[10] M. R. Bailey, D. T. Blackstock, R. O. Cleveland, and L. A. Crum, ''Comparison of electrohydraulic lithotripters with rigid and pressure-release ellipsoidal reflectors. II. Cavitation fields,'' J. Acoust. Soc. Am. **106**, 1149–1160 (1999).
[11] S. G. Conti, P. Roux, D. A. Demer, and J. de Rosny, ''Measurements of the total scattering and absorption cross-sections of the human body,'' J. Acoust. Soc. Am. **114**, 2357 (2003).
[12] R. Garmier and J.-P. Barriot, ''Ellipsoidal harmonic expansions of the gravitational potential: theory and application,'' Celest. Mech. Dyn. Astron. **79**, 235–275 (2001).
[13] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[14] A. D. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications* (McGraw-Hill, New York, 1981).
[15] P. Moon and D. E. Spencer, *Field Theory Handbook* (Springer-Verlag, Berlin, 1961).
[16] G. B. Arfken, *Mathematical Methods for Physicists* (Academic, San Diego, 1966).
[17] F. M. Arscott, P. J. Taylor, and R. V. M. Zahar, ''On the numerical construction of ellipsoidal wave functions,'' in Math. Comput. **40**, 367–380 (1983).

Modeling three-dimensional elastic wave propagation in circular cylindrical structures using a finite-difference approach

Daniel Gsell,[a] Tobias Leutenegger,[b] and Jürg Dual[c]
Institute of Mechanical Systems, ETH Zürich, Swiss Federal Institute of Technology, CH-8092 Zürich, Switzerland

(Received 20 March 2001; revised 25 May 2002; accepted 22 September 2003)

Wave propagation along circular cylindrical structures is important for nondestructive-testing applications and shocks in tubes. To simulate elastic wave propagation phenomena in such structures the governing equations in cylindrical coordinates are solved numerically. To reduce the required amount of computer memory and the computational time, the stress components are eliminated in the equilibrium equations. In the resulting coupled partial differential equations, in which only the three displacement components are involved, the derivatives with respect to spatial coordinates and time are approximated using second order central differences. This leads to the present new approach, which is both accurate and efficient. In order to obtain a stable scheme the displacements must be allocated on a staggered grid. The von Neumann stability analysis is performed and the result is compared with an existing empirical criterion. Mechanical energies are observed in order to validate the finite-difference code. Since no material damping or energy dissipation is taken into account in the equations of motion, the total energy must remain constant over time. Only negligible variations are observed during long-term simulations. Dispersion relations are used to check the physical behavior of the waves calculated with the proposed finite-difference method: Theoretically calculated curves are compared with values obtained by a spectrum estimation method, applied to the results of a simulation. [DOI: 10.1121/1.1625934]

PACS numbers: 43.20.Mv, 43.20.Bi [VWS] Pages: 3284–3293

I. INTRODUCTION

The extensive use of circular cylindrical geometries in engineering structures demands three-dimensional simulations of elastic wave propagation in tubes. The numerical treatment is an important and valuable tool to investigate and study the behavior of different wave modes. To develop and optimize nondestructive-testing (NDT) procedures based on structural waves, accurate data are necessary. Appropriate simulations can be used to substitute physical experiments. This allows thorough testing of signal processing algorithms used in the NDT procedures, e.g., determination of material properties in anisotropic structures.

To simulate wave propagation phenomena with wavelengths of the order of magnitude of the thickness of the structures, a three-dimensional model is necessary. Another application that requires a three-dimensional code is the study of the interaction of elastic waves with local inhomogeneities. For example, Leutenegger *et al.*[1] propose a NDT procedure for the detection of defects based on the present simulation code.

Schubert *et al.*[2] use the finite integration technique to describe the elastic axisymmetric wave propagation in cylindrical coordinates. The three-dimensional code of Chen and Chew[3] is based on a finite-difference scheme, in which central differences are used in space and time. Both methods use velocity and stress components which are distributed on a staggered grid, first introduced for the elastic equations by Madariaga.[4] They discretize the stress-equations of motion and the constitutive equations (with substituted kinematic relations) separately. This leads to nine explicit linear equations. In this paper the authors describe a three-dimensional code in cylindrical coordinates, based on the finite-difference-method (FDM). Instead of calculating the stress and displacement components, which is the common procedure found in literature, the present algorithm discretizes directly the displacement-equations of motion, leading to three explicit linear equations only. The three displacement components are still allocated on a staggered grid. The advantage of this new approach is that the elimination of the six stress components reduces the amount of variables and the required memory to approximately one third. Due to the fact that no stress components have to be calculated, around 25% less computational time is required.

The present method can be used to simulate wave propagation in thin-walled as well as thick-walled tubes and can be applied to any linear elastic material.

The orthotropic case is described by the authors in detail in Sec. II. Stress-free and displacement-fixed boundary conditions are discussed, and numerical results are given in Secs. III and IV. The exact von Neumann stability analysis[5,6] is performed and investigated numerically for the three-dimensional cylindrical case in Sec. V by the authors for the first time. The obtained results are then compared with the empirical stability criterion used by Chen and Chew.[3]

Since no material damping or dissipation is assumed in the implemented model, the total mechanical energy, i.e., the

[a]Electronic mail: daniel.gsell@alumni.ethz.ch
[b]Electronic mail: leutenegger@imes.mavt.ethz.ch
[c]Electronic mail: dual@imes.mavt.ethz.ch

sum of the kinetic and the strain energy, is invariant. This fact is used in Sec. VI to verify the validity of the code.

Ordinarily, analytical solutions of some simple cases are compared with the corresponding outputs of the numerical simulation algorithm to be confirmed. Here, a more general approach is applied. Dispersion relations are used to check the physical behavior of the simulated elastic waves in a broad frequency range (Sec. VII). The theoretically calculated dispersion curves are compared to the ones derived by processing numerical data, generated using the present simulation code. Good agreement is achieved.

All the simulations presented in this paper are performed on a state of the art personal computer.

II. DISCRETIZATION OF THE EQUATIONS OF MOTION

A. General material law

The FDM code described in this investigation can be applied to any material law having the general stress–strain relation

$$\tau_{lm} = C_{lmpq}\epsilon_{pq}, \tag{1}$$

where τ_{lm} and ϵ_{pq}, are the stress and strain components, respectively, and C_{lmpq} the stiffness elements. The Einstein's summation convention is used for repeated indices. For the case of cylindrically orthotropic material, as encountered in filament-wound composite shells, some elements of the stiffness matrix are zero according to Lekhnitskii[7] and Eq. (1) reduces to

$$\begin{pmatrix} \tau_{rr} \\ \tau_{\varphi\varphi} \\ \tau_{zz} \\ \tau_{r\varphi} \\ \tau_{rz} \\ \tau_{\varphi z} \end{pmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ \cdots & c_{22} & c_{23} & 0 & 0 & 0 \\ \cdots & \cdots & c_{33} & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & c_{44} & 0 & 0 \\ \cdots & \text{sym} & \cdots & \cdots & c_{55} & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & c_{66} \end{bmatrix} \cdot \begin{pmatrix} \epsilon_{rr} \\ \epsilon_{\varphi\varphi} \\ \epsilon_{zz} \\ \epsilon_{r\varphi} \\ \epsilon_{rz} \\ \epsilon_{\varphi z} \end{pmatrix}. \tag{2}$$

The principal orthotropy directions of the material coincide with the cylindrical coordinate system. The authors have implemented the orthotropic and isotropic material law in their FDM code. Numerical examples are given in Sec. IV. For simplicity, the basic ideas of the FDM algorithm are described for the case of isotropic material in the next section.

B. Isotropic material

For the isotropic case, the stiffness elements as functions of the Lamé constants λ and μ are given by

$$\begin{aligned} c_{11} &= c_{22} = c_{33} = \lambda + 2\mu, \\ c_{12} &= c_{13} = c_{23} = \lambda, \\ c_{44} &= c_{55} = c_{66} = 2\mu. \end{aligned} \tag{3}$$

The linear kinematic relations and the constitutive equations for a homogeneous linear elastic material[8] are used to express the stress components in terms of the displacement components. They are substituted into the equilibrium conditions expressed in cylindrical coordinates (r,φ,z). This leads to the following three coupled partial differential equations:

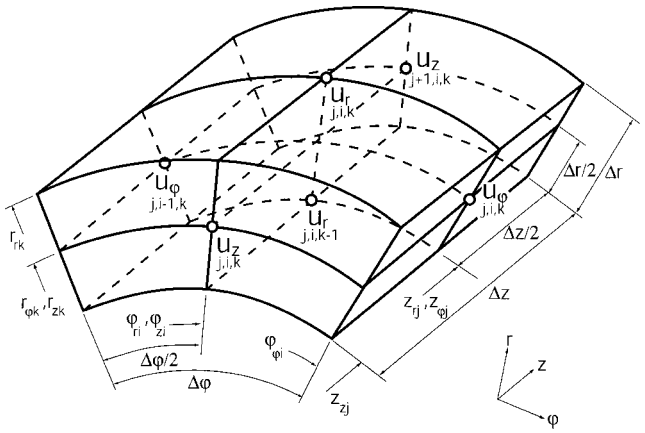


FIG. 1. Allocation of the three displacement components on a staggered grid cell.

$$\begin{aligned} \rho\frac{\partial^2 u_r}{\partial t^2} = &-\frac{(\lambda+2\mu)}{r^2}u_r + \mu\frac{\partial^2 u_r}{\partial z^2} - \frac{(\lambda+3\mu)}{r^2}\frac{\partial u_\varphi}{\partial\varphi} \\ &+ \frac{\mu}{r^2}\frac{\partial^2 u_r}{\partial\varphi^2} + \frac{(\lambda+2\mu)}{r}\frac{\partial u_r}{\partial r} + (\lambda+\mu)\frac{\partial^2 u_z}{\partial r\partial z} \\ &+ \frac{(\lambda+\mu)}{r}\frac{\partial^2 u_\varphi}{\partial r\partial\varphi} + (\lambda+2\mu)\frac{\partial^2 u_r}{\partial r^2}, \end{aligned} \tag{4}$$

$$\begin{aligned} \rho\frac{\partial^2 u_\varphi}{\partial t^2} = &-\frac{\mu}{r^2}u_\varphi + \mu\frac{\partial^2 u_\varphi}{\partial z^2} + \frac{(\lambda+3\mu)}{r^2}\frac{\partial u_r}{\partial\varphi} \\ &+ \frac{(\lambda+\mu)}{r}\frac{\partial^2 u_z}{\partial\varphi\partial z} + \frac{(\lambda+2\mu)}{r^2}\frac{\partial^2 u_\varphi}{\partial\varphi^2} + \frac{\mu}{r}\frac{\partial u_\varphi}{\partial r} \\ &+ \frac{(\lambda+\mu)}{r}\frac{\partial^2 u_r}{\partial r\partial\varphi} + \mu\frac{\partial^2 u_\varphi}{\partial r^2}, \end{aligned} \tag{5}$$

$$\begin{aligned} \rho\frac{\partial^2 u_z}{\partial t^2} = &\frac{(\lambda+\mu)}{r}\frac{\partial u_r}{\partial z} + (\lambda+2\mu)\frac{\partial^2 u_z}{\partial z^2} + \frac{(\lambda+\mu)}{r}\frac{\partial^2 u_\varphi}{\partial\varphi\partial z} \\ &+ \frac{\mu}{r^2}\frac{\partial^2 u_z}{\partial\varphi^2} + \frac{\mu}{r}\frac{\partial u_z}{\partial r} + (\lambda+\mu)\frac{\partial^2 u_r}{\partial r\partial z} + \mu\frac{\partial^2 u_z}{\partial r^2}, \end{aligned} \tag{6}$$

in which u_r, u_φ, and u_z denote the radial, tangential, and axial displacements, and ρ the mass density.

The partial differential Eqs. (4)–(6) depend on the three displacement components only and are discretized using second order central differences in time and space. The required amount of computer memory is reduced remarkably, as none of the six stress components are needed in the discretized equations.

Despite the fact that no stress components are involved, a staggered grid is used in order to obtain a stable scheme. Figure 1 shows the staggered grid cell with the positions of the three displacement components.

The derivatives with respect to r, φ, and z in Eqs. (4)–(6) are discretized and expressed in terms of values of the surrounding grid points. If a displacement component does

not exist at the position of interest, its value is calculated using linear interpolation. The discretization rules needed for the equation of motion in radial direction [Eq. (4)] are given in the appendix [Eq. (A1)] as an example. The discretization rules for Eqs. (5) and (6) are obtained accordingly.

The second derivatives of the displacements u_ξ (in which ξ represents either r, φ, or z) with respect to the time t are discretized using second order central differences in time

$$\frac{\partial^2 u_\xi}{\partial t^2} = \frac{1}{\Delta t^2}\left(\underset{j,i,k}{u_\xi^{n+1}} - 2\underset{j,i,k}{u_\xi^{n}} + \underset{j,i,k}{u_\xi^{n-1}} \right), \tag{7}$$

in which Δt is the time step and n denotes the index of the time

$$t_n = n\Delta t. \tag{8}$$

The indices j, i, and k define the axial, circumferential, and radial positions $z_{\xi j}$, $\varphi_{\xi i}$, and $r_{\xi k}$ of the three displacement components u_ξ on the grid (Fig. 1)

$$z_{\xi j} = j\Delta z + \delta_{\xi\varphi}\frac{\Delta z}{2} + \delta_{\xi r}\frac{\Delta z}{2}, \quad j = 0,1,...,j_{\max}, \tag{9}$$

$$\varphi_{\xi i} = i\Delta\varphi + \delta_{\xi\varphi}\frac{\Delta\varphi}{2}, \quad i = 0,1,...,i_{\max}, \tag{10}$$

$$r_{\xi k} = R - \frac{h}{2} + k\Delta r + \delta_{\xi r}\frac{\Delta r}{2}, \quad k = 0,1,...,k_{\max}, \tag{11}$$

with

$$\Delta z = \frac{l}{j_{\max}}, \quad \Delta\varphi = \frac{2\pi}{(i_{\max}+1)}, \quad \text{and} \quad \Delta r = \frac{h}{k_{\max}}. \tag{12}$$

Wall thickness, radius of the midplane, and length of the tube are denoted by h, R, and l, respectively. The Kronecker symbols $\delta_{\xi\varphi}$ and $\delta_{\xi r}$ in Eqs. (9)–(11) lead to different results for the coordinates of each displacement component, as a result of half-cell shifts that occur due to the staggered grid.

Substituting the discretization rules into Eqs. (4)–(6) and solving for

$$\underset{j,i,k}{u_r^{n+1}}, \quad \underset{j,i,k}{u_\varphi^{n+1}}, \quad \text{and} \quad \underset{j,i,k}{u_z^{n+1}}$$

yields the following linear equations:

$$\underset{j,i,k}{u_r^{n+1}} = -\underset{j,i,k}{u_r^{n-1}} + \underset{k}{\mathbf{a}_r^T} \cdot \underset{j,i,k}{\mathbf{u}_s^{n}}, \tag{13}$$

$$\underset{j,i,k}{u_\varphi^{n+1}} = -\underset{j,i,k}{u_\varphi^{n-1}} + \underset{k}{\mathbf{a}_\varphi^T} \cdot \underset{j,i,k}{\mathbf{u}_s^{n}}, \tag{14}$$

$$\underset{j,i,k}{u_z^{n+1}} = -\underset{j,i,k}{u_z^{n-1}} + \underset{k}{\mathbf{a}_z^T} \cdot \underset{j,i,k}{\mathbf{u}_s^{n}}, \tag{15}$$

where the superscript T refers to the transposed vectors. The elements of the vectors $\mathbf{a}_r$, $\mathbf{a}_\varphi$, and $\mathbf{a}_z$ are functions of material properties, discretization parameters, and radius $r_{\xi k}$. The vector $\underset{j,i,k}{\mathbf{u}_s^{n}}$ contains displacement components of the grid region $j\pm1$, $i\pm1$, $k\pm1$ at the time step n. Since the vector elements of $\mathbf{a}_r$, $\mathbf{a}_\varphi$, and $\mathbf{a}_z$ depend on the radial coordinate $r_{\xi k}$, they must be calculated separately for each radial layer of the grid. Equations (13)–(15) represent three uncoupled, explicit expressions of the displacements of each grid point at the time step $n+1$, using the displacement of this grid point at the time steps $n-1$ and n, and the values of the surrounding grid points at the time step n. As an example the expanded Eq. (13) is given in the Appendix B [Eq. (B1)] for the isotropic case.

For the orthotropic case, the only difference occurs in the elements of the vectors $\mathbf{a}_r$, $\mathbf{a}_\varphi$, and $\mathbf{a}_z$, which are now functions of the nine different stiffness elements c_{11} to c_{66}. Equation (B2) in Appendix B is the expanded Eq. (13) for the orthotropic case.

III. BOUNDARY CONDITIONS

The finite-difference approach is based on the use of the values of neighboring grid points. The grid points describing the structural boundaries are lacking such neighbors. In the present case, this means that some of the necessary grid components of the vector $\underset{j,i,k}{\mathbf{u}_s^{n}}$ are not defined. Therefore, the concept of a fictitious layer next to the surfaces outside the structure is introduced. Displacement values at grid points in the fictitious layer must be calculated according to the corresponding boundary conditions. As examples, stress-free and displacement-fixed boundaries are discussed below.

A. Stress-free boundary condition

The inner as well as the outer surface of the tubes studied in the present paper are stress free. Thus, the following stress components are zero:

$$\tau_{rr}\left(R \pm \frac{h}{2}\right) = \tau_{r\varphi}\left(R \pm \frac{h}{2}\right) = \tau_{rz}\left(R \pm \frac{h}{2}\right) = 0. \tag{16}$$

These stress components can be described using Hooke's law and the kinematic relations. The derivatives of the displacements with respect to the spatial coordinates are approximated with central differences. This leads to

$$\underset{j,i,k}{\tau_{r\xi}^{n}} = \underset{k}{\mathbf{e}_\xi^T} \cdot \underset{j,i,k}{\mathbf{u}_s^{n}}, \tag{17}$$

in which the vector $\mathbf{e}_\xi$ is a function of the material properties, the discretization parameters, and the radius $r_{\xi k}$. The vector $\underset{j,i,k}{\mathbf{u}_s^{n}}$ contains the displacement components of the grid region $j\pm1$, $i\pm1$, $k\pm1$ at the time step n.

For example, the grid is chosen in a way that the components $\underset{j,i,0}{u_z}$ and $\underset{j,i,0}{u_\varphi}$ are located at the inside surface of the tube. To calculate these displacements at the next time step, fictitious displacement components are necessary. For example, the displacement $\underset{j,i,-1}{u_r^{n}}$, which is allocated outside of the structure, is calculated by substituting Eq. (16) into Eq. (17). This results in

$$u_{r\,j,i,-1}^{n} = e_1\, u_{r\,0\,j,i,1}^{n} + e_2\, u_{\varphi\,0\,j,i,-1,0}^{n} + e_3\, u_{\varphi\,0\,j,i+1,0}^{n} + e_4\, u_{z\,0\,j-1,i,0}^{n} + e_5\, u_{z\,0\,j+1,i,0}^{n}. \tag{18}$$

Analogously, all fictitious components are determined. Due to the concept of fictitious layers, Eqs. (13)–(15) can still be used to calculate the components at the boundaries.

B. Displacement-fixed boundary condition

To fulfill the displacement-fixed boundary condition, all displacements at the boundary must vanish. Depending on the chosen grid, some of the displacement components are situated directly at the boundary and can therefore be set to zero. The missing components outside the structure are calculated using linear interpolation, in a way that the values at the boundaries become zero.

For example, the tube is fixed at the ends, and the grid is set such that the components $u_{z\,0,i,k}$ are located at the boundaries. The fictitious component $u_{\varphi\,-1,j,k}^{n}$ is necessary to calculate $u_{\varphi\,0,i,k}^{n+1}$ and is given by

$$u_{\varphi\,-1,i,k}^{n} = -\,u_{\varphi\,0,j,k}^{n}. \tag{19}$$

All other components are determined accordingly.

IV. NUMERICAL RESULTS

To demonstrate the capabilities of the present FDM code, two numerical examples are given below.

For both cases five sine cycles of a particular frequency (center frequency) are multiplied by a Hanning window and chosen as the forcing function.

The structure can be excited by either driving displacement or stress components at the source. In the first case, the desired displacement components are simply given as a function of time. To achieve a stress source at the surface, the corresponding displacement components in the fictitious layer must be defined, since all stress components are eliminated. The necessary fictitious displacement values are calculated by solving Eq. (17).

A. Point source on an orthotropic tube

An orthotropic tube of radius $R=0.02$ m, thickness $h=0.002$ m, length $l=0.5$ m, and density $\rho=1549$ kg/m^3 is used as the test specimen. The computational domain consists of 8, 140, and 500 grid cells in the radial, circumferential, and axial direction, respectively. A filament-would carbon fiber reinforced Epoxy matrix is chosen as material. Reasonable stiffness coefficients have been suggested by Dual[9] and are given below in N/m^2,

$$\begin{aligned} &c_{11}=1.03\times10^{10}, \quad c_{22}=1.03\times10^{10}, \quad c_{33}=1.34\times10^{11},\\ &c_{12}=3.27\times10^{9}, \quad c_{13}=5.58\times10^{9}, \quad c_{23}=5.58\times10^{9},\\ &c_{44}=6.6\times10^{9}, \quad c_{55}=9.8\times10^{9}, \quad c_{66}=1.0\times10^{10}. \end{aligned} \tag{20}$$

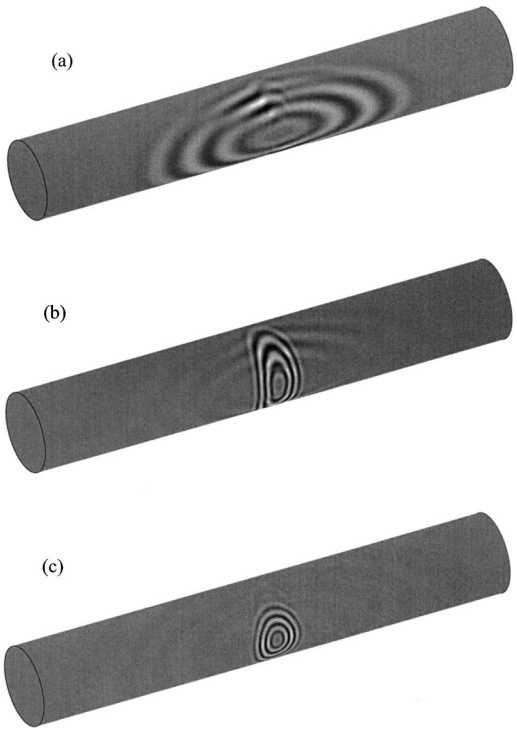


FIG. 2. Axial (a), tangential (b), and radial (c) displacements at the midplane of an orthotropic tube at 20 μs after the initiation of the excitation. A stress source with a center frequency of 250 kHz is implemented at one point. The radius is enlarged for better visualization.

In this simulation, the three surface stress components at one point are given as a function of time. These stress values, which are chosen equivalent in this case, are used to calculate the three driving displacement components at one single grid point of the fictitious layer. The excitation function of the stress source has a center frequency of 250 kHz.

Figure 2 shows the axial (a), tangential (b), and radial displacements (c) of the traveling wave modes 20 μs after the initiation of the excitation. Different wave speeds in the various directions of the traveling waves for the three displacement components occur. This is the characteristic behavior for an orthotropic material.

B. Axisymmetric excitation of an isotropic tube

An aluminum tube (radius $R=0.016$ m, thickness $h=0.002$ m, length $l=0.3$ m, density $\rho=2700$ kg/m^3, and the Lamé constants $\lambda=5.25\times10^{10}$ N/m^2 and $\mu=2.71\times10^{10}$ N/m^2) is excited axisymmetrically. In the radial and axial directions, 20 and 500 cells are used, respectively. Due to the axisymmetry, the number of circumferential grid cells is irrelevant. In this simulation the axial displacements at the left end are given as a function of time, while the other components remain fixed. The excitation function has a center frequency of 400 kHz.

Figure 3 shows the radial displacement at two different times. Figure 3(a) displays the tube at 15 μs, right after the excitation is over, while Fig. 3(b) shows the radial components after 40 μs. Two wave modes travel at different wave speeds. This occurs because the excitation frequency is between the first and second cutoff frequency of axisymmetric modes.

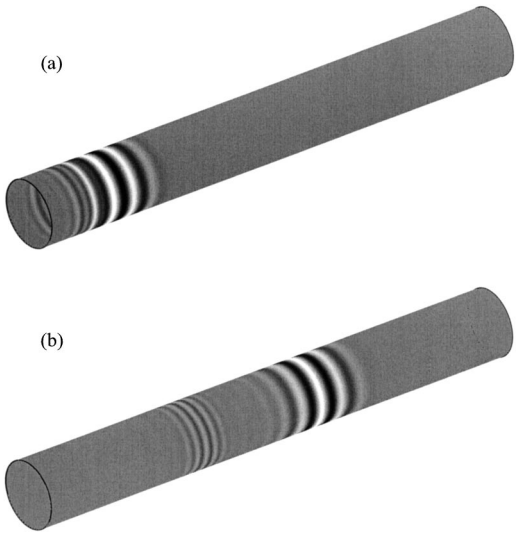


FIG. 3. Radial displacement at the outer surface of an aluminum tube 15 μs (a) and 40 μs (b) after the initiation of the excitation. The structure is excited axisymmetrically at the left end with a center frequency of 400 kHz. For better visualization, the radius is enlarged.

In Fig. 4 the radial (a) and axial (b) displacements of the slower wave mode at 40 μs are plotted versus the tube thickness. Since the cross sections do not remain plane at higher frequencies and the radial displacement is not constant over the thickness of the tube, the best way for an accurate simulation of the wave propagation is the three-dimensional theory.

V. STABILITY

Different stability criteria must be satisfied in order to obtain a stable and accurate FDM code. The wave with the shortest wavelength $\lambda_{\min}$ propagating along the structure must be sampled in space with sufficient resolution.[2,10] This determines the largest dimension of a grid cell, leading to the following criterion for a three-dimensional staggered grid in cylindrical coordinates:

$$\frac{\lambda_{\min}}{\max(\Delta r, \Delta\varphi r_{\max}, \Delta z)} > 8. \tag{21}$$

The other critical value is the size of the time step Δt. The maximal value of Δt to achieve stable conditions depends on material properties, geometry, and spatial discretization parameters. To determine the critical size of the time

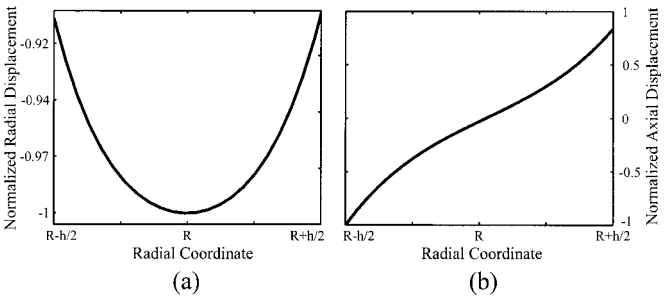


FIG. 4. Radial (a) and axial (b) displacements of the slower wave mode of Fig. 3(b). The displacements are normalized and plotted versus the radial coordinate.

step Δt, the von Neumann stability analysis[5,6] is applied to the present three-dimensional FDM code. The major steps are given below.

The harmonic wave ansatz

$$\underset{j,i,k}{u_\xi^n} = \hat{u}_\xi^n e^{I(k\Delta r k_r + i\Delta\varphi k_\varphi + j\Delta z k_z)}, \tag{22}$$

in which I represents the imaginary unit, and k_r, k_φ, and k_z are the components of the wave vector $\mathbf{k}$, is substituted into Eqs. (13)–(15). This leads to

$$\mathbf{u}^{n+1} = -\mathbf{I}\cdot\mathbf{u}^{n-1} + \mathbf{A}\cdot\mathbf{u}^{n}, \tag{23}$$

with $\hat{\mathbf{u}} = \{\hat{u}_r, \hat{u}_\varphi, \hat{u}_z\}^T$ and $\mathbf{I}$ is the 3×3-identity matrix. The matrix $\mathbf{A}$ depends on discretization, material, and geometrical parameters and the wave vector $\mathbf{k}$. Equation (23) can also be written as

$$\mathbf{w}^{n+1} = \mathbf{G}\cdot\mathbf{w}^{n}, \tag{24}$$

where

$$\mathbf{w}^{n} = \begin{Bmatrix} \mathbf{u}^{n} \\ \mathbf{u}^{n-1} \end{Bmatrix} \quad \text{and} \quad \mathbf{G} = \begin{bmatrix} \mathbf{A} & -\mathbf{I} \\ \mathbf{I} & \mathbf{0} \end{bmatrix}.$$

The 6×6 matrix $\mathbf{G}$ is called the amplification matrix.

Strikwerda[11] proved, that the following condition for all eigenvalues Λ_m of the amplification matrix $\mathbf{G}$ is required to guarantee stability:

$$|\Lambda_m| \leqslant 1 + g\Delta t, \quad m = 1,...,6, \tag{25}$$

in which g is a constant independent of Δt, Δr, $\Delta\varphi$, Δz, and $\mathbf{k}$. To the authors knowledge the constant g cannot be calculated and its value is usually not specified in literature.

Equation (25) is the general von Neumann condition[12] and must be satisfied, if the amplification matrix $\mathbf{G}$ is a function of the discretization parameters. If $\mathbf{G}$ is independent of Δt, Δr, $\Delta\varphi$, and Δz, the restricted von Neumann condition

$$|\Lambda_m| \leqslant 1 \tag{26}$$

must be fulfilled.

Fellinger *et al.*[6] solved Eq. (26) analytically for the case of a three-dimensional code for isotropic material in Cartesian coordinates (x_1, x_2, x_3). They obtained the following criterion for the time step:

$$\Delta t \leqslant \frac{1}{c_p \sqrt{\frac{1}{\Delta x_1^2} + \frac{1}{\Delta x_2^2} + \frac{1}{\Delta x_3^2}}}, \quad \text{with } c_p = \sqrt{\frac{\lambda + 2\mu}{\rho}}. \tag{27}$$

The same result was derived by Taflove.[13]

Chen and Chew[3] used the result of Taflove and Fellinger [Eq. (27)] to get the following empirical requirement for cylindrical coordinates:

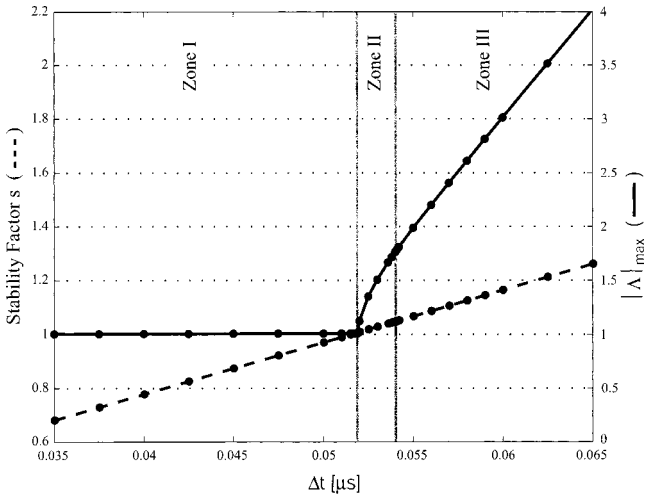


FIG. 5. Comparison of the stability factors (dashed line) and the largest absolute eigenvalue $|\Lambda|_{\max}$ (solid line) for different time steps Δt.

$$s = \Delta t c_{\max} \sqrt{\frac{1}{\Delta r^2} + \frac{1}{(r_{\min}\Delta\varphi)^2} + \frac{1}{\Delta z^2}} \leqslant 1, \tag{28}$$

where s is called the stability factor.

The performance of Eq. (28) is investigated by the present authors by comparing the numerically calculated eigenvalues of the amplification matrix **G** with the corresponding values s.

Figure 5 shows the comparison between the stability factor s and the largest absolute eigenvalue $|\Lambda|_{\max}$ of the amplification matrix **G** for an aluminum tube. While geometrical ($R = 0.016$ m, $h = 0.002$ m, and $l = 0.08$ m), material ($\rho = 2700$ kg/m^3, $\lambda = 5.25 \times 10^{10}$ N/m^2, and $\mu = 2.71 \times 10^{10}$ N/m^2), and spatial discretization parameters (6, 120, and 100 cells in the radial, circumferential, and axial directions, respectively) are held constant, the time step Δt is varied. For each time step, the largest eigenvalue $|\Lambda|_{\max}$ (which is always larger than one) and the stability factor s are calculated (black dots) and are plotted in Fig. 5. Then the simulations with the corresponding parameters are run and observed visually to see, whether they are stable or not. As excitation all three displacement components at one point are driven with a center frequency of 200 kHz.

The region of Fig. 5 can be divided into three zones. Zone I is the area in which the stability factor s is smaller than one and the absolute eigenvalues are slightly larger than one and approximately constant. In this region stability is always guaranteed.

In zone II the eigenvalues increase remarkably and s is slightly larger than one. While the slope of the stability factor curve is approximately constant, an obvious change in the gradient of the eigenvalues can be observed. The FDM scheme is still stable in region II.

In zone III stability cannot be achieved any longer. There is no remarkable change in the stability factor curve or the eigenvalue curve at the border between zone II and III.

Figure 5 shows that stability can even be achieved if the stability factor s is slightly larger than one. This was also noticed by Chen and Chew,[3] who experienced that their criterion [Eq. (28)] could be relaxed. The reason is that their stability condition is based on the result of Fellinger[6] and Taflove,[13] who used the restricted von Neumann criterion [Eq. (26)] instead of the general criterion [Eq. (25)].

However one must be careful when using stability factors larger than one. Whereas the limit between zone I and II is always at the time step where s equals one, the border between zone II and III depends on material, geometrical, and discretization parameters. Therefore, no exact upper limit of s can be determined to assure stability.

The authors have used Eq. (28) extensively as a stability criterion in their simulations and experienced that stability is always guaranteed for values of $s \leqslant 1$.

VI. ENERGY

Since the implemented model is nondissipative, a reasonable method to test the validity of the FDM code is the observation of the mechanical energy. For a linear elastic material, the total energy remains constant after the excitation of the structure, because the system is conservative.

The total energy E is given by

$$E = K + U, \tag{29}$$

in which K is the kinetic and U the strain energy,[14]

$$K = \frac{1}{2}\int_V \rho \frac{du_\xi}{dt}\frac{du_\xi}{dt} dV, \tag{30}$$

$$U = \frac{1}{2}\int_V \tau_{\xi\eta}\epsilon_{\xi\eta} dV. \tag{31}$$

The Einstein's summation convention is implied for repeated indices.

A. Kinetic energy

In a first step the discrete kinetic energy is considered to be constant within a grid cell

$$\underset{j,i,k}{dK} = \frac{\rho}{2} \underset{j,i,k}{v_\xi^m}\ \underset{j,i,k}{v_\xi^m}\ \underset{j,i,k}{dV}. \tag{32}$$

The volume for one cell is

$$\underset{j,i,k}{dV} = ((r_{rk})^2 - (r_{rk-1})^2)\frac{\Delta\varphi}{2}\Delta z. \tag{33}$$

and $\underset{j,i,k}{v_\xi^m}$ are the velocity components at the point $(r_{\varphi k}, \varphi_{ri}, z_{rj})$ in the middle of a cell (the index m defines components located at this middle point). The velocities at the grid points are calculated using central differences in time

$$\overset{n}{\underset{j,i,k}{v_\xi}} = \frac{1}{2\Delta t}\left(\overset{n+1}{\underset{j,i,k}{u_\xi}} - \overset{n-1}{\underset{j,i,k}{u_\xi}}\right). \tag{34}$$

By using linear interpolation of these values, the components $\underset{j,i,k}{v_r^m}$, $\underset{j,i,k}{v_\varphi^m}$, and $\underset{j,i,k}{v_z^m}$, are determined.

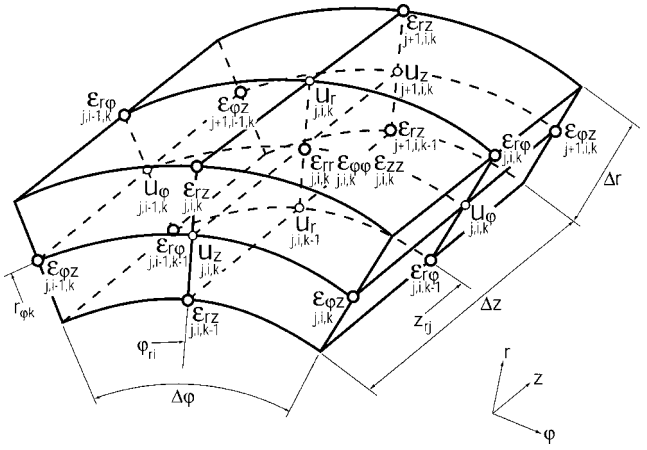


FIG. 6. Allocation of the strain components on the grid.

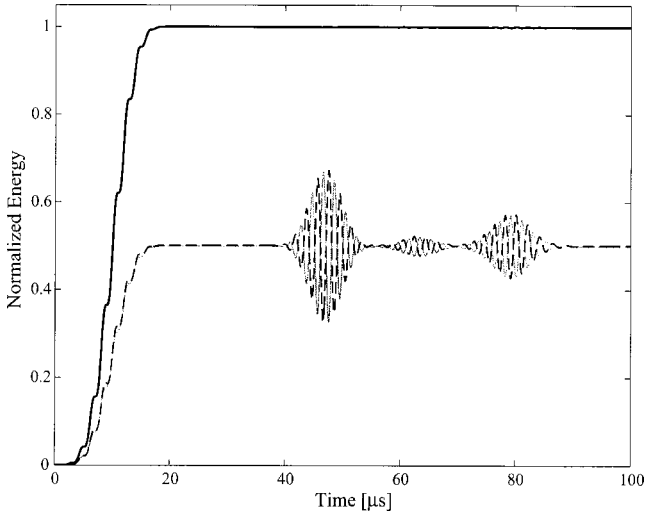


FIG. 7. In an aluminum tube, axisymmetric waves are excited at 250 kHz. Total energy (black solid line), strain energy (gray solid line), and kinetic energy (black dashed line) are plotted versus time.

B. Strain energy

For the same cell the discrete strain energy (constant within the grid cell) is approximated as

$$dU_{j,i,k} = \tfrac{1}{2}\tau^{m}_{\xi\eta\,j,i,k}\,\epsilon^{m}_{\xi\eta\,j,i,k}\,dV_{j,i,k}. \tag{35}$$

The strain components $\epsilon_{\xi\eta}$ at the positions shown in Fig. 6 are calculated with the kinematic relations and central differences in space domain.

The strain components $\epsilon^{m}_{\xi\eta\,j,i,k}$ are derived by linear interpolation of the known values. The stress components $\tau^{m}_{\xi\eta\,j,i,k}$ are calculated using the stress–strain relations [Eq. (1)].

C. Total energy

The summation over all cells of the discrete mechanical energies leads to the total energy E of the structure. This yields

$$E = \sum_{j}\sum_{i}\sum_{k}\left(dK_{j,i,k} + dU_{j,i,k}\right). \tag{36}$$

Figure 7 shows the normalized total energy, the strain energy, and the kinetic energy as a function of time, calculated during a simulation. In this case, an aluminum tube with $R = 0.016$ m, $h = 0.002$ m, and $l = 0.9$ m is used as specimen (same material data as in the axisymmetric example in Sec. IV). An axisymmetric displacement excitation in radial and axial direction is chosen 0.2 m from the left end. The center frequency of the forcing function (5 sine cycles in a Hanning window) is set to 250 kHz, resulting in two propagating wave modes. The boundary conditions are displacement fixed at both ends. The cell sizes in the radial and axial directions are $\Delta r = 0.2$ mm and $\Delta z = 0.2$ mm.

The total energy remains constant once the waves are excited (Fig. 7). Energy interchanges between strain and kinetic energy can be observed in Fig. 7 and in its enlargement (Fig. 8). The first and third changeovers are attributed to the reflections of the two different wave modes, traveling at different speeds, at the left boundary of the tube. The second interchange occurs due to the crossing of the faster (and already reflected) wave mode with the slower one. The small ripples in the total energy in Fig. 7 (around 80 μs) are due to the approximation that the energy is assumed constant within a grid cell. However, the ripples vanish as the grid cells are chosen smaller. More accurate results can be established, if the energy is approximated by a linear distribution (in r, φ, and z) of the displacements within the cell.

The authors have performed several long-term simulations to test, whether the total energy increases, decreases, or remains constant.

As an example, an aluminum tube ($R = 0.016$ m, $h = 0.002$ m, $l = 0.3$ m, $\rho = 2700$ kg/m^3, $\lambda = 5.25\times10^{10}$ N/m^2, and $\mu = 2.71\times10^{10}$ N/m^2) is driven by a displacement point source in all three directions. The spatial discretization parameters are 8, 140, and 450 cells in the radial, circumferential, and axial directions, respectively. The total energies E_1 and E_2 are calculated at two different times t_1 and t_2. In this example, the time difference is 4.235 ms which corre-

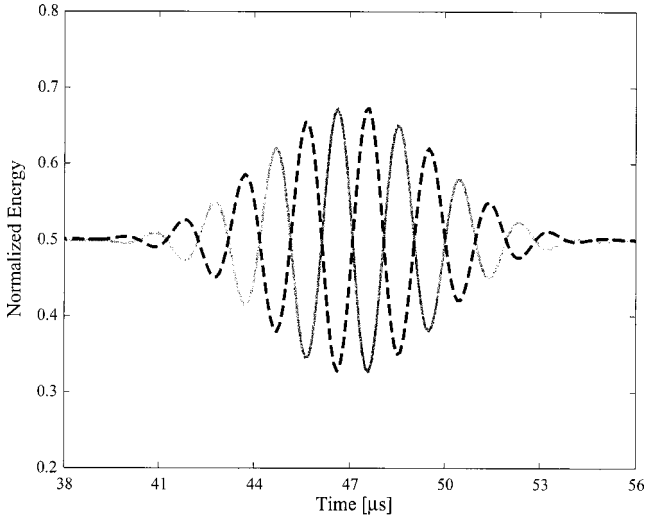


FIG. 8. Interchange between the strain energy (gray solid line), and the kinetic energy (black dashed line). The figure shows an enlargement of the first reflected wave mode of Fig. 7.

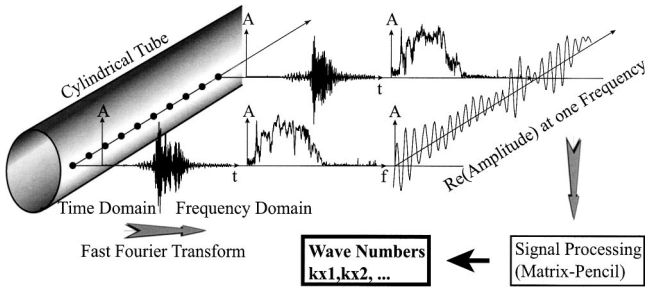


FIG. 9. Visualization of the major steps in the signal processing algorithm to determine the dispersion relation.

sponds to 122 000 time steps. The relative difference ΔE_{21} with respect to E_1 amounts only to -0.012%, which indicates a negligible decrease of the total energy.

Results of the same order of magnitude are obtained for several simulations with different material, geometrical, excitational, and discretization parameters. The observation of the total mechanical energy proves to be a useful criterion to demonstrate the validity of the finite-difference method presented.

VII. VALIDATION USING THE DISPERSION RELATION

Dispersion relations are used to check the physical behavior of the proposed FDM scheme. They are like a fingerprint of a structure, including all phenomena of the mechanical wave propagation.

On the one hand, dispersion relations can be calculated from the equations of motion according to Ramsakaya.[15] On the other hand, they can be estimated from an experiment using the matrix-pencil algorithm published by Hua.[16] The method is described in the papers of Vollmann *et al.*[17] and Gsell *et al.*[18] The major steps are outlined below and in Fig. 9.

The displacement data are recorded versus time at several locations along a surface line parallel to the axis of the structure. A fast Fourier transform is applied to each measured data set in order to obtain the complex amplitude spectrum. A new data vector is created with the complex amplitudes of all points at one discrete frequency. A signal processing algorithm is applied to determine the wave numbers at the chosen frequency. Here the matrix-pencil algorithm is used. The algorithm is able to estimate frequencies (or wave numbers) of multiple sinusoids in noisy signals. To derive the dispersion curves, the procedure is applied to each frequency.

To validate the present FDM code, displacement data are generated using a simulation. In a next step the signal processing algorithm described above is applied to this data to estimate the dispersion curves. In Fig. 10 they are compared with the theoretically calculated curves. An orthotropic tube of $R=0.014\,75$ m, $h=0.0015$ m, and $l=1$ m with the same material parameters as the point source example in Sec. IV is used as a test specimen. A broadband excitation function with a frequency range of 5–250 kHz is chosen in a way that the first axisymmetric and the first three nonaxisymmetric wave modes in the radial, circumferential, and axial directions are excited.

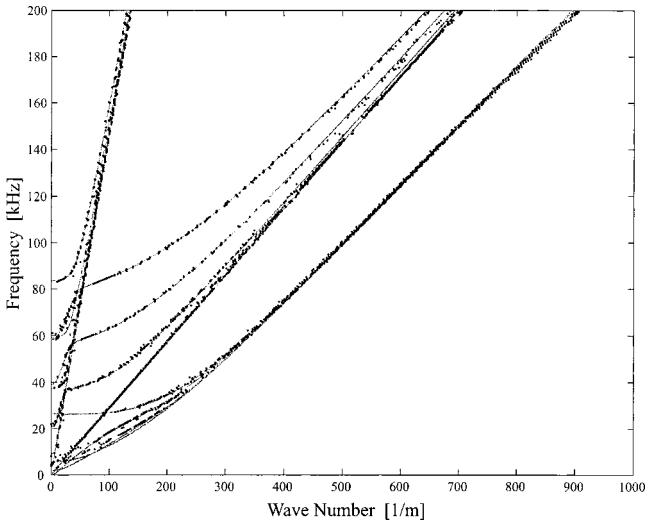


FIG. 10. Comparison between the theoretically calculated dispersion curves (gray lines) and the estimated values (black dots) from a numerical experiment.

Good agreement is found between the two methods. In the case of higher wave numbers one can see slight divergences. This fact is attributed to insufficient sampling of the shortest wavelength. In the region of lower wave numbers, deviations occur due to estimation errors of the signal processing algorithm in the case of large wavelength.

VIII. CONCLUSIONS

The present FDM code is a valuable tool for visualizing and enhancing the understanding of the wave propagation phenomena in elastic tubes.

Since the displacement equations of motion are discretized directly, no stress components must be calculated with this new algorithm. This reduces the memory requirement to approximately one-third and leads to a decrease of computational time.

The exact von Neumann stability analysis, which is investigated numerically for three-dimensional cylindrical coordinates for the first time, shows that the empirical result of Ref. 3 is a useful criterion for the limitation of the time step Δt.

Observation of the total mechanical energy of different simulations, which were implemented as nondissipative systems, showed only negligible deviations. This is another important finding which proves the validity of the code. Instead of using analytical solutions for simple cases to validate the capability of the algorithm to describe the physics appropriately, a new and more general approach is proposed. Therefore, theoretically calculated dispersion curves are compared with the estimated values from a numerical experiment. The achieved results coincide and verify the algorithm.

The code can be used to create numerical data, which are helpful to investigate signal processing algorithms used in nondestructive-testing procedures, e.g., neural networks. The study of the interaction of elastic waves at local inhomogeneities (e.g., the scattering at part-through cracks) is another application of the three-dimensional code.

APPENDIX A: DISCRETIZATION RULES

Central differences for the equation of motion in radial direction [Eq. (4)]:

$$\frac{\partial^2 u_r}{\partial t^2} = \frac{1}{\Delta t^2}\left({}^{n+1}_{j,i,k}u_r - 2\,{}^{n}_{j,i,k}u_r + {}^{n-1}_{j,i,k}u_r \right),$$

$$\frac{\partial^2 u_r}{\partial r^2} = \frac{1}{\Delta r^2}\left({}^{n}_{j,i,k+1}u_r - 2\,{}^{n}_{j,i,k}u_r + {}^{n}_{j,i,k-1}u_r \right),$$

$$\frac{\partial^2 u_r}{\partial \varphi^2} = \frac{1}{\Delta \varphi^2}\left({}^{n}_{j,i+1,k}u_r - 2\,{}^{n}_{j,i,k}u_r + {}^{n}_{j,i-1,k}u_r \right),$$

$$\frac{\partial^2 u_r}{\partial z^2} = \frac{1}{\Delta z^2}\left({}^{n}_{j+1,i,k}u_r - 2\,{}^{n}_{j,i,k}u_r + {}^{n}_{j-1,i,k}u_r \right),$$

$$\frac{\partial u_r}{\partial r} = \frac{1}{2\Delta r}\left({}^{n}_{j,i,k+1}u_r - {}^{n}_{j,i,k-1}u_r \right), \tag{A1}$$

$$\frac{\partial u_\varphi}{\partial \varphi} = \frac{1}{2\Delta \varphi}\left({}^{n}_{j,i,k+1}u_\varphi + {}^{n}_{j,i,k}u_\varphi - {}^{n}_{j,i-1,k+1}u_\varphi - {}^{n}_{j,i-1,k}u_\varphi \right),$$

$$\frac{\partial^2 u_\varphi}{\partial r \partial \varphi} = \frac{1}{\Delta r \Delta \varphi}\left({}^{n}_{j,i,k+1}u_\varphi - {}^{n}_{j,i,k}u_\varphi - {}^{n}_{j,i-1,k+1}u_\varphi + {}^{n}_{j,i-1,k}u_\varphi \right),$$

$$\frac{\partial^2 u_z}{\partial r \partial z} = \frac{1}{\Delta r \Delta z}\left({}^{n}_{j+1,i,k+1}u_z - {}^{n}_{j+1,i,k}u_z - {}^{n}_{j,i,k+1}u_z + {}^{n}_{j,i,k}u_z \right),$$

$$\frac{\partial u_z}{\partial z} = \frac{1}{2\Delta z}\left({}^{n}_{j+1,i,k+1}u_z + {}^{n}_{j+1,i,k}u_z - {}^{n}_{j,i,k+1}u_z - {}^{n}_{j,i,k}u_z \right).$$

APPENDIX B: FULL FORM OF EQ. (13)

Equation (13) to calculate the radial displacement at the next time step for the isotropic case:

$$\begin{aligned}
{}^{n+1}_{j,i,k}u_r = &- {}^{n-1}_{j,i,k}u_r + \frac{\Delta t^2}{\rho}\left(\left(\frac{2\rho}{\Delta t^2} - \frac{2\mu}{\Delta z^2} - \frac{2(\lambda+2\mu)}{\Delta r^2} - \frac{\lambda+2\mu}{r_{rk}^2} - \frac{2\mu}{r_{rk}^2\Delta\varphi^2} \right) {}^{n}_{j,i,k}u_r + \frac{(2r_{rk}-\Delta r)(\lambda+2\mu)}{2r_{rk}\Delta r^2}\,{}^{n}_{j,i,k-1}u_r \right. \\
&+ \frac{(2r_{rk}+\Delta r)(\lambda+2\mu)}{2r_{rk}\Delta r^2}\,{}^{n}_{j,i,k+1}u_r + \frac{\mu}{r_{rk}^2\Delta\varphi^2}\,{}^{n}_{j,i-1,k}u_r + \frac{\mu}{r_{rk}^2\Delta\varphi^2}\,{}^{n}_{j,i+1,k}u_r + \frac{\mu}{\Delta z^2}\,{}^{n}_{j-1,i,k}u_r + \frac{\mu}{\Delta z^2}\,{}^{n}_{j+1,i,k}u_r + \frac{\lambda+\mu}{\Delta r\Delta z}\,{}^{n}_{j,i,k}u_z - \frac{\lambda+\mu}{\Delta r\Delta z}\,{}^{n}_{j,i,k+1}u_z \\
&- \frac{\lambda+\mu}{\Delta r\Delta z}\,{}^{n}_{j+1,i,k}u_z + \frac{\lambda+\mu}{\Delta r\Delta z}\,{}^{n}_{j+1,i,k+1}u_z + \frac{2r_{rk}(\lambda+\mu)+\Delta r(\lambda+3\mu)}{2r_{rk}^2\Delta r\Delta\varphi}\,{}^{n}_{j,i-1,k}u_\varphi + \frac{-2r_{rk}(\lambda+\mu)+\Delta r(\lambda+3\mu)}{2r_{rk}^2\Delta r\Delta\varphi}\,{}^{n}_{j,i-1,k+1}u_\varphi \\
&\left. + \frac{-2r_{rk}(\lambda+\mu)-\Delta r(\lambda+3\mu)}{2r_{rk}^2\Delta r\Delta\varphi}\,{}^{n}_{j,i,k}u_\varphi + \frac{2r_{rk}(\lambda+\mu)-\Delta r(\lambda+3\mu)}{2r_{rk}^2\Delta r\Delta\varphi}\,{}^{n}_{j,i,k+1}u_\varphi \right).
\end{aligned} \tag{B1}$$

Equation (13) to calculate the radial displacement at the next time step for the orthotropic case:

$$\begin{aligned}
{}^{n+1}_{j,i,k}u_r = &- {}^{n-1}_{j,i,k}u_r + \frac{\Delta t^2}{\rho}\left(\left(\frac{2\rho}{\Delta t^2} - \frac{c_{55}}{\Delta z^2} - \frac{2c_{11}}{\Delta r^2} - \frac{c_{22}}{r_{rk}^2} - \frac{c_{44}}{r_{rk}^2\Delta\varphi^2} \right) {}^{n}_{j,i,k}u_r + \frac{(2r_{rk}-\Delta r)c_{11}}{2r_{rk}\Delta r^2}\,{}^{n}_{j,i,k-1}u_r + \frac{(2r_{rk}+\Delta r)c_{11}}{2r_{rk}\Delta r^2}\,{}^{n}_{j,i,k+1}u_r \right. \\
&+ \frac{c_{44}}{2r_{rk}^2\Delta\varphi^2}\,{}^{n}_{j,i-1,k}u_r + \frac{c_{44}}{2r_{rk}^2\Delta\varphi^2}\,{}^{n}_{j,i+1,k}u_r + \frac{c_{55}}{2\Delta z^2}\,{}^{n}_{j-1,i,k}u_r + \frac{c_{55}}{2\Delta z^2}\,{}^{n}_{j+1,i,k}u_r + \frac{(2r_{rk}-\Delta r)c_{13}+\Delta r c_{23}+r_{rk}c_{55}}{2r_{rk}\Delta r\Delta z}\,{}^{n}_{j,i,k}u_z \\
&- \frac{(2r_{rk}+\Delta r)c_{13}-\Delta r c_{23}+r_{rk}c_{55}}{2r_{rk}\Delta r\Delta z}\,{}^{n}_{j,i,k+1}u_z - \frac{(2r_{rk}-\Delta r)c_{13}+\Delta r c_{23}+r_{rk}c_{55}}{2r_{rk}\Delta r\Delta z}\,{}^{n}_{j+1,i,k}u_z \\
&+ \frac{(2r_{rk}+\Delta r)c_{13}-\Delta r c_{23}+r_{rk}c_{55}}{2r_{rk}\Delta r\Delta z}\,{}^{n}_{j+1,i,k+1}u_z + \frac{(2r_{rk}+\Delta r)c_{44}+2\Delta r c_{22}+4r_{rk}c_{12}}{4r_{rk}^2\Delta r\Delta\varphi}\,{}^{n}_{j,i-1,k}u_\varphi \\
&- \frac{(2r_{rk}-\Delta r)c_{44}-2\Delta r c_{22}+4r_{rk}c_{12}}{4r_{rk}^2\Delta r\Delta\varphi}\,{}^{n}_{j,i-1,k+1}u_\varphi - \frac{(2r_{rk}+\Delta r)c_{44}+2\Delta r c_{22}+4r_{rk}c_{12}}{4r_{rk}^2\Delta r\Delta\varphi}\,{}^{n}_{j,i,k}u_\varphi \\
&\left. + \frac{(2r_{rk}-\Delta r)c_{44}-2\Delta r c_{22}+4r_{rk}c_{12}}{4r_{rk}^2\Delta r\Delta\varphi}\,{}^{n}_{j,i,k+1}u_\varphi \right).
\end{aligned} \tag{B2}$$

[1] T. Leutenegger and J. Dual, ''Detection of defects in cylindrical structures using a time reverse method and a finite-difference approach,'' Ultrasonics **40**, 721–725 (2002).
[2] F. Schubert, A. Pfeiffer, and B. Köhler, ''The elastodynamic finite integration technique for waves in cylindrical geometries,'' J. Acoust. Soc. Am. **104**, 2604–2614 (1998).
[3] Y.-H. Chen and W. C. Chew, ''A three-dimensional finite difference code for the modeling of sonic logging tools,'' J. Acoust. Soc. Am. **103**, 702–712 (1998).
[4] R. Madariaga, ''Dynamics of an expanding circular fault,'' Bull. Seismol. Soc. Am. **66**, 639–666 (1976).
[5] P. D. Lax and R. D. Richtmyer, ''Survey of the stability of linear finite difference equations,'' Commun. Pure Appl. Math. **IX**, 267–293 (1956).
[6] P. Fellinger, R. Marklein, K. J. Langenberg, and S. Klaholz, ''Numerical modeling of elastic wave propagation and scattering with EFIT-elastodynamic finite integration technique,'' Wave Motion **21**, 47–66 (1995).
[7] S. G. Lekhnitskii, *Theory of Elasticity of an Anisotropic Body* (Mir, Moscow, 1981).
[8] K. F. Graff, *Wave Motion in Elastic Solids* (Oxford University Press, London, 1975).
[9] J. Dual, ''Quantitative nondestructive evaluation using guided waves,'' in *Nondestructive Characterization of Materials*, edited by T. Kishi, T. Saito, C. Ruud, and R. Green, Proceedings of the 5th International Symposium on Nondestructive Characterization of Materials, 1061ff (Gordon and Breach, Montreux, Switzerland, 1991).
[10] J. Temple, ''Modelling the propagation and scattering of elastic waves in inhomogeneous anisotropic media,'' J. Phys. D **21**, 859–874 (1988).
[11] J. C. Strikwerda, *Finite Difference Schemes and Partial Differential Equations* (CRC Press, LLC, Florida, 1999).
[12] V. G. Ganzha and E. V. Vorozhtsov, *Computer-Aided Analysis of Difference Schemes for Partial Differential Equations* (Wiley, New York, 1996).
[13] A. Taflove, *Computational Electrodynamics—The Finite-Difference Time-Domain Method* (Artech House Inc., Norwood, 1995).
[14] J. D. Achenbach, *Wave Propagation in Elastic Solids* (Elsevier Science Publishers, Amsterdam, 1984).
[15] E. I. Ramsakaya and N. A. Shul'ga, ''Propagation of nonaxisymmetrical elastic waves in an orthotropic hollow cylinder,'' Sov. Appl. Mech. **19**, 748–752 (1983).
[16] Y. Hua and T. K. Sarkar, ''Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise,'' IEEE Trans. Acoust., Speech, Signal Process. **38**, 814–824 (1990).
[17] J. Vollmann, R. Breu, and J. Dual, ''High-resolution analysis of the complex wave spectrum in a cylindrical shell containing a viscoelastic medium. Part II. Experimental results versus theory,'' J. Acoust. Soc. Am. **102**, 909–920 (1997).
[18] D. Gsell, D. Profunser, and J. Dual, ''Measurement of the dispersion relation of guided nonaxisymmetric waves in filament-wound cylindrical structures,'' Ultrasonics **38**, 517–521 (2000).

Transient propagation in media with classical or power-law loss

Richard S. C. Cobbold,[a] Nikolai V. Sushilov, and Adam C. Weathermon
Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Ontario M5S 3G9, Canada

(Received 1 June 2004; revised 30 September 2004; accepted 1 October 2004)

This paper addresses the problem of small-signal transient wave propagation in media whose absorption coefficient obeys power-law frequency dependence, i.e., $\alpha \propto \omega^n$. Our approach makes use of previously derived relations between the absorption and dispersion based on the Kramers–Kronig relations. This, combined with a recently obtained solution to a causal convolution wave equation enable expressions to be obtained for one-dimensional transient propagation when n is in the range $0<n<3$. For $n=2$, corresponding to no dispersion, straightforward analytical expressions are obtained for a δ-function and a sinusoidal step function sources and these are shown to correspond to relations previously derived. For other values of n, the effects of dispersion are accounted for by using Fourier transforms. Examples are used to illustrate the results for normal and anomalous dispersive media and to examine the question as to the conditions under which the effects of dispersion should be accounted for, especially for wideband ultrasound pulses of the type used in *B*-mode tissue imaging. It is shown that the product of the attenuation and total propagation path can be used as a criterion for judging whether dispersion needs to be accounted for. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1823271]

PACS numbers: 43.20.Bi, 43.20.Hq, 43.80.Cs [TDM] Pages: 3294–3303

I. INTRODUCTION

Experimentally, it has been found that the attenuation of many media is proportional to the nth power of the frequency. Both scattering and absorption can contribute to this behavior. For classical viscous loss media such as water, scattering is considered to be negligible and it has been found that $n \approx 2$ over a very wide range of frequencies.[1] Interest in accounting for the effects of absorption and accompanying dispersion for other types of media has partially arisen from the need to properly account for its effects in ultrasound diagnostic and therapeutic applications. In fact, it is well established that many soft biological tissues have an attenuation characteristic that can be well represented,[2] over a limited frequency extent, by n in the range $1 \leqslant n < 2$.

Our primary purpose is to describe a method for accounting for power-law absorption characteristic for transient signals and to illustrate how dispersive attenuation can influence the propagation of such signals. Using a recently obtained solution to a causal convolution wave equation[3] and previously derived relations between the absorption and dispersion based on the Kramers–Kronig relations, expressions are obtained for one-dimensional transient propagation when n is in the range $0<n<3$.

The final portion of this introduction presents a brief overview of the equation that describes propagation in classical viscous loss media and its transient solution in one dimension. Sections II and III provide a description of a new and more general approach when n is in the range $0<n<3$. Analytic solutions for $n=1$ and 2 (the nondispersive case) are considered in Secs. IV and V and a physical explanation is provided for the manner in which the transient component propagates. General dispersive propagation equations are obtained in Secs. VI and VII, and examples are used to illustrate propagation in normal and anomalous dispersive media, as well as to examine the question as to the conditions under which the effects of dispersion need to be included.

Background

The problem of predicting the manner in which transient plane-waves propagate into a continuous medium exhibiting classical viscous loss has been studied for many years. The small-signal equation characterizing the propagation was originally derived and published by Stokes[4] in 1845. In one dimension it can be written as

$$\frac{1}{c_0^2}\frac{\partial^2 v}{\partial t^2} - \frac{\partial^2 v}{\partial x^2} = \chi \frac{\partial}{\partial t}\left(\frac{\partial^2 v}{\partial x^2}\right), \tag{1}$$

where v is the particle velocity, x is distance, t is time, c_0 is the small signal propagation speed in the absence of viscous loss ($\chi=0$), and χ is a constant for the medium that expresses the product of the adiabatic compressibility and viscous losses. It is a third-order homogeneous linear equation based on the small-signal assumption.

Although the steady-state solution of this equation is well known, having been obtained by Stokes[4] and subsequently by Stefan[5] in 1866, transient problems have proven rather more difficult to address. It is well established that approximate solutions to such problems do not satisfy causality in the strict sense, i.e., a propagated pulse does not have a sharp front but extends asymptotically to plus and minus infinity, similar to solutions of the second order diffusion equation.[6,7] Exact transient solutions have been obtained using the Laplace transform. Of particular importance is the work of Hanin[8] in 1957 who obtained a solution to (1) for a

[a]Electronic address: cobbold@ecf.utoronto.ca

δ-function source in terms of power series expansions. More recently, using Laplace transforms Ludwig and Levin[9] also obtained a δ-function solution which was expressed in terms of two real integrals suitable for numerical evaluations for a wide range of conditions. Exact solutions have also been obtained by Norwood[10] for a step function, a δ-function, a decaying exponential and a sinusoidal step, the results for which were expressed in terms of double summations.

Approximation methods have also been used to obtain solutions for small absorption coefficients. Blackstock,[11] in an important contribution published in 1967, considered the one-dimensional (1D) problem in which a sinusoidal source was abruptly switched on. By assuming the absorption to be small, specifically that $\alpha c_0 \ll \omega$, where α is the absorption coefficient, he showed that (1) could be reduced to the following second order equation:

$$\frac{\alpha c_0}{\omega^2}\frac{\partial^2 v}{\partial t^2} - c_0\frac{\partial v}{\partial x} - \frac{\partial v}{\partial t} = 0. \tag{2}$$

This equation can be transformed to the well-known parabolic diffusion or heat conduction equation by changing the variable to $\tau = t - x/c_0$. For the assumed boundary and initial conditions, relatively conventional techniques were then be used to obtain a solution.

If $\alpha = \alpha_0\omega^2$, where α_0 is an absorption factor, the earlier approximation can be written as $\alpha_0 c_0 \omega \ll 1$, from which it can be seen that the validity of the solution is limited to locations where the spectral components are such that $\omega \ll 1/(\alpha_0 c_0)$. For example, substituting the values for water ($\alpha_0 = 6.4\times10^{-16}$ Np/[m (rad/s)^2], $c_0 = 1500$ m/s) gives: $\omega \ll 1.04\times10^{12}$ rad/s. On the other hand, for carbon tetrachloride[12] ($\alpha_0 = 1.357\times10^{-14}$ Np/[m (rad/s)^2], $c_0 = 938$ m/s) yields $\omega \ll 7.85\times10^{10}$ rad/s. Ludwig and Levin[9] noted that close to a transient source, where the frequency components become higher, the approximation breaks down thereby limiting the range of validity to both sufficient distances from the source and small absorption.

For the step sinusoidal problem posed by Blackstock[11] the boundary conditions for the particle velocity v can be expressed as

$$v(0,t) = H(t)\, v_0 \sin(\omega_0 t), \quad \text{and} \quad v(x,t) \to 0 \quad \text{as} \quad x \to \infty, \tag{3a}$$

and the initial conditions by,

$$v(x,t) = \partial v(x,t)/\partial t = 0 \quad \text{for} \quad t < 0, \tag{3b}$$

where v_0 is the velocity amplitude, $H(t)$ is the Heaviside step function, ω_0 is the angular frequency and propagation is assumed to be in the positive x-direction. The specification of both the velocity and its derivative for $t<0$ is often referred to as the Cauchy initial condition. However, trying to impose both conditions can result in a poorly posed problem depending on the differential equation. It may be better to impose only the first condition and let the system adjust itself to the proper derivative condition on the boundary.

Blackstock's exact solution[11] to the approximate second order equation that satisfies the initial condition $v(x,t) = 0$ for $t<0$, can be written as

$$v(x,t)/v_0 = e^{-\alpha_0 x\omega_0^2} H(t - x/c_0)\sin[\omega_0(t - x/c_0)] + \frac{1}{2}e^{-(t-x/c_0)^2/4\alpha_0 x} H(t) \times \mathrm{Im}[w(S_1) - w(S_2)], \tag{4a}$$

where $\alpha = \alpha_0\omega_0^2$ is the approximate absorption coefficient for classical viscous losses and

$$w(S) = e^{-S^2}\,\mathrm{erfc}(-jS),$$

$$S_1 = \omega_0\sqrt{\alpha_0 x} + j\left|t - \frac{x}{c_0}\right| \Big/ \sqrt{4\alpha_0 x}, \tag{4b}$$

$$S_2 = \omega_0\sqrt{\alpha_0 x} + j\left(t + \frac{x}{c_0}\right) \Big/ \sqrt{4\alpha_0 x}.$$

However, because of the approximation made in obtaining (2), it cannot be expected to be valid solution to (1) close to the source.

Blackstock[11] also derived a somewhat simpler noncausal solution, which is a good approximation for locations not too close to the source, though it does fail to satisfy the initial conditions. It is given by

$$v(x,t)/v_0 = e^{-\alpha_0 x\omega_0^2} H(t - x/c_0)\sin[\omega_0(t - x/c_0)] + \frac{1}{2}e^{-(t-x/c_0)^2/4\alpha_0 x}\,\mathrm{Im}[w(S_1)], \tag{4c}$$

which is the same as (4a) except for the absence of the $H(t)$ and $w(S_2)$ terms.

Both solutions predict that the wave-front propagation speed is infinite, which is of course is physically impossible. It has been suggested that the failure for both the classical viscous wave equation and the diffusion equation to properly account for the finite propagation speed arise from neglecting the particle inertia in response to an applied force. As a result, particles can acquire an arbitrary high velocity. The paradox of infinite propagation speed has been studied for at least 50 years and a number of phenomenological approaches have been used. Many of these yield a hyperbolic equation whose solutions are causal with a finite propagation speeds.[13] However, as discussed here and by others,[11] solutions to the parabolic equations do not provide sufficiently accurate descriptions of the behavior in regions of practical interest.

II. A NEW APPROACH FOR $0<n<3$

For many soft tissues the effects of attenuation over a limited frequency range can be well represented by a power law,[2] often with n close to 1. Although the contribution of scattering to the attenuation has not been well documented for many media, it is generally significantly smaller than absorption. Provided the effects of scattering can be ignored, dispersion equations, based on the Kramers–Kronig relations, can be used.

Recently[3] we obtained a solution describing wave propagation in a medium whose absorption coefficient obeys a power-law frequency dependence. This solution was shown

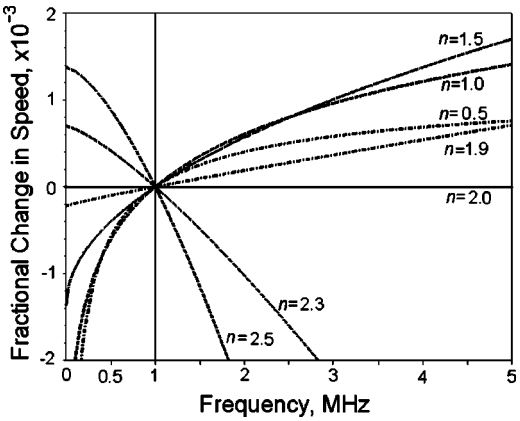


FIG. 1. The fractional change in phase propagation speed vs frequency, as predicted by (6) and (7), for different values of the exponent n. At the reference frequency of 1.0 MHz, $c_0=1500$ m/s. An (frequency) attenuation coefficient of 0.5 dB/[cm (MHz)^n] was assumed. This corresponds to an angular frequency coefficient of $\alpha_0=5.76/(2\pi\times10^6)^n$, m^{-1} (rad/s)^n.

to satisfy the causal time-domain wave equation as proposed by Szabo[14] for propagation in such a medium. It is given by

$$\phi(\mathbf{r},t)=\frac{1}{(2\pi)^4}\int_{-\infty}^{\infty}a(\omega)e^{-j[\mathbf{K}_r(\omega)\cdot\mathbf{r}-\omega t]}d\omega, \tag{5}$$

where, in order to simplify the presentation, the vector $\mathbf{K}_r(\omega)$ has been introduced whose components are all equal to $[\omega/c(\omega)-j\alpha(\omega)]/\sqrt{3}=K(\omega)/\sqrt{3}$. In addition, $a(\omega)$ is an arbitrary function which can be specified by choosing proper boundary conditions, $c(\omega)$ is the speed of sound at an angular frequency of ω and $\alpha(\omega)=\alpha_0|\omega|^n$, where n is a constant.

Making use of a generalized function (the tempered distribution), Szabo[15] obtained equations that relate $\alpha(\omega)$ and $c(\omega)$ for values of n in the range from 0 to 3. Similar relations, valid for positive frequencies, but based on the differential Kramers–Kronig relations, were subsequently obtained by Waters *et al.*[16,17] For both positive and negative frequencies the relations can be written as

$$\frac{1}{c(\omega)}=\frac{1}{c_0}-\frac{2}{\pi}\alpha_0\ln\left|\frac{\omega}{\omega_0}\right|,\quad \text{for } n=1, \tag{6}$$

$$\frac{1}{c(\omega)}=\frac{1}{c_0}+\alpha_0\tan\left(\frac{\pi n}{2}\right)(|\omega|^{n-1}-|\omega_0|^{n-1}),$$

$$\text{for } 0<n<1 \quad \text{and} \quad 1<n<3, \tag{7}$$

where ω_0 is a reference angular frequency at which $c(\omega)=c_0$. Recently, Waters *et al.*[18] pointed out that (7) reduces to (6) in the limit as $n\to1$.

It follows from (7) that if $n=2$, i.e., $\alpha(\omega)=\alpha_0\omega^2$, then $c(\omega)=c_0$, corresponding to the absence of any dispersion. For all other values of n dispersion is present, as is illustrated in Fig. 1 for a reference frequency of 1 MHz. The figure also assumes that for all values of n the attenuation at 1.0 MHz is equal to 5.76 m^{-1}: a value that corresponds to that typically seen in soft tissue.[19] It will be noted that for $n>2$ the speed decreases with increasing frequency and normal dispersion is said to be present, as is the case for cancellous bone.[20,21] For $n<2$ the opposite is true and the dispersion is said to be anomalous. The presence of dispersion increases the difficulty of obtaining an exact solution and, consequently, its effects are sometimes ignored. For narrow-band signals the effect of this simplification on the resulting transient wave is likely to be small.

An exact plane-wave harmonic solution of the classical viscous loss equation as given by (1), results in an absorption coefficient that can be expressed as a power-law series, the first term of which is quadratic in frequency, i.e., $\alpha(\omega)\approx\alpha_0\omega^2$. In addition, there is a small anomalous dispersion term. Because the derivation of (5) assumed that the propagation medium can be exactly characterized by a power-law absorption coefficient, it is not an exact solution to the classical viscous loss equation given by (1). However, for small losses it can be shown that (5) is an approximate solution of the classical equation. Nonetheless, it is an exact solution of the approximate viscous loss wave equation as expressed by (2).

We will show that for the special case of $n=2$ the solution given by (4c) for the problem expressed by (3) can obtained from the general solution given by (5). It will also be shown that solutions for other values of n in the range $0<n<3$, can be obtained either analytically (by neglecting dispersion) or exactly, by using Fourier transforms.

III. ONE-DIMENSIONAL PROPAGATION

For 1D propagation in a medium whose losses correspond to a power law in the range $0<n<3$, by following the same steps that were used to obtain (5), it can be shown that the velocity potential is given by[3]

$$\phi(x,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}a(\omega)e^{-j[\omega/c(\omega)-j\alpha(\omega)]x+j\omega t}d\omega. \tag{8}$$

Noting that the particle velocity and the velocity potential are related by $v=-\partial\phi/\partial x$, this can be written as

$$v(x,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}j\left[\frac{\omega}{c(\omega)}-j\alpha(\omega)\right]a(\omega)\times e^{-j[\omega/c(\omega)-j\alpha(\omega)]x+j\omega t}d\omega. \tag{9}$$

It can be seen that (9) can be written in the form of an inverse Fourier transform, i.e.:

$$v(x,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}v(x,\omega)e^{j\omega t}d\omega, \tag{10}$$

having a kernel given by

$$v(x,\omega)=ja(\omega)\left(\frac{\omega}{c(\omega)}-j\alpha(\omega)\right)e^{-jx(\omega/c(\omega)-j\alpha(\omega))}=v(0,\omega)e^{-jx(\omega/c(\omega)-j\alpha(\omega))}, \tag{11}$$

in which

$$v(0,\omega)=ja(\omega)\left(\frac{\omega}{c(\omega)}-j\alpha(\omega)\right).$$

If this result is substituted into (9), the particle velocity can be expressed as

$$\nu(x,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\nu(0,\omega)e^{-j[\omega/c(\omega)-j\alpha(\omega)]x+j\omega t}d\omega, \quad (12)$$

which shows that when the source frequency spectrum is known, the particle velocity at an observation point x can be calculated.

IV. PROPAGATION IN MEDIA FOR $n=2$

There are two ways of proceeding to solve the problem posed in (3). The first consists of determining the δ-function response and then, by convolving this with the source function to obtain the response. The second consists of determining the frequency spectrum of the source and then evaluating (12).

A. The impulse response

The impulse response can be obtained from (12) by noting that the response corresponds to $\nu(0,\omega)=1$, and using (7) with $n=2$, and $c(\omega)=c_0$. These steps enable the δ-response to be written as

$$\nu_\delta(x,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-j[\omega/c_0-j\alpha_0\omega^2]x+j\omega t}d\omega$$

$$=\frac{e^{-(t-x/c_0)^2/4x\alpha_0}}{2\sqrt{\pi x\alpha_0}}=\frac{e^{-\tau^2/4x\alpha_0}}{2\sqrt{\pi x\alpha_0}}, \quad (13)$$

which is a Gaussian function in retarded time $\tau=t-x/c_0$. In fact, a similar result was obtained by Blackstock [see Eq. (53) in Ref. 11] from the classical viscous loss differential equation by using a Laplace transform method. His derivation assumed the observation point to be at a large distance from the source.

The Gaussian function fails to satisfy causality and predicts that an infinitesimal portion of the wave-front propagates with an infinite propagation speed. Figure 2(a) shows the normalized particle velocity response at four different observation points, when the attenuation coefficient is assumed to be 100 times greater than that of water at 20 °C, i.e., $\alpha_0=6.4\times10^{-14}$ Np/[m (rad/s)^2]. Such a value is of the same order of magnitude as fluids such as glycerin. The initial condition, $\nu(x,t)=0$ for $t<0$, i.e., the causality requirement, corresponds to negative values of retarded time such that $\tau<-x/c_0$. For example, at a distance of 10 μm from the source ($x/c_0=6.67$ ns), from Fig. 2(a) it can be seen that at the retarded time of $\tau<-6.7$ ns the response is negligibly small. However, for smaller distances, such as 2 μm ($x/c_0=3$ ns) this is no longer the case. Consequently, for distances not too close to the source and/or a higher attenuation, the failure to satisfy causality has no significant effect on the computed result.

An alternative approach to that of Ludwig and Levin[9] for obtaining the impulse response of (1) is described in the Appendix. It expresses the result as an inverse Fourier transform, which can be evaluated using standard fast Fourier transform software. Figure 2(b) shows the results for the same conditions as in Fig. 2(a). It can be seen that at greater distances from the source, the response is very close to a symmetric Gaussian function, but that for small distances the response becomes asymmetric. Numerical results indicate that as the observation point approaches the source, the peak response approaches zero (retarded time) and the wave form tends toward a δ-function, as expected. Our numerical results also indicate that the response drops to zero for $\tau=-x/c_0$, i.e., $t=0$, making the δ-function solution causal. This appears to be contrary to prior suggestions that (1) is a parabolic equation and therefore, its solutions are noncausal.[13] However, the wave-front propagation speed remains infinite, i.e., a small portion of the response reaches any location x at a time $t>0$.

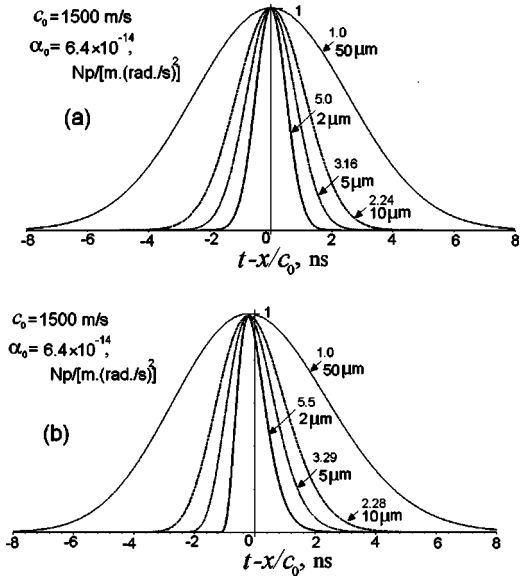


FIG. 2. Normalized δ-function response vs retarded time $\tau=t-x/c_0$, assuming a medium for which $\alpha_0=6.4\times10^{-14}$ Np/[m (rad/s)^2]. Each curve represents the waveform seen at a given location: the four waveforms in each graph correspond to the same set of distances from the source. The numbers above each distance indicate the relative amplitude normalization constant used. Shown in (a) is the approximate response, as calculated from (13). (b) Shows the exact δ-function response of (1). Note the asymmetry in the response for small distances.

By convolving (13) with (3), it can be shown that the resulting equation is identical to that obtained in the next subsection.

B. Using the source spectrum

In the second approach for solving the problem posed by (3), the frequency spectrum of the source function can be used in (12). For the boundary and initial conditions specified by (3b), $\nu(0,\omega)$ can be expressed as

$$\nu(0,\omega)=\int_{-\infty}^{\infty}\nu(0,t)e^{-j\omega t}dt$$

$$=\nu_0\int_{-\infty}^{\infty}\sin(\omega_0 t)H(t)e^{-j\omega t}dt. \quad (14)$$

By substituting this into (12), noting that $c(\omega)=c_0$ and changing the order of integration results in

$$\nu(x,t)=\frac{\nu_0}{2\pi}\int_{-\infty}^{\infty}\sin(\omega_0 t')H(t')\times\left(\int_{-\infty}^{\infty}e^{-\alpha_0 x\omega^2-j\omega(x/c_0-t+t')}d\omega\right)dt'. \quad (15)$$

From Ref. 22 [Eq. (3.323.2)], the first integral can be evaluated as

$$\int_{-\infty}^{\infty}e^{-\alpha_0 x\omega^2-j(x/c_0-t+t')\omega}d\omega=\sqrt{\frac{\pi}{\alpha_0 x}}e^{-(x/c_0-t+t')^2/4\alpha_0 x},$$

which, when substituted into (15), yields

$$\nu(x,t)=\frac{\nu_0}{\sqrt{4\pi\alpha_0 x}}\int_0^{\infty}\sin(\omega_0 t')e^{-(x/c_0-t+t')^2/4\alpha_0 x}dt'. \quad (16)$$

Again, using Ref. 22 [Eq. (3.897.1)], this second integral can be evaluated enabling the earlier equation to be written as

$$\nu(x,t)=\frac{\nu_0}{2}e^{-\alpha_0 x\omega_0^2}\left[\sin(\omega_0\tau)+\frac{1}{2j}e^{-j\omega_0\tau}\times\operatorname{erf}\left(\frac{-\tau+2j\omega_0\alpha_0 x}{2\sqrt{\alpha_0 x}}\right)+\frac{1}{2j}e^{j\omega_0\tau}\times\operatorname{erf}\left(\frac{\tau+2j\omega_0\alpha_0 x}{2\sqrt{\alpha_0 x}}\right)\right]. \quad (17)$$

If the relations given in (4b) are used, it can be shown that (17) is identical to Blackstock's solution as expressed by (4c). Since Blackstock's solution is an exact solution to the approximate wave equation as expressed by (2), our solution, as given by (17), is also an exact solution.

An alternative form for (17) that is convenient for separating the transient and steady-state components can be obtained by observing that $\operatorname{erf}(-X)=-\operatorname{erf}(X)$, and noting that $X-\bar{X}=2j\,\mathrm{Im}(X)$, where the complex conjugate is denoted by the overscore. These changes enable (17) to be expressed as

$$\nu(x,t)=\frac{\nu_0}{2}e^{-\alpha_0 x\omega_0^2}\left[\sin(\omega_0\tau)+\frac{1}{2j}\left[e^{j\omega_0\tau}\times\operatorname{erf}\left(\frac{\tau+2j\omega_0\alpha_0 x}{2\sqrt{\alpha_0 x}}\right)-\overline{e^{j\omega_0\tau}\operatorname{erf}\left(\frac{\tau+2j\omega_0\alpha_0 x}{2\sqrt{\alpha_0 x}}\right)}\right]\right],$$

or

$$\nu(x,t)=\frac{\nu_0}{2}e^{-\alpha_0 x\omega_0^2}\left[\sin(\omega_0\tau)+\mathrm{Im}\left[e^{j\omega_0\tau}\operatorname{erf}\left(\frac{\tau+2j\omega_0\alpha_0 x}{2\sqrt{\alpha_0 x}}\right)\right]\right]. \quad (18)$$

If this equation is rewritten as

$$\frac{\nu(x,t)}{\nu_0}=\underbrace{H(\omega_0\tau)\sin(\omega_0\tau)e^{-\alpha_0 x\omega_0^2}}_{\text{steady-state (forced response)}}+\underbrace{\frac{e^{-\alpha_0 x\omega_0^2}}{2}\left[\mathrm{Im}\left[e^{j\omega_0\tau}\operatorname{erf}\left(\frac{\omega_0\tau+2j\alpha_0 x\omega_0^2}{2\sqrt{\alpha_0 x\omega_0^2}}\right)\right]-\sin(\omega_0\tau)\operatorname{signum}(\omega_0\tau)\right]}_{\text{transient (natural response)}}, \quad (19)$$

where $\operatorname{signum}(y)=+1$ for $y>0$ and $=-1$ for $y<0$, the steady-state and the transient components can be identified. These are illustrated in Fig. 3 together with the total response for a particular observation point. In practical terms, at a frequency of 5 MHz and if the attenuation coefficient is again assumed to be 100 times greater than that of water, i.e., $\alpha_0=6.46\times10^{-14}$ Np/[m (rad/s)^2], then $\alpha_0 x\omega_0^2=1$ corresponds to $x=1.57$ cm. As pointed out by Blackstock,[11] the portion of the transient term ahead of the steady-state response, corresponding to negative values of τ, is responsible for the precursor.

Figure 4 shows the normalized particle velocity as a function of the normalized retarded time $\omega_0(t-x/c_0)$, at three different distances from the source for the same conditions as Fig. 3. For $\alpha_0 x\omega_0^2=1$ it can be seen that the precursor has a normalized retarded time duration of less than -5. For $c_0=1500$ m/s and at a frequency of 5 MHz, this corresponds to less than 0.8 of a cycle out of more than 50. With increasing distance from the source (higher values of $\alpha_0 x\omega_0^2$) the amplitude of the transient component, increases relative to the steady-state component: a feature that was reported by Blackstock.[11] As noted earlier, the presence of a response prior to a time of $-x/c_0$ results from the failure of the approximate solutions to satisfy the initial conditions.

Figure 5 shows that for $\alpha_0\omega_0^2 x>1$, the amplitude of the transient component falls off much more gradually relative to the steady-state amplitude, which diminishes as $\exp(-\alpha_0 x\omega_0^2)$. Also shown is the variation of the transient full-width-half-maximum (FWHM), which increases as a function of distance. Experimental measurements reported in

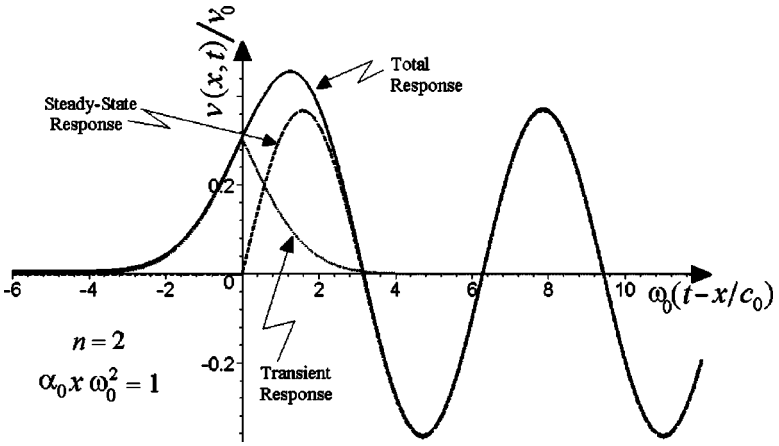


FIG. 3. The normalized particle velocity waveform is shown in response to a transient sinusoidal signal that propagates in a waterlike medium. The response consists of the sum of the steady-state and transient components as expressed by (19).

1970 by Moffett and Beyer[23] over a limited range of depths have confirmed this behavior. At first sight, it might seem that both features of the transient behavior are at variance with the underlying physics: an issue that does not appear to have been previously addressed and which is considered next.

C. Characteristics of the transient component

Perhaps the simplest way of obtaining an intuitive grasp of the transient component behavior is to examine how frequency dependent attenuation affects a Gaussian function that roughly approximates the characteristics of the transient component. For example, at the depth corresponding to $\alpha_0\omega^2 x=1$ in Fig. 3, the transient waveform can be approximated by

$$S(x,t)=S_0 e^{-(t-x/c_0)^2/4x\sigma}/\sqrt{2\pi x\sigma},$$

where the constants σ and S_0 are chosen to achieve a reasonable match. Its Fourier transform is given by $s(x,\omega)=\mathfrak{I}\{S(x,t)\}=S_0 e^{-j\omega(x/c_0)-\sigma\omega^2 x}$. Each frequency component will be subject to an attenuation of $\exp(-\alpha_0\omega^2 x)$ so that the resulting waveform is given by

$$S_\alpha(x,t)=\mathfrak{I}^{-1}\{S_0 e^{-j\omega x/c_0-\sigma x\omega^2}e^{-\alpha_0\omega^2 x}\}$$
$$=S_0 e^{-(t-x/c_0)^2/4x(\sigma+\alpha_0)}/\sqrt{4\pi x(\sigma+\alpha_0)}.$$

Consequently, the peak value, which occurs at a time of x/c_0, falls off as $1/\sqrt{x}$ rather than exponentially, and the FWHM increases with depth. Thus, the lower frequencies present in the transient waveform suffer less attenuation than the higher frequencies, a fact that becomes more apparent as the wave propagates to greater depths. This establishes a physical basis for the behavior of the transient component relative to the steady-state component as seen in Fig. 5.

V. NONDISPERSIVE PROPAGATION IN MEDIA FOR $n=1$

As noted earlier, this case is of particular interest since it corresponds to the attenuation behavior often regarded as a good approximation for soft tissue. For the same boundary and initial conditions as previously assumed, (12) and (14) yield

$$v(x,t)=\frac{v_0}{2\pi}\int_{-\infty}^{\infty}\sin(\omega_0 t')H(t')$$
$$\times\left(\int_{-\infty}^{\infty}e^{-\alpha(\omega)x-j\omega[t'-t+x/c(\omega)]}d\omega\right)dt'. \quad (20)$$

For $n=1$, the dispersion relation given by (6) enables (20) to be expressed as

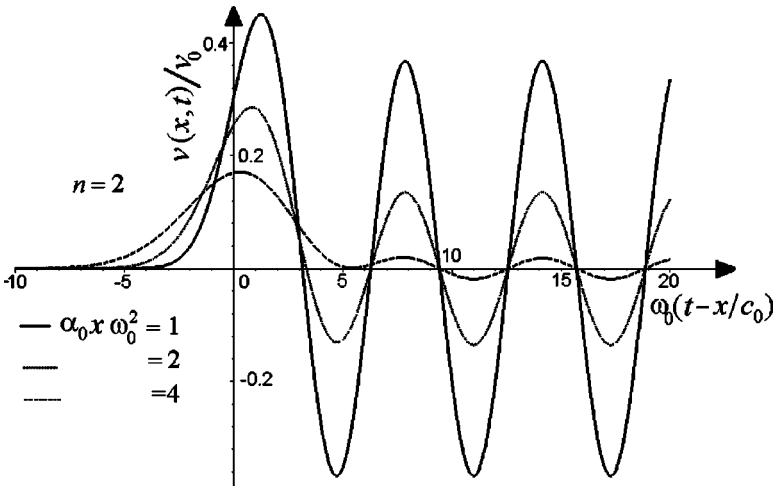


FIG. 4. The normalized particle velocity waveform caused by a sinusoidal transient at $x=0$ is illustrated. The waveforms were calculated from the approximate equation, i.e., (4c) or (17) or (19), for three different observation distances.

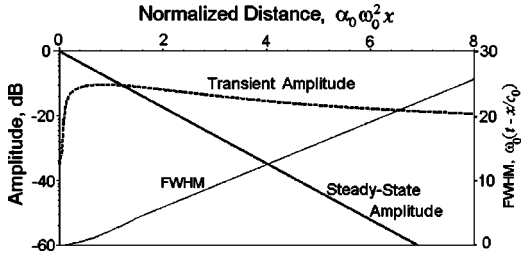


FIG. 5. The peak value of the transient component as obtained from (19) is shown as a function of the normalized distance from the source. Also shown are the variations of the steady-state amplitude component and the transient term FWHM.

$$\nu(x,t)=\frac{\nu_0}{2\pi}\int_{-\infty}^{\infty}\sin(\omega_0 t')H(t')\times\left(\int_{-\infty}^{\infty}e^{-\alpha_0 x|\omega|-j\omega(t'-t+x/c_0+2x/\pi\ln|\omega/\omega_0|)}d\omega\right)dt'. \tag{21}$$

To proceed further in an analytic manner it is necessary to assume the dispersion of the medium is negligible, enabling (21) to be simplified to

$$\nu(x,t)=\frac{\nu_0}{2\pi}\int_{-\infty}^{\infty}\sin(\omega_0 t')H(t')\times\left(\int_{-\infty}^{\infty}e^{-\alpha_0|\omega|x-j\omega(t'-\tau)}d\omega\right)dt', \tag{22}$$

where $\tau=t-x/c_0$. Now the integral in parenthesis can be rewritten as

$$\begin{aligned}&\int_{-\infty}^{0}e^{-\alpha_0|\omega|x}[\cos\omega(t'-\tau)-j\sin\omega(t'-\tau)]d\omega\\&\quad+\int_{0}^{\infty}e^{-\alpha_0|\omega|x}[\cos\omega(t'-\tau)-j\sin\omega(t'-\tau)]d\omega\\&=2\int_{0}^{\infty}e^{-\alpha_0|\omega|x}\cos\omega(t'-\tau)d\omega=\frac{2\alpha_0 x}{(\alpha_0 x)^2+(t'-\tau)^2}.\end{aligned} \tag{23}$$

Substituting (23) into (21) yields

$$\nu(x,t)=\frac{\nu_0\alpha_0 x}{\pi}\int_0^{\infty}\frac{\sin(\omega_0 t')}{(\alpha_0 x)^2+(t'-\tau)^2}dt'. \tag{24}$$

If it is recalled that $y=\omega_0(t-x/c_0)$, the integral in the earlier equation can be written as

$$\begin{aligned}&\int_0^{\infty}\frac{\sin(\omega_0 t')}{(\alpha_0 x)^2+(t'-\tau)^2}dt'\\&=\frac{1}{4\alpha_0 x}\{i\pi[\cosh(B+iy)-\cosh(B-iy)]\\&\quad+2\mathrm{Shi}(B+iy)\cosh(B+iy)+2\mathrm{Shi}(B-iy)\\&\quad\times\cosh(B-iy)-2\mathrm{Ci}(-y+iB)\sinh(B+iy)\\&\quad\times-2\mathrm{Ci}(-y-iB)\sinh(B-iy)\},\end{aligned} \tag{25}$$

where $\mathrm{Shi}(\cdot)$ and $\mathrm{Ci}(\cdot)$ are the hyperbolic sine and cosine integrals,[24] respectively, and where $B=\alpha_0\omega_0 x$. Substituting (25) into (24) and making use of the symmetry properties of earlier functions results in

$$\frac{\nu(x,t)}{\nu_0}=-\frac{1}{2}\sin(y)\sinh(B)+\frac{1}{\pi}\mathrm{Re}[\mathrm{Shi}(D)\cosh(D)-\mathrm{Ci}(jD)\sinh(D)], \tag{26}$$

where $D=B+jy=\alpha_0\omega_0 x+j\omega_0(t-x/c_0)$. In order to check that (26) gives the correct boundary conditions, we take the limit as $x\to 0$. This yields

$$\begin{aligned}\frac{\nu(0,t)}{\nu_0}&=\sin(\omega_0 t)-\frac{1}{\pi}\mathrm{Re}[j\mathrm{Si}(\omega_0 t)\cos(\omega_0 t)\\&\quad-j\mathrm{Ci}(\omega_0 t)\sin(\omega_0 t)]\\&=\sin(\omega_0 t)\left\{1+\frac{1}{\pi}\mathrm{Re}[j\mathrm{Ci}(\omega_0 t)]\right\}\\&=H(t)\sin(\omega_0 t),\end{aligned} \tag{27}$$

where $\mathrm{Si}(y)$ is the sine integral.[24]

An alternative approach is based on velocity impulse response. For $n=1$, the impulse response can be obtained from (12) as

$$\nu_\delta(x,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{j\omega[t-x/c_0]-\omega x\alpha_0}d\omega.$$

But

$$\int_{-\infty}^{\infty}e^{-\alpha_0|\omega|x+j\omega\tau}d\omega=2\int_0^{\infty}e^{-\alpha_0|\omega|x}\cos(\omega\tau)d\omega=\frac{2\alpha_0 x}{(\alpha_0 x)^2+\tau^2},$$

so that

$$\nu_\delta(x,t)=\frac{\alpha_0 x/\pi}{(\alpha_0 x)^2+(t-x/c_0)^2}.$$

By convolving this with $\nu(0,t)=H(t)\nu_0\sin(\omega_0 t)$, it can be shown that at the observation point the waveform is given by

$$\begin{aligned}\nu(x,t)=\frac{j}{2\pi}\{&-\pi\cos(-y+j\omega_0\alpha_0 x)\\&+\pi\cos(y+j\omega_0\alpha_0 x)\\&-2\mathrm{Si}(y+j\omega_0\alpha_0 x)\cos(y+j\omega_0\alpha_0 x)\\&+2\mathrm{Ci}(y+j\alpha_0\omega_0 x)\sin(y+j\omega_0\alpha_0 x)\\&-2\mathrm{Si}(-y+j\omega_0\alpha_0 x)\cos(-y+j\omega_0\alpha_0 x)\\&+2\mathrm{Ci}(y-j\omega_0\alpha_0 x)\sin(-y+j\omega_0\alpha_0 x)\},\end{aligned}$$

which can be shown to be identical to (26).

The steady-state, transient and total response are shown in Fig. 6 for a given spatial distance. Comparison with Fig. 4

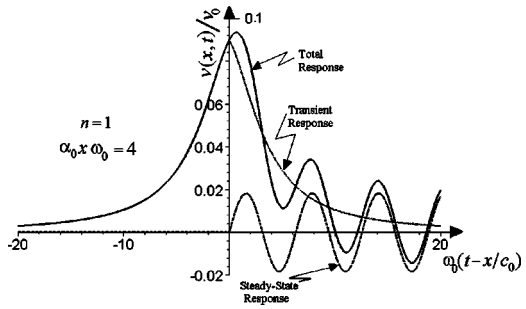


FIG. 6. The normalized velocity waveform caused by a transient sinusoidal source is shown for propagation in a tissue-like medium ($n=1$). The total response and its components were calculated from (26).

for $\alpha_0\omega_0^2 x=4$ shows that the transient amplitude is less, though the sinusoidal amplitudes are identical, as expected. By comparison of Figs. 7 and 5, the transient amplitude for $n=1$ decays more rapidly with distance than for $n=2$. Moreover the FWHM increases less rapidly with distance than for $n=2$. The effects of the frequency dependent attenuation on the transient component, as discussed earlier, provides a physical explanation of these changes.

VI. GENERAL DISPERSIVE SOLUTION

Of considerable practical interest is the causal solution for n in the range for $0<n<3$. By substituting (7) into (12) and noting that $\nu(0,\omega)$ is the Fourier transform of $\nu(0,t)$, the velocity can be expressed as

$$\nu(x,t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\nu(0,t')\left\{\int_{-8}^{\infty}\exp\left\{-j\omega x \times\left[\frac{1}{c_0}+\alpha_0\tan\left(\frac{\pi n}{2}\right)(|\omega|^{n-1} -|\omega_0|^{n-1})\right]-\alpha_0 x|\omega|^n+j\omega(t-t')\right\}d\omega\right\}dt'.$$

This can be rewritten as

$$\frac{\nu(x,t)}{\nu_0}=\Im^{-1}\left\{\Im[\nu(0,t)]\exp\left(-\alpha_0 x|\omega|^n-j\omega x \times\left\{\frac{1}{c_0}+\alpha_0\tan\left(n\frac{\pi}{2}\right)[|\omega|^{n-1}-|\omega_0|^{n-1}]\right\}\right)\right\}. \tag{28}$$

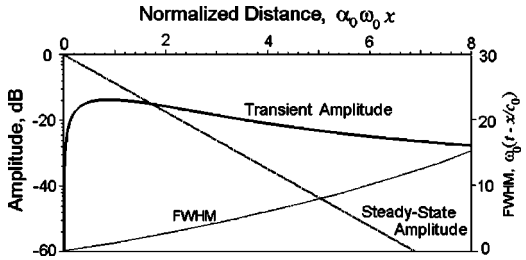


FIG. 7. Peak values of the transient components, obtained from (26) relative to the steady-state amplitude $\exp(-\alpha_0 x\omega_0)$, are shown as a function of the normalized distance from the source. Also shown is the variation of the transient term FWHM as a function of distance.

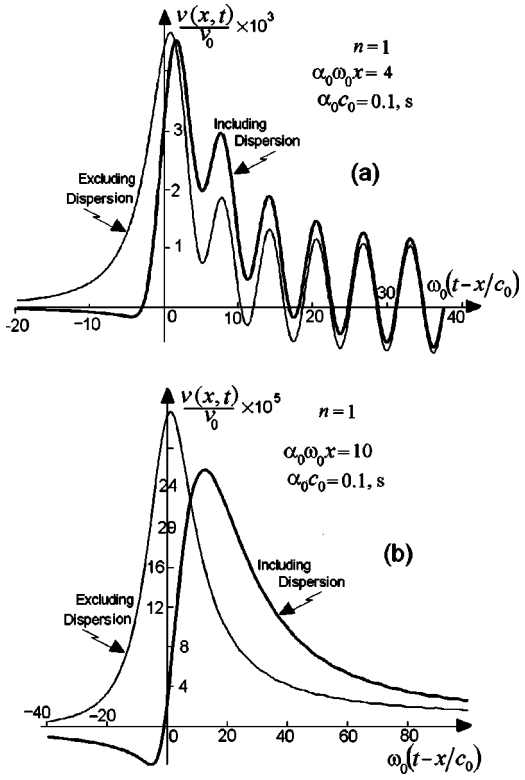


FIG. 8. The effects of speed dispersion are shown for propagation of a sinusoidal transient signal into a medium whose attenuation is proportional to the frequency ($n=1$). (a) $\alpha_0\omega_0 x=4$. (b) $\alpha_0\omega_0 x=10$. In (b), if $c_0=1500$ m/s then $\alpha_0=6.666\times10^{-5}$ Np/(m rad/s), i.e., 36.3 dB/(MHz m), so that for an excitation frequency of 100 kHz, $\alpha_0\omega_0 x=10$ corresponds to an observation point 23.9 cm from the source.

When $n=1$ a more convenient computational form can be obtained from (6) as

$$\frac{\nu(x,t)}{\nu_0}=\Im^{-1}\left\{\Im[\nu(0,t)] \times\exp\left(-\alpha_0 x|\omega|-j\omega x\left[\frac{1}{c_0}-\frac{2}{\pi}\alpha_0\ln\left|\frac{\omega}{\omega_0}\right|\right]\right)\right\}. \tag{29}$$

Numerical results

Three sets of results will be used to show the effects of dispersion. Figure 8 shows the time-retarded response for a sinusoidal step-function in a medium with $n=1$ at two different depths: $x=5$ cm in Fig. 8(a) and $x=30$ cm in Fig. 8(b). The panels show that that the inclusion of dispersion causes the waveform to be delayed, corresponding to anomalous dispersion, and that the delay increases with depth. As expected, the effective duration of the dispersive transient component also increases with depth due to the decrease in center frequency as the wave propagates through the medium. It can also be seen that the amplitude of the transient component is slightly reduced when dispersion is included compared to the nondispersive case.

Of particular practical interest is the effect of dispersion on the transmission of a Gaussian pulse in a tissue-like medium, similar to those used in B-mode imaging systems. Figure 9(a) shows that for a pulse with a bandwidth of 75%,

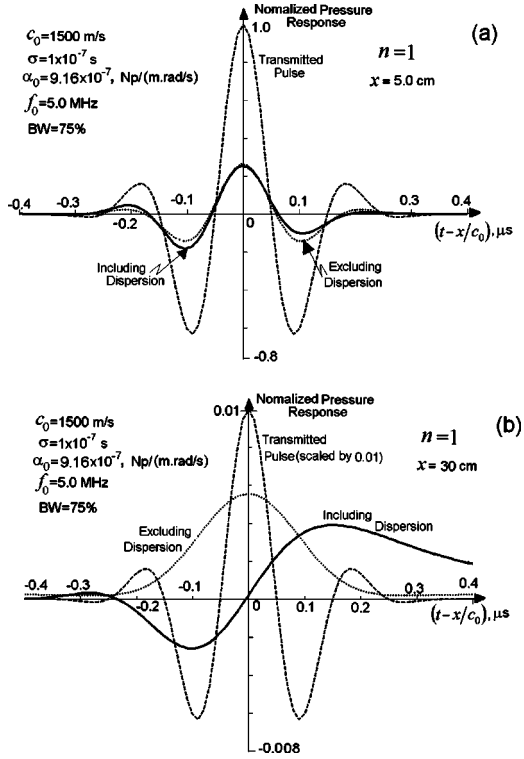


FIG. 9. The effects of dispersion on the propagation of a Gaussian pulse with a 75% bandwidth into a tissue-like medium ($n=1.0$) are demonstrated. The received pulse, when dispersion is either ignored or accounted for, are shown for depths of (a) 5.0 and (b) 30 cm. The time-retarded received pressure waveform is normalized with respect to the transmitted pulse.

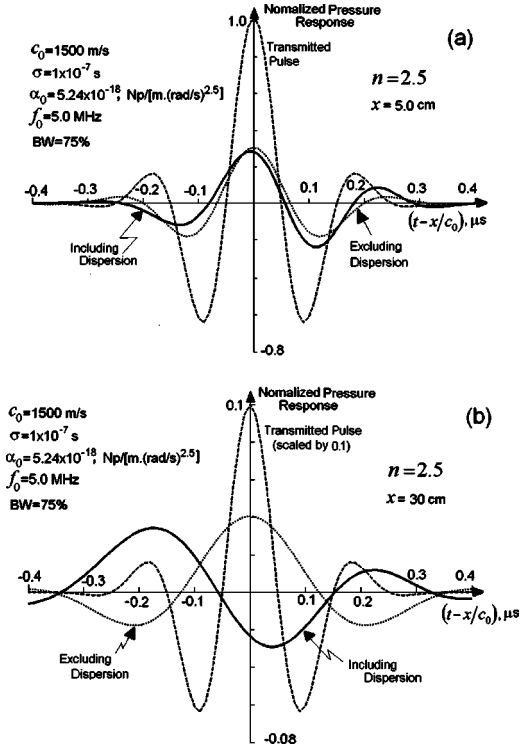


FIG. 10. The effects of dispersion on the propagation of a Gaussian pulse with a 75% bandwidth into a medium with $n=2.5$ are illustrated. The transmitted pulse and the received pulse, when dispersion is either ignored or accounted for, are shown for depths of: (a) 5.0 and (b) 30 cm. The time-retarded received pressure waveform is normalized with respect to the transmitted pulse.

propagating in a power-law medium characterized by $n=1$, attenuation does not cause significant distortion of the pulse shape (ignoring changes in amplitude) when the two-way propagation distance is 5 cm. Additionally, Fig. 9(a) shows that dispersion does not significantly alter the pulse shape at this distance. In contrast, when the two-way path distance is 30 cm, the cumulative effects of dispersion are more pronounced, and two distinct changes can be observed in the resulting waveform shown in Fig. 9(b). First, the dispersionless model produces a symmetrical waveform about the time axis, whereas accounting for dispersion results in asymmetry in the waveform due to the different propagation speeds on the various frequency components of the wave. Second, the moment of the velocity response occurs later in time, as expected for an anomalous dispersive medium ($n<2$). In the case of normal dispersion ($n=2.5$), Fig. 10(a) shows similar results for the 5 cm case (the corresponding attenuation coefficient has been adjusted to give roughly equal peak amplitude levels). In Fig. 10(b), the distortion of the resulting waveform is similar to that of the previous case, however, the direction of time offset is reversed, as would be expected when propagating in a normal dispersive medium ($n>2$).

It is seen from (28) that the contribution of the dispersion term to the signal is proportional to the dimensionless parameter αx. This term incorporates the attenuation coefficient, the frequency, and the propagation distance, and can be understood as an aggregate parameter representing the total attenuation experienced by a given wave propagating through a medium. When two identical waves propagate through different media for a different propagation distance, the distorted waves will be identical if the corresponding values of $\alpha(\omega)x$ are equal. The value of this term may serve as criteria to estimate approximately whether effects of dispersion should be included into consideration. For example, for a propagation distance of 5 cm (Figs. 9(a) and 10(a)) and using the given values of α_0 and ω_0, we obtain $\alpha x \approx 1.4$. When the propagation distance is increased to 10 cm, $\alpha x \approx 8.6$ (Figs. 9(b) and 10(b)). Therefore, the effects of dispersion are noticeable when $\alpha x \approx 1.4$ and the effects become substantial when $\alpha x \approx 8.6$.

VII. CONCLUSIONS

We have shown that our general solution as given by (5) can be used to calculate the effects of dispersive attenuation on the propagation of an acoustic field in media whose attenuation is a power law function of frequency. For $n=2$, dispersion is absent, and exact analytical expressions are obtained for the impulse response and response to a sinusoidal step. This condition corresponds closely to propagation in a classical viscous medium whose behavior is generally assumed to be governed by (1). Comparisons are made between the exact impulse response of (1) and that for a medium whose exponent is exactly 2. As noted by others, for a sinusoidal step function source the transient component of the response has an amplitude that increases with depth, relative to the steady-state amplitude. An intuitive explanation is provided based on the frequency content of the transient and

steady-state solutions. For tissue-like media with $n \approx 1$ whose characteristics exhibit anamolous dispersion, the effects of dispersion are studied for both a sinusoidal step and a Gaussian source. A dimensionless parameter αx has been shown to represent the overall effects of attenuation and dispersion. Results indicate that for $\alpha x \leqslant 1.4$, the effect of dispersion is not significant and can be ignored, while for large values such as $\alpha x \approx 8.6$, the effects of dispersion significantly alter the pulse shape. Our results indicate that for a wideband Gaussian modulated sinusoidal source, the effects of dispersion, combined with attenuation can have a profound effect on the shape of the pulse when propagating through a tissue-like medium.

ACKNOWLEDGMENTS

The authors thank the National Science and Engineering Research Council of Canada and the Canadian Institute of Health Research for the financial support that enabled this research to be conducted.

APPENDIX

Our purpose is to show that the impulse response can be expressed as an inverse Fourier transform. The Fourier transform of (1) can be written as

$$\frac{\partial^2 \nu(x,\omega)}{\partial x^2} + \frac{\omega^2 \nu(x,\omega)}{c_0^2(1+j\omega\chi)} = 0,$$

which has a particular solution of

$$\nu(x,\omega) = \nu(0,\omega) e^{\pm j\omega x/c_0\sqrt{(1+j\omega\chi)}},$$

where the source frequency spectrum is denoted by $\nu(0,\omega)$. Consequently, if the source consists of a δ-function at $x=0$ and $t=0$, i.e., $\nu(0,\omega)=1$, the resulting particle velocity is given by

$$\nu_\delta(x,t) = \Re e\{\Im^{-1}[e^{\pm j\omega x/c_0\sqrt{(1+j\omega\chi)}}]\}.$$

The time response at various locations were calculated from this expression and found to be in agreement with those presented by Ludwig and Levin.[9] If we assume that the absorption is small and the observation point is sufficiently distant from the source so that the high-frequency components of the spectrum are small, it can be shown that the earlier equation reduces to (13).

[1] J. M. M. Pinkerton, "The absorption of ultrasonic waves in liquids and its relation to molecular constitution," Proc. Phys. Soc. London, Sect. B **62**, 129–141 (1949).
[2] J. C. Bamber, "Ultrasonic properties of tissues," in *Ultrasound in Medicine*, edited by F. A. Duck, A. C. Baker, and H. C. Starritt (Inst. of Physics Pub., Bristol, 1998), Chap. 4, pp. 57–88.
[3] N. V. Sushilov and R. S. C. Cobbold, "Frequency-domain wave equation and its time-domain solutions in attenuating media," J. Acoust. Soc. Am. **115**, 1431–1436 (2004).
[4] G. G. Stokes, "On the theories of the internal friction of fluids in motion, and of the equilibrium and motion of elastic solids," Trans. Cambridge Philos. Soc. **8**, 287–319 (1845).
[5] J. Stefan, "Über den Einfluß der inneren Reibung in der Luft auf die Schallbewegung," (On the effect of inner friction on the propagation of sound in air) Sitzungber, Akad. Wien, Math.-Naturwiss. Kl. **53**, 529–527 (1866).
[6] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Part 1, Sec. 7.4.
[7] G. Barton, *Elements of Green's Functions and Propagation: Potentials, Diffusion and Waves* (Oxford University Press, Oxford, 1989), Sec. 8.3.2.
[8] M. Hanin, "Propagation of an aperiodic wave in a compressible viscous medium," J. Math. Phys. **37**, 234–249 (1957).
[9] R. Ludwig and P. L. Levin, "Analytical and numerical treatment of pulsed wave propagation into a viscous fluid," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **42**, 789–792 (1995).
[10] F. R. Norwood, "Propagation of transient sound signals into a viscous fluid," J. Acoust. Soc. Am. **44**, 450–457 (1968).
[11] D. T. Blackstock, "Transient solution for sound radiated into a viscous fluid," J. Acoust. Soc. Am. **41**, 1312–1319 (1967).
[12] M. B. Moffett and R. T. Beyer, "Transient effects in the propagation of a sound pulse in a viscous liquid," J. Acoust. Soc. Am. **47**, 1241–1249 (1970).
[13] P. M. Jordan, M. R. Meyer, and A. Puri, "Causal implications of viscous damping in compressible fluid flows," Phys. Rev. E **62**, 7918–7926 (2000).
[14] T. L. Szabo, "Time domain wave equation for lossy media obeying a frequency power law," J. Acoust. Soc. Am. **96**, 491–500 (1994).
[15] T. L. Szabo, "Causal theories and data for acoustic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **97**, 14–24 (1995).
[16] K. R. Waters, M. S. Hughes, J. Mobley, G. H. Brandenburger, and J. G. Miller, "On the applicability of Kramers–Kronig relations for ultrasonic attenuation obeying a frequency power law," J. Acoust. Soc. Am. **108**, 556–563 (2000).
[17] K. R. Waters, M. S. Hughes, G. H. Brandenburger, and J. G. Miller, "On a time-domain representation of the Kramers–Kronig dispersion relations," J. Acoust. Soc. Am. **108**, 2114–2119 (2000).
[18] K. R. Waters, M. S. Hughes, J. Mobley, and J. G. Miller, "Differential forms of the Kramers-Kronig dispersion relations," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **50**, 68–76 (2003).
[19] F. A. Duck, *Physical Properties of Tissue. A Comprehensive Reference Book* (Academic, London, 1990).
[20] K. A. Wear, "Measurement of phase velocity and group velocity in human calcaneus," Ultrasound Med. Biol. **26**, 641–646 (2000).
[21] S. Chaffai, F. Padilla, G. Berger, and P. Laugier, "*In vitro* measurement of the frequency-dependent attenuation in cancellous bone between 0.2 and 2 MHz," J. Acoust. Soc. Am. **108**, 1281–1289 (2000).
[22] I. S. Gradstein and I. M. Ryzhik, *Tables of Integrals, Series and Products*, 4th ed. (Academic, New York, 1965).
[23] M. B. Moffett and R. T. Beyer, "Transient effects in the propagation of a sound pulse in a viscous liquid," J. Acoust. Soc. Am. **47**, 1241–1249 (1970).
[24] *Handbook of Mathematical Functions*, edited by M. Abramovitz and I. A. Stegun (Dover, New York, 1965), Chap. 7.

The complex equivalent source method for sound propagation over an impedance plane

Martin Ochmann[a)]
TFH Berlin—University of Applied Sciences, Fachbereich Mathematik, Physik und Chemie, Luxemburger Strasse 10, 13353 Berlin, Germany

(Received 19 May 2004; revised 24 September 2004; accepted 29 September 2004)

The sound field caused by a monopole source above an impedance plane can be calculated by using a superposition of equivalent point sources located along a line in the mirror space below the plane. Originally, such an approach for representing the half-space Green's function was described by Sommerfeld at the beginning of the last century, in order to treat half-space problems of heat conduction. However, the representation converges only for masslike impedances and cannot be used for the more important case of reflecting planes with springlike surface impedances. The singular part of the line integral can be transformed into a Hankel function, which shows that surface waves are contained in the whole solution. Unfortunately, this representation suffers from the lack of validity at certain receiver points and from restrictions on wave number and impedance range to ensure the necessary convergence. The main idea of the present method is to use also a superposition of equivalent point sources, but to allow that these sources can be located at complex source points. The corresponding form of the half-space Green's function is suitable for both masslike and springlike surface impedances, and can be used as a cornerstone for a boundary element method. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1819504]

PACS numbers: 43.20.Bi, 43.20.Fn, 43.28.Fp, 43.20.Rz [ANN] Pages: 3304–3311

I. INTRODUCTION

In 1909 and in later years, Sommerfeld[1–3] investigated the electromagnetic field of a point source in a half-space bounded by an infinite plane with extended reaction. Since that time the problem of determining the corresponding half-space Green's function and the associated acoustical problem was treated by many researchers. Review articles including comprehensive lists of references can be found in Refs. 4, 5, and 6. If the half-space Green's function is known, it can be incorporated in a boundary element method (BEM);[7–12] consequently, there is no need to model the infinite plane itself, too. However, the BEM requires that the Green's function be differentiated in the normal direction (once or twice) and evaluated many times during the BEM calculation. Therefore, the expression for the Green's function should be constructed as simply as possible, in order to lead to a fast and efficient numerical method, especially if the radiation or scattering from complex structures consisting of many thousand boundary elements is considered. On the other hand, if an equivalent source method[13] is used for calculating the sound radiation from vibrating bodies above impedance planes, it is necessary to derive not only the field of monopole sources over infinite planes, but also the field of multipole sources of arbitrary order (cf. Ref. 13 or 14). For this purpose, the aim of the present paper is to derive a representation of the half-space Green's function G, which fulfills three achievements. First, G should be easily constructed by a superposition of simple sources, which could be computed numerically in an efficient way. Second, the formula for G should not contain singularities at certain receiver points, other restrictive conditions for ensuring suitable convergence behavior, or complicated mathematical expressions, for example, contour integrals in the complex plane. Third, it should be possible to generalize G from monopole scattering to multipole scattering in a straightforward manner. In addition, it would be desirable for the physical meaning of all terms in the representation of G to become as evident as possible.

To achieve these purposes, two methods were combined: the original ''ancient'' approach of Sommerfeld (see Ref. 3, Chap. 13), where a superposition of point sources for treating problems of heat conduction is used, and the method of complex source points, which was recently proposed in acoustical literature for the investigation of sound propagation, too (see Refs. 15–17). The combination of both methods leads to a powerful expression for G. Using only point sources with real source points results in a representation of G, which only converges for masslike boundary conditions. However, if the point sources are allowed to have complex locations, the expression obtained for G is valid for masslike as well as for springlike impedances, and satisfies all three requirements postulated above. In the following, the method will be called the *complex equivalent source method* (for short: CESM).

In a subsequent paper, the CESM will be generalized, in order to derive an expression for the reflected sound field of an arbitrary multipole over an impedance plane.

Further work will be devoted to a detailed numerical investigation of the CESM together with its incorporation in the BEM.

a)Electronic mail: ochmann@tfh-berlin.de

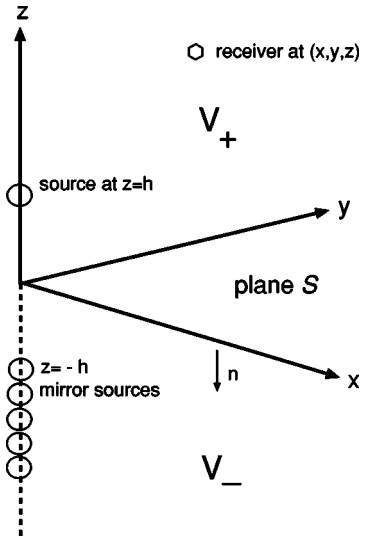


FIG. 1. Geometry of the half-space problem.

II. DESCRIPTION OF THE PROBLEM

A monopole source in the frequency domain at the source location $\boldsymbol{y}=(x_s,y_s,z_s)=(0,0,h)$ in the three-dimensional space is given by

$$g(\boldsymbol{x},\boldsymbol{y})=\exp(ikR)/(4\pi R),\quad R=\sqrt{x^2+y^2+(z-h)^2},\quad (1)$$

where $\boldsymbol{x}=(x,y,z)$ is the receiver position, $k=\omega/c$ is the wave number with angular frequency ω, and c is the speed of sound. All time-varying quantities should obey the time dependence $\exp(-j\omega t)$ with $j=\sqrt{-1}$. It is well-known that $g(\boldsymbol{x},\boldsymbol{y})$ is the free-space Green's function and solves the inhomogeneous Helmholtz equation

$$\Delta g+k^2g=-\delta(x,y,z-h),\quad (2)$$

where Δ is the Laplace operator. As shown in Fig. 1, it is now assumed that an infinite reflecting plane S is located at $z=0$. The plane separates the whole space into the upper half-space $V_+:=\{(x,y,z)|z>0\}$ and the lower half-space $V_-:=\{(x,y,z)|z<0\}$. On the plane S, a normal impedance boundary condition is prescribed

$$\frac{p}{v_n}=Z\quad \text{on } S.\quad (3)$$

Here, p is the complex sound pressure, v_n the normal velocity, and Z the normal impedance, where the unit normal $n=(0,0,-1)$ is directed into the exterior of the acoustic domain V_+. Since the gradient of the pressure is given by

$$\frac{\partial p}{\partial n}=j\omega\rho v_n\quad \text{on } S,\quad (4)$$

the boundary condition (3) can be written as

$$\frac{\partial p}{\partial n}-\gamma p=0\quad \text{on } S.\quad (5)$$

Here, ρ is the fluid density, $\partial/\partial n$ is the derivative in the direction of the normal n, and $\gamma=jk/Z_0$ with the normalized acoustical impedance $Z_0=Z/(\rho c)$. Now, we are looking for a solution of the Helmholtz equation (2) together with the boundary condition (5) in the half-space V_+. Such a solution is also called a half-space Green's function.

III. THE HALF-SPACE GREEN'S FUNCTION FOR A MASSLIKE IMPEDANCE PLANE

A. The equivalent source method of Sommerfeld

In Ref. 3, Chap. 3, paragraph 13, Sommerfeld treated half-space problems of heat conduction by writing the total thermal field as a superposition of an original heat source, a mirror source, and a line integral combined of single thermal sources placed at the z axis below the mirror source as shown in Fig. 1.

By using just the same approach for the corresponding acoustical problem, we try to find a solution of Eqs. (2) and (5) in the form

$$G(\boldsymbol{x},\boldsymbol{y})=\hat{g}(h)+A\hat{g}(-h)+\int_{-\infty}^{-h}a(\eta)\hat{g}(\eta)\,d\eta,\quad (6)$$

with the unknown coefficient A and the unknown amplitude function $a(\eta)$ of the mirror sources along the z axis. For brevity, the free-space Green's function (1) is written as

$$\hat{g}(h):=g\left(\begin{pmatrix}x\\y\\z\end{pmatrix},\begin{pmatrix}0\\0\\h\end{pmatrix}\right)=\frac{1}{4\pi}\frac{e^{jk\sqrt{x^2+y^2+(z-h)^2}}}{\sqrt{x^2+y^2+(z-h)^2}}.\quad (7)$$

Thus, $\hat{g}(-h)$ is the mirror source of $\hat{g}(h)$ with respect to the plane S. The second and third terms on the right-hand side of Eq. (6) represent the field reflected from the plane S. Clearly, $G(\boldsymbol{x},\boldsymbol{y})$ is a solution of the Helmholtz equation (2) with only one singularity at the source point $(0,0,h)$. All other singularities at points $(0,0,z)$ with $z\leqslant -h$ are located in the nonphysical lower half-space V_-. $G(\boldsymbol{x},\boldsymbol{y})$ can be interpreted as a representation of the half-space Green's function by a superposition of equivalent point sources. Therefore, this way of constructing the Green's function can be considered as a variant of the equivalent source method or the source simulation technique.[13,14]

To satisfy the boundary condition (5) on the surface S and determining A and $a(\eta)$, the normal derivative

$$\frac{\partial G}{\partial n}=-\frac{\partial G}{\partial z},\quad (8)$$

has to be calculated. By performing the differentiation with respect to z and setting $z=0$, one obtains from Eq. (6)

$$4\pi\frac{\partial}{\partial z}G(\boldsymbol{x},\boldsymbol{y})\bigg|_{z=0}=\left(\frac{jk-1/r_h}{r_h}\right)\frac{\exp(jkr_h)}{r_h}(A-1)h-\int_{-\infty}^{-h}a(\eta)\frac{\partial}{\partial\eta}(\exp(jkr_\eta)/r_\eta)\,d\eta,\quad (9)$$

using

$$\frac{\partial\hat{g}(\eta)}{\partial z}=-\frac{\partial\hat{g}(\eta)}{\partial\eta},\quad (10)$$

which follows from definition (7). Here, we have set

$$r_h^2 = x^2 + y^2 + h^2. \tag{11}$$

Special attention must be paid to the line integral: the differentiation can be shifted to the amplitude function $a(\eta)$ by performing a partial integration

$$I := \int_{-\infty}^{-h} a(\eta) \frac{\partial}{\partial \eta} (\exp(jkr_\eta)/r_\eta) d\eta$$
$$= \frac{a(-h)}{r_h} \exp(jkr_h) - \int_{-\infty}^{-h} \frac{da(\eta)}{d\eta} (\exp(jkr_\eta)/r_\eta) d\eta, \tag{12}$$

assuming that the contribution from the lower limit vanishes

$$\frac{a(\eta)}{r_\eta} \exp(jkr_\eta) \xrightarrow[\eta \to -\infty]{} 0. \tag{13}$$

In the following, we will see that condition (13) is valid only if the impedance Z is masslike. Substituting $G(\boldsymbol{x},\boldsymbol{y})$ into the boundary condition (5) and taking into account Eqs. (8)–(12), we obtain

$$-\left(\frac{jk - 1/r_h}{r_h}\right)(A-1)h \frac{\exp(jkr_h)}{r_h} + \frac{a(-h)}{r_h} \exp(jkr_h)$$
$$- \int_{-\infty}^{-h} \frac{da(\eta)}{d\eta} \exp(jkr_\eta)/r_\eta d\eta$$
$$\times (-\gamma) \left\{ (1+A) \frac{\exp(jkr_h)}{r_h} + \int_{-\infty}^{-h} a(\eta) \frac{\exp(jkr_\eta)}{r_\eta} d\eta \right\}$$
$$= 0. \tag{14}$$

As a first step, for satisfying Eq. (14), $A=1$ is chosen. Second, in order to enforce that both integrals and the remaining two terms become zero, two equations remain

$$a(-h) - 2\gamma = 0, \tag{15a}$$

$$da/d\eta + \gamma a(\eta) = 0. \tag{15b}$$

The solution of Eqs. (15) is easily obtained by

$$a(\eta) = 2\gamma \exp(-\gamma(\eta + h)). \tag{16}$$

Hence, by substituting Eq. (16) into Eq. (6), the final result can be expressed as

$$G(\boldsymbol{x},\boldsymbol{y}) = \hat{g}(h) + \hat{g}(-h) + 2e^{-\gamma h} \int_{-\infty}^{-h} e^{-\gamma\eta} \hat{g}(\eta) d\eta, \tag{17}$$

or in a more explicit form

$$G(\boldsymbol{x},\boldsymbol{y}) = \frac{e^{jkr(h)}}{4\pi r(h)} + \frac{e^{jkr(-h)}}{4\pi r(-h)}$$
$$+ 2\gamma e^{-\gamma h} \int_{-\infty}^{-h} e^{-\gamma\eta} \frac{e^{jkr(\eta)}}{4\pi r(\eta)} d\eta, \tag{18}$$

with $r(h) = \sqrt{x^2 + y^2 + (z-h)^2}$.

If the time dependence $\exp(j\omega t)$ is used, the half-space Green's function is simply given by the complex conjugate $G^*(\boldsymbol{x},\boldsymbol{y})$ of Eq. (17) or (18), where it should be noted that

$$\gamma^* = -jk/Z_0^*.$$

Here, the complex conjugate impedance Z_0^* corresponds to the normalized impedance of the problem formulated with respect to the time dependence $\exp(j\omega t)$.

B. Discussion of the result (18) for the half-space Green's function

For $\gamma = 0$ [see Eq. (5)], the plane S becomes perfectly rigid, and Eq. (18) gives the expected solution, which is a superposition of the original source and the corresponding mirror source

$$G(\boldsymbol{x},\boldsymbol{y}) = \frac{e^{jkr(h)}}{4\pi r(h)} + \frac{e^{jkr(-h)}}{4\pi r(-h)}. \tag{19}$$

The perfectly soft condition $p=0$ on S must be treated separately by reformulating the boundary condition (5) in a more general way as

$$\alpha \frac{\partial p}{\partial n} + \beta p = 0 \quad \text{on } S, \tag{20}$$

and letting $\alpha \to 0$. Then, by substituting Eq. (6) into this perfectly soft condition, one obtains

$$G(\boldsymbol{x},\boldsymbol{y})|_{z=0} = (A+1) \frac{e^{jkr_h}}{4\pi r_h} + \int_{-\infty}^{-h} a(\eta) \frac{e^{jkr_\eta}}{4\pi r_\eta} d\eta = 0,$$

which is now satisfied for $A=-1$ and $a(\eta)=0$ and leads to the exact half-space Green's function for the pressure-released case.

However, due to the definition of $\gamma = jk\rho c/Z$, the line integral in Eq. (18) converges to a finite value for a real wave number k only if

$$\mathrm{Re}\{\gamma\} < 0 \Leftrightarrow \mathrm{Im}\{Z\} < 0, \tag{21}$$

where Re{ } or Im{ } denotes the real or imaginary part of the quantity in brackets, respectively. Therefore, the normal impedance of the reflecting plane is enforced to be masslike. Also, by substituting Eq. (16) into Eq. (13), it is obvious that the necessary condition (13)

$$2\gamma e^{-\gamma(\eta+h)} \frac{e^{jkr_\eta}}{r_\eta} \xrightarrow[\eta \to -\infty]{} 0, \tag{22}$$

is satisfied only for masslike impedances. On the other hand, it has been known for many years that surfaces with springlike impedances can lead to the occurrence of surface waves. This topic was extensively discussed in the literature[4–6,18] starting with the work of Sommerfeld.[1–3] The approach (6) seems inadequate to model such surface waves, which may contribute to the divergent behavior of the line integral. However, the imaginary part of the ground impedance (the reactance) is positive for widely used models: For example, Embleton stated that ''for porous ground this imaginary component of the impedance is a compliance, like an elastic cushion'' (see Ref. 6, p. 38). As a second example, the model of Delaney and Bazley (see Ref. 4, p. 74) predicts a positive reactance. This gives the motivation to extend formula (18) to the case of springlike surface impedances with positive reactance in Sec. IV.

C. Comparison with known solutions

Originally, Sommerfeld analyzed the propagation of radio waves above the earth. For this reason, he developed the solution into an integral over Bessel or Hankel functions.[1–3] Such an approach is nowadays called a Hankel transformation.[19] By transferring the electromagnetic to the corresponding acoustical impedance case, the solution of Sommerfeld [Ref. 3, p. 228, Eq. (9)] can be written (please note that Sommerfeld's constant μ_E/n^2 is equal to $-\gamma$ in our notation)

$$G(\boldsymbol{x},\boldsymbol{y}) = \frac{e^{jkr(h)}}{4\pi r(h)} + \frac{e^{jkr(-h)}}{4\pi r(-h)} + \frac{2\gamma}{4\pi}\int_0^\infty J_0(\lambda r)e^{-\mu(z+h)}\frac{\lambda\, d\lambda}{\mu(\mu-\gamma)}, \tag{23}$$

where $r^2=x^2+y^2$, $\mu=\sqrt{\lambda^2-k^2}$, J_0 is the Bessel function of order zero, and $\mathrm{Re}\{\mu\}>0$ must be chosen. For demonstrating the agreement between Eq. (23) and the present solution (18), we use the so-called Sommerfeld identity (see Ref. 15), originally derived by Sommerfeld (see Ref. 3, p. 222 or p. 229)

$$\frac{e^{jk\sqrt{r^2+(z+\eta)^2}}}{\sqrt{r^2+(z+\eta)^2}} = \int_0^\infty J_0(\lambda r)e^{-\mu(z+\eta)}\frac{\lambda\, d\lambda}{\mu}. \tag{24}$$

By multiplying Eq. (24) with $\exp(\gamma\eta)$ and integrating over η from h to infinity, we get

$$\int_h^\infty \frac{e^{jk\sqrt{r^2+(z+\eta)^2}+\gamma\eta}}{\sqrt{r^2+(z+\eta)^2}}\,d\eta = e^{h\gamma}\int_0^\infty J_0(\lambda r)e^{-\mu(z+h)}\frac{\lambda\, d\lambda}{\mu(\mu-\gamma)}, \tag{25}$$

by using

$$\int_h^\infty e^{-(\mu-\gamma)\eta}d\eta = \frac{1}{(\mu-\gamma)}e^{-(\mu-\gamma)h}, \tag{26}$$

because the contribution from the upper limit disappears for masslike impedances with $\mathrm{Re}\{\gamma\}<0$ ($\mathrm{Re}\{\mu\}>0$ is presupposed). By introducing expression (25) into Sommerfeld's representation (23), we obtain

$$G(\boldsymbol{x},\boldsymbol{y}) = \frac{e^{jkr(h)}}{4\pi r(h)} + \frac{e^{jkr(-h)}}{4\pi r(-h)} + \frac{2\gamma}{4\pi}e^{-\gamma h}\int_h^\infty \frac{e^{jk\sqrt{r^2+(z+\eta)^2}+\gamma\eta}}{\sqrt{r^2+(z+\eta)^2}}\,d\eta, \tag{27}$$

which is exactly the same result as presented in Eq. (18), if the variable of integration is changed from η to $-\eta$.

Similar to Sommerfeld, Habault and Filippi[20] also used the Hankel transform pair for investigating sound propagation over a homogeneous ground. With the ground characterized by a locally reacting surface, the authors obtained expression Eq. (36) in Ref. 20. By taking into account that sign $(\hat{\varsigma})=-1$ ($\hat{\varsigma}:=\mathrm{Im}\{Z_0\}$ in Ref. 20) for a masslike surface impedance, and that the right-hand side of the Helmholtz equation (2) is not a negative but a positive Dirac delta function in Ref. 20, it is easy to see that the representation of Habault and Filippi is identical with the present result (18).

A third interesting comparison can be made by studying the paper of Li *et al.*[10] They used a totally different technique, in order to derive the half-space Green's function by expanding it in a power series of the parameter γ. Again, for masslike impedances, their result is in full agreement with formula (18). This can be seen by considering Eq. (24) of Ref. 10: By switching to the present notation, changing the variable of integration from η to $-\eta$, and dividing by 4π due to the different right-hand sides of the Helmholtz equation (2), one gets the exact complex conjugate Green's function $G^*(\boldsymbol{x},\boldsymbol{y})$ (cf. the end of Sec. III A). Thus, the results are identical, since Li *et al.* used the time dependence $\exp(j\omega t)$.

Then, Li *et al.* incorporated the half-space Green's function into a boundary element formulation. By using Gaussian–Laguerre quadrature for evaluating the improper integral numerically, their BEM results agreed very well with analytical solutions. These findings could be confirmed by a recent paper,[21] where G was also used for developing a BEM MATLAB code. However, for springlike impedances the derivation of Li *et al.*[10] was not fully successful, since only an approximate Green's function could be determined.

In conclusion, the equivalent source method of Sommerfeld leads to the exact same result as derived by several authors using different methods. This may be helpful for a deeper understanding due to a statement of the mathematical giant Gauss, who said that ''a variety of methods normally contributes very much to the explanation of things which are somewhat more difficult to understand'' (quoted from Ref. 22, translation from German by the author).

IV. THE HALF-SPACE GREEN'S FUNCTION FOR A SPRINGLIKE IMPEDANCE PLANE

A. The Hankel term solution

How can the present method be extended to the treatment of sound propagation over springlike impedance planes? For a positive reactance $\mathrm{Re}\,\gamma>0$, condition (13) is violated. Hence, the partial integration can not be performed as described in Sec. III A. An obvious approach to overcome this difficulty is to split the integral appearing in Eq. (6) and, hence, in the boundary condition (14) as follows:

$$\hat{I} := = \int_{-\infty}^{-h} a(\eta)\frac{e^{jkr(\eta)}}{r(\eta)}d\eta = \int_{-\infty}^{\infty} a(\eta)\frac{e^{jkr(\eta)}}{r(\eta)}d\eta - \int_{-h}^{\infty} a(\eta)\frac{e^{jkr(\eta)}}{r(\eta)}d\eta =: I_1 - I_2.$$

Thus, the Green's function (6) can be expressed as

$$G(x,y) = \frac{1}{4\pi}\left(\frac{e^{jkr(h)}}{r(h)} + A\frac{e^{jkr(-h)}}{r(-h)} - I_2 + I_1\right). \tag{28}$$

Now, we first apply the boundary condition (5)

$$R(G) := \frac{\partial G}{\partial n} - \gamma G = -\frac{\partial G}{\partial z} - \gamma G = 0 \quad \text{on } S,$$

to the first three terms in Eq. (28) separately

$$R\left(\frac{e^{jkr(h)}}{r(h)}+A\frac{e^{jkr(-h)}}{r(-h)}-I_2\right)=0 \quad \text{on } S.$$

Proceeding along the same lines as before, we again obtain $A=1$ and expression (16)

$$a(\eta)=2\gamma\exp(-\gamma(\eta+h)).$$

Condition (13) changes to

$$\frac{a(\eta)}{r_\eta}\exp(jkr_\eta)\xrightarrow[\eta\to+\infty]{}0, \tag{29}$$

since now the contribution from the upper limit of the integral I_2 must vanish. Clearly, condition (29) is fulfilled for a positive reactance [cf. Eq. (21)]. Consequently, the integral I_1 takes the form

$$I_1=2\gamma\int_{-\infty}^{\infty}e^{-\gamma(\eta+h)}\frac{e^{jk\sqrt{x^2+y^2+(z-\eta)^2}}}{\sqrt{x^2+y^2+(z-\eta)^2}}d\eta. \tag{30}$$

To satisfy the boundary condition completely, it must be shown that the improper integral I_1 exists and that it fulfills the boundary condition $R(I_1)=0$ on S automatically, too. For a proof, we introduce the transformation $\eta_1:=\eta-z$. Then, I_1 can be written as

$$\begin{aligned}I_1&=\int_{-\infty}^{\infty}a(\eta_1+z)\frac{e^{jk\sqrt{x^2+y^2+\eta_1^2}}}{\sqrt{x^2+y^2+\eta_1^2}}d\eta_1\\&=2\gamma\int_{-\infty}^{\infty}e^{-\gamma(\eta_1+z+h)}\frac{e^{jk\sqrt{x^2+y^2+\eta_1^2}}}{\sqrt{x^2+y^2+\eta_1^2}}d\eta_1.\end{aligned} \tag{31}$$

Now, it is no longer necessary to perform a partial integration [cf. Eq. (12)], since the variable z is only contained in the expression $a(\eta_1+z)$. Therefore, we get the equivalence

$$R(I_1)=0\Leftrightarrow\frac{\partial a(\eta_1+z)}{\partial z}+\gamma a(\eta_1+z)=0,$$

which is automatically satisfied for $a(\eta)$ given by Eq. (16). Hence, I_1 fulfills the boundary condition.

Strictly speaking, the integral converges only if condition (29) for $\eta\to\pm\infty$ is satisfied. At a first glance, that seems to be impossible. On the other hand, the integral can be solved analytically by using the generalized formula of Heine,[23] which for $\nu=0$ gives

$$\begin{aligned}&\pi jH_0^{(1)}[(z^2-\varsigma^2)^{1/2}]\\&=\int_{-\infty}^{\infty}e^{jz\cosh t+j\varsigma\sinh t}dt \quad \text{for } \operatorname{Im}\{z\pm\varsigma\}>0.\end{aligned} \tag{32}$$

Here, $H_0^{(1)}$ is the Hankel function of the first kind of order zero.

For that reason, the two coordinate transformations

$$\tau=(\eta-z)/r \quad \text{and} \quad \tau=\sinh u, \tag{33}$$

have been applied, and the integral Eq. (30) becomes

$$I_1=2\gamma e^{-\gamma(z+h)}\int_{-\infty}^{\infty}e^{-jkr[(1/Z_0)\sinh u-\cosh u]}du. \tag{34}$$

Recall that $r=\sqrt{x^2+y^2}$.

Now, the integral is in an appropriate form for applying the formula of Heine (32), which transforms the integral in (34) into a Hankel function. The result is

$$I_1=2\gamma e^{-\gamma(h+z)}(\pi j)H_0^{(1)}\left(kr\sqrt{1-\frac{1}{Z_0^2}}\right), \tag{35}$$

which is valid for

$$\operatorname{Im}\left\{kr\left(1\mp\frac{1}{Z_0}\right)\right\}>0. \tag{36}$$

For a real wave number k, condition (36) is always violated. Nevertheless, the complete solution for springlike impedances

$$\begin{aligned}G(\boldsymbol{x},\boldsymbol{y})=&\frac{e^{jkr(h)}}{4\pi r(h)}+\frac{e^{jkr(-h)}}{4\pi r(-h)}+\frac{1}{2}j\gamma e^{-\gamma(h+z)}\\&\times H_0^{(1)}\left(kr\sqrt{1-\frac{1}{Z_0^2}}\right)-2\gamma e^{-\gamma h}\\&\times\int_{-h}^{\infty}e^{-\gamma\eta}\frac{e^{jkr(\eta)}}{4\pi r(\eta)}d\eta,\end{aligned} \tag{37}$$

which is obtained by substituting Eqs. (28), (16), and (35) into Eq. (6), is in full agreement with results found in the literature:

Sommerfeld (see Ref. 1 or Ref. 3, Chap. 6, p. 233) obtained the Hankel function by performing a contour integral and applying the residue theorem. He called the third term in Eq. (37) the surface wave contribution and the fourth term the volume wave (cf. Ref. 1, Chaps. 5 and 6). He stated (see Ref. 1, pp. 695–696) that such a representation of the scattered field is only valid in the region $r^2=x^2+y^2>0$. In the vicinity of the z axis, the third and the fourth terms contained singularities. These singularities must cancel each other out for $r\to0$, but they destroy the usefulness of formula (37) for a numerical evaluation. Especially, Eq. (37) cannot be incorporated into a boundary element method due to this restricted validity, which seems to be the main drawback of representation (37).

Looking again at Eq. (36) of Habault and Filippi[20] and choosing a positive reactance, so that $\operatorname{sign}\hat{\varsigma}=+1$ (in Ref. 20) follows, Eq. (37) is obtained. Surprisingly, in their derivation no problems of convergence are mentioned and no additional conditions [like (36)] emerge.

B. Superposition of equivalent sources with complex source points

Representations (18) and (37) of the half-space Green's function have the advantage that their physical meaning can easily be interpreted as a superposition of acoustical sources. However, for the more important springlike impedance, the equivalent sources in the last term of Eq. (37) are located in the lower half-space $V_-(-h\leqslant z<0)$ as well as in the acoustical domain $V_+(0\leqslant z<\infty)$. The latter leads to singularities in the vicinity of the z axis for $z>0$, too. In order to find an expression better suited for numerical purposes, we use a completely different approach.

In several papers (see for example, Refs. 15–17), the idea appeared to use a discrete sum of simple acoustic sources with complex source points for the modeling of

sound propagation. This idea was originally introduced by Deschamps[24] and then enhanced by Felsen and other authors (see, for example, Refs. 25–28). They pointed out that spherical waves with complex source locations could be used to model the propagation of Gaussian beams. For example, if the complex location $(x,y,z)=(0,0,jb)$ with real b is used as source point of a monopole, then the corresponding field approximately behaves like a Gaussian beam in the paraxial region near the z axis (details can be found in Ref. 25). On the other hand, these complex sources are exact solutions of the Helmholtz equation.

The keystone of the approach presented in this paper is to use the source function

$$\hat{g}(-h+j\varsigma)=\frac{1}{4\pi}\frac{e^{jk\sqrt{x^2+y^2+(z+h-j\varsigma)^2}}}{\sqrt{x^2+y^2+(z+h-j\varsigma)^2}} \quad \text{for } -\infty<\varsigma\leqslant 0 \ (\varsigma \text{ real}), \tag{38}$$

instead of

$$\hat{g}(-h+\eta)=\frac{1}{4\pi}\frac{e^{jk\sqrt{x^2+y^2+(z+h-\eta)^2}}}{\sqrt{x^2+y^2+(z+h-\eta)^2}} \quad \text{for } -\infty<\eta\leqslant 0, \tag{39}$$

in the line integral of Eq. (6). Therefore, instead of using a superposition of monopole sources along the negative z axis [starting at the mirror point $(x,y,z)=(0,0,-h)$], we start again at the same mirror point, but now source functions are added along an ''imaginary z axis'' at the points $z=-h+j\varsigma$. Thus, the approach for the half-space Green's function is now [cf. Eq. (6)]

$$G(\boldsymbol{x},\boldsymbol{y})=\hat{g}(h)+A\hat{g}(-h)+\int_{-\infty}^{0}b(\varsigma)\hat{g}(-h+j\varsigma)d\varsigma. \tag{40}$$

By substituting expression (40) into the impedance boundary condition and proceeding just in the same way as described in Sec. III A, we get

$$A=1 \quad \text{and} \quad b(\varsigma)=2j\gamma e^{-j\gamma\varsigma}. \tag{41}$$

It follows that the entire half-space Green's function can be written as

$$G(\boldsymbol{x},\boldsymbol{y})=\hat{g}(h)+\hat{g}(-h)+\frac{j\gamma}{2\pi}\int_{-\infty}^{0}\frac{e^{jk\sqrt{x^2+y^2+(z+h-j\varsigma)^2}}}{\sqrt{x^2+y^2+(z+h-j\varsigma)^2}}e^{-j\gamma\varsigma}d\varsigma. \tag{42}$$

The change from real to complex source points has the enormous advantage that the line integral is now convergent for springlike surface impedances, too. This can be proven by considering the kernel

$$K(x,y,z;\varsigma)=\frac{b(\varsigma)}{\sqrt{x^2+y^2+(z+h-j\varsigma)^2}}\times\exp(jk\sqrt{x^2+y^2+(z+h-j\varsigma)^2}), \tag{43}$$

of the integral. The first factor

$$b(\varsigma)=2j\gamma e^{-j\gamma\varsigma}=2j\gamma e^{k\varsigma/Z_0}, \tag{44}$$

decays exponentially for real wave numbers k and $\varsigma\rightarrow-\infty$, if

$$\mathrm{Re}\left\{\frac{1}{Z_0}\right\}>0\Leftrightarrow\mathrm{Re}\{Z_0\}>0. \tag{45}$$

Such an assumption is true if the surface impedance possesses a small real part, which is a reasonable assumption, since acoustical propagation always involves dissipative processes. However, Eq. (45) is not a necessary condition, because the second factor of the kernel (43) also decays exponentially, thereby allowing pure imaginary surface impedances, too. This fact can be verified by observing that the square-root symbol $\sqrt{\ }$ in Eq. (43) must represent the root with a positive real part, in order to fulfill the radiation condition for the time dependency $\exp(-j\omega t)$. If we now choose receiver points at the z axis $(x=y=0)$ for simplicity, we obtain for the exponential function in Eq. (43) $(z>0, h>0)$

$$e^{jk\sqrt{(z+h-j\varsigma)^2}}=e^{+jk(z+h)}e^{k\varsigma}\xrightarrow[\varsigma\rightarrow-\infty]{}0. \tag{46}$$

Consequently, the improper integral in Eq. (42) exists for all kinds of surface impedances. It converges much faster then the one in representation (18), since the corresponding source function (39) is a rapidly oscillating function for $\eta\rightarrow-\infty$, whereas the source function (38) decays exponentially.

C. Discussion of the main result Eq. (42)

The question arises whether a similar formula for the half-space Green's function can be found in the related literature. The answer is yes. In Ref. 29, Nobile and Hayek investigated a similar problem and started by using the Hankel transform. In the middle of their derivation, they obtained a line integral [Eq. (7) in Ref. 29]. Taking in to account the nomenclature of Nobile and Hayek, changing their variable of integration from η to $-\varsigma$, and dividing by 4π, their result exactly agrees with the representation of the half-space Green's function (42) and, thus, can be interpreted as a superposition of sources with complex source points, too.

Also, in the electromagnetic literature, line integrals over source points with complex source locations were considered, in order to treat half-space problems as pointed out by one of the referees: In Ref. 30, Lindell and Alanen obtained a solution by applying a two-dimensional Fourier transform to the basic equations and by expressing the reflection coefficient as the Laplace transform of a certain function. The final result is a line integral containing Bessel functions. Convergence of the Bessel functions can only be achieved if the image current is located in complex space, too.

Some care is needed if the time dependence $\exp(j\omega t)$ is used. In this case, the sources in the line integral must be located along the z axis at points with

$$z=-h-j\varsigma \quad \text{for} \quad -\infty<\varsigma\leqslant 0, \tag{47}$$

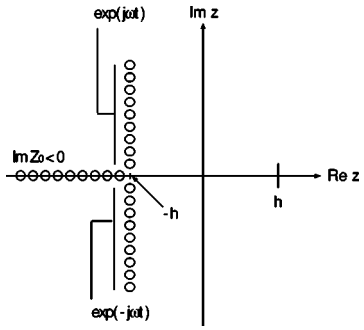


FIG. 2. Location of equivalent sources in the complex z plane depending on surface impedance and time dependency.

to ensure that the kernel of the integral decays exponentially for all kinds of surface impedances. Again, the root which yields a positive real part must be taken in order to satisfy the radiation condition at infinity. A graphical illustration of how to position the source points in the complex z plane for different time dependences is given in Fig. 2.

The complex equivalent sources behave in a similar manner as surface waves, which explains the close relationship between formulas (37) and (42). The square root in Eq. (38) becomes zero if and only if

$$z=-h,\ \ x^2+y^2=\varsigma^2.$$

Whereas the ordinary monopole source becomes singular at a single point (point source), the complex point source shows a different behavior: It becomes singular on circles around the z axis at $z=-h$ with radius ς (circle or ring sources), and for $|\varsigma|\to\infty$ the radius of these circles of singularity increases more and more like a growing circular surface wave. Figure 3 shows the development of a complex point source with decreasing values of ς. Also, the close relationship of complex equivalent sources with Gaussian beams can be easily seen and is described in detail by Felsen,[25] for example.

V. CONCLUSIONS

The complex source representation for the half-space Green's function has the following advantages:

(1) It can be used for positive as well as for negative reactances of the surface impedance. In all cases, the improper line integral converges, since the kernel consists of complex image sources which decay exponentially.
(2) No singularities occur in the acoustic domain V_+, since the singularities of the complex image sources are located on circles around the mirror point in the plane $z=-h$.
(3) The representation (42) is valid for all receiver positions — even if the receiver is placed directly above or below the source in contrast to the Hankel presentation in formula (37). Therefore, it can be incorporated into a boundary element code, in order to take into account the influence of the impedance plane automatically without the need of preparing a finite-element model of the plane, too.

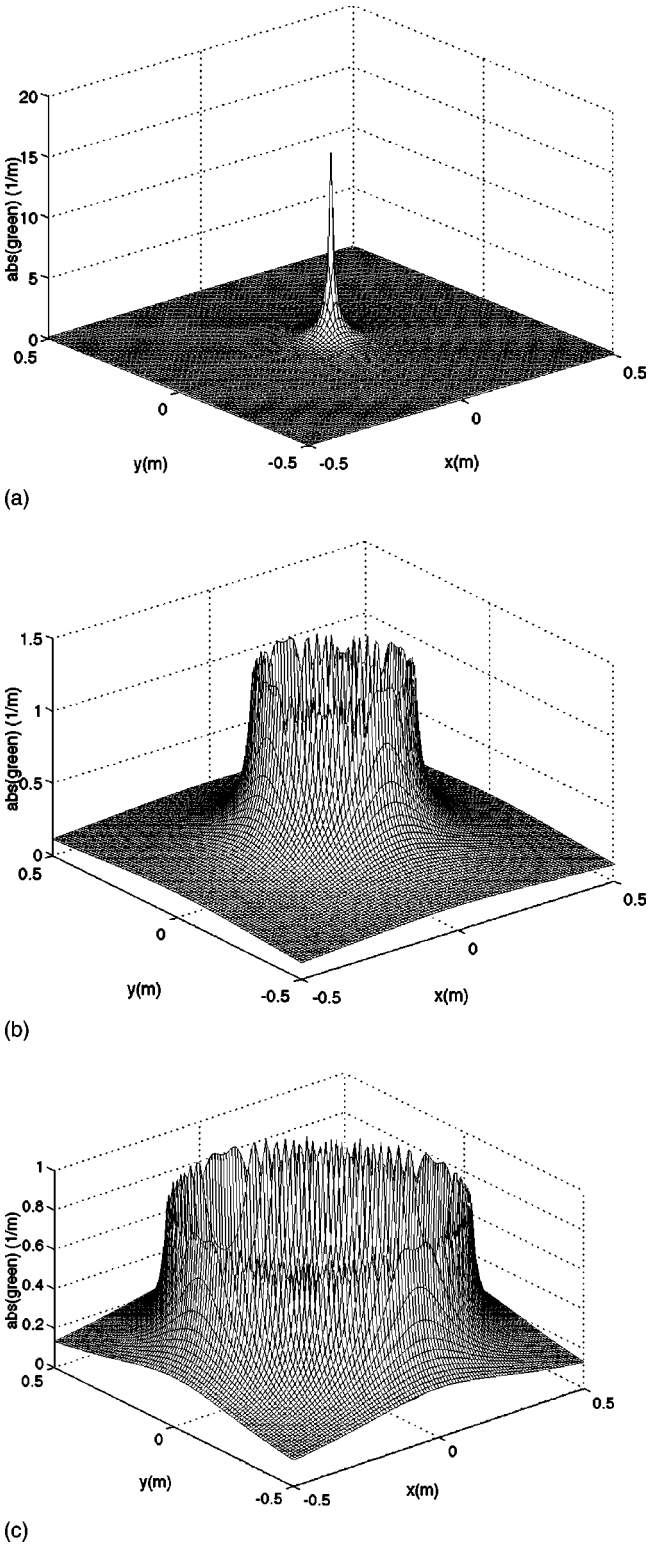


FIG. 3. Complex modulus $|\hat{g}|=\mathrm{abs(green)}$ of the Green's function (38) for decreasing values of the imaginary z location ς over the xy plane; $z+h=0.005$ m and $f=420$ Hz; (a) $\varsigma=0$ m; (b) $\varsigma=-0.2$ m, (c) $\varsigma=-0.4$ m.

(4) All terms in the solution (42) have a clear physical meaning and can be interpreted as contributions from acoustical sources.
(5) The approach can be generalized in a straightforward manner for treating the reflection of multipole sources of arbitrary order from infinite impedance planes. These results will be published in the near future.

Work is in progress to implement the half-space Green's function numerically and to incorporate it into a three-dimensional program code. Especially, the speed of convergence for different numerical set ups will be tested.

ACKNOWLEDGMENTS

The author is very grateful to Professor Dr. F. P. Mechel (Grafenau, Germany) for drawing his attention to the interesting problem how to calculate the reflection of spherical waves from infinite planes. The author would like to thank Haike Brick for numerous stimulating discussions concerning particularly numerical aspects of evaluating the half-space Green's function. Also, I am grateful to Berndt Zeitler from Technical University of Berlin for looking over the manuscript.

[1] A. Sommerfeld, ''Über die Ausbreitung der Wellen in der drahtlosen Telegraphie,'' (''The propagation of waves in wireless telegraphy''), Ann. Phys. (Leipzig) **28**, 665–736 (1909).
[2] A. Sommerfeld, ''Über die Ausbreitung der Wellen in der drahtlosen Telegraphie'' (''The propagation of waves in wireless telegraphy''), Ann. Phys. (Leipzig) **81**, 1135–1153 (1926).
[3] A. Sommerfeld, *Partielle Differentialgleichungen der Physik (Partial Differential Equations in Physics)* (Harri Deutsch, Thun-Frankfurt/M, 1978).
[4] K. Attenborough, ''Sound propagation close to the ground,'' Annu. Rev. Fluid Mech. **34**, 51–82 (2002).
[5] F. P. Mechel, *Schallabsorber Band. I*, Sound Absorbers Vol. I (Hirzel, Stuttgart, 1989), Chap. 13, pp. 513–666.
[6] T. F. W. Embleton, ''Tutorial on sound propagation outdoors,'' J. Acoust. Soc. Am. **100**, 31–48 (1996).
[7] *Boundary Element Acoustics: Fundamentals and Computer Codes*, Advances in Boundary Element Series, edited by T. W. Wu (WIT, Computational Mechanics, Southampton, Boston, 2000).
[8] *Boundary Elements in Acoustics, Advances and Applications*, Advances in Boundary Element Series, edited by O. von Estorff (WIT, Computational Mechanics, Southampton, Boston, 2000).
[9] A. F. Seybert and B. Soenarko, ''Radiation and scattering of acoustic waves from bodies of arbitrary shape in a three-dimensional half space,'' ASME J. Vib., Acoust., Stress, Reliab. Des. **110**, 112–117 (1988).
[10] W. L. Li, T. W. Wu, and A. F. Seybert, ''A half-space boundary element method for acoustic problems with a reflecting plane of arbitrary impedance,'' J. Sound Vib. **171**, 173–184 (1994).
[11] L. A. de Lacerda, L. C. Wrobel, H. Power, and W. J. Mansur, ''A novel boundary integral formulation for three-dimensional analysis of thin acoustics barriers over an impedance plane,'' J. Acoust. Soc. Am. **104**, 671–678 (1998).
[12] C. Granat, M. Ben Tahar, and T. Ha-Duong, ''Variational formulation using integral equations to solve sound scattering above an absorbing plane,'' J. Acoust. Soc. Am. **105**, 2557–2564 (1999).
[13] M. Ochmann, ''The source simulation technique for acoustic radiation problems,'' Acustica **81**, 512–527 (1995).
[14] M. Ochmann, ''The full-field equations for acoustic radiation and scattering,'' J. Acoust. Soc. Am. **105**, 2574–2584 (1999).
[15] Y. L. Li and M. J. White, ''Near-field computation for sound propagation above ground—using complex image theory,'' J. Acoust. Soc. Am. **99**, 755–760 (1996).
[16] J. A. Fawcett, ''Complex-image approximations to the half-space acousto-elastic Green's function,'' J. Acoust. Soc. Am. **108**, 2791–2795 (2000).
[17] J. A. Fawcett, ''A method of images for a penetrable acoustic waveguide,'' J. Acoust. Soc. Am. **113**, 194–204 (2003).
[18] R. Raspet and G. E. Baird, ''The acoustic surface wave above a complex impedance ground,'' J. Acoust. Soc. Am. **85**, 638–640 (1989).
[19] B. Davies, *Integral Transforms and Their Applications*, 3rd ed. (Springer, New York Berlin Heidelberg, 2002).
[20] D. Habault and P. J. T. Filippi, ''Ground effect analysis: Surface wave and layer potential representations,'' J. Sound Vib. **79**, 529–550 (1981).
[21] H. Brick, M. Ochmann, Eine BEM-Toolbox zur Berechnung der Schallabstrahlung schwingender Strukturen vor Hindernissen und über absorbierendem Boden (''A BEM-Toolbox for calculation of sound radiation from vibrating structures in front of barriers and above absorbing ground''), on CD-ROM: Aachen DAGA 2003, Fortschritte der Akustik (ISBN 3-9808659-0-8), available from Deutsche Gesellschaft für Akustik, Fachbereich Physik, Universitaet Oldenburg, 26111 Oldenburg, Germany), 542–543, 2003.
[22] C. F. Gauss, *Disquisitiones Arithmeticae* (Lispiae in commissis apud Gerh. Fleischer Iun. 1801; German translation by H. Maser (Chelsea, New York, 1965), Art. 55.
[23] W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*, 3rd ed. (Springer, Berlin Heidelberg New York, 1966), Sec. 3.6.2, p. 82.
[24] G. A. Deschamps, ''Gaussian beams as a bundle of complex rays,'' Electron. Lett. **7**, 684–685 (1971).
[25] L. B. Felsen, ''Complex-source-point-solutions of the field equations and their relation to the propagation and scattering of Gaussian beams,'' in Instituto Nazionale di Alta Matematica, Symposia Matematica (Academic, London, 1976), **18**, pp. 40–56 (1976).
[26] L. B. Felsen, ''Evanescent waves,'' J. Opt. Soc. Am. **66**, 751–760 (1976).
[27] E. Heyman and L. B. Felsen, ''Complex-source pulsed-beam fields,'' J. Opt. Soc. Am. A **6**, 806–817 (1989).
[28] R. W. Ziolkowski, ''Exact solutions of the wave equation with complex source locations,'' J. Math. Phys. **26**, 861–863 (1985).
[29] M. A. Nobile and S. I. Hayek, ''Acoustic propagation over an impedance plane,'' J. Acoust. Soc. Am. **78**, 1325–1336 (1985).
[30] I. V. Lindell and E. Alanen, ''Exact image theory for the Sommerfeld half-space problem,'' IEEE Trans. Antennas Propag. **32**, Part I: Vertical magnetic dipole, **32**, 126–133 (1984); Part II: Vertical electric dipole, **32**, 841–847 (1984); Part III: General formulation, **32**, 1027–1032 (1984).

Modifications of acoustic modes and coupling due to a leaning wall in a rectangular cavity

Y. Y. Li and L. Cheng[a)]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China

(Received 1 January 2004; revised 24 September 2004; accepted 30 September 2004)

Acoustic modes and the coupling characteristics of a rectangular-like cavity with a slight geometrical distortion introduced through a leaning wall are investigated in this paper. A pressure variation index is proposed to quantify the global changes in acoustic modes caused by the inclination of the wall. Effects on the coupling between acoustic modes and structural modes are investigated using coupling coefficients. Numerical results show a simple relationship between the distortion effect and the acoustic wavelength. The effect is most significant when the distortion approaches the half wavelength. Compared with a rectangular enclosure, the existence of the leaning wall gives rise to a much more effective coupling between the structure and the enclosure. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1823331]

PACS numbers: 43.20.Ks, 43.20.Tb [MO] Pages: 3312–3318

I. INTRODUCTION

The study of sound radiation by a vibrating structure into an enclosure has received a great deal of attention for years. A comprehensive modal-based theoretical framework for interior sound field simulation was developed in the early work of Dowell *et al.*[1] and Fahy.[2] Since then, a large amount of effort has been devoted to investigating the vibro-acoustic behavior of such systems,[3,4] Given an excitation on the structure, structural vibration radiates sound into the enclosure through its coupling with acoustic modes. Therefore an accurate characterization of the sound-structure interaction plays a key role in the prediction of acoustic field. The interplay between the structure and the enclosure is usually characterized by the structural-acoustic modal coupling coefficient, which is a measure of the spatial match between structure modes and cavity modes. The coupling analysis can be easily done for cavities with simple geometry,[5,6] due to the existence of analytical modal solutions.[7] This exercise turns out to be very useful in many aspects,[8,9] especially in revealing useful physical insights to lead subsequent sound control strategies.[10–13]

Literature review shows that most of previous work dealt with rectangular or cylindrical enclosure. Although the use of such regular-shaped cavities with perfect geometry greatly simplifies the modeling procedure, one of the direct consequences of such assumption is that, due to the symmetry in both the structure and the enclosure, the structural-acoustic coupling occurs in a very selective way, which physically means that a structure mode can only be coupled to a small portion of acoustic modes to warrant an effective sound radiation.[10,12] In practice, however, slight imperfection in geometry always exists, which may affect the acoustic mode shapes and, consequently, bring drastic changes in the coupling nature. As a first step, it is necessary to comprehend how acoustic modes are altered. Literature survey shows that few works on the irregular-shaped cavity have been reported. One of the plausible reasons is that, classical modal-based methods rely on the availability of acoustic modes, which cannot be analytically known in the presence of geometry irregularity. For a long time, numerical methods, such as the finite element method[14] and the boundary element method,[15] have been adopted to deal with the problem. The development of acoustoelastic method[16] and the Green function method[17] made it possible to handle the irregular shaped cavities in a semi-analytical way. Both methods were then improved by the authors, who proposed the ''combined integro-modal (CIM) approach,'' in which the cavity was discretized into a series of subcavities, and the acoustic pressure was decomposed either over a modal basis of regular subcavities or over that of the bounding cavities in the case of irregular-shaped boundaries.[18,19] Comparisons between theoretical or other existing results and the presented numerical solutions showed excellent agreement.[20]

The purpose of this paper is to investigate possible changes in acoustic modes and coupling characteristics due to the introduction of a leaning wall in a rectangular-like cavity. The whole system is modeled using the CIM approach. A pressure variation index is defined to quantify the global change in acoustic modes caused by the wall inclination. The tendency plots of index reveal the relationship between the distortion effect and the wavelength of acoustic modes involved. The impact on the structural-acoustic coupling is also examined.

II. FORMULATION

As shown in Fig. 1, the cavity under investigation is a rectangular-like cavity with one leaning wall (a trapezoidal enclosure). The enclosure has a volume V_0 (cavity with solid lines) surrounded by a surface S_0 which is acoustically rigid. A small angle α, which defines the degree of the inclination of the leaning wall S_1, is used to represent the geometry

a)Author to whom correspondence should be addressed; electronic address: mmlcheng@polyu.edu.hk

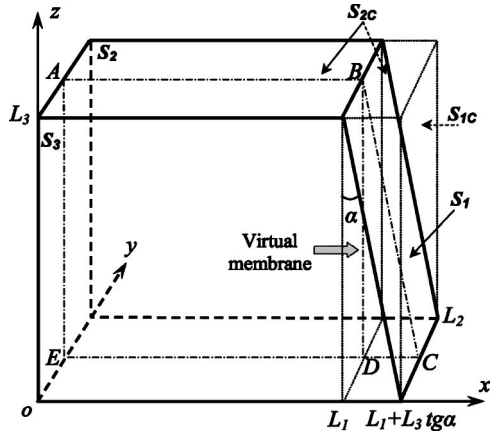


FIG. 1. Geometry and coordinate system of a rectangularlike cavity with one leaning wall.

distortion of the cavity, as opposed to its rectangular counterpart. The acoustic pressure Ψ inside the enclosure can be expressed in the form of wave equation

$$(\nabla^2+k^2)\Psi=0 \tag{1}$$

with the Neumann boundary condition

$$(\partial\Psi/\partial\mathbf{n})_{S_0}=0, \tag{2}$$

where k is the wave number and $\mathbf{n}$ the normal direction towards outside. In parallel, a rectangular bounding cavity, which encloses the trapezoidal enclosure V_0 and occupies a volume V_c $((L_1+L_3\times tg\alpha)\times L_2\times L_3)$ with a surface S_c, is constructed (dashed in Fig. 1). Inside the bounding cavity, the Green's function $G(r,r_0)$ satisfies the following inhomogeneous Helmholtz equation with a point source:

$$(\nabla^2+K^2)G(r,r_0)=-\delta(r-r_0), \tag{3a}$$

$$(\partial G(r,r_0)/\partial\mathbf{n})_{S_C}=0, \tag{3b}$$

where $\delta(r-r_0)$ is the Dirac delta function. $G(r,r_0)$ can be expressed in terms of normal modes φ_p of the bounding cavity V_c as

$$G(r,r_0)=\sum_{l,m,n}\frac{\varphi_{lmn}(r)\varphi_{lmn}(r_0)}{(k_{lmn}^2-k^2)V_R\wedge_{lmn}}, \tag{4}$$

where

$$\varphi_{lmn}=\cos\left(\frac{l\pi x}{L_1+L_3\cdot tg\alpha}\right)\cos\left(\frac{m\pi y}{L_2}\right)\cos\left(\frac{n\pi z}{L_3}\right), \tag{5a}$$

with k_{lmn} and $\wedge_{lmn}$ being, respectively, the wave number and the generalized acoustic mass of the lmn-th mode, viz.

$$k_{lmn}^2=\left(\frac{l\pi}{L_1+L_3\cdot tg\alpha}\right)^2+\left(\frac{m\pi}{L_2}\right)^2+\left(\frac{n\pi}{L_3}\right)^2, \tag{5b}$$

$$\wedge_{lmn}=\frac{1}{V_c}\int_{V_c}\varphi_{lmn}(r)\varphi_{lmn}(r)dv. \tag{5c}$$

According to the CIM approach,[18,20] the acoustic pressure inside V_0 is decomposed on the basis of normal modes φ_{lmn} of the bounding cavity as

$$\Psi=\sum_{l,m,n}b_{lmn}\varphi_{lmn},\quad l=1,...,L;\quad m=1,...,M;\quad n=1,...N, \tag{6}$$

where b_{lmn} are the unknown coefficients to be determined; (L,M,N) the numbers of the terms to be kept after the truncation of the series. Combining Eqs. (1)–(4) with (6) and using the orthogonality property of mode shapes lead to the following eigenvalue equation:

$$(k^2-k_{lmn}^2)b_{lmn}=\sum_{i,j,k}b_{ijk}n_{lmn,ijk}(\alpha), \tag{7}$$

where

$$n_{lmn,ijk}(\alpha)=\int\int_{S_0}\varphi_{ijk}\frac{\partial\varphi_{lmn}}{\partial\mathbf{n}}ds. \tag{8}$$

For the cavity V_0, $\partial\varphi_{lmn}/\partial\mathbf{n}=0$ holds at all walls except on the leaning wall S_1. In light of the relationship between variables x and z:

$$x=L_1+(L_3-z)\cdot tg\alpha, \tag{9}$$

one has

$$\frac{\partial\varphi_{lmn}}{\partial\mathbf{n}}=-\cos\frac{m\pi y}{L_2}\left(\frac{l\pi\cos\alpha}{L_1+L_3tg\alpha}\times\sin\frac{l\pi x}{L_1+L_3tg\alpha}\cos\frac{n\pi z}{L_3}+\frac{n\pi\sin\alpha}{L_3}\cos\frac{l\pi x}{L_1+L_3tg\alpha}\sin\frac{n\pi z}{L_3}\right). \tag{10}$$

Substituting Eq. (10) into (8) and then integrating over S_1, $n_{lmn,ijk}(\alpha)$ can be expressed as

$$n_{lmn,ijk}(\alpha)=\frac{[1-|\mathrm{sign}(m-j)|]\pi L_2}{16}\cdot\left\{\frac{l\cos\alpha}{L_1+L_3tg\alpha}[c_1(a_1-a_2+a_3-a_4)+|\mathrm{sign}(l-i)|\cdot c_2(a_5-a_6+a_7-a_8)]+\frac{n\sin\alpha}{L_3}[c_1(a_1+a_2+a_3+a_4)+c_2(a_5+a_6+a_7+a_8)]\right\}, \tag{11}$$

where $(c_1, c_2, a_1, ..., a_8)$ are a set of coefficients which can be calculated for a given α. sign is a symbol function defined as

$$\mathrm{sign}(x)=\begin{cases}1, & x>0\\ 0, & x=0\\ -1, & x<0\end{cases}.$$

It is clear that $n_{lmn,ijk}(\alpha)$ vanishes when $\alpha=0$, showing the orthogonality property of the modes when there is no distortion. Rearranging Eq. (7) in the matrix form gives

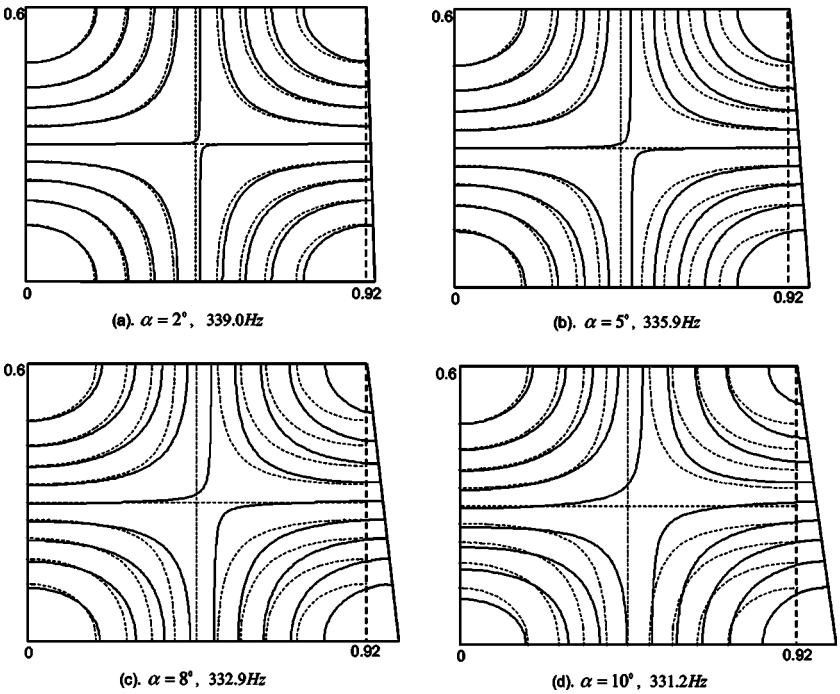


FIG. 2. Contour of the iso-pressure of the acoustic mode (1,0,1) over the cross area $ABCE$.

$$\begin{bmatrix} n_{000,000}(\alpha)-(k^2-k_{000}^2) & \cdots & n_{000,LMN}(\alpha) \\ \cdots & \cdots & \cdots \\ n_{LMN,000}(\alpha) & \cdots & n_{LMN,LMN}(\alpha)-(k^2-k_{LMN}^2) \end{bmatrix} \begin{Bmatrix} b_{000} \\ \vdots \\ b_{LMN} \end{Bmatrix} = \mathbf{0}. \tag{12}$$

The solution of Eq. (12) yields b_{lmn} for constructing acoustic pressure (or acoustic mode shape) inside the cavity from Eq. (6).

In order to quantify changes in acoustic modes, a variation index is defined as

$$J_{lmn}(\alpha) = \int_S \Delta\Psi_{lmn}^2(\alpha)ds, \tag{13}$$

where $\Delta\Psi_{lmn}(\alpha)$ is the residual mode shape over a given surface S:

$$\Delta\Psi_{lmn}(\alpha) = \Psi_{lmn}(\alpha) - \Psi_{lmn}(\alpha=0). \tag{14a,b}$$

Obviously, $J_{lmn}(\alpha)$ represents the global change in the lmn-th mode over the surface caused by the wall inclination.

III. NUMERICAL SIMULATION AND DISCUSSIONS

The formulation described in Sec. II is implemented. The dimension of the cavity is set as $L_1 \times L_2 \times L_3 = 0.92 \times 0.15 \times 0.6$ m^3. Since the inclination is introduced parallel to y-axis, discussions will focus on those modes with order (l,m,n) with $m=0$. In the following discussions, the term "(l,m,n) mode" will be used to designate a pair of acoustic modes before and after α is introduced. It should be mentioned that although this notation has clear physical meaning when $\alpha=0$, it is loosely used for the cavity with a leaning wall for the sake of convenience. In the latter case, it simply stands for a mode evolving from the (l,m,n) mode ($\alpha=0$) due to the wall inclination. The matching of the pair is ensured by carefully checking the mode shape (pressure distribution) of each mode during the calculation.

The truncation of the decomposition series [Eq. (6)] is a main factor affecting the accuracy of the calculation. A careful convergence analysis was carried out by following the procedure detailed in our previous work.[18] Roughly speaking, (L,M,N) was gradually increased until no noticeable changes in the calculated results were observed. For the present configuration, the series is truncated up to 60, 3, and 10 in L, M, and N, respectively.

A. Analysis of acoustic modes

Changes in acoustic pressure distribution due to the variation of α are first investigated by choosing one typical mode (1,0,1). Figure 2 shows the contour plot of the acoustic pressure over the cross area $ABCE$ ($y=L_2/2$ in Fig. 1) with different inclination angles. The dashed are the iso-pressure lines when $\alpha=0°$ taken as the nominal case. The pressure amplitude is normalized to the maximal pressure value in the cavity. It can be seen that the distortion has no significant influence on natural frequencies, due to the large wavelength of the mode with respect to the distortion. With the increase of α from 2° to 10°, the mode shape deviates gradually from its nominal case. This change in acoustic pressure is quantified by the residual mode shape $\Delta\Psi_{101}(\alpha)$ in Fig. 3. It can be seen that:

- For the case of $\alpha=2°$, change in the acoustic mode is observed, and the maximal difference is 4% compared to

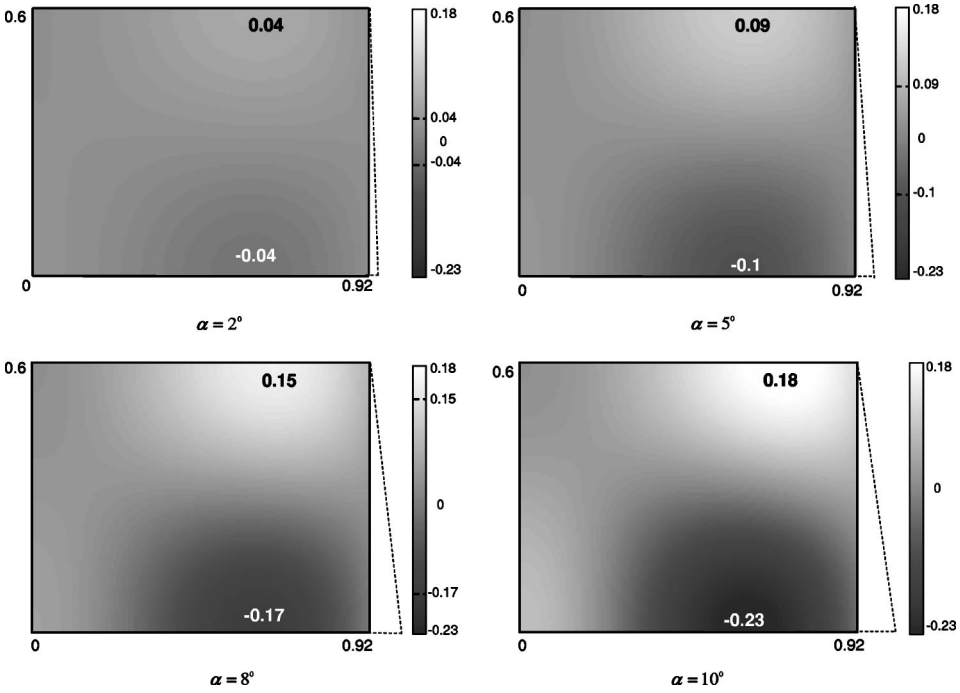


FIG. 3. Residual acoustic pressure $\Delta\Psi_{101}(\alpha)$.

the nominal one. With the increase of α, this change becomes more significant, reaching 23% for $\alpha=10°$.

- The perfect symmetry of mode shape at edges AB and DE in the case of $\alpha=0°$ is altered due to the distortion.
- There is a maximum pressure change area at edge AB for each configuration, ranging from 0.04 to 0.18. Its location gradually moves towards the leaning wall of the enclosure when α increases. For the same token, an even larger variation area (from 0.04 to 0.23) can be observed at the bottom part of the cavity (edge DE), with the same moving tendency as shown at edge AB. In comparison, there is little change on the surface opposite to the leaning wall. Therefore, the effect of α on acoustic pressure variation is mainly on the two walls adjacent to the inclined wall rather than on the opposite one, with the maximal variation appearing at the area close to the leaning wall.

The global variation of a number of selected acoustic modes is quantified using the variation index $J_{lmn}(\alpha)$. Figure 4 illustrates the tendency curve of $J_{l00}(\alpha)$ for a number of selected modes with α varying from 0° to 30°. It can be seen that

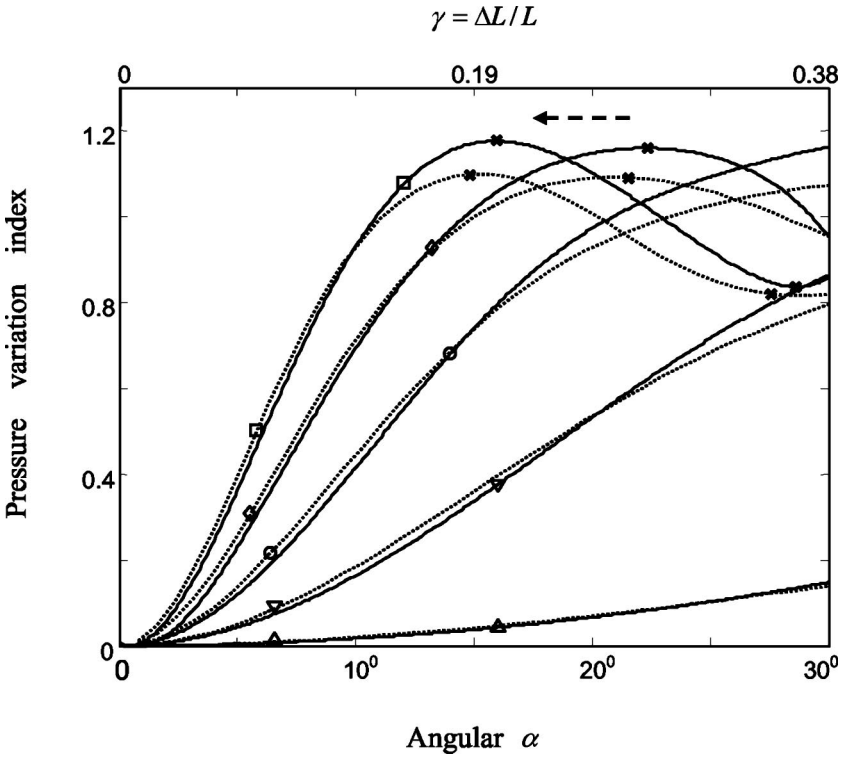


FIG. 4. Pressure variation index. Solid line: acoustic modes (l, 0,0) with $0° \leqslant \alpha \leqslant 30°$ at S_2; dashed: 1D duct modes with γ varying from 0 to 0.38. △: $l=1$; ▽: $l=3$; ○: $l=5$; ◇: $l=7$; □: $l=9$.

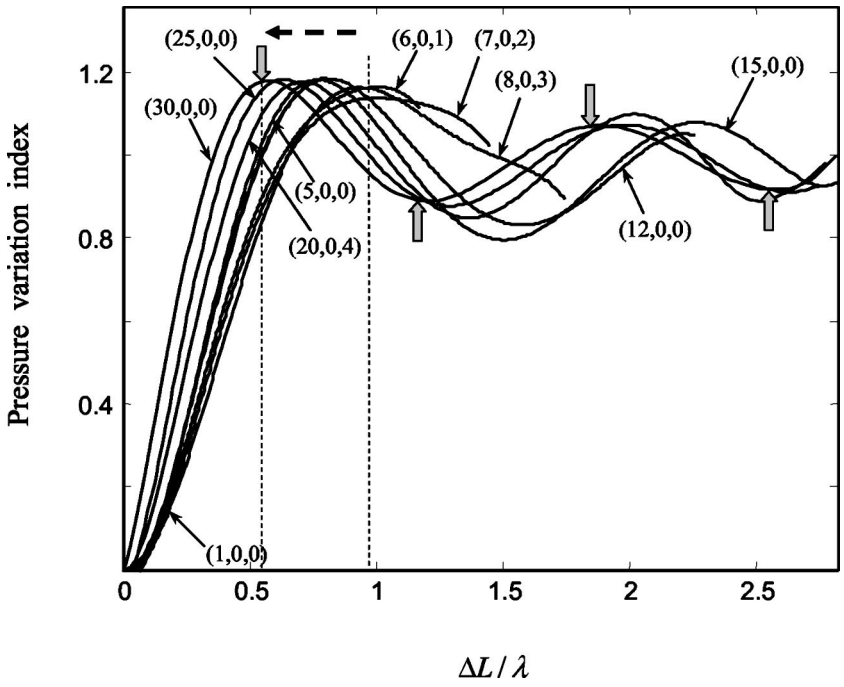


FIG. 5. Pressure variation index $J_{lmn}(\alpha)$ at S_2 vs $\Delta L/\lambda$ with $0° \leqslant \alpha \leqslant 30°$ ($\Delta L = L_3 \cdot tg\,\alpha, L_3 = 0.6$).

- For a small α, the variation index $J_{l00}(\alpha)$ increases with the increase of order l. That is, a certain distortion will have a more sensitive effect on pressure variation for high-order modes than for low-order ones.
- Within the range of interest, no extremum of $J_{l00}(\alpha)$ is observed for low-order modes. However, peaks appear with the increase of l. For example, $J_{l00}(\alpha)$ reaches its maximum at $\alpha \approx 22°$ for mode (7,0,0) and at $\alpha \approx 16°$ for mode (9,0,0). With further increase of l, multiple peaks emerge.

The earlier results show that there are critical values of α for each mode in which the variation is the most remarkable. There might be a relationship between these critical values and the wavelength of modes in question. Quantifying this possible relationship can help estimate the impact of distortion on one particular mode of interest.

Since α mainly affects the acoustic modes in x-direction, an auxiliary study was performed using a one-dimensional (1D) duct with a length L to understand the observed phenomena. Some details are given in the Appendix. Using the duct theory,[21] it can be found that maximum alteration in the acoustic pressure occurs at those locations ΔL satisfying

$$\Delta L|_{i=l-1,l-2,\ldots} = \frac{\lambda l}{2}\left(\frac{l}{i}-1\right) = \left(\frac{l}{i}-1\right)L, \quad \lambda = \frac{2L}{l}, \quad (15)$$

and the *first* one is located at

$$\Delta L|_{i=l-1} = \frac{\lambda}{2}\cdot\frac{l}{l-1}. \quad (16)$$

Clearly, the number and locations of extrema depend on the wavelength λ and mode order l. For the *first* one, an increase in l makes $\Delta L/\lambda$ approaching 0.5, implying that for high-order modes, the change in acoustic modes is the most remarkable when $\Delta L \to \lambda/2$. Since higher-order modes have shorter wavelength, a slight distortion can therefore lead to a significant variation of mode shapes. Results are also plotted in Fig. 4 and compared with previous ones for the cavity, showing a strong similarity.

This observation is further verified in Fig. 5 using the cavity defined before, which illustrates the variation of $J_{lmn}(\alpha)$ ($0° \leqslant \alpha \leqslant 30°$) with respect to $\Delta L/\lambda$ ($\Delta L = L_3 \cdot tg\,\alpha$). A large variety of modes are included in the figure to check the observation made earlier. It should be mentioned that, due to the difference in wavelength of different modes, the variation ranges of $\Delta L/\lambda$ are not the same for all modes considered. The highest mode involved is 30 in x-direction of the cavity, for which *four* extremum are detected at $\Delta L/\lambda = 0.55$, 1.15, 1.8, and 2.54, respectively (marked with arrows in the figure). This result is in consistent with the prediction given by Eq. (15), which are $\Delta L/\lambda|_{i=29,\ldots,26} = 0.52$, 1.07, 1.67, and 2.3. In addition, with the increase of mode order l, the *first* extremum $\Delta L/\lambda|_{i=l-1}$ indeed convergences to 0.5.

The effect of the inclination on different walls is also examined. Two walls, i.e., S_2 which is adjacent to the leaning wall and S_3 which is opposite to it are taken as example, and the pressure variation indices for a number of selected modes are compared in Fig. 6. Apparently, the distortion has much greater effect on S_2 than S_3.

B. Vibroacoustic coupling analysis

Replacing the top wall of the enclosure by a simply-supported flexible panel, a coupling analysis is conducted to quantify the effects of the wall inclination on the structural-acoustic coupling feature. The commonly used coupling coefficient $L(\alpha)$ is defined as the integral of the product between ij-th structure mode ϕ_{ij} and the lmn-th cavity mode $\Psi_{lmn}(\alpha)$ over the vibrating surface A_f:

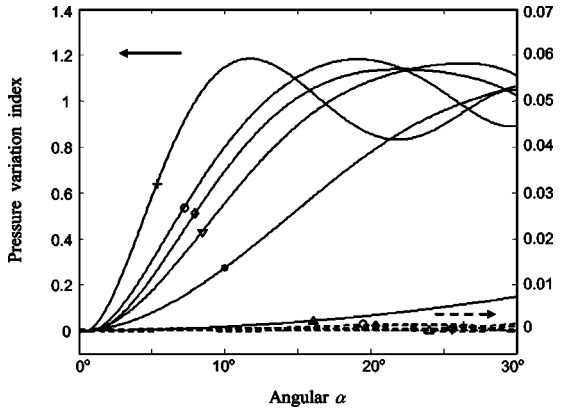


FIG. 6. Pressure variation index $J_{lmn}(\alpha)$ with $0° \leqslant \alpha \leqslant 30°$ at S_2 (solid line) and S_3 (dashed). △:(1,0,0); *: (4,1,2); ▽: (6,0,1); ◇: (7,0,2); ○: (8,0,2); +: (12,0,0).

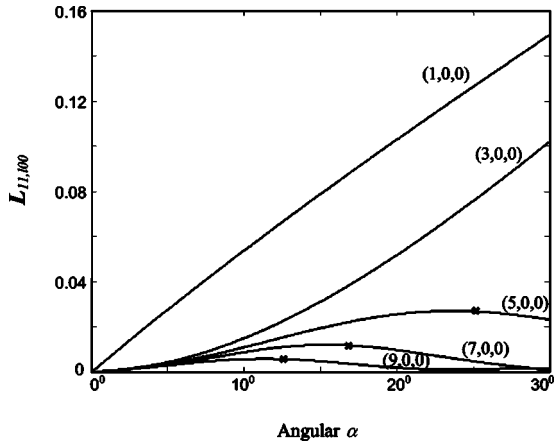


FIG. 8. Effect of α on the coupling coefficient $L_{11,l00}$ ($l=1,3,...,9$).

$$L_{ij,lmn}(\alpha)=\frac{1}{A_f}\int \phi_{ij}\Psi_{lmn}(\alpha)ds. \tag{17}$$

Figure 7 compares the magnitude of the coupling coefficients when $\alpha=0$ and $\alpha=10°$ using sixteen acoustic modes ($l=0,...,3; m,n=0,1$) and nine structural modes ($i,j=1,..,3$). It can be seen that, when $\alpha=0$, only a few acoustic modes are coupled to each structural mode (denoted by a star in Fig. 7). In fact, any symmetric/anti-symmetric modes of the panel, with respect to the center, would not be coupled to an acoustic mode if the latter is anti-symmetrical/symmetrical in one of the two directions parallel to the panel surface. The distortion of the wall ($\alpha=10°$) greatly increases the number of the coupled modes, denoted by circles in Fig. 7. Comparing the two cases, the coupling strength between the originally coupled modes are not significantly altered, as judged by the closeness of the star-circle pairs in Fig. 7. The additional coupling caused by the inclination of the wall, however, can reach a relatively high level in some cases. For example, $L_{31,300}(0)=0$ between the (3,1) mode and the (3,0,0) mode while $L_{31,300}(10°)=0.15$ (marked with an arrow in the figure), attaining 37% of the maximum values of $L_{ij,lmn}(\alpha)$ ($L_{11,100}(\alpha)=0.4053$). Apparently, a strong coupling is created due to the distortion of the enclosure.

Taking the structural mode (1,1) as an example, variations of the coupling strength with respect to α are examined in Fig. 8. The *five* acoustic modes are the same as the ones previously used in Fig. 4. For each particular acoustic mode, the general tendency of the curve is somehow similar to its counterpart in Fig. 4, which implies that a maximum alteration of the pressure would most likely also lead to a significant change in its coupling strength to the lower-order structural modes. This agreement becomes, however, less consistent, for higher-order structural modes (not shown). Therefore, the criterion previously established to quantify the pressure variation can be roughly used to predict the coupling strength between an acoustic mode with a lower-order structural mode.

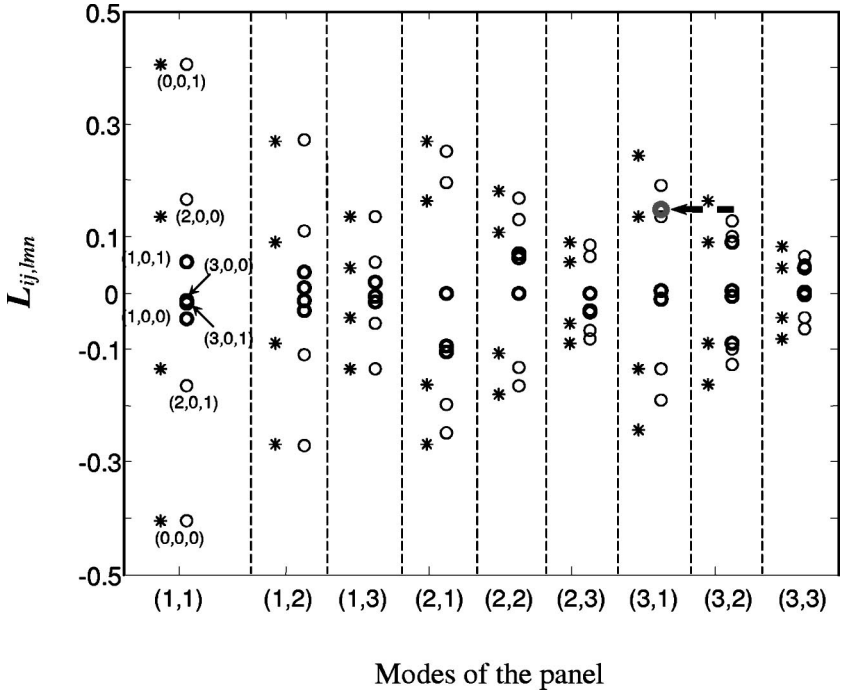


FIG. 7. Coupling coefficient $L_{ij,lmn}$ between acoustic modes and structural modes of a simply-supported panel. *: $\alpha=0$ and ○: $\alpha=10°$ ($i,j=1,...,3$; $l=0,...,3$; $m,n=0,1$).

IV. CONCLUSIONS

Acoustic modes and the coupling characteristics of a rectangular-like cavity with a slight geometrical distortion introduced through a leaning wall are investigated in this paper. A pressure variation index is proposed to quantify the global change in acoustic modes. The coupling coefficient is used to measure the effect caused by the distortion. Numerical simulations are performed to find out the relationships between the variation index, the coupling coefficient and the distortion. The following conclusions can be drawn.

(1) The pressure distribution inside the cavity is sensitive to geometrical changes. The most affected areas move to the leaning wall when the distortion is getting larger.

(2) A simple relationship between the distortion effect and the acoustic wavelength involved is established. Whether a given distortion is important depends on the wavelength of the acoustic modes. For a lower-order mode, a small distortion has no apparent influence due to its large acoustic wavelength. For high-order modes, however, the effect is apparent and intensifies when the distortion approaches the half wavelength.

(3) The distortion affects more the two adjacent walls than the opposite one.

(4) The inclination of the wall gives rise to a much more effective coupling between the structure and the enclosure. The most affected modes are these pairs which are not initially coupled when $\alpha=0$. The coupling strength between an acoustic mode and a lower-order structural mode roughly follows the tenancy stated in (2).

ACKNOWLEDGMENTS

This project was supported by the Hong Kong Polytechnic University Research Grant No. (G-U031) and the Research Grants Committee of HKSAR (Grant No. PolyU 5165/02E).

APPENDIX: RELATIONSHIP BETWEEN THE DISTORTION AND THE WAVELENGTH OF A ONE-DIMENSIONAL DUCT

The l-th mode shape of a duct with a length L satisfies $\Psi_l=\cos(l\pi x/L)$, and the corresponding residual mode shape when the length varies from L to $L(1+\gamma)$ can be written as

$$\Delta\Psi_l(\gamma)=\cos\left(\frac{l\pi x}{L(1+\gamma)}\right)-\cos\left(\frac{l\pi x}{L}\right),\quad \gamma=\frac{\Delta L}{L}. \tag{A1}$$

Using the pressure variation index defined in Eq. (13):

$$\begin{aligned} J_l(\varepsilon)&=\int_0^L \Delta\Psi_l^2(\varepsilon)dx \\ &=L\left(1+\frac{\sin(2\pi l\varepsilon)}{4\pi l\varepsilon}+\frac{1}{l\pi}\left(\frac{(-1)^{l+1}}{1+\varepsilon}+\frac{(-1)^l}{1-\varepsilon}\right)\sin(l\pi\varepsilon)\right), \\ \varepsilon&=\frac{1}{1+\gamma}. \end{aligned} \tag{A2}$$

$J_l(\varepsilon)$ reaches its extrema when $\partial J_l(\varepsilon)/\partial\varepsilon=0$, i.e.:

$$\begin{aligned} &\frac{1}{4\varepsilon^2}\sqrt{1+(2\pi l\varepsilon)^2}\sin(\operatorname{arctg}(2\pi l\varepsilon)-2\pi l\varepsilon) \\ &\quad+\frac{2}{1-\varepsilon^2}\sqrt{\left(\frac{1+\varepsilon^2}{1-\varepsilon^2}\right)^2+(\pi l\varepsilon)^2}\sin\left(l\pi\varepsilon+\operatorname{arctg}\frac{l\pi\varepsilon(1-\varepsilon^2)}{1+\varepsilon^2}\right)=0. \end{aligned} \tag{A3}$$

A simplification of Eq. (A3) results in an approximate solution of ε, viz., $\varepsilon=i/l$ $(i=1,...,n)$. Thus, one has

$$\Delta L|_{i=l-1,l-2,...}=\frac{\left(\frac{l}{i}-1\right)\lambda l}{2}=\left(\frac{l}{i}-1\right)L. \tag{A4}$$

where λ is the wavelength of the l-th mode and $\lambda=2L/l$.

[1] E. H. Dowell, G. F. Gorman, and D. A. Smith, "Acoustoelasticity: General theory, acoustic natural modes and forced response to sinusoidal excitation, including comparisons with experiment," J. Sound Vib. **52**, 519–541 (1977).
[2] F. J. Fahy, *Sound and Structural Vibration* (Academic, England, 1987).
[3] J. Pan, "Analysis of low frequency acoustic response in a damped rectangular enclosure," J. Sound Vib. **223**, 543–566 (1999).
[4] J. H. Wu, H. L. Chen, and W. B. An, "A method to predict sound radiation from a plate-ended cylindrical shell excited by an external force," J. Sound Vib. **237**, 793–803 (2000).
[5] L. P Franzoni and D. B. Bliss, "A discussion of modal uncoupling and an approximate closed-form solution for wealy-coupled systems with application to acoustics," J. Acoust. Soc. Am. **103**, 1923–1932 (1998).
[6] K. S. Sum and J. Pan, "On acoustic and structural modal cross-couplings in plate-cavity systems," J. Acoust. Soc. Am. **107**, 2021–2038 (2000).
[7] P. M. Morse and K. U. Ingard, *Theoretical Acoustics* (McGraw-Hill, New York, 1968).
[8] R. F. Keltie and H. Peng, "The effects of modal coupling on the acoustic power radiation from panels," J. Vibr. Acoust. **109**, 48–54 (1987).
[9] K. A. Cunefare, "Effect of modal interaction on sound radiation from vibrating structures," AIAA J. **30**, 2819–2828 (1992).
[10] L. Cheng and J. Nicolas, "Radiation of sound into a cylindrical enclosure from a point-driven end plate with general boundary conditions," J. Acoust. Soc. Am. **91**, 1504–1513 (1992).
[11] L. F. Peretti and E. H. Dowell, "Asymptotic modal-analysis of a rectangular acoustic cavity excited by wall vibration," AIAA J. **30**, 1191–1198 (1992).
[12] J. Pan, "The forced response of an acoustic-structural coupled system," J. Acoust. Soc. Am. **91**, 949–956 (1992).
[13] M. Kronast and M. Hildebrandt, "Vibro-acoustic modal analysis of automobile body cavity noise," Sound Vib. **34**, 20–23 (2000).
[14] M. Petyt, J. Lea, and G. H. Koopman, "A finite element method for determining the acoustic modes of irregular shaped cavities," J. Sound Vib. **45**, 497–502 (1976).
[15] C. A. Brebbia, J. C. F. Telles, and L. C. Wrobel, *Boundary Element Techniques* (Springer, New York, 1984).
[16] C. F. Chao, E. H. Dowell, and D. B. Bliss, "Modal analysis of interior noise field," Vib. Conf. Design Eng. 1981.
[17] G. P. Succi, "The interior acoustic field of an automobile cabin," J. Acoust. Soc. Am. **81**, 1688–1694 (1987).
[18] J. Missaoui and L. Cheng, "A combined integro-modal approach for predicting acoustic properties of irregular-shaped cavities," J. Acoust. Soc. Am. **101**, 3313–3321 (1997).
[19] J. Missaoui and L. Cheng, "Vibroacoustic analysis of a finite cylindrical shell with internal floor partition," J. Sound Vib. **226**, 101–123 (1999).
[20] E. Anyunzoghe and L. Cheng, "Improved integro-modal approach with pressure distribution assessment and the use of overlapped cavities," Appl. Acoust. **63**, 1233–1255 (2002).
[21] P. M. Morse and M. Foshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953)

Attenuation of ultrasonic waves in rolled metals

Liyong Yang and Joseph A. Turner[a)]
Department of Engineering Mechanics, Center for Materials Research and Analysis, W317.4 Nebraska Hall, University of Nebraska—Lincoln, Lincoln, Nebraska 68588-0526

(Received 12 March 2004; revised 31 August 2004; accepted 1 September 2004)

Scattering of ultrasonic waves in polycrystals with texture is studied in this article. The attenuations of the three wave modes are determined as a function of dimensionless frequency and propagation direction, respectively, for given orientation distribution coefficients (ODCs). The calculation is done in the case of a statistically orthorhombic sample made up of cubic crystallites. The wave propagation and scattering model is formulated by the Dyson equation using an anisotropic Green's function approach. Within the limits of the first-order smoothing approximation, the Dyson equation is solved in the spatial Fourier transform domain. The results presented are shown to be directional dependent, frequency dependent, and especially dependent on the texture coefficients (ODCs) for the quasilongitudinal and two quasishear waves. The theoretical results presented may be used to improve the understanding of the microstructure during recrystallization processes. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1810236]

PACS numbers: 43.20.Bi, 43.20.Gp, 43.35.Cg [YHB] Pages: 3319–3327

I. INTRODUCTION

Metals and alloys are made of crystallite grains whose characteristics and arrangements can be changed by the application of heat processing, such as annealing. Microstructural parameters of metals determine the macroscopic mechanical properties of a material and include the grain size, grain shape, and the orientation of the grains, or texture, and their distribution in the microstructure. Ultrasonic waves propagating in such aggregates lose energy due to scattering from the granular microstructure of these materials. This scattering is often characterized by the attenuation of the medium. In general, the attenuation and wave velocity are dependent on the grain size, shape, and on the particular orientation distributions of the grains. If the grains are randomly oriented such that the medium is statistically isotropic, these propagation properties are independent of direction. However, the scattering attenuation and wave velocity are a function of the propagation direction if the grains have a preferred orientation. The preferred orientation of grains, or texture, is best quantitatively described by the orientation distribution function (ODF) defining a probability density function, which is usually expanded in a series of generalized spherical harmonics (Roe, 1965, 1966; Bunge, 1982). Often, most metallic materials with preferred orientation of grains display anisotropy of material properties. Therefore, knowledge of the anisotropic nature of the wave propagation and scattering in textured materials such as attenuation and velocity is critical for use with ultrasonic nondestructive techniques. Such information will provide valuable insight for modeling the microstructure of such complex materials during processing.

The scattering of elastic waves by grains of polycrystals has received considerable attention. The most recent contributions for cubic symmetry with uniformly distributed orientations of grains were made by Hirsekorn (1982, 1983), Stanke (1984), and Weaver (1990). The problem of wave propagation and scattering in the case of polycrystalline grains with an aligned [001] axis has been examined by Ahmed and Thompson (1996) and Turner (1999). In that particular case, the average medium is statistically transversely isotropic. Ahmed and Thompson (1992, 1996) also studied correlations defined by both equiaxed grains and grains with elongation.

During the recrystallization process of metals, such as annealing, the microstructure may contain grains having preferred crystallographic orientations. For rolling texture, there are three orthogonal axes of symmetry which are defined as the rolling, transverse, and normal directions. Thus, the material properties of this specific case may be assumed orthorhombic due to the feature of the preferred orientation. Hirsekorn (1985) also was one of the first to investigate the wave scattering in polycrystals of cubic symmetry with rolling texture as a function of frequency by using the perturbation approach. She then extended her theory to determine the directional dependence of the phase velocities and attenuations of the three wave types under the same assumption with fiber texture (Hirsekorn, 1986). Her discussions were restricted to waves propagating in the direction of an axis of symmetry of the texture. The general formalism of the waves propagating in any direction through polycrystalline metals with rolling texture, however, has not yet been reported. The detailed wave velocities of the three wave types, inclusive of the quasilongitudinal and two quasishear waves, have been discussed elsewhere (Sayers, 1982; Johnson, 1985; Hirao *et al.*, 1987; Li and Thompson, 1990) under the assumption of orthorhombic-cubic symmetry.

In this article, the more sensitive ultrasonic parameter, scattering attenuation, is studied for waves propagating in any direction through such textured media. The wave propagation and scattering model is formulated using the Dyson

[a)]Author to whom correspondence should be addressed. Electronic mail: jaturner@unl.edu

equation studied by Frisch (1968) and Weaver (1990) which is easily solved in the spatial Fourier transform domain within the limits of the first-order smoothing approximation (FOSA) or Keller (Karal and Keller, 1964) approximation. The problem is studied here using the anisotropic Green's dyadic, an approach not used previously for textured materials. The attenuations of the three wave types are calculated numerically as a function of dimensionless frequency and propagation direction, respectively, for given orientation distribution coefficients (ODCs) using the derived expressions. The resulting attenuations are shown to be directional dependent, frequency dependent, and dependent on the texture coefficients (ODCs) for the quasilongitudinal and two quasishear waves. The theoretical results presented may be used to improve the understanding of the microstructure during the recrystallization process. In addition, the present formulation may be used to study diffuse ultrasonic problems in a straightforward manner. Although the present model is for the case of orthorhombic-cubic symmetry, the formalism can be easily modified to apply to other given symmetry cases.

In the next section, the preliminary elastodynamics of elastic wave propagation and scattering is introduced in terms of an anisotropic Green's dyadic. The formalism of the attenuation is then developed for the anisotropic case of orthorhombic-cubic symmetry.

II. WAVE PROPAGATION AND SCATTERING MODEL

The equation of motion for the elastodynamic response of an infinite, linear-elastic material to deformation is given in terms of the Green's dyadic by

$$\{-\delta_{jk}\rho\partial_t^2+\partial x_i C_{ijkl}(\mathbf{x})\partial x_l\}G_{k\alpha}(\mathbf{x},\mathbf{x}';t) = \delta_{j\alpha}\delta^3(\mathbf{x}-\mathbf{x}')\,\delta(t), \tag{1}$$

where $\delta^3(\mathbf{x}-\mathbf{x}')$ is the three-dimensional spatial Delta function. The second-order Green's dyadic, $G_{k\alpha}(\mathbf{x},\mathbf{x}';t)$, defines the response at location $\mathbf{x}$ in the kth direction to a unit impulse at location $\mathbf{x}'$ in the αth direction. The moduli are considered to vary spatially and density is assumed uniform throughout. In the case of orthorhombic-cubic symmetry, the moduli $\mathbf{C}$ are supposed to be spatially heterogeneous and have the form $C_{ijkl}(\mathbf{x})=C^0_{ijkl}+\delta C_{ijkl}(\mathbf{x})$. The material properties might have global anisotropy such that the mean moduli are not necessarily isotropic. The covariance of the moduli is characterized by an eighth-rank tensor

$$\langle \delta C_{ijkl}(\mathbf{x})\,\delta C_{\alpha\beta\gamma\delta}(\mathbf{x}')\rangle = \Xi^{\alpha\beta\gamma\delta}_{ijkl}\eta(\mathbf{x}-\mathbf{x}'). \tag{2}$$

The spatial and tensorial parts of the above covariance, Ξ and η, are assumed independent. The correlation function η is also assumed a function of the difference between two vectors, $\mathbf{x}-\mathbf{x}'$. This assumption implies that the medium is statistically homogeneous.

The mean response, $\langle\mathbf{G}\rangle$, is governed by the Dyson equation (Weaver, 1990; Frisch, 1968)

$$\langle G_{i\alpha}(\mathbf{x},\mathbf{x}')\rangle = G^0_{i\alpha}(\mathbf{x},\mathbf{x}') + \int\int G^0_{i\beta}(\mathbf{x},\mathbf{y})M_{\beta j}(\mathbf{y},\mathbf{z}) \times\langle G_{j\alpha}(\mathbf{z},\mathbf{x}')\rangle d^3y\,d^3z. \tag{3}$$

In Eq. (3), the quantity $\mathbf{G}^0$ is the bare Green's dyadic defined as the ensemble average response of the medium without fluctuations, namely, the solution to Eq. (1) when $\delta C_{ijkl}(\mathbf{x})=0$. The second order tensor $\mathbf{M}$ is the mass or self-energy operator. The Dyson equation, Eq. (3), is easily solved in the Fourier transform domain under the assumption of statistical homogeneity. The assumption of statistical homogeneity ensures that $\mathbf{G}^0$, $\mathbf{M}$, and $\langle\mathbf{G}\rangle$ are functions of a single wave vector in Fourier space. The Dyson equation is then transformed and solved to give the result for $\langle\mathbf{G}(\mathbf{p})\rangle$ of the form

$$\langle\mathbf{G}(\mathbf{p})\rangle=[\mathbf{G}^0(\mathbf{p})^{-1}-\tilde{\mathbf{M}}(\mathbf{p})]^{-1}, \tag{4}$$

where $\tilde{\mathbf{M}}$ is the spatial transform of the self-energy. The self-energy $\mathbf{M}$ can be written as an expansion in powers of moduli fluctuations. To first order (Frisch, 1968; Karal and Keller, 1964) $\mathbf{M}$ is expressed as (Weaver, 1990)

$$M_{\beta j}(\mathbf{y},\mathbf{z}) \approx\left\langle \frac{\partial}{\partial y_\alpha}\delta C_{\alpha\beta\gamma\delta}(\mathbf{y})\frac{\partial}{\partial y_\delta}G^0_{\gamma k}(\mathbf{y},\mathbf{z})\frac{\partial}{\partial z_i}\delta C_{ijkl}(\mathbf{z})\frac{\partial}{\partial z_l}\right\rangle. \tag{5}$$

Such an approximation is assumed valid if the fluctuations, $\delta\mathbf{C}$, are not too large. The components of $\tilde{\mathbf{M}}$ are employed to calculate the attenuation of the three wave modes. Further details of the scattering theory can be reviewed by the reader in the articles of Karal and Keller (1964), Frisch (1968), Stanke and Kino (1984), Weaver (1990), and Turner (1999).

The medium of oriented grains with rolling texture has orthorhombic symmetry. When ultrasonic waves propagate in such media, the phase velocity and the associated polarization vector are determined by the Christoffel equation. The dispersion relations for the mean response are then given by the solution of the Dyson equation, Eq. (4), as

$$g_\beta(\mathbf{p})=[g^0_\beta(\mathbf{p})^{-1}-m_\beta(\mathbf{p})]^{-1} =[\omega^2-p^2c_\beta^2-m_\beta(\mathbf{p})]^{-1}, \tag{6}$$

for each wave type, β, quasilongitudinal (qP) and two quasishear ($qS1$ and $qS2$) waves. The expressions for the dispersion relations of the mean response are written

$$\omega^2-p^2c_\beta^2-m_\beta(\mathbf{p})=0, \tag{7}$$

which is solved for the wave vector $\mathbf{p}$. The attenuation of each wave type is given by the imaginary part of $\mathbf{p}$. The explicit expressions of the attenuation can be determined using an approximation valid below the high-frequency geometric optics limit [$m_\beta(\mathbf{p})\approx m_\beta((\omega/c_\beta)\hat{\mathbf{p}})$] (Stanke and Kino, 1984; Weaver, 1990). This approximation allows the imaginary part of $\mathbf{p}$ to be calculated directly from Eq. (7). Thus, the attenuations of the three wave types are calculated as

$$\alpha_\beta(\hat{\mathbf{p}})=-\frac{1}{2\omega c_\beta(\hat{\mathbf{p}})}\,\mathrm{Im}\,m_\beta\left(\frac{\omega}{c_\beta}\hat{\mathbf{p}}\right). \tag{8}$$

The attenuations for the three wave types, which are each defined in Eq. (8), are finally given in the general form (Turner, 1999)

$$\alpha_\beta(\hat{\mathbf{p}})=\frac{1}{c_\beta^3(\hat{\mathbf{p}})}\left\{\frac{\pi}{4}\int d^2\hat{s}\,\frac{\omega^4}{c_{qS1}^5(\hat{\mathbf{s}})}\,\tilde{\eta}\left(\frac{\omega}{c_\beta(\hat{\mathbf{p}})}\hat{\mathbf{p}}-\frac{\omega}{c_{qS1}(\hat{\mathbf{s}})}\hat{\mathbf{s}}\right)^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\mathbf{v}_1}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\mathbf{v}_1}\right.$$
$$+\frac{\pi}{4}\int d^2\hat{s}\,\frac{\omega^4}{c_{qP}^5(\hat{\mathbf{s}})}\,\tilde{\eta}\left(\frac{\omega}{c_\beta(\hat{\mathbf{p}})}\hat{\mathbf{p}}-\frac{\omega}{c_{qP}(\hat{\mathbf{s}})}\hat{\mathbf{s}}\right)\Xi^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\mathbf{v}_2}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\mathbf{v}_2}$$
$$\left.+\frac{\pi}{4}\int d^2\hat{s}\,\frac{\omega^4}{c_{qS2}^5(\hat{\mathbf{s}})}\,\tilde{\eta}\left(\frac{\omega}{c_\beta(\hat{\mathbf{p}})}\hat{\mathbf{p}}-\frac{\omega}{c_{qS2}(\hat{\mathbf{s}})}\hat{\mathbf{s}}\right)\Xi^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\mathbf{v}_3}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\mathbf{v}_3}\right\},\tag{9}$$

where K is defined as the polarization for the wave type β (1, 2, or 3 for wave types $qS1$, qP, and $qS2$, respectively). In the above equation, the integrals are over the unit sphere, which is defined by unit vector $\hat{\mathbf{s}}$. The direction $\hat{\mathbf{p}}$ defines the propagation direction, $\hat{\mathbf{s}}$ is the scattered direction, and $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ are defined as the polarization directions. The dependence of the vectors $\hat{\mathbf{u}}$ on $\hat{\mathbf{p}}$ and of $\hat{\mathbf{v}}$ on $\hat{\mathbf{s}}$ is implicit. The argument of $\tilde{\eta}$ is the difference between the incoming and outgoing propagation directions. The inner products on the covariance of the moduli fluctuations are given in terms of four unit vectors. In the next section, the correlation function is specified.

III. CORRELATION FUNCTION

As shown in Eq. (2), the tensorial and spatial contributions of the material covariance are assumed independent. The spatial correlations are characterized by η. Here, it is assumed that η has an exponential form

$$\eta(\mathbf{r})=e^{-r/L}.\tag{10}$$

The correlation length, L, is of the order of the grain radius in polycrystals. In general, a simple exponential form of the spatial correlation function is not exact for polycrystals with rolling texture for which there is grain elongation. For elongated grains, a more general spatial correlation function must be used (Ahmed and Thompson, 1992). The influence of this choice of correlation function on the attenuations is left as a subject of future investigations. In Fourier transform space, the correlation function is then given by

$$\tilde{\eta}(\mathbf{q})=\frac{L^3}{\pi^2(1+L^2q^2)^2}.\tag{11}$$

The forms of the attenuation given above contain the difference of two vectors, $\tilde{\eta}(\mathbf{q})=\tilde{\eta}([\omega/c_1(\theta)]\hat{\mathbf{p}}-[\omega/c_2(\theta')]\hat{\mathbf{s}})$ as the argument for covariance in Eq. (2). Now the correlation functions $\tilde{\eta}_{\beta-\gamma}(\hat{\mathbf{p}},\hat{\mathbf{s}})$ are considered. If the three nondimensional frequencies are then defined as $x_\beta=\omega L/c_\beta$, using the expression of the spatial Fourier transform of the correlation function in Eq. (11), the functions $\tilde{\eta}_{\beta-\gamma}(\hat{\mathbf{p}},\hat{\mathbf{s}})$ are then expressed in terms of the above dimensionless quantities as

$$\tilde{\eta}_{\beta-\gamma}(\hat{\mathbf{p}},\hat{\mathbf{s}})=\frac{L^3}{\pi^2(1+x_\beta^2(\phi)+x_\gamma^2(\phi')-2x_\beta(\phi)x_\gamma(\phi')\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2},\tag{12}$$

for the incoming wave type β and outgoing wave type γ. The inner product, $\hat{\mathbf{p}}\cdot\hat{\mathbf{s}}=\cos\phi\cos\phi'\sin\theta\sin\theta'+\sin\phi\sin\phi'\sin\theta\sin\theta'+\cos\theta\cos\theta'$, if the unit vectors $\hat{\mathbf{p}}$ and $\hat{\mathbf{s}}$ are generally defined by $\hat{\mathbf{p}}=\mathbf{x}_1\cos\phi\sin\theta+\mathbf{x}_2\sin\phi\sin\theta+\mathbf{x}_3\cos\theta$ and $\hat{\mathbf{s}}=x_1\cos\phi'\sin\theta'+\mathbf{x}_2\sin\phi'\sin\theta'+\mathbf{x}_3\cos\theta'$. The angles θ, ϕ and θ', ϕ' are respectively defined as Euler angles in a general coordinate system. The form of the eighth-rank tensor, $\Xi_{ijkl}^{\alpha\beta\gamma\delta}$, is discussed next for rolling texture made up of cubic crystallites.

IV. COVARIANCE AND ATTENUATION

To calculate the attenuations, the relevant inner products on the covariance of the moduli fluctuations are required. The covariance of the moduli fluctuations is represented by an eighth-rank tensor which is given explicitly by

$$\Xi^{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}}_{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}}=\Xi_{\alpha\beta\gamma\delta}^{ijkl}\hat{u}_\beta\hat{u}_k\hat{p}_\alpha\hat{p}_l\hat{s}_i\hat{s}_\delta\hat{v}_\gamma\hat{v}_j.\tag{13}$$

For polycrystals of cubic symmetry, the eighth-rank covariance, $\Xi_{ijkl}^{\alpha\beta\gamma\delta}$, is written as

$$\Xi_{ijkl}^{\alpha\beta\gamma\delta}=\langle C_{ijkl}C_{\alpha\beta\gamma\delta}\rangle-\langle C_{ijkl}\rangle\langle C_{\alpha\beta\gamma\delta}\rangle$$
$$=\kappa^2\left\langle\sum_{n=1}^{3}a_{in}a_{jn}a_{kn}a_{ln}\sum_{n=1}^{3}a_{\alpha n}a_{\beta n}a_{\gamma n}a_{\delta n}\right\rangle$$
$$-\kappa^2\left\langle\sum_{n=1}^{3}a_{in}a_{jn}a_{kn}a_{ln}\right\rangle$$
$$-\left\langle\sum_{n=1}^{3}a_{\alpha n}a_{\beta n}a_{\gamma n}a_{\delta n}\right\rangle,\tag{14}$$

where the brackets, $\langle\ \rangle$, denote an ensemble average over all orientations of grains, and $\kappa=C_{11}^0-C_{12}^0-2C_{44}^0$ is the single-crystal anisotropy factor. If the polycrystal is of orthorhombic-cubic symmetry, only certain terms are nonzero. An example term necessary for calculating the attenuations is presented in the Appendix. Details of the other nonzero terms may be found elsewhere (Yang, 2003). For the second term in Eq. (14), the results are given in the details of other articles (Sayers, 1982; Johnson, 1985; Hirao *et al.*, 1987; Li and Thompson, 1990).

The forms of the attenuations presented in Eq. (9) require various inner products on the covariance tensor. These inner products have the general form of $\Xi^{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}}_{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}}$, where the vectors $\hat{\mathbf{p}}$ and $\hat{\mathbf{s}}$, respectively, represent the incoming and outgoing propagation directions. The vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$ are vectors defining the polarization directions of the particular waves. While waves propagate in arbitrary directions, the polarization vectors are found by the Christoffel equation. Substituting the correlation function, Eq. (12), and the inner products into Eq. (9), the resulting dimensionless attenuations are given in the form

$$\alpha_\beta(\hat{\mathbf{p}})L=\frac{x_\beta^4 c_\beta(\hat{\mathbf{p}})}{2\rho^2}\times\left\{\int_{4\pi}\frac{\Xi^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}(\hat{\mathbf{p}},\hat{\mathbf{s}})}{(1+x_\beta^2(\hat{\mathbf{p}})+x_{qS1}^2(\hat{\mathbf{s}})-2x_\beta(\hat{\mathbf{p}})x_{qS1}(\hat{\mathbf{s}})\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2}\frac{1}{c_{qS1}^5(\hat{\mathbf{s}})}d^2\hat{s}\right.$$
$$+\int_{4\pi}\frac{\Xi^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}(\hat{\mathbf{p}},\hat{\mathbf{s}})}{(1+x_\beta^2(\hat{\mathbf{p}})+x_{qP}^2(\hat{\mathbf{s}})-2x_\beta(\hat{\mathbf{p}})x_{qP}(\hat{\mathbf{s}})\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2}\frac{1}{c_{qP}^5(\hat{\mathbf{s}})}d^2\hat{s}$$
$$\left.+\int_{4\pi}\frac{\Xi^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}(\hat{\mathbf{p}},\hat{\mathbf{s}})}{(1+x_\beta^2(\hat{\mathbf{p}})+x_{qS2}^2(\hat{\mathbf{s}})-2x_\beta(\hat{\mathbf{p}})x_{qS2}(\hat{\mathbf{s}})\hat{\mathbf{p}}\cdot\hat{\mathbf{s}})^2}\frac{1}{c_{qS2}^5(\hat{\mathbf{s}})}d^2\hat{s}\right\},\tag{15}$$

where K has the same definition as discussed in Eq. (9). It should be noted that these inner products have units of κ^2. In the long wavelength Rayleigh limit, $x_\beta \ll 1$, Eq. (15) can be simplified as

$$\alpha_\beta(\hat{\mathbf{p}})L/x_\beta^4=\frac{c_\beta(\hat{\mathbf{p}})}{2\rho^2}\times\left\{\int_{4\pi}\frac{\Xi^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_1}}{c_{qS1}^5(\hat{\mathbf{s}})}d^2\hat{s}\right.$$
$$\left.+\int_{4\pi}\frac{\Xi^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_2}}{c_{qP}^5(\hat{\mathbf{s}})}d^2\hat{s}+\int_{4\pi}\frac{\Xi^{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}_{\cdots\hat{\mathbf{u}}_K\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}_3}}{c_{qS2}^5(\hat{\mathbf{s}})}d^2\hat{s}\right\}.\tag{16}$$

In Eq. (16), the dimensionless attenuation has been normalized by the fourth power of the dimensionless frequency for the respective wave type.

In Eqs. (15) and (16) the inner products, $\Xi^{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}}_{\cdots\hat{\mathbf{u}}\hat{\mathbf{p}}\hat{\mathbf{s}}\hat{\mathbf{v}}}$, do not have simple analytical forms for arbitrary propagation direction in this orthorhombic-cubic case. Thus, these results must be calculated numerically. In the next section, example numerical results and discussions are presented.

V. NUMERICAL RESULTS AND DISCUSSIONS

Numerical results are now presented for a 70% rolled steel plate. The material constants of a single crystal and the texture coefficients of the orientation distribution function with respect to the generalized spherical functions are given by (Bunge, 1982)

$$C_{11}^0=2.37\times10^{11}\ \text{Pa},\quad C_{12}^0=1.41\times10^{11}\ \text{Pa},$$
$$C_{44}^0=1.16\times10^{11}\ \text{Pa},\quad \rho=7850\ \text{kg/m}^3,\tag{17}$$

and

$$c_4^{00}=-1.47,\quad c_4^{20}=0.46,$$
$$c_4^{40}=0.50,\quad c_6^{00}=2.69,$$
$$c_6^{20}=-1.20,\quad c_6^{40}=0.46,$$
$$c_6^{60}=-0.14,\quad c_8^{00}=-0.07,\tag{18}$$
$$c_8^{20}=0.29,\quad c_8^{40}=-0.45,$$
$$c_8^{60}=-0.47,\quad c_8^{80}=-0.22.$$

The orientation distribution coefficients (ODCs) in Bunge's notation c_l^{mn} must be converted into those in Roe's notation W_{lmn}, which are used in this discussion. In order to carry out the calculations for the attenuations, Eq. (15), numerical methods are employed. The procedure of numerical methods for calculating the wave attenuations is now described.

First, using the Christoffel equation, the eigenvalue-eigenvector problem is solved for a given wave propagation direction and scattering direction. Second, the covariance of the moduli fluctuations is calculated by Eq. (14). Next, using the known covariance and eigenvectors, the inner products of each wave type are calculated numerically. Finally, the double integration is implemented numerically by the extended trapezoidal method. Here, examples are presented to describe important features of the wave attenuations for several propagation directions. The examples are generated using the methods discussed above. Since the orthorhombic symmetry has three mutually orthogonal planes of symmetry, all calculations are made for $0°\leqslant\phi\leqslant90°$ and $0°\leqslant\theta\leqslant90°$.

Convergence of the numerical integration was examined first. Wave attenuations of each wave type were examined for waves propagating in the rolling direction, that is $\phi=0°$ and $\theta=90°$, and at a dimensionless frequency $x_{qS1}=1.0$. The results show fast convergence for each wave mode in numerical integrations using the extended trapezoidal method (Yang, 2003). In order to achieve a balance between efficiency and accuracy, the number of intervals in the integra-

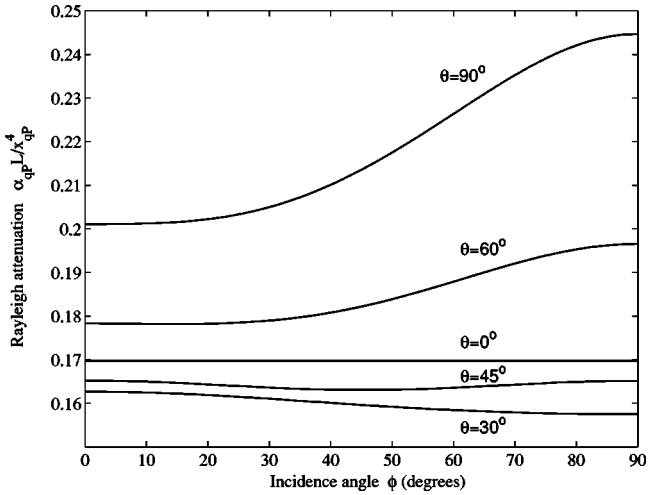


FIG. 1. Rayleigh attenuation, $\alpha_{qP}L/x_{qP}^4$, as a function of propagation direction for qP waves using the specified ODCs.

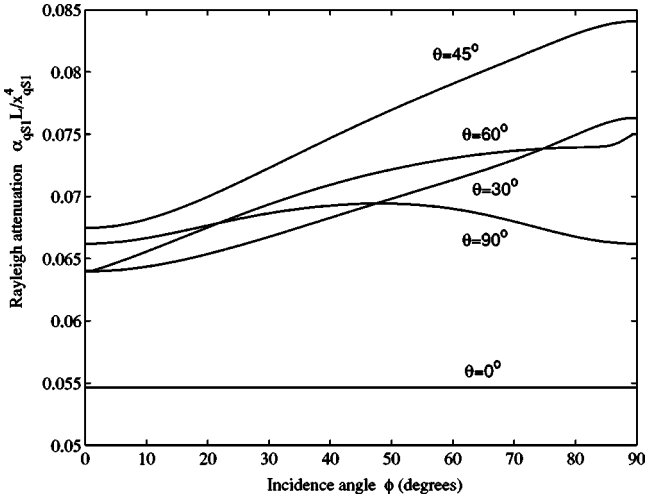


FIG. 2. Rayleigh attenuation, $\alpha_{qS1}L/x^4_{qS1}$, as a function of propagation direction for $qS1$ waves using the specified ODCs.

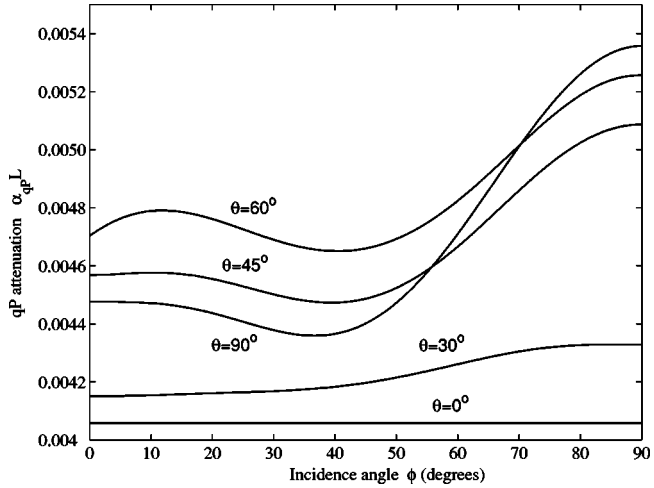


FIG. 4. Directional dependence of the normalized qP attenuation, $\alpha_{qP}L$, for frequency $x_{qS1}=1.0$.

tion $N=20$ is chosen for calculating the attenuations for each wave mode in the results shown here.

First, the attenuations within the Rayleigh limit are calculated using Eq. (16). It is known that the attenuations depend on the fourth power of frequency in the Rayleigh regime. Thus, the normalized Rayleigh attenuation $\alpha_\beta L/x^4_\beta$ of each wave mode is shown with the angular dependence in Figs. 1–3 for various propagation directions, respectively. It is observed that in this specific case the attenuations of each wave mode are considerably dependent on the wave propagation direction. For waves propagating in different directions, the curves of the attenuations have smoothly changing shapes. The percent variations of the qP, $qS1$, and $qS2$ attenuations in terms of polar angle θ shown in Figs. 1–3 are about 25–40%, 20–50%, and 15%, respectively. The variation of the $qS2$ attenuations is more uniform than that of the others.

Outside the Rayleigh regime, the attenuation results are calculated using the complete integrals, Eq. (15). The directional dependence of the attenuation is presented first for a given dimensionless frequency, $x_{qS1}=1.0$. Figure 4 shows the normalized quasilongitudinal wave (qP) attenuation, $\alpha_{qP}L$, as a function of azimuthal direction ϕ for various polar angles θ. It is seen that the attenuation is dependent on the propagation direction as expected. Here, the attenuation variation with respect to polar angle θ is around 15–30%. The results for the normalized shear wave ($qS1$ and $qS2$) attenuations are presented in Figs. 5 and 6, respectively. The directional dependence on the propagation direction for these attenuations is also significant. The percent variation of the $qS1$ and $qS2$ attenuations in terms of polar angle θ is about 10–50%. These results may be contrasted with the results in the Rayleigh limit. Comparisons of the Rayleigh attenuations with attenuations outside the Rayleigh regime show that the tendency of variation is quite different with each due to the effect of frequency. In Fig. 4, the qP wave attenuation is observed to have the maximum at $\phi=90°$ for given angles θ. In Fig. 5, the curves of the $qS1$ wave attenuations have smoothly changing shapes. Figure 6 shows that for propagation at polar angle $\theta=30°$, 45°, and 60°, the maximum attenuation is about $\phi=45°$, and at polar angle $\theta=90°$, there is a minimum attenuation approximately at $\phi=45°$. Further-

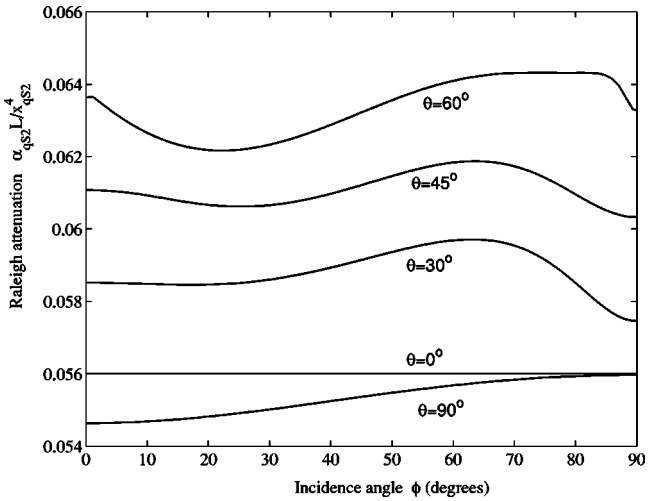


FIG. 3. Rayleigh attenuation, $\alpha_{qS2}L/x^4_{qS2}$, as a function of propagation direction for $qS2$ waves using the specified ODCs.

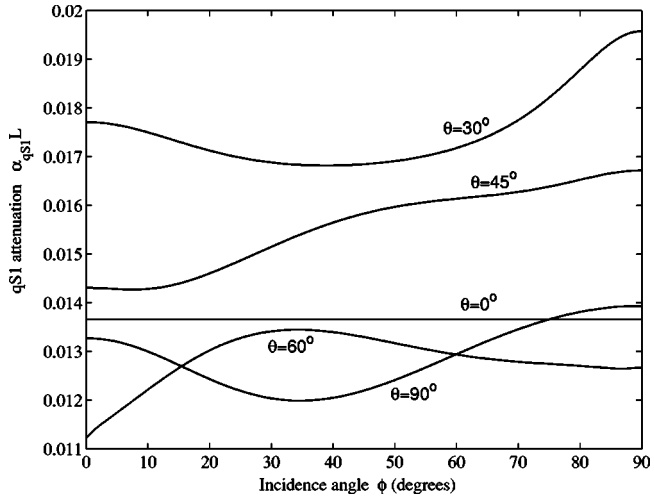


FIG. 5. Directional dependence of the normalized $qS1$ attenuation, $\alpha_{qS1}L$, for frequency $x_{qS1}=1.0$.

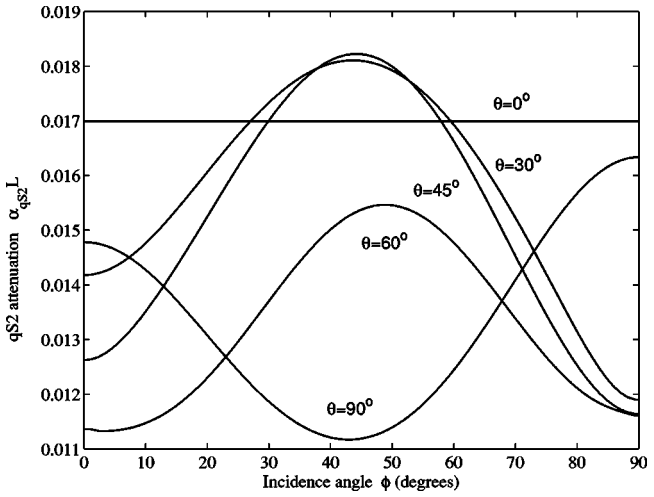


FIG. 6. Directional dependence of the normalized $qS2$ attenuation, $\alpha_{qS2}L$, for frequency $x_{qS1}=1.0$.

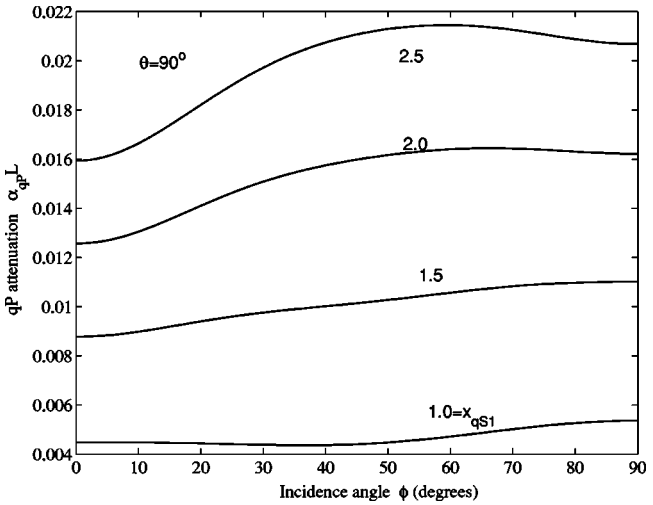


FIG. 8. Angular dependence of the normalized qP attenuation, $\alpha_{qP}L$, for various frequencies, x_{qS1}, at polar angle $\theta=90°$.

more, the asymmetry is observed in Fig. 6 for various polar angles.

Next, results are presented for the normalized attenuation as a function of aximuthal direction ϕ for four different frequencies at given polar angle $\theta=90°$. The normalized shear wave ($qS1$) attenuations, $\alpha_{qS1}L$, are shown in Fig. 7 for $\theta=90°$ for normalized frequencies $x_{qS1}=1.0$, 1.5, 2.0, and 2.5. It is seen that the attenuation curves show a similar shape with increasing frequency for the respective polar angle. The results for the normalized qP attenuations, $\alpha_{qP}L$, are shown in Fig. 8 for the same frequencies. Figure 9 shows the normalized $qS2$ attenuations, $\alpha_{qS2}L$, as a function of propagation direction for the same four frequencies. It is observed that the attenuation curves show a similar shape with increasing frequency for each wave type as well. All curves of the attenuations have smoothly changing shapes for various frequencies. Figure 9 shows that at polar angle $\theta=90°$ there is a minimum attenuation at $\phi=45°$ for various frequencies. There is no symmetry to be observed with increasing frequency as well in Fig. 9.

Finally, results are presented for the normalized attenuations as a function of frequency for several propagation directions. In Figs. 10–12, the normalized attenuations of the three wave modes are plotted versus dimensionless frequency, x_{qS1}, for propagation directions along the rolling, normal, and transverse directions, respectively. For the example considered here, Fig. 11 shows the $qS1$ wave attenuation for propagation in the rolling, normal, and transverse directions with polarization in the normal, transverse, and rolling directions, respectively. Figure 12 shows the $qS2$ wave attenuation propagated in the rolling, normal, and transverse directions and polarized in the transverse, rolling, and normal directions, respectively. It is observed that there is a transition region as the dimensionless frequency increases. Thus, the relative order of the attenuation is switched in such a transition region for the three wave modes, respectively. The attenuations increase with the fourth power of frequency in the low frequency limit. After a transition region, the attenuations scale with the square of frequency as expected. Moreover, the normalized attenuation of each wave type is plotted versus normalized frequency, x_{qS1}, for propagation within the x_1-x_2 plane for various

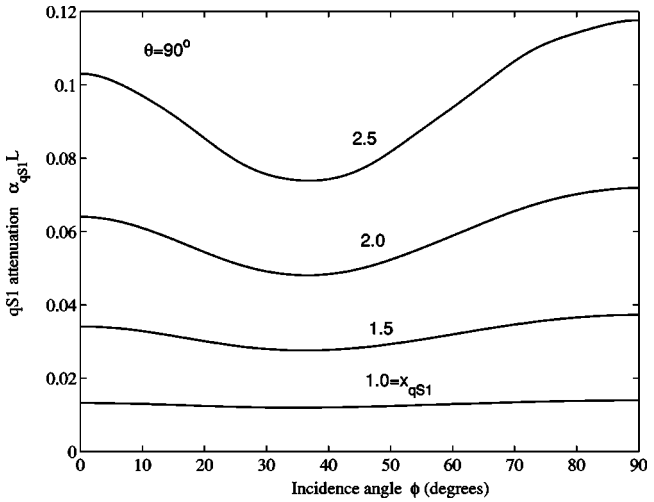


FIG. 7. Angular dependence of the normalized $qS1$ attenuation, $\alpha_{qS1}L$, for various frequencies, x_{qS1}, at polar angle $\theta=90°$.

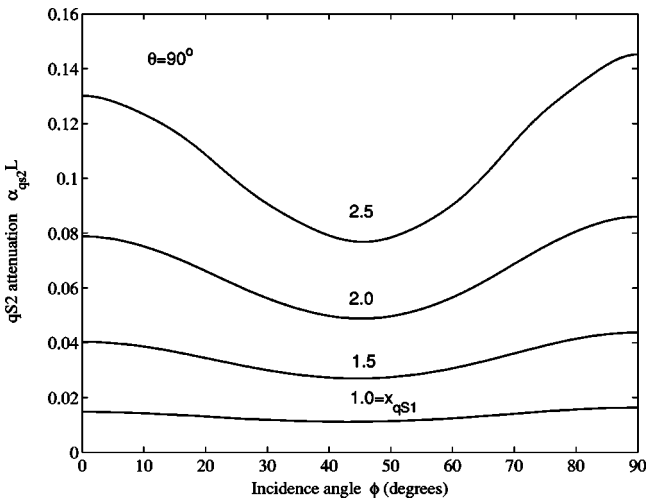


FIG. 9. Angular dependence of the normalized $qS2$ attenuation, $\alpha_{qS2}L$, for various frequencies, x_{qS1}, at polar angle $\theta=90°$.

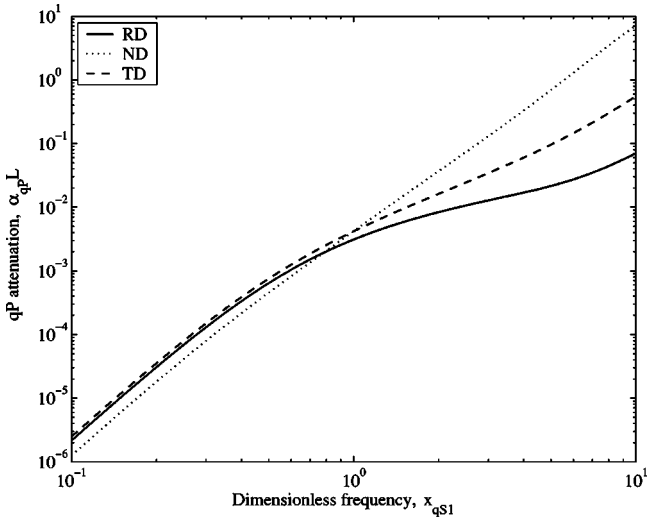


FIG. 10. Normalized qP attenuation, $\alpha_{qP}L$, as a function of dimensionless frequency, x_{qS1}, for waves propagating in the rolling (RD), normal (ND) and transverse (TD) directions.

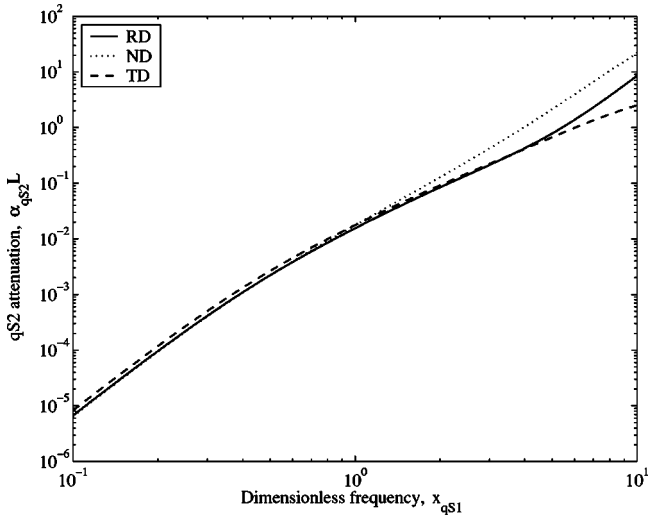


FIG. 12. Normalized $qS2$ attenuation, $\alpha_{qS2}L$, as a function of dimensionless frequency, x_{qS1}, for waves propagating in the rolling (RD), normal (ND) and transverse (TD) directions.

azimuthal angles in Figs. 13–15. In the Rayleigh regime, it is observed that the curves for each wave type mainly show the same shape. The variation of the attenuation of each wave mode is about 10%. Outside the Rayleigh regime, the attenuation of each wave mode in the rolling direction is about ten times higher than that for propagation at $\phi=30°$, 45°, and 60°. This feature is thought to be the result of the weak texture for the case discussed here.

The generalized attenuation results presented in this article for arbitrary propagation direction suggest that new ultrasonic techniques for characterization of texture coefficients may be possible. Further study is necessary to unravel the complex relations between the ODCs and the angular and frequency dependence of the attenuations. Attenuation measurements could ultimately be inverted such that the ODCs may be determined. However, such an approach must be optimized by choosing measurement directions that are the most sensitive to the desired ODC. Thus, theoretical developments such as this one will provide a firm basis for directing new experiments. Eventual process control of recrystallization that is quantitative will require modeling-directed experimental methods. In addition to theoretical research, numerical methods will also be necessary for progress to be made. This work must also be expanded to include other factors important to textured materials, such as grain size distribution and grain shape.

VI. SUMMARY

In this article, the scattering of elastic waves in polycrystalline materials with texture was discussed. The ensemble average response of the elastic waves is governed by the Dyson equation within the limits of first-order smoothing approximation. In contrast with previous work, here an anisotropic Green's dyadic approach was used. In order to calculate the attenuations, the relevant inner products on the covariance of the effective moduli fluctuations were derived

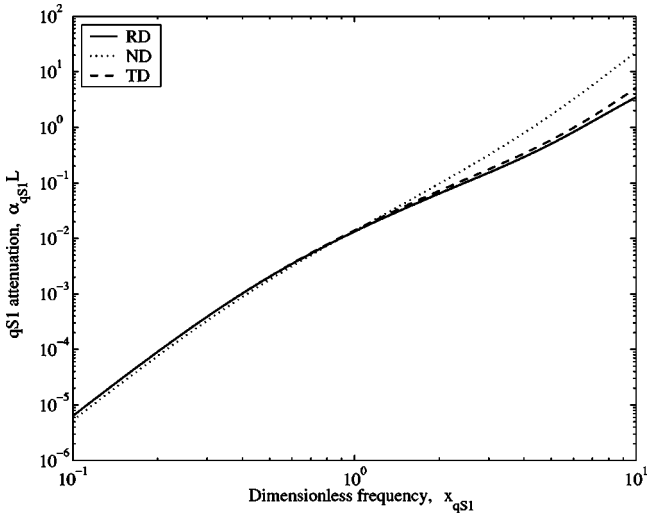


FIG. 11. Normalized $qS1$ attenuation, $\alpha_{qS1}L$, as a function of dimensionless frequency, x_{qS1}, for waves propagating in the rolling (RD), normal (ND) and transverse (TD) directions.

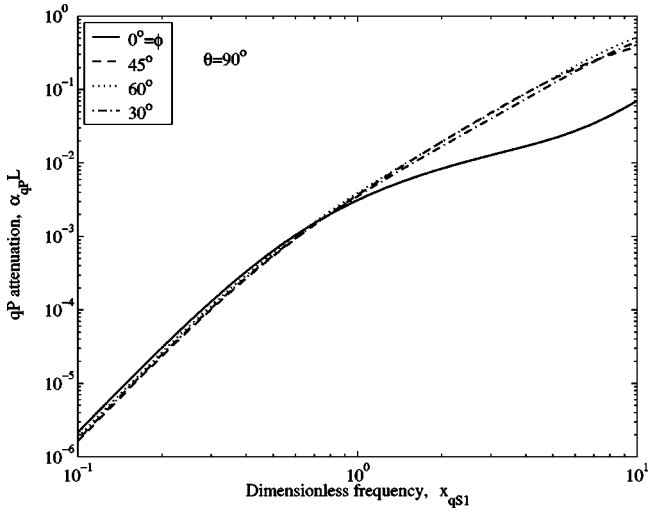


FIG. 13. Normalized qP attenuation, $\alpha_{qP}L$, as a function of dimensionless frequency, x_{qS1}, for waves propagating within the x_1-x_2 plane.

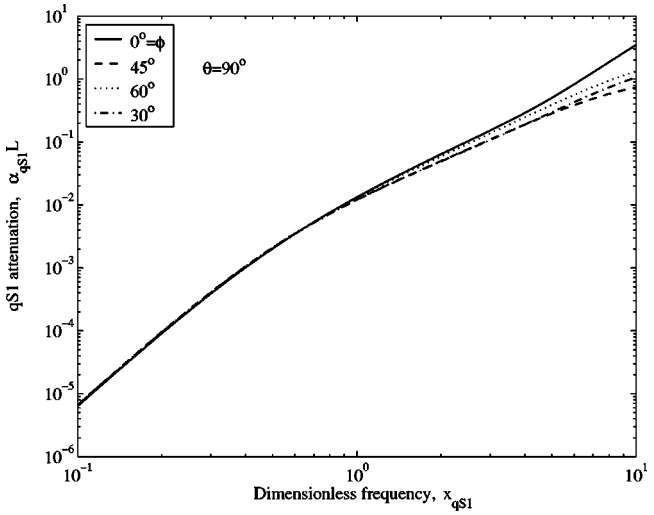


FIG. 14. Normalized $qS1$ attenuation, $\alpha_{qS1}L$, as a function of dimensionless frequency, x_{qS1}, for waves propagating within the x_1-x_2 plane.

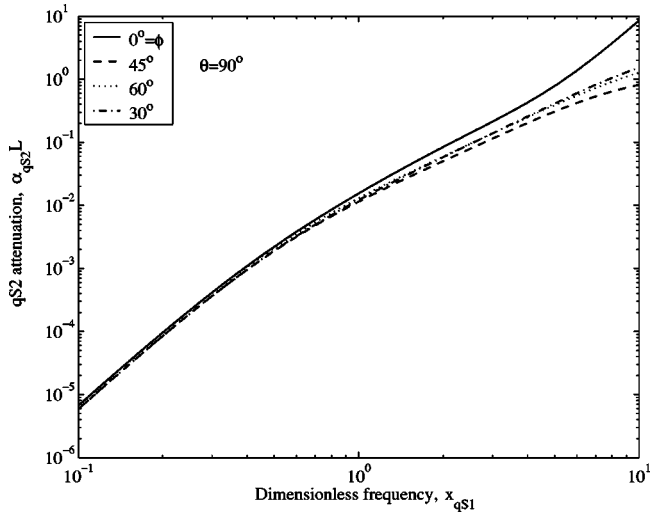


FIG. 15. Normalized $qS2$ attenuation, $\alpha_{qS2}L$, as a function of dimensionless frequency, x_{qS1}, for waves propagating within the x_1-x_2 plane.

in polycrystals of cubic symmetry with rolling texture. Compact expressions for the attenuations of the quasilongitudinal and two quasishear waves were then presented in terms of integrations over the scattered directions. In general, attenuations for each wave type are dependent on frequency, wave velocity, and wave propagation direction, as well as texture coefficients which are the expansion coefficients of the orientation function with respect to the generalized spherical functions. The results show that the attenuations of each wave type can be comprehensively affected by those parameters. The general formulation is also directly related to backscattering problems. The simple form of the results makes them particularly useful for nondestructive testing and materials characterization research. To use ultrasonic techniques for monitoring texture during processing, the relationships between ultrasonic parameters such as ultrasonic attenuation and materials texture must be investigated. If one knows the relationships between the ODCs and the ultrasonic attenuation, the texture coefficients can be inverted from ultrasonic attenuation measurements. The ultrasonic attenuations of sample specimens can then be measured during annealing such that the ODCs can be determined during processing.

ACKNOWLEDGMENTS

The financial support of the U.S. Department of Energy (Grant No. DE-FG02-01ER45890) is gratefully acknowledged. We also thank James C. Foley, R. Bruce Thompson, and David B. Rehbein at Ames Laboratory for their discussions.

APPENDIX

For polycrystals of cubic symmetry, the nonzero terms of the eighth-rank covariance, $\Xi_{ijkl}^{\alpha\beta\gamma\delta}$ in Eq. (14), are determined. The first term within the brackets is given as

$$\left\langle \sum_{n=1}^{3} a_{in}a_{jn}a_{kn}a_{ln} \sum_{n=1}^{3} a_{\alpha n}a_{\beta n}a_{\gamma n}a_{\delta n} \right\rangle$$

$$= \frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}\int_{-1}^{+1}\left(\sum_{n=1}^{3} a_{in}a_{jn}a_{kn}a_{ln} \right)\left(\sum_{n=1}^{3} a_{\alpha n}a_{\beta n}a_{\gamma n}a_{\delta n} \right) w(\xi,\psi,\varphi)\,d\xi\,d\psi\,d\varphi$$

$$= 4\pi^2\Bigg\{ W_{000}T_{000} + W_{400}\left[T_{400} + \frac{5}{\sqrt{70}}(T_{404} + T_{40\bar{4}}) \right] + W_{420}\left[T_{420} + T_{42\bar{0}} + \frac{5}{\sqrt{70}}(T_{424} + T_{42\bar{4}} + T_{424} + T_{42\bar{4}}) \right]$$

$$+ W_{440}\left[T_{440} + T_{44\bar{0}} + \frac{5}{\sqrt{70}}(T_{444} + T_{44\bar{4}} + T_{444} + T_{44\bar{4}}) \right] + W_{600}\left[T_{600} - \frac{\sqrt{14}}{2}(T_{604} + T_{60\bar{4}}) \right]$$

$$+ W_{620}\left[T_{620} + T_{62\bar{0}} - \frac{\sqrt{14}}{2}(T_{624} + T_{62\bar{4}} + T_{624} + T_{62\bar{4}}) \right] + W_{640}\left[T_{640} + T_{64\bar{0}} - \frac{\sqrt{14}}{2}(T_{644} + T_{64\bar{4}} + T_{644} + T_{64\bar{4}}) \right]$$

$$+ W_{660}\left[T_{660} + T_{66\bar{0}} - \frac{\sqrt{14}}{2}(T_{664} + T_{66\bar{4}} + T_{664} + T_{66\bar{4}}) \right] + W_{800}\left[T_{800} + \frac{\sqrt{154}}{33}(T_{804} + T_{80\bar{4}}) + \frac{\sqrt{1430}}{66}(T_{808} + T_{80\bar{8}}) \right]$$

$$+W_{820}\left[T_{820}+T_{82\bar{0}}+\frac{\sqrt{154}}{33}(T_{824}+T_{82\bar{4}}+T_{824}+T_{82\bar{4}})+\frac{\sqrt{1430}}{66}(T_{828}+T_{82\bar{8}}+T_{828}+T_{82\bar{8}})\right]$$

$$+W_{840}\left[T_{840}+T_{84\bar{0}}+\frac{\sqrt{154}}{33}(T_{844}+T_{84\bar{4}}+T_{844}+T_{84\bar{4}})+\frac{\sqrt{1430}}{66}(T_{848}+T_{84\bar{8}}+T_{848}+T_{84\bar{8}})\right]$$

$$+W_{860}\left[T_{860}+T_{86\bar{0}}+\frac{\sqrt{154}}{33}(T_{864}+T_{86\bar{4}}+T_{864}+T_{86\bar{4}})+\frac{\sqrt{1430}}{66}(T_{868}+T_{86\bar{8}}+T_{868}+T_{86\bar{8}})\right]$$

$$+W_{880}\left[T_{880}+T_{88\bar{0}}+\frac{\sqrt{154}}{33}(T_{884}+T_{88\bar{4}}+T_{884}+T_{88\bar{4}})+\frac{\sqrt{1430}}{66}(T_{888}+T_{88\bar{8}}+T_{888}+T_{88\bar{8}})\right]\Bigg\},$$

where the w is orientation distribution function (ODF). It can be expanded in a series of generalized spherical harmonics, with the coefficients W_{lmn} defining the orientation distribution coefficients (ODCs). The notations T_{lmn} are defined as

$$T_{lmn}=\frac{1}{4\pi^2}\int_0^{2\pi}\int_0^{2\pi}\int_{-1}^{+1}\left(\sum_{n=1}^{3}a_{in}a_{jn}a_{kn}a_{ln}\right)\left(\sum_{n=1}^{3}a_{\alpha n}a_{\beta n}a_{\gamma n}a_{\delta n}\right)\times Z_{lmn}(\xi)e^{-im\psi}e^{-in\varphi}\,d\xi\,d\psi\,d\varphi.$$

If the polycrystal is of orthorhombic-cubic symmetry, an example term ($i=j=k=l=1$ and $\alpha=\beta=\gamma=\delta=1$) is given as follows:

$$\left\langle\sum_{n=1}^{3}a_{n1}^4\sum_{m=1}^{3}a_{m1}^4\right\rangle=\frac{41}{105}+\frac{2232\sqrt{2}\pi^2}{5005}W_{400}-\frac{2976\sqrt{5}\pi^2}{5005}W_{420}-\frac{1448\sqrt{35}\pi^2}{5005}W_{440}-\frac{12\sqrt{26}\pi^2}{1001}W_{600}+\frac{64\sqrt{2730}\pi^2}{15015}W_{620}$$
$$-\frac{128\sqrt{91}\pi^2}{5005}W_{640}+\frac{64\sqrt{6006}\pi^2}{15015}W_{660}+\frac{56\sqrt{34}\pi^2}{7293}W_{800}-\frac{32\sqrt{1190}\pi^2}{12155}W_{820}+\frac{32\sqrt{1309}\pi^2}{12155}W_{840}-\frac{32\sqrt{14586}\pi^2}{36465}W_{860}$$
$$+\frac{16\sqrt{12155}\pi^2}{12155}W_{880}.$$

The remaining nonzero terms necessary for the calculation are given in detail elsewhere (Yang, 2003).

Ahmed, S., and Thompson, R. B. (**1992**). ''Effects of preferred grain orientation and grain elongation on ultrasonic wave propagation in stainless steel,'' in *Review of Progress in Quantitative NDE*, edited by D. O. Thompson and D. E. Chimenti (Plenum, New York), Vol. 11, pp. 1999–2006.

Ahmed, S., and Thompson, R. B. (**1996**). ''Propagation of elastic waves in equiaxed stainless-steel polycrystals with aligned [001] axes,'' J. Acoust. Soc. Am. **99**, 2086–2096.

Bunge, H. J. (**1982**). *Texture Analysis in Materials Science: Mathematical Methods* (Butterworths, London).

Frisch, U. (**1968**). ''Wave propagation in random media,'' in *Probabilistic Methods in Applied Mathematics*, edited by A. T. Barucha-Reid (Academic, New York), Vol. 1, pp. 75–198.

Hirao, M., Aoki, K., and Fukuoka, H. (**1987**). ''Texture of polycrystalline metals characterized by ultrasonic velocity measurements,'' J. Acoust. Soc. Am. **81**, 1434–1440.

Hirsekorn, S. (**1982**). ''The scattering of ultrasonic waves by polycrystals,'' J. Acoust. Soc. Am. **72**, 1021–1031.

Hirsekorn, S. (**1983**). ''The scattering of ultrasonic waves by polycrystals. II. Shear waves,'' J. Acoust. Soc. Am. **73**, 1160–1163.

Hirsekorn, S. (**1985**). ''The scattering of ultrasonic waves in polycrystalline materials with texture,'' J. Acoust. Soc. Am. **77**, 832–843.

Hirsekorn, S. (**1986**). ''Directional dependence of ultrasonic propagation in textured polycrystals,'' J. Acoust. Soc. Am. **79**, 1269–1279.

Johnson, G. C. (**1985**). ''Acoustoelastic response of a polycrystalline aggregates with orthotropic texture,'' J. Appl. Mech. **52**, 659–663.

Karal, F. C., and Keller, J. B. (**1964**). ''Elastic, electromagnetic, and other waves in a random medium,'' J. Math. Phys. **5**, 537–547.

Li, Y., and Thompson, R. B. (**1990**). ''Relations between elastic constants C_{ij} and texture parameters for hexagonal materials,'' J. Appl. Phys. **67**, 2663–2665.

Roe, R.-J. (**1965**). ''Description of crystallite orientation in polycrystalline materials. III. General solution to pole figure inversion,'' J. Appl. Phys. **36**, 2024–2031.

Roe, R.-J. (**1966**). ''Inversion of pole figures for materials having cubic crystal symmetry,'' J. Appl. Phys. **37**, 2069–2072.

Sayers, C. M. (**1982**). ''Ultrasonic velocities in anisotropic polycrystalline aggregates,'' J. Phys. D: Appl. Phys. **15**, 2157–2167.

Stanke, F. E., and Kino, G. S. (**1984**). ''A unified theory for elastic wave propagation in polycrystalline materials,'' J. Acoust. Soc. Am. **75**, 665–681.

Turner, J. A. (**1999**). ''Elastic wave propagation and scattering in heterogeneous, anisotropic media: Textured polycrystalline materials,'' J. Acoust. Soc. Am. **106**, 541–552.

Weaver, R. L. (**1990**). ''Diffusion of ultrasound in polycrystals,'' J. Mech. Phys. Solids **38**, 55–86.

Yang, L. (**2003**). ''Scattering of elastic waves in statistically anisotropic media,'' Ph.D. thesis, University of Nebraska.

A theoretical study of special acoustic effects caused by the staircase of the El Castillo pyramid at the Maya ruins of Chichen-Itza in Mexico

Nico F. Declercq[a] and Joris Degrieck
Soete Laboratory, Department of Mechanical Construction and Production, Faculty of Engineering, Ghent University, Sint Pietersnieuwstraat 41, 9000 Gent, Belgium

Rudy Briers
KATHO (RENO dept), Sint Jozefstraat 1, 8820 Torhout, Belgium

Oswald Leroy
Interdisciplinary Research Center, KULeuven Campus Kortrijk, E. Sabbelaan 53, 8500 Kortrijk, Belgium

(Received 21 February 2003; revised 18 February 2004; accepted 3 May 2004)

It is known that a handclap in front of the stairs of the great pyramid of Chichen Itza produces a chirp echo which sounds more or less like the sound of a Quetzal bird. The present work describes precise diffraction simulations and attempts to answer the critical question what physical effects cause the formation of the chirp echo. Comparison is made with experimental results obtained from David Lubman. Numerical simulations show that the echo shows a strong dependence on the kind of incident sound. Simulations are performed for a (delta function like) pulse and also for a real handclap. The effect of reflections on the ground in front of the pyramid is also discussed. The present work also explains why an observer seated on the lowest step of the pyramid hears the sound of raindrops falling in a water filled bucket instead of footstep sounds when people, situated higher up the pyramid, climb the stairs.
[DOI: 10.1121/1.1764833]

PACS numbers: 43.20.El, 43.20.Fn, 43.28.En [DKW] Pages: 3328–3335

I. INTRODUCTION

During the post meeting tour of the first PanAmerican/Iberian meeting on Acoustics that was held in Cancun (Mexico) in 2002 (hereafter called 'the post meeting tour'), the participants were shown that there are plenty of interesting sound effects that occur at Chichen-Itza. Chichen Itza is a Maya ruin where, besides the famous ''ball court,''[1] there is a pyramid (El Castillo) that produces a sound echo, in response to a handclap, which sounds like the chirp of a Quetzal bird. This effect has been one of the major subjects during plenty of talks given by David Lubman[2–5] and others.[6–8] Lubman has stressed the fact that the Quetzal bird chirp is actually caused by Bragg scattering. However, there has never been presented an actual simulation of the effect, except for some heuristic simulations based on the ray theory[2–5] or a heuristic approach for the case of incidence at 45° measured from the normal to the surface.[8–12] In what follows, a full diffraction simulation is presented of the echo, based on a (time-) delta function like handclap and also a real handclap, based on the physical parameters of the staircase of the Pyriamid at Chichen-Itza and based on the monofrequent single homogeneous plane wave diffraction theory of Claeys *et al.*,[9,10] which is a simplified case of the inhomogeneous plane wave diffraction theory.[11] The present work describes the first simulations of a spherical sound pulse, based on that monofrequent pure plane wave diffraction theory.[9,10] Furthermore, it is for the first time that the theory has been applied to audio frequencies.

Before presenting this development, it is of cultural importance to stress the fact that some people believe that the Quetzal bird chirp echo is caused by accident and others believe that it is caused as a consequence of the Pyramid builders' purpose. Nevertheless, it is known that the Quetzal bird has played a very important role in Mayan culture, which is probably due to the fact that Mayans originally lived for many centuries in the forest before getting involved in the construction of cities and religious sites. However, what is sure about this pyramid is that it certainly functioned as a great solar calendar. For example a large serpent is built on one side that causes special light effects around the time of spring and fall equinox. This serpent is culturally connected to the Quetzal bird (as can be seen on a Mayan glyph from the Dresden Codex), whence the generation of a Quetzal bird echo might not be a real coincidence. It is also known that an echo in Mayan culture represents a spirit. However, it must also be notified that a Quetzal bird echo also occurs at other Pre-Columbian sites and Ancient Mexican ruins.[12] Furthermore the first author encountered similar effects as in Chichen Itza at two religious sites in Sri Lanka. There, the short concrete staircase, that enables people to take a bath in the Menik Ganga river at the religious site of Katharagama, produces the low frequency sound of quacking ducks in response to a handclap. Furthermore high frequency echoes occur on the immense staircase leading to the religious site of Sri Pada (Adam's peak). Nevertheless, the ef-

[a]Electronic mail: NicoF.Declercq@UGent.be

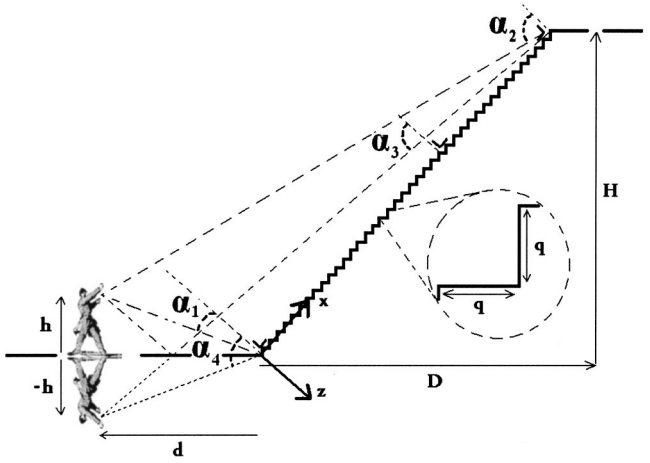


FIG. 1. Depiction of the pyramid's staircase with and observer in front of it.

fects in Sri Lanka are probably a coincidence and are not a result of purposely construction.

The last part of this paper is devoted to the less known fact that an observer seated on the lowest stair step of the great pyramid at Chichen Itza, hears pulses that sound like raindrops falling in a water filled bucket, when other people are climbing the pyramid higher up. This phenomenon (hereafter called ''raindrop effect''), has been observed by the first author and by a student fellow Cécile Goffaux during the post meeting tour. Since the ''rain god'' plays a very important role in the Yucatan Mayan culture, this finding might be an impetus for future cultural studies.

II. THEORETICAL DEVELOPMENT OF THE ECHO SIMULATION

The staircase is seen as a periodically corrugated (infinite) surface, being sawtooth shaped (see Fig. 1). This is only true within the interval of the physical staircase. This infinite mathematical model is matched to reality by modeling a handclap not by a truly spherical wave, but by a wave that only contains propagation directions from the emitter directly to the staircase within the angular interval $[\alpha_1,\alpha_2]$ that assures impingement on the staircase and within the interval $[\alpha_3,\alpha_4]$ if, in addition, reflections on the ground are considered as well. Hence, the handclap is only spherical if observed on the staircase. Whatever sound patterns are emitted to areas outside of the considered intervals is unimportant for the present study. The vectors **d**, **h**, **D**, and **H** are defined in Fig. 1. For $\mathbf{e}_x$ and $\mathbf{e}_z$ being unit vectors along the *x*, respectively, *z* direction, straightforward geometrical considerations result in

$$\alpha_1=\arccos\frac{(\mathbf{h}+\mathbf{d})\cdot\mathbf{e}_z}{\sqrt{h^2+d^2}}, \tag{1}$$

$$\alpha_2=\arccos\frac{(\mathbf{h}+\mathbf{d})\cdot\mathbf{e}_z}{|\mathbf{h}+\mathbf{d}-\sqrt{D^2+H^2}\mathbf{e}_x|}, \tag{2}$$

$$\alpha_3=\arccos\frac{(-\mathbf{h}+\mathbf{d})\cdot\mathbf{e}_z}{|-\mathbf{h}+\mathbf{d}-\sqrt{D^2+H^2}\mathbf{e}_x|}, \tag{3}$$

$$\alpha_4=\arccos\frac{(-\mathbf{h}+\mathbf{d})\cdot\mathbf{e}_z}{\sqrt{h^2+d^2}}, \tag{4}$$

with

$$\begin{pmatrix} d_x \\ d_z \end{pmatrix}=\begin{pmatrix} \cos\xi & \sin\xi \\ -\sin\xi & \cos\xi \end{pmatrix}\begin{pmatrix} |d| \\ 0 \end{pmatrix}, \tag{5}$$

$$\begin{pmatrix} h_x \\ h_z \end{pmatrix}=\begin{pmatrix} \cos\xi & \sin\xi \\ -\sin\xi & \cos\xi \end{pmatrix}\begin{pmatrix} 0 \\ |h| \end{pmatrix}, \tag{6}$$

and

$$\xi=\arctan\frac{H}{D}+\frac{\pi}{2}. \tag{7}$$

The diffraction theory of Claeys *et al.* that is applied here can be found in the literature.[9,11] Nevertheless, some characteristics of that theory are outlined below. The theory is based on the decomposition of the diffracted acoustic field into pure plane waves, which is essentially only allowed whenever the Lipmann conditions[9,11] are fulfilled, stating that the incident wave length must be of the same order of magnitude as the corrugation period and that the corrugation height must not exceed the incident wave length. If these conditions do not hold, then errors will occur in the description of the sound field within the corrugation. Elsewhere the errors will be small, except when the Lipmann conditions are seriously violated of course. Basically, each of the reflected and transmitted wave fields are decomposed into a series of plane waves, each plane wave of order *m* having a wave vector

$$\mathbf{K}^m=k_x^m\mathbf{e}_x+k_z^m\mathbf{e}_z, \tag{8}$$

with

$$k_x^m=k_x^{inc}+m\frac{2\pi}{\sqrt{2}q}, \tag{9}$$

and k_z^m determined by k_x^m, the material properties of the considered medium and the dispersion relation $k^2=\omega^2/v^2$, omega being the angular frequency and v being the plane wave velocity. The sign of k_z^m is chosen such, as to fulfill the necessity of plane waves to propagate away from the interface and, whenever k_z^m is purely imaginary, the amplitude must decay away from the interface. The continuity conditions demand continuity of normal stress and normal particle displacements on each spot of the pyramid's staircase. It can be found in Claeys *et al.*[9,11] that this leads to a set of equations that is periodical in *x*, whence the discrete Fourier transform can be applied, resulting in an equal number of equations and unknown amplitudes of all diffracted orders. It can also be found in Claeys *et al.*[9,11] that this discrete infinite set of equations and unknowns can be chopped to a square linear matrix equation that can be solved by a computer.

III. NUMERICAL RESULTS AND DISCUSSION

The following parameters are chosen such as to match the physical reality of the reported experiments[13] at 10 m in front of the pyramid (see Fig. 1). The observer's height is

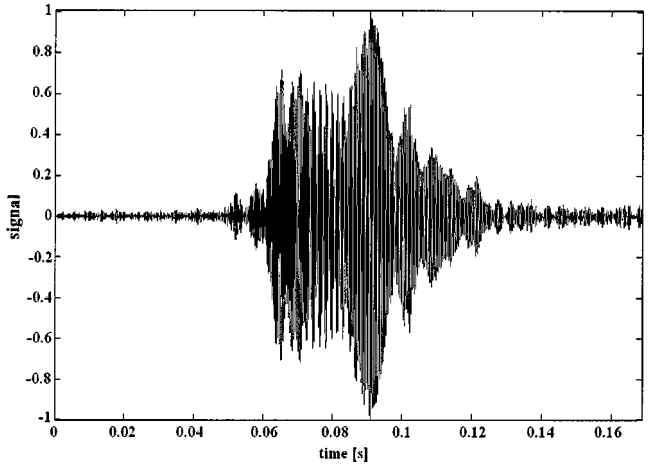


FIG. 2. Normalized calculated direct echo coming from a delta pulse.

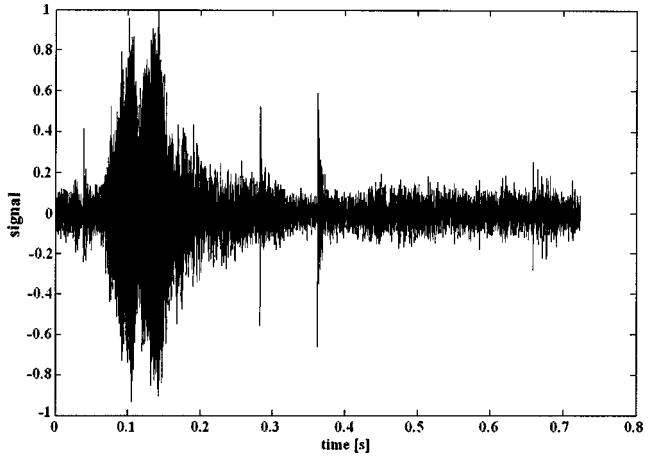


FIG. 3. Normalized recorded signal produced by a Quetzal bird in the forest.

chosen $h=1.80$ m, the observers distance $d=10$ m, the pyramid's dimensions $D=23.84$ m, $H=24.02$ m, $q=0.263$ m. It then follows from (1)–(4) that $\alpha_1=35.01°$, $\alpha_2=78.15°$, $\alpha_3=82.22°$, and $\alpha_4=55.42°$. The material properties in the humid Yucatan air have been taken as $\rho=1.1466$ kg/m^3 for the density and $v=343$ m/s for the sound velocity. Those for the limestone[14] staircase have been taken as $\rho=2000$ kg/m^3 for the density, $v_l=4100$ m/s for the longitudinal wave velocity and $v_s=2300$ m/s for the shear wave velocity. Damping has not been taken under consideration. For the parameters just given, the Lipmann conditions are given as follows: For frequencies lower than 1844 Hz, the numerical simulations will be perfect. For frequencies higher than 1844 Hz, there will be small errors in the description of the sound field within the corrugation, but not elsewhere. For very high frequencies, say more than 5000 Hz, errors may also occur in the prediction of the sound field outside of the corrugation, i.e., in the air and where the observer is situated. The errors gradually grow for higher frequencies and are due to ''shadow zones'' and neglecting internal reflection within the stairs.

A. Direct echo coming from a delta pulse

Within the angular interval $[\alpha_1,\alpha_2]$, the incident sound is considered to be spherical and contains 500 frequencies equally distributed between 500 and 3000 Hz. All incident plane waves have the same amplitude regardless of their direction and frequency. The former is necessary to produce the spherical wave, the latter is needed to produce a delta function like handclap. The spherical wave is modeled by 300 plane waves propagating along equally distributed angles within the interval $[\alpha_1,\alpha_2]$. There is no serious violation of the Lipmann conditions. Only for frequencies above 1844 Hz can there be some errors in the sound field description within the corrugation, but that is not of significant importance here because we are only interested in effects at the observer's position.

In Fig. 2, the calculated echo as a function of time is given, corresponding to an incident spherical pulse. This signal looks very clean, i.e., there is not too much noise outside of the echo, and is somewhat similar to the normalized plot in Fig. 3 of the actual sound of a real Quetzal bird in the forest. The latter signal was downloaded in*.wav format from the website of David Lubman.[13] The few delta function like peaks in the middle of that latter plot are the result of cracks that can be heard in the recorded sound file and are probably due to wood creaks in the bird's biotope. Figure 4 shows a normalized plot of the pyramid's echo and is obtained from a*.wav file that was also downloaded from Lubman's website.[13] This signal is far from clean. This is primarily due to low frequency noise coming from the interaction of wind with the microphone. Since it is almost impossible to compare sound signals in time–space, it is necessary to study sonograms or spectrograms of the obtained signals.[15] A sonogram depicts the amplitude as a function of time ''t'' and as a function of frequency ''f.'' It is obtained by a time limited Fourier transform. Here, we used a gaussian window of 0.002 s width. The sonograms are plotted by means of a gamma correction of 2. If the recorded sound is truly and solely an echo that comes from diffraction on the staircase, some patterns that will be mathematically described now, may appear in the sonogram. From (9) an $-m$'th order echo may appear if the following relation holds:

$$k_x^{\mathrm{inc}}=m\frac{\pi}{\sqrt{2}q}. \tag{10}$$

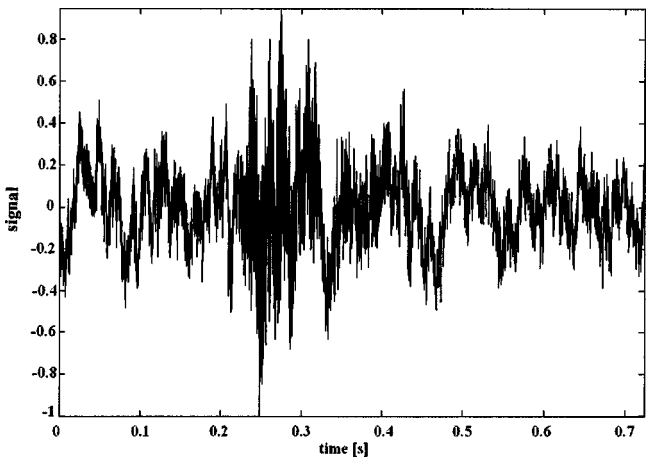


FIG. 4. Normalized recorded signal of the echo coming from the pyramid.

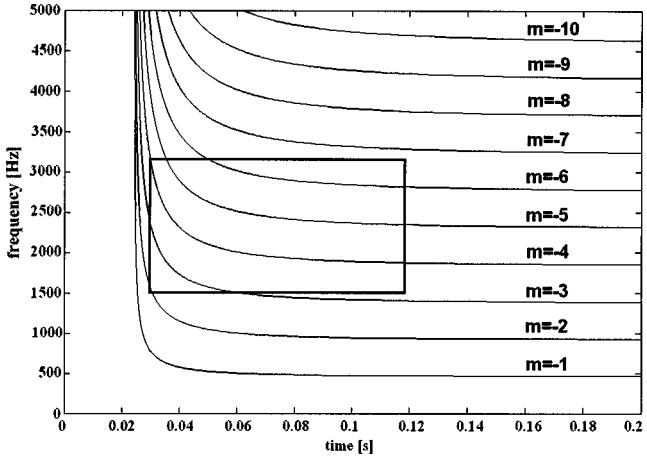


FIG. 5. Bragg diffraction lines on a sonogram. The sonogram shows information as a function of time (horizontal axis) and frequency (vertical axis). The square window is a reference window that represents the same time–frequency values in each sonogram in this report.

If this is combined with the dispersion relation, the angle, as a function of the frequency at which the echo may appear, can be calculated. If a ray-consideration is then applied, the time delay as a function of each angle, taking into account the wave speed in air, can also be obtained. This ultimately results in:

$$t(-m,f) = \frac{|d_z+h_z|}{v\cos \mathrm{Re}\left[\frac{\pi}{2}-\arctan\frac{\sqrt{\left(\frac{2\pi f}{v}\right)^2-\left(m\frac{\pi}{\sqrt{2}q}\right)^2}}{m\frac{\pi}{\sqrt{2}q}}\right]}. \tag{11}$$

In Fig. 5 the curves that are represented by (11) are depicted by means of a sonogram. In all sonograms that are presented here, the vertical axis represents the frequency in the range from 0 Hz (bottom) to 5000 Hz (top). The horizontal axis always spans a range of 0.2 s. However, the instant values on the horizontal axis do not always range from 0 to 0.2 s. It is only the difference between the right side of the horizontal axis and the left side that is 0.2 s. This is of course due to the fact that sound recordings contain no information about the absolute values of the start of recording and the end of recording. However, in order to compare the different sonograms that are presented here, we have taken into account physical considerations like the presence of the handclap in the recordings of Lubman[13] or the knowledge of the time-origin in our calculations, to draw a time–frequency window on each of the presented sonograms that is absolutely the same in each sonogram. This window will therefore function as the reference window for the discussions below. The absolute position of the window is chosen as to contain the relevant information that is present in Fig. 6, which is the sonogram that corresponds to the calculated echo of Fig. 2. This sonogram shows almost the same structure as the one of

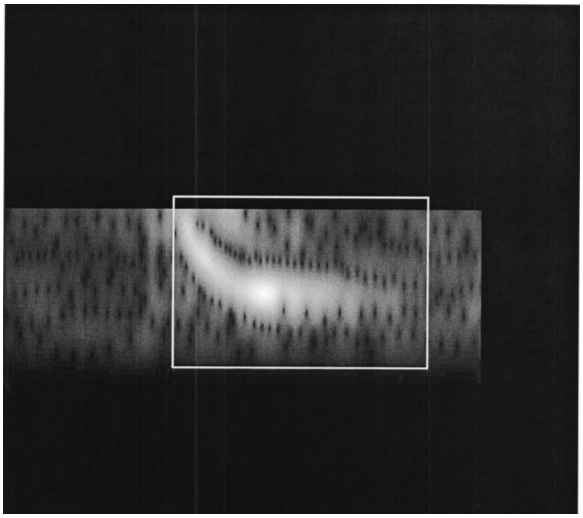


FIG. 6. Sonogram of the calculated direct echo coming from a delta pulse. The axes are equal to those of Fig. 5.

Fig. 7, which corresponds to the recorded Quetzal bird chirp in the woods (see also Fig. 3). The only important difference is the frequency at which the patterns appear and their duration. The actual bird chirps at lower frequencies than the calculated pyramid's echo. The authors do not know how a young Quetzal bird sounds like, but perhaps the resemblance would then be better. If Fig. 5 is compared with Fig. 6, it is noticed that even though the classical grating equation predicts the possibility of elevated amplitude lines in the sonogram, not all lines are associated with a relevant amplitude if the continuity conditions are also taken into account (see Fig. 6). However, the elevated amplitude patterns that do appear correspond more or less to the lines of Fig. 5. Especially there is a strong appearance of the $m=-4$ or $m=-5$ back reflected sound. The fact that it is not simple to decide which order is actually determining the elevated amplitudes is probably due to the interference of several plane waves because the incident sound is spherical. This is slightly in contrast with the assumption of Lubman[5] that the Bragg-orders can

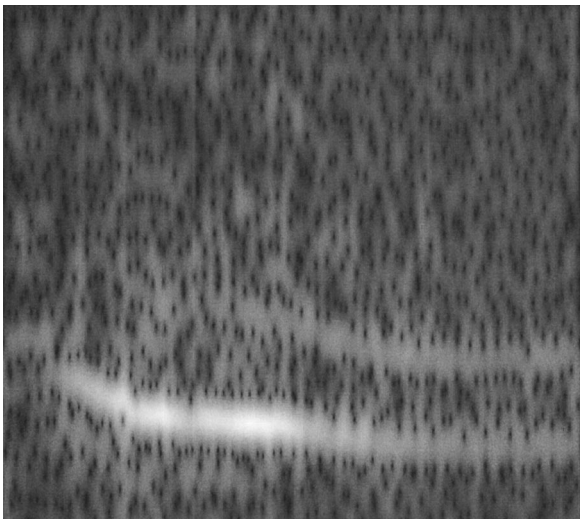


FIG. 7. Sonogram of the recorded Quetzal bird chirp in the forest. The axes are equal to those of Fig. 5.

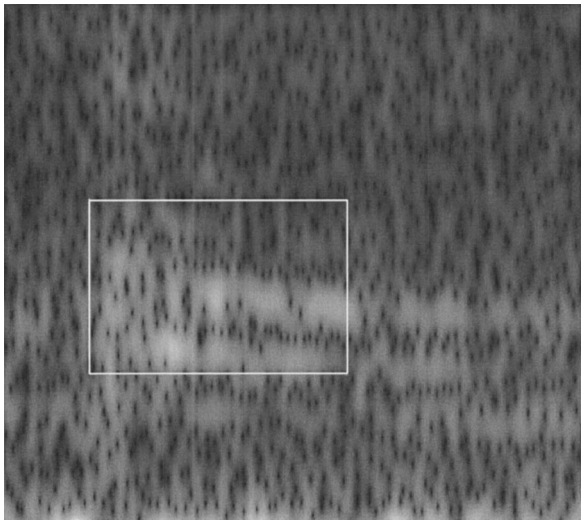


FIG. 8. Sonogram of the recorded echo coming from the pyramid. The vertical axis is equal as in Fig. 5, the horizontal axis spans the same time interval length. The reference window is situated at the same time/frequency values as in Fig. 5.

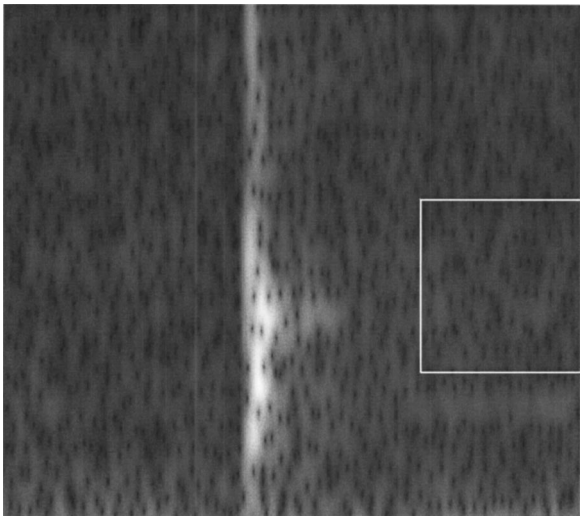


FIG. 9. Sonogram of the recorded (and mathematically isolated) handclap. Same comments on the axes as in Fig. 8.

be well seen in the sonogram of the recorded echo. In order to examine this contradiction, we have calculated the sonogram that actually corresponds to the recorded pyramid's echo of Fig. 4. The result is shown in Fig. 8. Within the reference time/frequency window, the same pattern can be found more or less (if you look through the noise) as in Fig. 6. However, Fig. 8 shows that it is absolutely not for certain that all patterns that are noticeable would correspond to the lines of Fig. 5. There is even something more obscure, which is the presence of ''patterns'' outside the reference window. If these were simply coming from Bragg diffraction, they would also appear in Fig. 6, where not only the mathematical grating equation is taken into account, but also the continuity conditions. Since they do not appear in Fig. 6 (or have an amplitude which is too small to be noticed), it can already be concluded that these patterns cannot simply be the result of pure Bragg diffraction and that an extra effect must be involved.

B. Direct echo coming from a handclap

The answer to the critical question as to what then actually causes these patterns can be revealed if one considers Fig. 9. The latter figure depicts the sonogram of the handclap taken from the recordings of Lubman[13] and being isolated from the echo of the same recording. A handclap is actually far from a delta function, because not all frequencies have the same amplitude. Actually, the handclap contains several frequency bands. For this purpose we have also simulated the echo resulting from a real handclap instead of a pulse. The handclap itself (as taken from Lubman[13]), which takes 0.02 s and must be followed by 0.18 s of silence in order to get a realistic time window of 0.2 s, needs to be represented by 4096 frequencies in between 5 and 25 000 Hz. Because of the amount of RAM memory needed and due to a limited CPU speed, taking into account all these frequencies in our diffraction procedure would result in a calculation time that exceeds the lifetime of our high speed computer.

This, together with the fact that the higher the frequencies, the more seriously Lipmann's conditions are violated, and a trade off between handclap reproducibility and calculation time, led to the decision to reduce the number of frequencies to 1968 in between 400 and 10 240 Hz. Taking into account higher frequencies would have violated Lipmann's conditions and would have taken us too much time. Consideration of only frequencies up to 5000 Hz led to an incident handclap that didn't sound right and led to an echo that did not at all correspond with reality. The reason of the latter effect is that a complicated handclap is much harder to deal with than the pulse of last section. Whereas a frequency chop for a pulse results in a new pulse that is quickly followed by a period of silence within the 2 s time window of interest, a frequency chop for a handclap results in unnegligible noise following the handclap, which is too strong if only frequencies up to 5000 Hz are considered. This noise, which is less important if frequencies up to 10 240 Hz are taken into account, is also diffracted and due to time shifts may even overlap with neighboring time windows after diffraction. Therefore the numerical echo (as can be seen in Fig. 10), corresponding with an incident numerical handclap with fre-

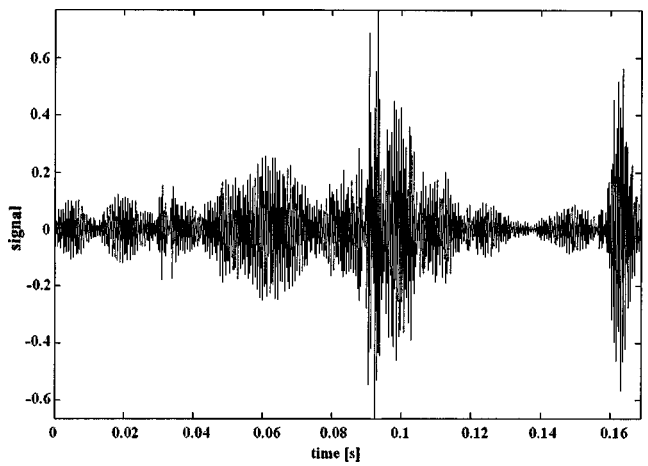


FIG. 10. Calculated direct echo coming from a handclap.

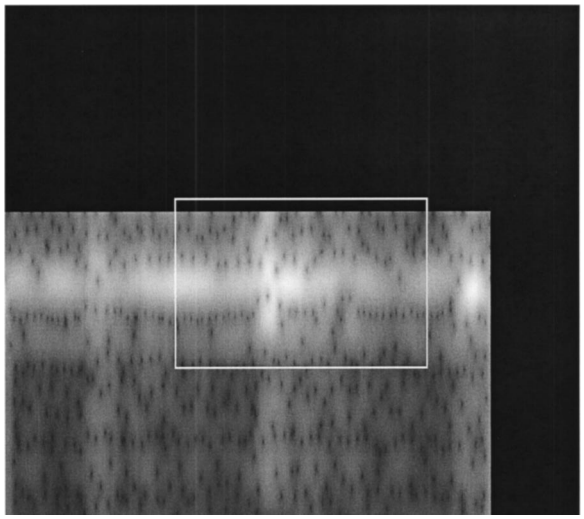


FIG. 11. Spectrogram of the calculated direct echo coming from a handclap.

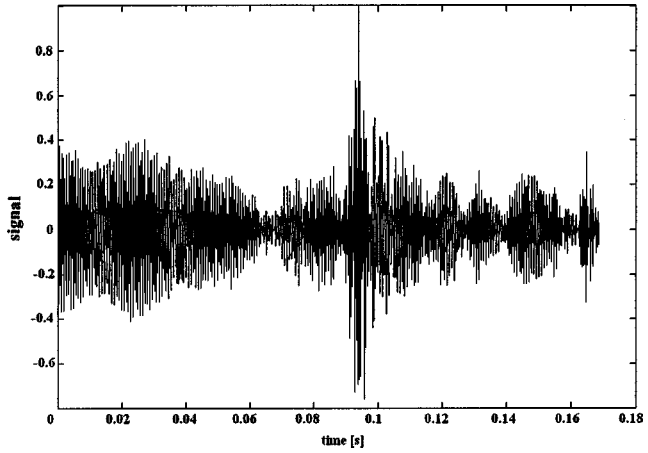


FIG. 12. Calculated direct and indirect echo coming from a handclap.

quencies higher than 10 240 Hz neglected, is, contrary to physical experiments, not limited in time. In Fig. 10, for reasons of calculation time limitations, we have, just as in the previous calculations, considered the results for all applied plane waves at all applied frequencies, but we have only taken into account 1024 positions of time within the interval of interest for reproducing the result. This means that a time limited Fourier transform cannot extract frequencies higher than the sampling frequency of 3034 Hz. However, if we take a look at the sonogram in Fig. 11, which corresponds with the numerical signal in Fig. 10 and is made just like all previous sonograms, we can see 4 frequency bands instead of only 2 in Fig. 6. Even more important is that they coincide with the experimentally measured frequency bands of Fig. 8. Therefore, even if, because of computer limitations, a true temporal description cannot be obtained, still what the frequencies are concerned the simulation reproduces the experimental result obtained by Lubman.[13] This proves that the lower two frequency bands in the experiments are mainly caused by the nature of the handclap and not as much by the diffraction process itself. In other words the echo is a function of the kind of incident sound.

C. Direct and indirect echo coming from a handclap

In Sec. III A we discussed the echo coming from a pulse and showed that the presence of 4 frequency bands in the reflected sound instead of 2 was probably due to the kind of incident sound. In Sec. III B this statement was proved by simulating the echo coming from the handclap in the experiments.[13] Yet another important question that needs to be resolved is the influence of the ground in front of the stairs of the pyramid. Up until now we have neglected this effect. We now consider the extreme condition where the ground is a perfect reflector. Hence sound coming from the handclap is not only propagating strait to the pyramid, but is also reflected on the ground before propagating towards the pyramid. Furthermore sound reflected from the pyramid may be received after strait propagation from the stairs or may again be reflected by the ground before being received. Therefore, the received signal G consists of 4 parts:

(i) G_1: Sound traveled directly to the pyramid and being received directly;
(ii) G_2: Sound traveled directly to the pyramid and being received after being reflected by the ground;
(iii) G_3: Sound being reflected by the ground before having traveled to the pyramid and being received directly;
(iv) G_4: Sound being reflected by the ground before having traveled to the pyramid and being received after being reflected by the ground.

We call the person in front of the pyramid ''person'' and his mirror image (see Fig. 1) the ''mirror person.'' The ground is replaced by a mathematical mirror plane. Mathematically G_1 is emitted by the person and again received by the person. G_2 is emitted by the person and received by the mirror person. G_3 is emitted by the mirror person and received by the person. G_4 is emitted and received by the mirror person. By filling in the correct coordinates of the person ($\mathbf{d}+\mathbf{h}$) and the mirror person ($\mathbf{d}-\mathbf{h}$), simulation is again possible of each signal. Then

$$G = G_1 + G_2 + G_3 + G_4 . \tag{12}$$

The result of G can be seen in Figs. 12 and 13. Again these figures cannot really tell anything about the temporal distribution of the frequencies, nevertheless it is seen that the ground has no influence on the presence or absence of the 4 frequency bands. In the future it would be great if someone would do some experiments at the pyramid by placing a reflector or an absorber in front of the staircase in order to see what effect it has on the received echo.

IV. EXPLANATION OF THE RAINDROP EFFECT

If people are climbing the pyramid, their shoes produce sound pulses containing all frequencies. Even though such pulses are more complicated, we model them here by means of a superposition of normally incident pure plane waves. Figure 14 shows the amplitude of the reflection coefficient of the zero order and the -1st order as a function of the frequency. Since we are only interested in understanding the raindrop effect, we focus in Fig. 15 on the frequency zone

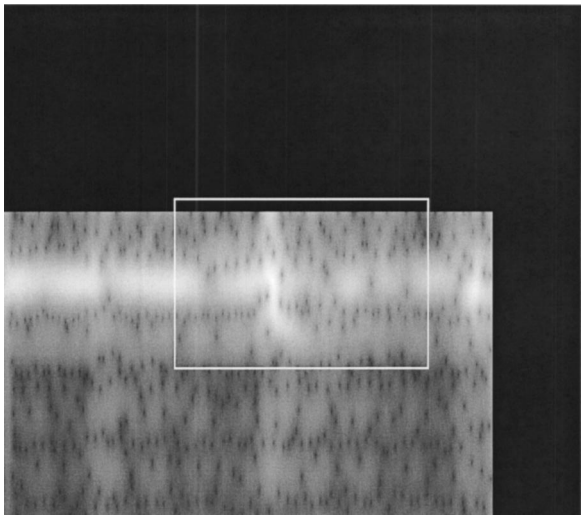


FIG. 13. Spectrogram of the calculated direct and indirect echo coming from a handclap.

where the −1st order reflected sound undergoes a transition from evanescent to bulk waves. That happens at a frequency f given by

$$f = \frac{v}{\sqrt{2}q} = 919.57 \text{ Hz}. \tag{13}$$

In addition it can be verified with what has been explained above that this transition zone fulfills the Lipmann conditions whence the validity of the numerical calculations cannot be cast doubt on.

On the right side of the transition frequency in Fig. 15, the −1st order reflected sound is as important, regarding its amplitude, as the zero order reflected sound. Furthermore, in Fig. 16, the propagation direction (measured from the pyramid's surface) of the −1st order reflected sound is depicted as a function of the frequency. On the right of and close to the transition frequency, the −1st order diffracted sound travels almost parallel to the pyramid's surface. Now, since that sound is bulk in nature (not evanescent) and since it has

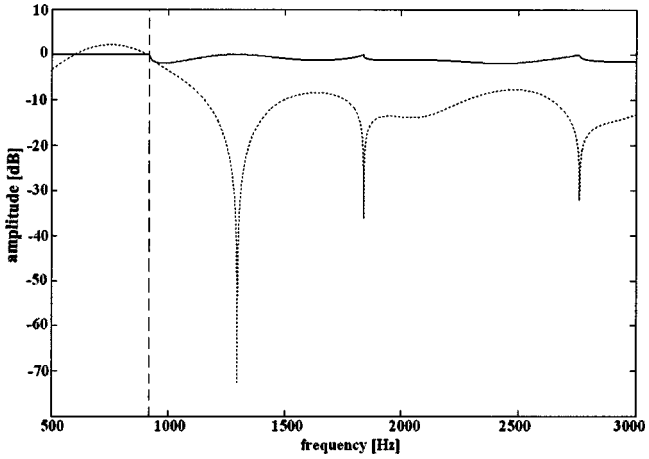


FIG. 14. The zero order reflection coefficient (solid line) and the −1st order reflection coefficient (dotted line) as a function of the frequency, for normal incident sound on the pyramid. The left side of the dashed line corresponds to evanescent −1st order reflected waves, while the right side corresponds to bulk −1st order reflected waves.

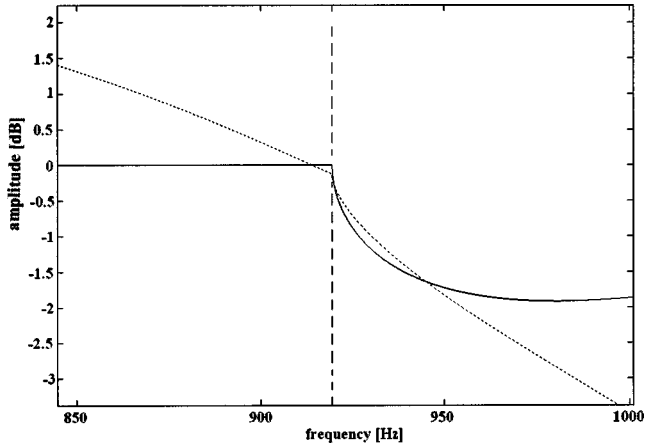


FIG. 15. Close up of Fig. 14.

a considerable amplitude (see Fig. 15), it is actually hearable for the observer seated on the lowest stair step. The observed frequency range is limited since (see Fig. 16) only a limited bunch of frequencies produce sound that can reach the observer's ear, which is situated at small angles from the pyramid's surface. Frequencies between 920 and 1000 Hz indeed sound like the main frequency that is present in the bunch of frequencies generated by a raindrop falling in a bucket filled with water.

V. CONCLUDING REMARKS

It is shown that the echo that is produced by the pyramid consists of diffracted sound coming from the staircase. The echo is formed by a process which is connected with Bragg reflection, but more effects are as important as well, such as the continuity conditions on the stairs and the frequency pattern of the incident sound. Therefore we would be pleased if someone could do some extended experiments in front of the pyramid in order to measure the echo as a function of the incident sound. We would not be surprised if the use of drums or timber wood to produce sound pulses would result in a better echo. The model also showed that the ground in front of the pyramid has no influence on the reflected frequency bands. Nevertheless it could not be shown what the

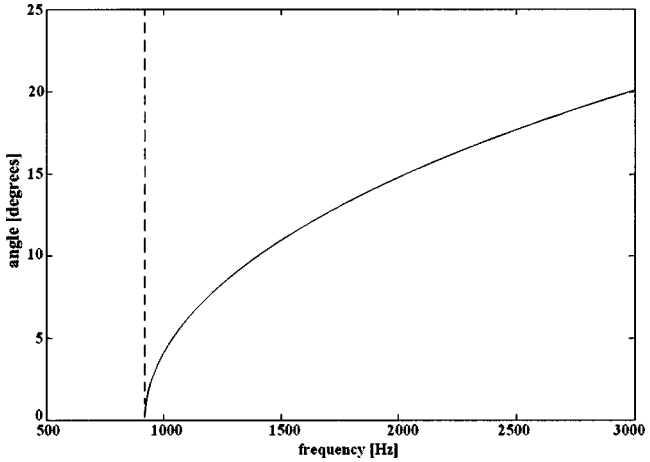


FIG. 16. The propagation angle of the −1st order reflected sound as a function of the frequency, measured from the pyramid's surface.

temporal effect is. It could elongate the echo or shorten it depending on the reflective properties of the ground. It would also be interesting to test the effect of the sound speed in air on the produced echo. This speed can vary in the dry season and wet season and can also vary with temperature. It is also explained how an observer seated on the lowest stair step may hear ''raindrops'' falling in a water filled bucket when other people are climbing the upper stairs.

ACKNOWLEDGMENTS

The authors are thankful to ''The Flemish Institute for the Promotion of the Scientific and Technological Research in Industry (I.W.T.)'' for sponsoring this research. The experimental work of David Lubman and the spread of his data through the internet is sincerely acknowledged. We hope that this theoretical work will form an impetus for him to do even more such interesting experiments in the near future. The first author strongly acknowledges the inspiring interdisciplinary contacts with fellow students, research colleagues and seniors from all over the world, during the First Pan American/Iberian Meeting on Acoustics in Cancun-Mexico in December 2002. Furthermore we express our sincere gratitude to the editor's and reviewer's comments and their encouragement to perform, besides the calculations for the direct pulse, also the simulations for the exact handclap and also for the effect of the ground in front of the pyramid.

[1] Wayne Van Kirk, ''The accidental (acoustical) tourist,'' J. Acoust. Soc. Am. **112**(5), 2284 (2002).
[2] David Lubman, ''Acoustical features of two Mayan monuments at Chichen Itza: Accident or design,'' J. Acoust. Soc. Am. **112**(5), 2285 (2002).
[3] David Lubman, ''Singing stairs,'' Sci. News (Washington, D. C.) **155**, 44 (1999).
[4] David Lubman, ''Mayan acoustics: Of rainbows and resplendent quetzals,'' J. Acoust. Soc. Am. **106**(4), 2228 (1999).
[5] David Lubman, ''Archaeological acoustic study of chirped echo from the Mayan pyramid at Chichén Itzá,'' J. Acoust. Soc. Am. **104**(3), 1763 (1998).
[6] Bijal P. Trivedi, ''Was Maya Pyramid Designed to Chirp Like a Bird?'' National Geographic Today, Dec. 6, 2002.
[7] Fernando J. Elizondo-Garza, ''Quetzal or not Quetzal, that is the question... on the stairs of the Castillo monument in Chichen Itza,'' J. Acoust. Soc. Am. **112**(5), 2285 (2002).
[8] Jorge Carrera and Sergio Beristain, ''Theoretical interpretation of a case study: Acoustic resonance in an archeological site,'' J. Acoust. Soc. Am. **112**(5), 2285 (2002).
[9] J.-M. Claeys and O. Leroy, ''Diffraction of plane waves by periodic surfaces,'' Rev. Cethedec **72**, 183–193 (1982).
[10] J. M. Claeys, Oswald Leroy, Alain Jungman, and Laszlo Adler, ''Diffraction of ultrasonic waves from periodically rough liquid–solid surface,'' J. Appl. Phys. **54**(10), 5657–5662 (1983).
[11] Nico F. Declercq, Joris Degrieck, Rudy Briers, and Oswald Leroy, ''A theoretical elucidation for the experimentally observed backward displacement of waves reflected from an interface having superimposed periodicity,'' J. Acoust. Soc. Am. **112**(5), 2414 (2002); Nico F. Declercq, Joris Degrieck, Rudy Briers, and Oswald Leroy, ''Theoretical verification of the backward displacement of waves reflected from an interface having superimposed periodicity,'' Appl. Phys. Lett. **82**(15), 2533–2534 (**2003**); Nico F. Declercq, Joris Degrieck, Rudy Briers, and Oswald Leroy, ''Theory of the backward beam displacement on periodically corrugated surfaces and its relation to leaky Scholte-Stoneley waves,'' in press with J. Appl. Phys.
[12] Sergio Beristain, Cecilia Coss, Gabriela Aquino, and Jose Negrete, ''Tonal response on the stairway of the main pyramid at La Ciudela, Teotihuacan archeological site,'' J. Acoust. Soc. Am. **112**(5), 2285 (2002); paper 3aAA4 in Proceedings of the first PanAmerican/Iberian meeting on acoustics, 2002.
[13] David Lubman: http://www.ocasa.org/MayanPyramid.htm (site visited on 12/13/2002).
[14] Jacques R. Chamuel and Gary H. Brooke, ''Transient Scholte wave transmission along rough liquid–solid interfaces,'' J. Acoust. Soc. Am. **83**(4), 1336–1344 (1988).
[15] Leon Cohen, *Time-Frequency Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1995).

On the acoustic modes in a cylindrical duct with an arbitrary wall impedance distribution

L. M. B. C. Campos and J. M. G. S. Oliveira
Secção de Mecânica Aeroespacial, ISR, Instituto Superior Técnico, 1049-001 Lisboa Codex, Portugal

(Received 11 August 2003; revised 19 July 2004; accepted 3 September 2004)

The present paper considers the propagation of sound in a cylindrical duct, with a wall section of finite length covered by an acoustic liner whose impedance is an arbitrary function of position. The cases of (i) uniform wall impedance, and wall impedance varying along the (ii) circumference or (iii) axis of the duct, or (iv) both simultaneously, are explicitly considered. It is shown that a nonuniform wall impedance couples modes with distinct azimuthal l or axial m wave numbers, so that their radial wave numbers k can no longer be calculated separately for each pair (m,l). The radial wave numbers are the roots of an infinite determinant, in the case when the wall impedance varies either (i) circumferentially or (ii) radially. If the wall impedance varies (iv) both radially and circumferentially, then the radial wave numbers are the roots of a doubly infinite determinant, i.e., an infinite determinant in which each term is an infinite determinant. The infinite determinants specifying the radial wave numbers are written explicitly for sound in a cylindrical nozzle with a uniform axial flow, in which case the radial eigenfunctions are Bessel functions; the method of calculation of the radial wave numbers applies equally well to a cylindrical nozzle with shear flow and/or swirling flows, with the Bessel functions replaced by other eigenfunctions. The radial wave numbers are calculated by truncation of the infinite determinants, for several values of the aspect ratio, defined as the ratio of length to diameter. It is shown that a nonuniform wall impedance will give rise to additional modes compared with a uniform wall impedance. The radial wave numbers specify the eigenfrequencies for the acoustic modes in the duct; the imaginary parts of the eigenfrequencies specify the decay of the sound field with time, and thus the effectiveness of the acoustic liner. [DOI: 10.1121/1.1812308]

PACS numbers: 43.20.Mv, 43.20.Ks, 43.20.Bi [MO] Pages: 3336–3347

I. INTRODUCTION

One of the most widespread methods of noise reduction in the inlets (Horowitz *et al.*, 1982; Meyer, 1996) and exhausts of jet engines (Savkar, 1975; Koch, 1977b; Rienstra, 1984; Martinez, 1988; Snakovska *et al.*, 1996; Snakowska and Idczak, 1997; Joseph and Morfey, 1999; Hocter, 1999), including fan, compressor, and turbine ducts, is the use of acoustic liners. These liners are represented most simply by an impedance condition (Rawlins, 1978; Campos, 1978a; Koch and Möhring, 1983; Koch, 1977a; Howe, 1983; Rienstra, 1985; Bies *et al.*, 1991), which can be applied locally if the impedance is uniform. The ever more stringent noise standards, demanding higher performance from acoustic liners, have led to the consideration of nonuniform (Watson, 1984; Fuller, 1984a,b; Vaidya, 1985; Regan and Eaton, 1999), e.g., circumferentially and/or longitudinally segmented liners, for which the wall impedance is an arbitrary function of position. Such a nonuniform distribution of impedance over the duct walls has a global rather than local effect on the acoustic natural frequencies, or eigenfrequencies, in the duct. These frequencies are generally complex, implying that the sound field decays with time. The very purpose of using segmented acoustic liners is to optimize the impedance distribution, so as to obtain the fastest possible decay in time of the eigenfunctions, as specified by their complex eigenfrequencies. The aim of the present paper is to demonstrate the effect of arbitrary distributions of nonuniform wall impedance, on sound in nozzles containing an axial mean flow by calculating the corresponding normal modes and eigenfrequencies.

The specific application chosen for detailed analysis (Sec. II) is a cylindrical nozzle containing an uniform axial flow. It is well known that in this case (Sec. II A) the eigenfunctions consist of four factors: (i) sinusoidal functions in the circumferential direction, with integer wave number m; (ii) for a duct of finite length, sinusoidal functions also in the longitudinal direction, with integer wave number l; (iii) for finite acoustic pressure on the duct axis, Bessel coefficients of order m, having as argument the radial distance r multiplied by a radial wave number k; (iv) sinusoidal functions in time, with frequencies ω determined by k, l, m. Since the boundary condition at the duct wall determines the radial wave number k_n for each l, m, it follows that it also specifies the eigenfrequencies ω_{lmn}. The present method also applies to the acoustics of ducts containing sheared (Campos and Kobayashi, 2000; Campos *et al.*, 1999; Campos and Serrão, 1998) or swirling (Campos, 1993) mean flows, with the difference that the eigenfunctions may be more complicated (Campos, 1993, 2000b,a) than Bessel functions (Watson, 1944; Whittaker and Watson, 1927; Campos, 2003) Vortical mean flows can absorb sound (Munt, 1977; Campos, 1978b; Plumblee and Doak, 1979; Nelson and Morfey, 1981; Rienstra, 1984; Howe, 1995), in addition to the effect of the liner.

In the simplest case (Sec. II A) of a rigid wall boundary condition, the k_n are real, and ω is real or imaginary, corre-

sponding, respectively, to propagating or evanescent waves. In the case (Sec. II B) of a uniform complex impedance, both k and ω are generally complex, and thus the eigenfunctions decay with time. The calculation of the eigenfrequencies is nonlocal in the case of impedance varying either circumferentially (Sec. III A) or axially (Sec. III B), when it can be represented by a single Fourier series. A double Fourier series is needed if the wall impedance varies (Sec. III C) both axially and circumferentially. In the cases of nonuniform wall impedance (Secs. III A–III C), the complex eigenfrequencies are specified by the roots of infinite determinants, which can be determined approximately (Sec. IV A) by truncation and iteration methods. It is shown that a nonuniform wall impedance can give rise to additional eigenmodes relative to the case of uniform wall impedance (Sec. IV B). The discussion (Sec. V) indicates how the complex eigenfrequencies, and more specifically, their imaginary parts, which specify the decay of eigenmodes with time, may be used to maximize the attenuation of sound by optimizing the impedance distribution over the duct wall.

II. ACOUSTIC EIGENFUNCTIONS AND EIGENVALUES FOR SOUND IN A CYLINDRICAL NOZZLE

The solutions of the acoustic wave equation, without sources, in cylindrical coordinates and in presence of a uniform axial flow, are illustrated first for rigid wall (Sec. A) and uniform impedance (Sec. B) boundary conditions, before proceeding to the case of nonuniform linings (Sec. III).

A. Cylindrical nozzle with uniform axial flow

Consider the convected wave equation (Pierce, 1981; Campos, 1986) with sound speed c for the acoustic pressure p, using cylindrical coordinates (r,θ,z) for a nozzle containing an uniform axial flow with velocity U

$$\left\{\frac{1}{c^2}\left(\frac{\partial}{\partial t}+U\frac{\partial}{\partial z}\right)^2-\left[\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}+\frac{\partial^2}{\partial z^2}\right]\right\}\times p(r,\theta,z,t)=0. \tag{1}$$

The free solutions, in the absence of sources, are sought; these eigenfunctions can also be used to solve the wave equation forced by sound sources. Considering a lined section of length L, the acoustic fields are sinusoidal functions of the axial and circumferential coordinates, z and θ, with wave numbers, respectively, $2\pi l/L$ and m appearing in the corresponding Fourier series, and the time dependence is specified by a Fourier integral over the frequency ω

$$p(r,\theta,z,t)=\sum_{m=-\infty}^{+\infty}\mathrm{e}^{im\theta}\sum_{l=-\infty}^{+\infty}\mathrm{e}^{i2\pi lz/L}\times\int_{-\infty}^{+\infty}\mathrm{e}^{-i\omega t}P_{lm}(r,\omega)\mathrm{d}\omega. \tag{2}$$

The axial dependence of the acoustic field has been represented by a Fourier series over a section $0\leqslant z\leqslant L$ of the duct, to be able to match a similar Fourier series representation Eq. (32) of the nonuniform wall impedance varying axially. On substitution of Eq. (2) in Eq. (1), it is found that the radial dependence is specified by a Bessel equation of order m

$$r^2P''+rP'+(k^2r^2-m^2)P=0, \tag{3}$$

where prime denotes derivative with respect to r, and k, which plays the role of radial wave number, is defined by

$$k^2\equiv\frac{1}{c^2}\left(\omega-\frac{2\pi lU}{L}\right)^2-\left(\frac{2\pi l}{L}\right)^2. \tag{4}$$

The radial wave number k is distinguished from the axial wave number, which is represented by $2\pi l/L$.

The solution of the Bessel equation Eq. (3) of order m, which is finite (Watson, 1944; Campos, 1993) on the duct axis, is a Bessel function of integral order m

$$P_{lm}(r,\omega)=A_{lm}J_m(kr), \tag{5}$$

where A_{lm} is a constant. In the simplest case of a rigid wall boundary condition

$$\left.\frac{\partial p}{\partial r}\right|_{r=R}=0\Rightarrow 0=\left.\frac{\mathrm{d}P}{\mathrm{d}r}\right|_{r=R}=kJ'_m(kR), \tag{6}$$

the roots j_{mn} of the derivative J'_m of the Bessel functions specify the axial wave numbers k_{mn}

$$J'_m(j_{mn})=0\Rightarrow k_{mn}=\frac{j_{mn}}{R}, \tag{7}$$

and hence the eigenfrequencies ω_{lmn}

$$\omega_{lmn}=\frac{2\pi lU}{L}\pm c\sqrt{(2\pi l/L)^2+(j_{mn}/R)^2}. \tag{8}$$

The eigenfrequencies are specified by the axial, circumferential, and radial wave numbers, respectively, l, m, and k_{mn}, and have a Doppler shift due to the axial mean flow.

B. Wave modes with a complex impedance wall boundary condition

The eigenfrequencies are given in all cases by

$$\omega_{lmn}=\frac{2\pi lU}{L}\pm c\sqrt{(2\pi l/L)^2+k_{lmn}^2}, \tag{9}$$

where the radial wave number k is determined by the appropriate boundary condition, e.g., Eq. (7) in the case of a rigid wall, and can depend on l, m, and n. Note that the mean flow velocity U, which is axial, affects the wave frequency Eq. (9) as a Doppler shift $(2\pi l/L)U$, with $2\pi l/L$ playing the role of axial wave number. In all cases, including walls with complex impedance, uniform or nonuniform, the total acoustic pressure Eq. (2) is a superposition of eigenfunctions Eq. (5), *viz.*

$$p(r,\theta,z,t)=\sum_{m=-\infty}^{+\infty}\mathrm{e}^{im\theta}\sum_{l=-\infty}^{+\infty}\mathrm{e}^{i2\pi lz/L}\times\sum_{n=1}^{\infty}\mathrm{e}^{-i\omega_{lmn}t}A_{lmn}J_m(k_{lmn}r), \tag{10}$$

with amplitudes

$$A_{lmn}=\frac{1}{\pi L}\left[\left(1-\frac{m^2}{k_{lmn}^2}\right)\{J_m(k_{lmn}R)\}^2\right]\times\int_0^{2\pi}e^{-im\theta}d\theta\int_0^L e^{-i2\pi lz/L}dz\times\int_0^R rJ_m(k_{lmn}r)p(r,\theta,z,0)dr, \quad (11)$$

determined by the initial acoustic pressure $p(r,\theta,z,0)$ at time $t=0$. The arbitrary constants could also be determined by the acoustic pressure at the nozzle entrance for all time. Regardless of whether the mode amplitudes A_{lmn} in Eq. (11) are determined by (a) the initial acoustic pressure in the tube or (b) the boundary conditions at the tube end(s), this will not affect the calculation of eigenvalues, which is presented next. For general complex impedance wall boundary conditions, both k_{lmn} and ω_{lmn} are complex

$$\exp(-i\omega_{lmn}t)=\exp[-i\,\mathrm{Re}(\omega_{lmn})t]\exp[\mathrm{Im}(\omega_{lmn})t], \quad (12)$$

and thus the acoustic fields decay with time if $\mathrm{Im}(\omega_{lmn})<0$. Thus, the complex wall impedance, whether uniform or not, should be chosen so that the imaginary part of the eigenfrequency is negative, and as large in modulus as possible. This will be shown first for uniform wall impedance, before proceeding to one- and two-dimensional nonuniform impedance distributions over the duct wall.

The acoustic pressure p is related to the radial acoustic velocity v_r through the radial component of the momentum equation

$$\left(\frac{\partial}{\partial t}+U\frac{\partial}{\partial z}\right)v_r(r,\theta,z,t)+\frac{1}{\rho}\frac{\partial}{\partial r}p(r,\theta,z,t)=0, \quad (13)$$

or in terms of eigenfunctions by

$$iV_{lm}(r,\omega)=\frac{1}{\rho(\omega-2\pi lU/L)}\frac{dP_{lm}(r;\omega)}{dr}. \quad (14)$$

Using Eq. (9) together with the definition (Rayleigh, 1945) of acoustic impedance as minus the ratio of acoustic pressure and velocity spectra

$$\pm\left[\frac{i}{\rho c}\frac{1}{\sqrt{k^2+(2\pi l/L)^2}}\right]\frac{dP_{lm}(R;\omega)}{dr}=-V_{lm}(R;\omega)=\frac{P_{lm}(R;\omega)}{\overline{Z}}, \quad (15)$$

leads to

$$\frac{iZ}{\sqrt{1+(2\pi l/kL)^2}}J_m'(kR)=J_m(kR), \quad (16)$$

where prime denotes the derivative of the Bessel function with respect to its argument kr, and the dimensionless specific impedance Z is defined by

$$Z=\frac{\overline{Z}}{\rho c}; \quad (17)$$

note that the specific impedance Z equals the ratio of the liner impedance $\overline{Z}$ to the impedance ρc of a plane wave, and thus would be unity in the latter case. In the case of a rigid wall, or infinite impedance, then Eq. (16) simplifies to Eq. (6). In the case of finite uniform complex wall impedance $Z=\mathrm{const}$, the roots $n=1,2,...,\infty$ of Eq. (16) specify the radial wave numbers k_{lmn}, which are generally complex; thus, the eigenfrequencies Eq. (9) are also complex, and their imaginary part should be made as negative as possible to have Eq. (12) a marked decay of the acoustic field with time.

III. ACOUSTIC EIGENFREQUENCIES FOR ONE- AND TWO-DIMENSIONAL IMPEDANCE DISTRIBUTIONS

The consideration of nozzles with nonuniform wall impedance leads to acoustic eigenvalues which are determined by roots of an infinite determinant, in the case of one-dimensional impedance distribution, i.e., impedance varying circumferentially (Sec. A) or axially (Sec. B). In the case of a two-dimensional impedance distribution, i.e., impedance varying both axially and circumferentially, the acoustic eigenvalues are the roots (Sec. C) of a doubly infinite determinant, i.e., an infinite determinant whose elements are infinite determinants.

It is usual in duct acoustics to use complex axial wave numbers which, like the complex frequencies in Eq. (12), specify the axial decay of the modes. Two cases arise: (a) if the wall impedance is uniform or varies only circumferentially, then the acoustic fields are sinusoidal in the axial direction, and an axial wave number does exist; (b) if the wall impedance varies axially, regardless of whether or not it varies circumferentially, then the acoustic fields are specified by a Fourier series in the axial direction, with real axial wave numbers $2\pi l/L$, and it is necessary to consider complex frequencies as in Eq. (9). Thus, the general approach, which applies in both cases, is to use complex frequencies, as is done in the sequel.

A. Duct with a circumferentially nonuniform wall impedance distribution

The impedance wall boundary condition can be written in general

$$\tilde{p}(R,\theta,z;\omega)=\overline{Z}(\theta,z;\omega)\tilde{v}_r(R,\theta,z;\omega), \quad (18)$$

where $\overline{Z}$ may vary axially and circumferentially and where $\tilde{p}$, $\tilde{v}_r$ are the spectra of the acoustic pressure perturbation and radial velocity perturbation

$$p(r,\theta,z,t)=\int_{-\infty}^{+\infty}d\omega e^{-i\omega t}\tilde{p}(r,\theta,z;\omega), \quad (19a)$$

$$v_r(r,\theta,z,t)=\int_{-\infty}^{+\infty}d\omega e^{-i\omega t}\tilde{v}_r(r,\theta,z;\omega). \quad (19b)$$

The radial acoustic velocity is related to the pressure by the radial component of the linearized momentum equation Eq. (13) which, by use of Eqs. (19), can be written as

$$\rho\left(-i\omega+U\frac{\partial}{\partial z}\right)\tilde{v}_r+\frac{\partial\tilde{p}}{\partial r}=0. \quad (20)$$

Consider first a circumferentially varying but axially uniform wall-specific impedance $Z(\theta;\omega)$, which is an arbitrary periodic function with bounded variation, and thus can be represented by a Fourier series

$$Z(\theta;\omega)=\sum_{m'=-\infty}^{+\infty} Z_{m'}(\omega)\mathrm{e}^{im'\theta}, \tag{21}$$

with coefficients

$$Z_{m'}(\omega)=\frac{1}{2\pi}\int_0^{2\pi} Z(\theta)\mathrm{e}^{-im'\theta}\mathrm{d}\theta. \tag{22}$$

The coefficients $Z_{m'}$ may be interpreted as the amplitudes of the circumferential harmonics m' in Eq. (21) of the impedance.

From Eqs. (2) and (5), it follows that

$$\tilde{p}(r,\theta,z;\omega)=\sum_{m=-\infty}^{+\infty}\mathrm{e}^{im\theta}\sum_{l=-\infty}^{+\infty}\mathrm{e}^{i2\pi lz/L}\sum_{n=1}^{\infty}A_{lmn}J_m(kr), \tag{23}$$

and, due to the linearity of the problem, it can be assumed that the spectra of the radial velocity perturbation can also be represented by a Fourier series in z and θ

$$\tilde{v}_r(r,\theta,z;\omega)=\sum_{l=-\infty}^{+\infty}\mathrm{e}^{i2\pi lz/L}\sum_{m=-\infty}^{+\infty}\mathrm{e}^{im\theta}V_{lm}(r;\omega). \tag{24}$$

Substitution of Eqs. (23) and (24) in Eq. (20) leads to

$$V_{lm}(r;\omega)=\frac{1}{i\left(\frac{\omega}{c}-\frac{2\pi l}{L}M\right)\rho c}kA_{lmn}J'_m(kr), \tag{25}$$

where $M=U/c$ is the Mach number. This expression for V_{lm}, together with Eqs. (21), (23), and (24), can now be used in Eq. (18), yielding

$$\sum_{l=-\infty}^{+\infty}\mathrm{e}^{i2\pi lz/L}\left[\sum_{m=-\infty}^{+\infty}A_m\mathrm{e}^{im\theta}J_m(kR) \mp\sum_{m,m'=-\infty}^{+\infty}\frac{1}{i\alpha_l}A_mZ_{m'}\mathrm{e}^{i(m+m')\theta}J'_m(kR)\right]=0, \tag{26}$$

where Eq. (4) was used to eliminate ω. The parameter α_l is given by

$$\alpha_l\equiv\sqrt{1+(2\pi l/kL)^2}, \tag{27}$$

and A_{lmn} is denoted A_m for simplicity, since only the circumferential wave number m is summed in Eq. (26). Note that the Mach number of the axial mean flow which appears in Eq. (25) drops on account of Eq. (9) from the parameter α_l in Eq. (27); thus, the radial wave numbers, which are the roots of Eq. (26), are not affected by a uniform axial flow, although the frequency is, by the Doppler shift in Eq. (9). Each coefficient of the sum in l must vanish. Rearrangement of terms and redefinition of the summation indexes leads to

$$\sum_{m=-\infty}^{+\infty}\mathrm{e}^{im\theta}\left[A_mJ_m(kR)\mp\sum_{m'=-\infty}^{+\infty}\frac{Z_{m'}}{i\alpha_l}A_{m-m'}J'_{m-m'}(kR)\right]=0. \tag{28}$$

This equation must hold for arbitrary values of θ, which leads to

$$m\in\mathbb{Z}:\quad\sum_{m'=-\infty}^{+\infty}\left[J_{m'}(kR)\,\delta_{mm'}\mp\frac{Z_{m-m'}}{i\alpha_l}J'_{m'}(kR)\right]A_{m'}=0, \tag{29}$$

where $\delta_{mm'}$ is the Kronecker delta, i.e., 1 for $m'=m$ and 0 for $m'\neq m$, *viz.* the infinite identity matrix. This is a homogeneous system of linear equations in the amplitudes A_m of the eigenfunctions. Since the A_m for $m=0,\pm1,\pm2,...,\pm\infty$ cannot be all zero, the determinant of the coefficients in Eq. (29) must vanish

$$D\equiv\begin{vmatrix}
\ddots & \vdots & \vdots & \vdots & \vdots \\
\cdots & \left[J_{-1}(kR)\mp\frac{Z_0}{i\alpha_l}J'_{-1}(kR)\right] & \mp\frac{Z_{-1}}{i\alpha_l}J'_0(kR) & \mp\frac{Z_{-2}}{i\alpha_l}J'_1(kR) & \cdots \\
\cdots & \mp\frac{Z_1}{i\alpha_l}J'_{-1}(kR) & \left[J_0(kR)\mp\frac{Z_0}{i\alpha_l}J'_0(kR)\right] & \mp\frac{Z_{-1}}{i\alpha_l}J'_1(kR) & \cdots \\
\cdots & \mp\frac{Z_2}{i\alpha_l}J'_{-1}(kR) & \mp\frac{Z_1}{i\alpha_l}J'_0(kR) & \left[J_1(kR)\mp\frac{Z_0}{i\alpha_l}J'_1(kR)\right] & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots
\end{vmatrix}=0. \tag{30}$$

Note that D depends on k through the arguments kR of the Bessel functions and also through kL in α_l in Eq. (27), which also involves l. Also, if the impedance depends on the frequency $Z_l(\omega)$, then by use of (4) it can be expressed in terms of k. Thus, the roots of

$$D(kR,\alpha_l,Z_m)=0 \tag{31}$$

specify the radial wave numbers k_{ln}, which depend on l and are ordered by n. They in turn specify the eigenfrequencies ω_{ln} in Eq. (9), which in this case do not depend on m.

B. Nozzle with an axially varying wall impedance distribution

Consider next a complex wall impedance varying only in the axial direction but not circumferentially, which is represented by the Fourier series

$$Z(z;\omega)=\sum_{l'=-\infty}^{+\infty} Z^{l'}(\omega)\mathrm{e}^{i2\pi l' z/L}, \tag{32}$$

with coefficients given by

$$Z^{l'}(\omega)=\frac{1}{L}\int_0^L Z(z;\omega)\mathrm{e}^{-i2\pi l' z/L}\mathrm{d}z. \tag{33}$$

The coefficients $Z^{l'}$ may be interpreted as the amplitudes of the axial harmonics l' in Eq. (32) of the impedance.

In this case substitution of Eqs. (23), (24), (25), and (32) in Eq. (18) yields

$$\sum_{m=-\infty}^{+\infty}\mathrm{e}^{im\theta}\left[\sum_{l=-\infty}^{+\infty}\mathrm{e}^{i2\pi lz/L}A_{lm}J_m(kR)\right.$$
$$-\sum_{l'=-\infty}^{+\infty}\sum_{l=-\infty}^{+\infty}\mathrm{e}^{i2\pi(l+l')z/L}Z^{l'}$$
$$\left.\times\frac{k}{(\pm)i\sqrt{k^2-\left(\frac{2\pi l}{L}\right)^2}}A_{lm}J'_m(kR)\right]=0. \tag{34}$$

Each coefficient of the sum in m must vanish. Rearrangement of terms and redefinition of the summation indexes leads to

$$\sum_{l=-\infty}^{+\infty}\mathrm{e}^{i2\pi lz/L}\left\{\sum_{l'=-\infty}^{+\infty}\left[A_{lm}J_m(kR)\mp Z^{l-l'}\right.\right.$$
$$\left.\left.\times\frac{k}{i\sqrt{k^2-\left(\frac{2\pi l'}{L}\right)^2}}A_{l'm}J'_m(kR)\right]\right\}=0. \tag{35}$$

The coefficients of the Fourier series in l must now vanish. Using the Kronecker delta, these terms can be written as

$$\sum_{l'=-\infty}^{+\infty}A_{l'm}\left[\delta_{l'l}J_m(kR)\mp Z^{l-l'}\frac{1}{i\alpha_{l'}}J'_m(kR)\right]=0. \tag{36}$$

These are a set of equations (one for each value of l and m) from which the A_{lm} can be found. Not all the A_{lm} can vanish simultaneously. For Eqs. (36) to hold, it is necessary that the (infinite) determinant of the coefficients vanishes

$$E_m=\det\left\{\delta_{l'l}J_m(kR)\mp Z^{l-l'}\frac{1}{i\alpha_{l'}}J'_m(kR)\right\}=0. \tag{37}$$

The roots of Eq. (37) specify the radial wave numbers k_{mn} and hence, by Eq. (9), the eigenfrequencies ω_{mn}.

C. General case of impedance distribution varying both axially and circumferentially

In the most general case of complex wall impedance varying both axially and circumferentially, it is represented by a double Fourier series

$$Z(\theta,z;\omega)=\sum_{l'=-\infty}^{+\infty}\mathrm{e}^{i2\pi l' z/L}\sum_{m'=-\infty}^{+\infty}\mathrm{e}^{im'\theta}Z^{l'}_{m'}(\omega), \tag{38}$$

with coefficients

$$Z^{l'}_{m'}(\omega)=\frac{1}{2\pi L}\int_0^{2\pi}\mathrm{d}\theta\mathrm{e}^{-im'\theta}\int_0^L\mathrm{d}z\mathrm{e}^{-i2\pi l' z/L}Z(\theta,z;\omega). \tag{39}$$

The coefficients $Z^{l'}_{m'}$ may be interpreted as the amplitudes of the circumferential m' and axial l' harmonics of the impedance Eq. (38). Substitution of Eqs. (23), (24), (25), and (38) in Eq. (18) leads to

$$\sum_{m'=-\infty}^{+\infty}\sum_{l'=-\infty}^{+\infty}\left[J_{m'}(kR)\delta_{mm'}\delta_{ll'}\right.$$
$$\left.\mp\frac{1}{i\alpha_{l'}}Z^{l-l'}_{m-m'}J'_{m'}(kR)\right]A_{l'm'}=0, \tag{40}$$

for all m, $l\in\mathbb{Z}$. Since the A_{lm} with l, $m=0,\pm1,\pm2,...,\pm\infty$ cannot all vanish, the determinant of the coefficients must vanish

$$0=F(kR,\alpha_l,Z^l_m)=F(k_n)$$
$$=\det\left[J_{m'}(kR)\delta_{mm'}\delta_{ll'}\mp\frac{1}{i\alpha_{l'}}Z^{l-l'}_{m-m'}J'_{m'}(kR)\right], \tag{41}$$

thus specifying the radial wave numbers k_n and, through Eq. (9), the eigenfrequencies ω_n.

The condition Eq. (41) specifies the vanishing of a doubly infinite determinant, i.e., an infinite determinant in m' whose terms are infinite determinants in l', as follows: (i) take l fixed in Eq. (40), so that like Eq. (29) it specifies the infinite determinants similar to Eq. (30), *viz.*

$$D^{l,l'} \equiv \begin{vmatrix} \ddots & \vdots & \vdots & \vdots \\ \cdots & \left[J_{-1}(kR) + \dfrac{Z_0^{l-l'}}{i\alpha_{l'}} J'_{-1}(kR) \right] & \dfrac{Z_{-1}^{l-l'}}{i\alpha_{l'}} J'_0(kR) & \cdots \\ \cdots & \dfrac{Z_1^{l-l'}}{i\alpha_{l'}} J'_{-1}(kR) & \left[J_0(kR) + \dfrac{Z_0^{l-l'}}{i\alpha_{l'}} J'_0(kR) \right] & \cdots \\ \cdots & \dfrac{Z_2^{l-l'}}{i\alpha_{l'}} J'_{-1}(kR) & \dfrac{Z_1^{l-l'}}{i\alpha_{l'}} J'_0(kR) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix}, \tag{42}$$

of which Eq. (30) is $D^{0,0}=D$; (ii) the infinite determinant of Eq. (42) is the double infinite determinant Eq. (41)

$$F(kR,\alpha,Z_m^l) = \det[D^{l-l'}(kR,\alpha,Z_m^l)]. \tag{43}$$

This doubly infinite determinant could be calculated in the reverse way, by: (i) taking m fixed in Eq. (40), so that it specifies the infinite determinants similar to Eq. (37), *viz.*

$$E_{m,m'} = \det\left\{ J_{m'}(kR)\,\delta_{mm'}\delta_{ll'} - \frac{Z_{m-m'}^{l-l'}}{i\alpha_{l'}} J'_{m'}(kR) \right\} = 0, \tag{44}$$

of which Eq. (37) is the case $m'=m$, i.e., $E_m = E_{m,m}$; (ii) the infinite determinant of Eq. (44)

$$F(kR,\alpha_l,Z_m^l) = \det[E_{m,m'}(kR,\alpha_l,Z_m^l)], \tag{45}$$

is the double-infinite determinant Eq. (41). The evaluation of double-infinite determinants involves two truncations, one for each infinite determinant, as discussed in more detail elsewhere (Campos and Oliveira, 2004).

IV. COUPLING OF ACOUSTIC MODES THROUGH THE NONUNIFORM IMPEDANCE

In the case of uniform impedance, the infinite determinants are diagonal, and thus there is no coupling of the radial acoustic modes n for different circumferential m and axial m orders (Sec. A); the infinite determinants cease to be diagonal for nonuniform impedance, showing that a circumferentially/axially nonuniform impedance couples the radial modes n for different circumferential m/axial l orders respectively (Sec. C) besides splitting modes, so that a nonuniformly lined nozzle has more modes than a uniformly lined one.

A. Exact calculation of radial wave numbers and eigenfrequencies

As an example of the calculation of eigenvalues, consider the central term in Eq. (30), written in the form

$$i\alpha_l J_0(kR) + Z_0 J'_0(kR) = 0. \tag{46}$$

This expression involves two dimensionless parameters: radial wave number made dimensionless multiplying by the radius R

$$\kappa \equiv kR, \tag{47}$$

and the "aspect ratio" of the duct, i.e., the ratio of length to radius

$$\mu \equiv \frac{L}{R}. \tag{48}$$

Both appear in the parameter α_l defined by Eq. (27)

$$\alpha_l = \sqrt{1 + \left(\frac{2\pi l}{\mu\kappa} \right)^2}. \tag{49}$$

Thus, for each value of the aspect ratio of the nozzle μ, the dimensionless radial wave number is the root of

$$i\sqrt{1 + \left(\frac{2\pi l}{\mu\kappa} \right)^2} J_0(\kappa) - Z_0 J_1(\kappa) = 0, \tag{50}$$

where $J'_0(x) = -J_1(x)$ was used. For each of four aspect ratios $\mu=1$, 2, 5, 10, the wave numbers were searched in the interval

$$-10 < \mathrm{Re}(\kappa) < 10, \tag{51a}$$

$$-10 < \mathrm{Im}(\kappa) < 10, \tag{51b}$$

to exclude high-order modes.

As an example, the dimensionless radial eigenfrequencies [Eq. (47)] for a cylindrical nozzle with four aspect ratios [Eq. (48)], *viz.*

$$\mu = 1,2,5,10, \tag{52}$$

are indicated in Table I in the case of uniform impedance

$$l=1: \quad \text{case } 1: \quad Z_0 = 1+i, \tag{53}$$

and first circumferential mode $l=1$. In this case the eigenvalues are specified by the roots of the central term [Eq. (50)]

TABLE I. Eigenvalues for the dimensionless radial wave number Eq. (47), lying in the intervals Eq. (51a), Eq. (51b), for the central term Eq. (46) of the determinant Eq. (30), for each of the four values Eq. (52) of the aspect ratio Eq. (48), in the case 1 of uniform wall impedance Eq. (53).

$\mu=1$	$\mu=2$	$\mu=5$	$\mu=10$
$-9.522+i0.4161$	$-9.601+i0.3934$	$-9.625+i0.3845$	$-9.628+i0.3831$
$-6.266+i0.4261$	$-6.416+i0.3963$	$-6.466+i0.3780$	$-6.473+i0.3748$
$-2.807+i0.3562$	$-3.111+i0.4123$	$-3.270+i0.3662$	$-3.295+i0.3544$
$-1.979+i3.9733$	$-0.835+i2.2174$	$-0.219+i1.1000$	$-0.051+i0.6139$
$2.036+i0.2886$	$1.762+i0.4127$	$1.380+i0.5368$	$1.204+i0.6241$
$4.797+i0.4441$	$4.560+i0.4663$	$4.441+i0.4577$	$4.420+i0.4549$
$7.764+i0.4541$	$7.637+i0.4416$	$7.593+i0.4328$	$7.586+i0.4312$

of the determinant [Eq. (30)]. The radial wave numbers, shown in Table I, are ordered by increasing real part.

B. Additional eigenmodes due to nonuniform impedance distribution

In the preceding case of uniform wall impedance $0=Z_{\pm 1}=Z_{\pm 2}=\ldots$, the determinant Eq. (30) is diagonal and its roots correspond to the vanishing of the diagonal terms

$$m=0,1,2,...: \quad i\alpha_l J_{\pm m}(kR)+Z_0 J'_{\pm m}(kR)=0, \tag{54}$$

for all $m=0,\pm 1,\pm 2,...$, of which Eq. (50)≡Eq. (46) is just the case $m=0$. The first 13 eigenvalues are indicated in Tables II and III in the case 1 of uniform impedance. Case 2 introduces a circumferentially nonuniform wall impedance

$$\text{case } 2: \quad Z=Z_0+Z_1\cos\theta, \tag{55}$$

with amplitude of the first harmonic

$$\text{subcase } 2a: \quad Z_1=Z_{-1}=0.1+i0.1, \tag{56a}$$

$$\text{subcase } 2b: \quad Z_1=Z_{-1}=0.5+i0.5, \tag{56b}$$

larger in subcase 2b than in subcase 2a. In case 2 the determinant Eq. (30) is three banded, coupling all modes. Case 3 introduces a second harmonic in the circumferentially nonuniform impedance

$$\text{case } 3: \quad Z=Z_0+Z_1\cos\theta+Z_2\cos 2\theta, \tag{57}$$

larger in subcase 3b than in subcase 3a:

$$\text{subcase } 3a: \quad Z_2=Z_{-2}=0.01+i0.01, \tag{58a}$$

$$\text{subcase } 3b: \quad Z_2=Z_{-2}=0.1+i0.1, \tag{58b}$$

In case 3 the determinant in Eq. (30) is five-banded, coupling all modes more strongly than in case 2. Note that the amplitudes of the fundamental Z_0 and harmonics $Z_{\pm 1}$, $Z_{\pm 2}$,... of the impedance must be such that the real part of the impedance is positive $\mathrm{Re}(Z)>0$ at all positions $0\leqslant\theta<2\pi$.

The eigenvalues indicated in Tables II and III for cases 2 and 3 were calculated as the roots of the determinant Eq. (30) truncated as a 3×3 determinant, i.e., with the terms explicitly written in Eq. (30). The effects of nonuniform wall impedance on changing the eigenvalues are more marked in the subcases 2b and 3b in Table III than in the subcases 2a and 3a in Table II, because the harmonics of the impedance has a larger magnitude in modulus. The more striking feature of Tables II and III is the observation that a nonuniform wall impedance causes the appearance of additional eigenmodes: the 13 eigenvalues in the case 1 of uniform wall impedance split into 19 eigenvalues in cases 2 and 3 of nonuniform wall impedance. The appearance of these six additional eigenmodes in the same range is due to the interaction between the harmonics of the acoustic field and the harmonics of the nonuniform wall impedance.

The appearance of additional acoustic modes, as a consequence of the interaction with the circumferentially nonuniform impedance, can be explained as follows, starting from the determinant Eq. (30) truncated in the form $M\times M$ (*viz.* 3×3 in the case of Tables II and III): (i) in the case of uniform wall impedance, the $2M+1$ conditions Eq. (54) for $m=0,\pm 1,\pm 2,...,\pm M$ actually reduce to only $M+1$ condi-

TABLE II. Eigenvalues of Eq. (30) for the dimensionless radial wave number Eq. (47) lying in the intervals Eq. (51a), Eq. (51b) for each of the four values Eq. (52) of the aspect ratio Eq. (48) in case 1 of uniform wall impedance, case 2a of first Eq. (56a) and case 3a of second Eq. (58a) harmonic of the circumferentially nonuniform wall impedance.

$\mu=1$					
Case 1		Case 2a		Case 3a	
Re	Im	Re	Im	Re	Im
−9.521 94	0.416 13	−9.512 18	0.411 56	−9.512 22	0.411 63
−7.835 44	0.425 65	−7.835 44	0.425 65	−7.834 28	0.419 30
		−7.827 53	0.419 78	−7.828 58	0.426 24
−6.265 59	0.426 12	−6.258 12	0.419 51	−6.258 13	0.419 58
−4.434 63	0.425 55	−4.434 63	0.425 55	−4.439 30	0.420 33
		−4.432 60	0.418 77	−4.427 79	0.423 92
−2.807 34	0.356 20	−2.807 10	0.351 55	−2.807 08	0.351 56
−2.090 80	3.928 34	−2.090 80	3.928 34	−2.090 70	3.963 16
		−1.994 27	3.513 20	−1.995 74	4.458 95
−1.979 19	3.973 27	−1.994 04	4.478 11	−1.994 24	3.498 17
2.036 24	0.288 63	2.035 25	0.288 74	2.035 25	0.288 74
3.278 02	0.397 92	3.278 02	0.397 92	3.273 15	0.394 81
		3.278 66	0.392 75	3.283 59	0.395 82
4.796 96	0.444 07	4.799 57	0.437 25	4.799 59	0.437 27
6.193 78	0.461 83	6.193 78	0.461 83	6.194 34	0.454 25
		6.201 53	0.454 00	6.201 11	0.461 65
7.764 46	0.454 13	7.772 99	0.447 40	7.772 98	0.447 48
9.243 63	0.451 74	9.243 63	0.451 74	9.247 48	0.445 18
		9.254 57	0.446 09	9.250 81	0.452 80
		$\mu=2$			
−9.600 82	0.393 46	−9.590 35	0.391 06	−9.590 42	0.391 12
−7.944 93	0.398 51	−7.944 93	0.398 51	−7.941 48	0.393 92
		−7.935 37	0.395 27	−7.938 80	0.399 98
−6.416 07	0.396 37	−6.405 98	0.393 22	−6.406 04	0.393 28
−4.668 33	0.413 82	−4.668 33	0.413 82	−4.666 73	0.408 02
		−4.660 22	0.408 76	−4.661 72	0.414 67
−3.110 64	0.412 32	−3.102 79	0.406 23	−3.102 81	0.406 30
−1.156 69	2.092 95	−1.156 69	2.092 95	−1.152 16	2.110 02
		−1.064 16	1.985 13	−1.071 91	1.974 36
−0.835 04	2.217 43	−0.889 44	2.357 75	−0.886 81	2.351 78
1.761 76	0.412 69	1.760 51	0.410 33	1.760 52	0.410 33
2.980 94	0.496 23	2.980 94	0.496 23	2.981 61	0.487 72
		2.989 07	0.487 00	2.988 53	0.495 58
4.559 54	0.466 33	4.569 18	0.459 84	4.569 16	0.459 92
6.013 85	0.461 88	6.013 85	0.461 88	6.019 55	0.455 88
		6.026 79	0.457 23	6.021 13	0.463 42
7.637 41	0.441 64	7.648 94	0.437 72	7.648 92	0.437 80
9.147 66	0.438 13	9.147 66	0.438 13	9.153 61	0.433 44
		9.160 38	0.435 20	9.154 43	0.440 07
		$\mu=5$			
−9.624 88	0.384 52	−9.614 40	0.382 81	−9.614 48	0.382 86
−7.979 48	0.385 85	−7.979 48	0.385 85	−7.975 59	0.382 00
		−7.969 77	0.383 54	−7.973 65	0.387 51
−6.466 02	0.377 97	−6.455 84	0.376 20	−6.455 91	0.376 25
−4.757 14	0.383 74	−4.757 14	0.383 74	−4.754 01	0.379 57
		−4.748 18	0.380 88	−4.751 29	0.385 16
−3.269 49	0.366 15	−3.260 00	0.363 98	−3.260 06	0.364 03
−0.948 99	0.886 87	−0.948 99	0.886 87	−0.935 21	0.869 05
		−0.926 40	0.875 42	−0.932 08	0.893 72
−0.219 00	1.109 97	−0.214 90	1.122 06	−0.214 94	1.121 85
1.380 25	0.536 78	1.384 41	0.529 83	1.384 43	0.529 87
2.737 52	0.546 44	2.737 52	0.546 44	2.746 15	0.547 66
		2.754 89	0.539 45	2.746 32	0.538 46
4.440 81	0.457 73	4.453 48	0.454 08	4.453 40	0.454 17
5.941 18	0.452 40	5.941 18	0.452 40	5.948 15	0.454 62
		5.955 34	0.450 09	5.948 34	0.448 07
7.592 87	0.432 79	7.605 07	0.430 24	7.604 99	0.430 32
9.116 20	0.431 23	9.116 20	0.431 23	9.122 62	0.427 33
		9.129 21	0.429 35	9.122 77	0.433 43

TABLE II. (*Continued*.)

$\mu=1$					
Case 1		Case 2a		Case 3a	
Re	Im	Re	Im	Re	Im
		$\mu=10$			
−9.628 39	0.383 14	−9.617 92	0.381 52	−9.618 00	0.381 57
−7.984 56	0.383 82	−7.984 56	0.383 82	−7.980 63	0.380 08
		−7.974 85	0.381 65	−7.978 78	0.385 51
−6.473 46	0.374 87	−6.463 30	0.373 30	−6.463 37	0.373 35
−4.770 67	0.377 94	−4.770 67	0.377 94	−4.767 40	0.374 06
		−4.761 70	0.375 44	−4.764 96	0.379 43
−3.294 74	0.354 43	−3.285 37	0.352 87	−3.285 44	0.352 92
−1.018 18	0.595 14	−1.018 18	0.595 14	−1.017 09	0.586 21
		−1.008 67	0.588 83	−1.009 65	0.597 95
−0.050 92	0.613 92	−0.048 96	0.615 13	−0.048 99	0.615 12
1.204 00	0.624 06	1.213 67	0.614 30	1.213 69	0.614 39
2.677 44	0.557 63	2.677 44	0.557 63	2.686 30	0.559 41
		2.697 13	0.552 14	2.688 29	0.550 65
4.420 02	0.454 88	4.433 07	0.451 88	4.432 99	0.451 97
5.929 57	0.450 37	5.929 57	0.450 37	5.936 48	0.452 69
		5.943 86	0.448 47	5.936 91	0.446 35
7.586 10	0.431 23	7.598 36	0.428 91	7.598 28	0.428 98
9.111 52	0.430 10	9.111 52	0.430 10	9.117 99	0.426 31
		9.124 56	0.428 38	9.118 05	0.432 34

tions $m=0,1,2,...,M$, because the case $+m$ coincides with $-m$, e.g., there are in the range [Eq. (51a), Eq. (51b)] seven eigenmodes which are the roots of Eq. (50) indicated in Table I, plus six eigenmodes which are roots of Eq. (54) with $m=\pm 1$, for a total of 13 in the first column of Tables II and III; (ii) in the case of nonuniform wall impedance, the determinant Eq. (30) in general no longer has double roots, so the number of eigenmodes increases, i.e., the seven single roots and six double roots in the case of uniform wall impedance (first column of Tables II and III) give rise to $7+(6\times 2)=19$ distinct roots in the case of nonuniform wall impedance (remaining columns of Tables II and III).

C. Interaction of harmonics of impedance and harmonics of sound

Figure 1 shows the eigenvalues for the duct $L=R$ of aspect ratio unity $\mu=1$, in the cases a of weaker and b of stronger harmonics of the wall impedance. It is seen that in case: (a) the weak harmonics of the wall impedance almost does not change the eigenvalues; (b) the change of eigenvalues is small but more visible in the case of larger harmonics of the wall impedance. In both cases of small (a) or larger (b) harmonics of the wall impedance, they cause the appearance of additional eigenmodes, which did not exist for uniform wall impedance in the sense that they were double roots which became distinct single roots. This phenomenon could be called eigenmode splitting by the nonuniform wall impedance. The full set of 22 split eigenmodes is shown in Fig. 2 for the cases of two weak (3a) or stronger (3b) harmonics of the wall impedance and several aspect ratios, Eq. (52). It is seen that the wider duct $\mu=1$ or $L=R$ allow radial wave numbers with larger imaginary parts, and hence faster decay. As the duct becomes narrower, i.e., the diameter becomes smaller relative to the length, the radial wave numbers have smaller imaginary parts, implying a slower spatial decay. Given the initial amplitudes of each eigenmode, their decay specifies how effective the nonuniform impedance liner is in reducing the acoustic energy. An assessment of the noise reduction potential of a nonuniform acoustic liner, and its optimization, requires an analysis beyond the preceding remarks, including (Campos and Oliveira, 2004) evaluation for various modes of the way their spatial and/or temporal decay is affected by the impedance distribution.

The preceding analysis could be further elaborated in several directions. The radial eigenvalues calculated from the truncated 3×3 form of the infinite determinant Eq. (30) could be checked for accuracy comparing with larger 4×4 and 5×5 truncation; an alternative, less reliable measure of accuracy, would be to start with the 3×3 truncation of Eq. (30), and compare the roots of the upper and lower 2×2 determinants along the diagonal. In the case just discussed, when the impedance varies only circumferentially and not axially, there is an axial wave number, k_z, i.e., the discrete spectrum $\exp(i2\pi lz/L)$ could be replaced by a continuous spectrum $\exp(ik_z z)$, with $k_z=2\pi l/L$; in this case the frequency ω could be taken as real in Eq. (9) and the complex radial wave numbers k_{lmn} would specify complex axial wave numbers $(k_z)_{lmn}$ as roots of

$$\left|[(k_z)_{lmn}]^2+(k_{lmn})^2\right|^{1/2}=\omega-(k_z)_{lmn}U. \tag{59}$$

As in Eq. (12), the imaginary part of the axial wave number

$$\exp[i(k_z)_{lmn}z]=\exp\{iz\,\mathrm{Re}[(k_z)_{lmn}]\}\cdot\exp\{-z\,\mathrm{Im}[(k_z)_{lmn}]\}, \tag{60}$$

would specify the decay of the wave mode along the length of the duct.

One way to optimize the acoustic liner for sound attenuation is to maximize $\mathrm{Im}[(k_z)_{lmn}]$ in Eq. (60) or make $\mathrm{Im}(\omega_{lmn})$ in Eq. (12) as large a negative value as possible; the optimization could concern just one mode or a combination of them. If the liner impedance varies in the axial direction, then a longitudinal wave number k_z does not exist, and there is no alternative to the formulation Eq. (12) in terms of natural frequencies ω_{lmn}. In the case of sound generation by a source of frequency ω, all the eigenfunctions in Eq. (10) would be forced to oscillate at frequency ω, rather than ω_{lmn}, and thus the time dependence would be $e^{-i\omega t}$. The amplitudes A_{lmn} would depend on the amplitudes of the forcing, and also on the difference between the forcing frequency ω and the natural frequencies ω_{lmn}. If the forcing had constant amplitude, i.e., real frequency ω, since ω_{lmn} are complex for a nonuniform liner, there are phase shifts between the forcing and the response. A resonance $\omega=\omega_{lmn}$ would only be possible by a forcing decaying in time, such that the frequency $\mathrm{Re}(\omega)=\mathrm{Re}(\omega_{lmn})$ and rate of decay in time $\mathrm{Im}(\omega)=\mathrm{Im}(\omega_{lmn})$ would match those of a natural mode. In the case of a cylindrical annulus $r_1<r<r_2$, the solution Eq. (10) would be replaced by a linear combination of Bessel J_m and Neumann Y_m functions, with amplitudes A_{lmn} and B_{lmn}, respectively, which can be specified by applying boundary conditions in the z direction, e.g., at the duct entry $z=0$ and at the exit $z=L$ to the section with acoustic lining. Indepen-

TABLE III. As in Table II, in the case 2b of first Eq. (56b) and case 3b of second Eq. (58b) harmonic of the circumferentially nonuniform wall impedance with larger amplitudes.

$\mu=1$					
Case 1		Case 2a		Case 3a	
Re	Im	Re	Im	Re	Im
−9.521 94	0.416 13	−9.341 48	0.268 98	−9.343 16	0.281 60
−7.835 44	0.425 65	−7.835 44	0.425 65	−7.762 96	0.428 31
		−7.698 21	0.275 73	−7.752 00	0.288 88
−6.265 59	0.426 12	−6.134 25	0.269 32	−6.133 66	0.279 02
−4.434 63	0.425 55	−4.434 63	0.425 55	−4.444 46	0.305 15
		−4.394 18	0.290 23	−4.365 83	0.405 54
−2.807 34	0.356 20	−2.793 98	0.254 03	−2.791 93	0.256 73
−2.090 80	3.928 34	−2.090 80	3.928 34	−2.068 56	4.297 27
		−1.633 91	2.435 46	−1.613 60	2.369 21
−1.979 19	3.973 27	−0.452 88	6.228 45	−0.605 43	6.184 13
2.036 24	0.288 63	2.012 94	0.289 82	2.012 72	0.289 71
3.278 02	0.397 92	3.278 02	0.397 92	3.249 73	0.303 39
		3.296 37	0.285 94	3.334 31	0.374 99
4.796 96	0.444 07	4.844 59	0.304 90	4.847 72	0.308 99
6.193 78	0.461 83	6.193 78	0.461 83	6.270 09	0.456 70
		6.326 67	0.284 90	6.271 60	0.300 90
7.764 46	0.454 13	7.907 07	0.288 51	7.908 55	0.298 60
9.243 63	0.451 74	9.243 63	0.451 74	9.320 13	0.459 40
		9.435 03	0.281 91	9.376 94	0.297 75
$\mu=2$					
−9.600 82	0.393 46	−9.385 26	0.266 86	−9.389 99	0.281 73
−7.944 93	0.398 51	−7.944 93	0.398 51	−7.878 80	0.411 45
		−7.758 27	0.268 59	−7.814 68	0.281 13
−6.416 07	0.396 37	−6.214 73	0.264 19	−6.218 56	0.278 14
−4.668 33	0.413 82	−4.668 33	0.413 82	−4.598 10	0.419 53
		−4.523 46	0.272 75	−4.577 36	0.285 25
−3.110 64	0.412 32	−2.968 75	0.260 12	−2.969 23	0.270 53
−1.156 69	2.092 95	−1.156 69	2.092 95	−1.102 01	2.270 60
		−0.937 26	1.394 71	−0.955 37	1.348 71
−0.835 04	2.217 43	−0.195 72	3.119 65	−0.260 58	3.101 94
1.761 76	0.412 69	1.736 55	0.361 84	1.737 31	0.361 31
2.980 94	0.496 23	2.980 94	0.496 23	3.059 75	0.486 53
		3.122 25	0.293 62	3.065 17	0.312 49
4.559 54	0.466 33	4.718 55	0.295 76	4.720 04	0.306 74
6.013 85	0.461 88	6.013 85	0.461 88	6.092 08	0.474 66
		6.244 98	0.287 01	6.183 77	0.304 52
7.637 41	0.441 64	7.846 40	0.285 94	7.845 08	0.300 23
9.147 66	0.438 13	9.147 66	0.438 13	9.221 13	0.455 64
		9.390 38	0.283 55	9.328 55	0.300 02
$\mu=5$					
−9.624 88	0.384 52	−9.399 72	0.266 38	−9.405 65	0.281 88
−7.979 48	0.385 85	−7.979 48	0.385 85	−7.916 41	0.401 22
		−7.778 90	0.266 43	−7.836 32	0.278 67
−6.466 02	0.377 97	−6.245 11	0.262 70	−6.251 34	0.277 87
−4.757 14	0.383 74	−4.757 14	0.383 74	−4.694 13	0.396 52
		−4.576 09	0.264 22	−4.631 75	0.275 90
−3.269 49	0.366 15	−3.062 28	0.253 64	−3.068 70	0.267 80
−0.948 99	0.886 87	−0.948 99	0.886 87	−0.846 42	0.957 12
		−0.778 25	0.633 13	−0.843 73	0.613 91
−0.219 00	1.109 97	−0.051 78	1.251 76	−0.066 14	1.248 35
1.380 25	0.536 78	1.440 97	0.388 92	1.444 59	0.392 27
2.737 52	0.546 44	2.737 52	0.546 44	2.829 61	0.555 35
		3.024 94	0.308 67	2.958 41	0.333 15
4.440 81	0.457 73	4.664 82	0.294 97	4.663 84	0.310 09
5.941 18	0.452 40	5.941 18	0.452 40	6.017 19	0.472 84
		6.213 29	0.290 52	6.148 80	0.308 91
7.592 87	0.432 79	7.825 21	0.286 11	7.822 47	0.301 97
9.116 20	0.431 23	9.116 20	0.431 23	9.187 90	0.451 80
		9.375 43	0.284 76	9.312 12	0.301 44
$\mu=10$					
−9.628 39	0.383 14	−9.401 90	0.266 32	−9.408 01	0.281 90
−7.984 56	0.383 82	−7.984 56	0.383 82	−7.921 98	0.399 51
		−7.782 04	0.266 12	−7.839 62	0.278 30

TABLE III. (*Continued*.)

$\mu=1$					
Case 1		Case 2a		Case 3a	
Re	Im	Re	Im	Re	Im
−6.473 46	0.374 87	−6.249 92	0.262 48	−6.256 56	0.277 81
−4.770 67	0.377 94	−4.770 67	0.377 94	−4.709 02	0.391 50
		−4.584 81	0.262 84	−4.640 77	0.274 34
−3.294 74	0.354 43	−3.080 30	0.252 12	−3.088 14	0.266 61
−1.018 18	0.595 14	−1.018 18	0.595 14	−0.928 30	0.427 51
		−0.857 68	0.425 38	−0.926 45	0.621 94
−0.050 92	0.613 92	−0.011 75	0.627 66	−0.014 38	0.627 31
1.204 00	0.624 06	1.336 10	0.407 04	1.341 30	0.414 96
2.677 44	0.557 63	2.677 44	0.557 63	2.772 72	0.572 03
		3.005 31	0.315 01	2.936 13	0.341 58
4.420 02	0.454 88	4.655 71	0.295 31	4.654 17	0.311 20
5.929 57	0.450 37	5.929 57	0.450 37	6.005 06	0.472 02
		6.208 20	0.291 27	6.143 13	0.309 81
7.586 10	0.431 23	7.821 96	0.286 20	7.818 98	0.302 30
9.111 52	0.430 10	9.111 52	0.430 10	9.182 91	0.451 11
		9.373 17	0.284 97	9.309 64	0.301 69

dently from these possible additional developments, which may be the subject of future work, a number of conclusions can be drawn from the preceding analysis.

V. DISCUSSION

There is a fundamental difference between the acoustics of a nozzle with uniform (Sec. II) and nonuniform (Sec. III) wall impedance. In the case of uniform wall impedance every azimuthal $\exp(im\theta)$ mode can exist independently from the others, because the impedance is isotropic, e.g., the reflection of an isotropic mode $m=0$ from the wall is an isotropic mode. If the wall impedance varies circumferentially it is clear that an isotropic mode $m=0$ will be reflected as an anisotropic mode $m\neq 0$, or rather as a superposition of isotropic $m=0$ and anisotropic modes $m=1,2,\ldots$. Thus, a circumferentially varying impedance couples all circumferential modes $m=0,1,2,\ldots$. This is why the radial eigenfunctions are calculated separately for each m in the case [Eq. (54)] of uniform impedance

$$i\alpha J_m(kR)+Z_0 J_m'(kR)=0, \tag{61}$$

whereas in the case of circumferentially varying impedance the eigenvalues are the roots of an infinite determinant [Eq. (30)] coupling all m. The coupling coefficients $Z_{m'}$ are the amplitudes [Eq. (22)] of the harmonics of the circumferentially varying impedance expressed by Eq. (21) as a Fourier series. If the impedance were uniform $Z_0\neq 0=Z_{\pm 1}=Z_{\pm 2}=\cdots=Z_{\pm m'}$, then the infinite determinant Eq. (30) would reduce to its diagonal, and it would vanish if one of the terms of the diagonal was zero, which is precisely the condition Eq. (54), which is equivalent to Eq. (61). If the impedance were circumferentially nonuniform, but had only one harmonic Z_0, $Z_{\pm 1}\neq 0=Z_{\pm 2}=\cdots=Z_{\pm m}$, then the determinant Eq. (30) would be three-banded, and all azimuthal modes would already be coupled. As more harmonics is added to the impedance Z_0, $Z_{\pm 1},\ldots,Z_{\pm m}\neq 0=Z_{\pm m\pm 1}=\cdots$, the determinants have more bands, *viz.* $(2m+1)$ bands for m harmonics, and the coupling becomes more complex.

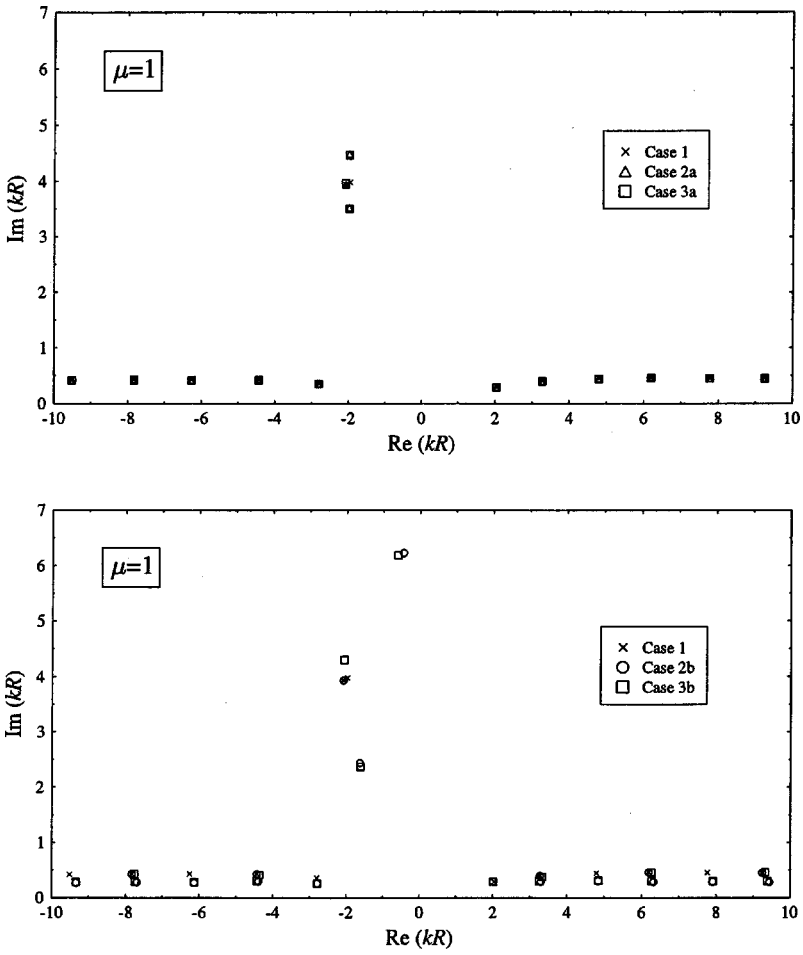


FIG. 1. For a cylindrical nozzle with length equal to the radius, eigenvalues of the dimensionless radial wave number Eq. (48), in the cases of wall impedance either uniform (×), with one harmonic (△), or with a second harmonic (□), distinguishing the case a Eq. (56a), Eq. (58a) of weak harmonics (top) from the case b Eq. (56b), Eq. (58b) of strong harmonics (bottom).

The preceding argument on circumferentially varying impedance (Sec. III A) would apply as well to an axially varying impedance (Sec. III B), replacing the infinite determinant Eq. (30) by Eq. (37). In the case of an impedance varying both circumferentially and axially (Sec. III C) the eigenvalues are roots of a ''doubly infinite'' determinant Eq. (44), i.e., an infinite determinant Eq. (45) whose elements are infinite determinants Eq. (44). It is clear the generalization to *n*-times-infinite determinant consists of a sequence of *n* ''nested'' infinite determinants which could be calculated in *n*! different ways. Multiple infinite determinants can be calculated approximately by truncating *n* times, at each stage. Of course attention should be paid to accumulation of errors. Returning to the acoustics of a nozzle, in the general case of impedance varying both axially and circumferentially, the coefficients Eq. (39) of the double Fourier series Eq. (38) act as coupling coefficients between the axial *l* and the azimuthal *m* modes, adding more ''bands'' to the inner and outer infinite determinants as more impedance harmonics is considered. This occurrence of infinite determinants goes beyond the earlier applications, e.g., in the Hill's solution of Mathieu's equation (Whittaker and Watson, 1927; Campos, 2000b) and solutions of linear differential equations with irregular singularities (Forsyth, 1902; Campos, 2000a), in that doubly infinite determinants also arise. The latter may be extended to *n*-times multiply infinite determinants, defined as a sequence of *n* infinite determinants, each having as terms another infinite determinant, the process being repeated *n* times. Since a nonuniform duct liner has two dimensions, in this application doubly infinite determinants arise at most.

The roots of these infinite determinants specify an infinite sequence of radial wave numbers k_{lmn}, and hence the corresponding eigenfrequencies ω_{lmn}. The method has been applied to uniform axial flow in a cylindrical duct, for which the radial eigenfunctions are Bessel functions Eq. (5); in the case of a cylindrical annulus this solution would be replaced by a linear combination of Bessel and Neumann functions $Y_m(kR)$. In the presence of shear or swirl (Campos and Kobayashi, 2000; Campos, Oliveira, and Kobayashi, 1999; Campos and Serrão, 1998; Tam and Auriault, 1998) in the mean flow, the eigenfunctions may no longer be Bessel functions, but the eigenfrequencies would still be roots of infinite determinants like Eq. (30), Eq. (37), and Eq. (41), with the Bessel functions replaced by the relevant eigenfunctions. In the actual calculations the infinite determinants have to be truncated, and a finite number of eigenfrequencies result; this is consistent with practical applications, where only the lowest few eigenfrequencies need be considered. The eigenfrequencies are ordered by their real parts, with their imaginary parts indicating the decay of the eigenmodes. Thus, given an initial set of eigenmodes and their amplitudes, it is possible to calculate how their acoustic energy decays with time. This may be used to maximize the acoustic attenuation, by optimizing the amplitudes $Z_{m'}$, $Z^{l'}$, $Z^{l'}_{m'}$ of the harmonics of the liner impedance, which specify the nonuniform impedance

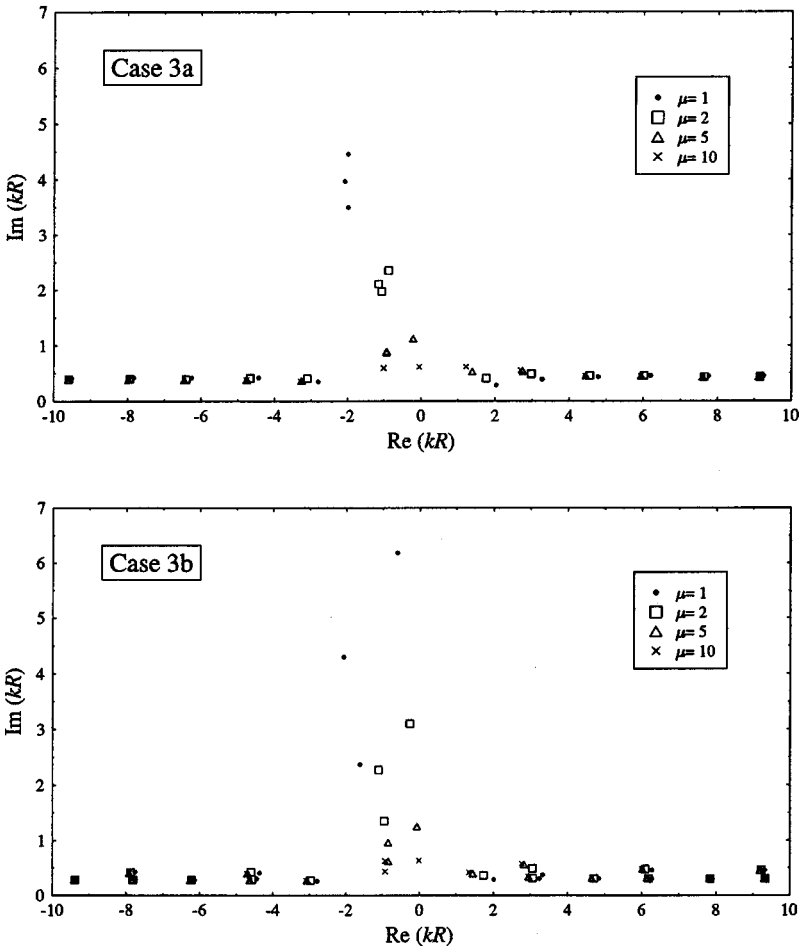


FIG. 2. Eigenvalues of the dimensionless radial wave number Eq. (47), for four values of the aspect ratio Eq. (48) of the duct, *viz.* $\mu=1$ (●), $\mu=2$ (□), $\mu=5$ (△), and $\mu=10$ (×), in case 3 of wall impedance with two harmonics Eq. (56b), Eq. (57), distinguishing between case 3a of weak Eq. (56a), Eq. (58a) harmonics (top) and case 3b of strong Eq. (58b) harmonics (bottom).

distribution over the duct wall. Several optimization criteria could be chosen. For example, the slowest decaying mode could be chosen, and the impedance optimized to maximize its decay. Since a nonuniform impedance causes mode coupling or splitting and thereby causes the appearance of additional modes, the most reliable optimization criterion should be to maximize the decay of the total acoustic energy in the duct, or the sound radiation in the far field outside the nozzle. The acoustic energy at the nozzle exit would depend on the pressure distribution at the inlet, and on the noise sources in the duct. These sound generation and radiation aspects of nonuniformly lined nozzles may be the subject of future work, expanding on the propagation theory presented here. The acoustic fields calculated in the present paper for a nonuniformly lined duct section can be matched to known solutions for rigid walled or uniformly lined ducts with straight or angled junctions.

Bies, D. A., Hansen, C. H., and Bridges, G. E. (**1991**). ''Sound propagation in rectangular and circular cross-section ducts with flow and bulk-reacting liner,'' J. Sound Vib. **146**(1), 47–80.

Campos, L. M. B. C. (**1993**). ''Fractional calculus of analytical and branched functions,'' in Kalia, R. N. (ed.), Recent Advances in the Fractional Calculus, pgs. 65–144 (Global Press).

Campos, L. M. B. C. (**2000a**). ''On the extended hypergeometric equation and functions of arbitrary degree,'' Int. Trans. Spec. Funct. **11**, 233–256.

Campos, L. M. B. C. (**2000b**). ''On the singularities and solutions of the extended hypergeometric equations,'' Int. Trans Spec. Funct. **9**, 99–120.

Campos, L. M. B. C. (**2003**). ''On the derivation of asymptotic expansions for special functions from the corresponding differential equations,'' Int. Trans. Spec. Funct. **12**, 227–236.

Campos, L. M. B. C. (**1978b**). ''The spectral broadening of sound by turbulent shear layers. Part 2: The spectral broadening of sound and aircraft noise,'' J. Fluid Mech. **89**, 723–749.

Campos, L. M. B. C. (**1986**). ''On waves in gases. Part 1: Acoustics of jets, turbulence, and ducts,'' Rev Mod. Phys. **58**(1), 117–182.

Campos, L. M. B. C., and Oliveira, J. M. G. S. (**2004**). ''On the optimization of acoustic liners in annular nozzles,'' J. Sound Vib. **275**, 557–576.

Campos, L. M. B. C., and Kobayashi, M. H. (**2000**). ''On the reflection and transmission of sound in a thick shear layer,'' J. Fluid Mech. **420**, 1–24.

Campos, L. M. B. C., Oliveira, J. M. G. S., and Kobayashi, M. H. (**1999**). ''On sound propagation in a linear flow,'' J. Sound Vib. **219**(5), 739–770.

Campos, L. M. B. C., and Serrão, P. G. T. A. (**1998**). ''On the acoustics of an exponential boundary layer,'' Philos. Trans. R. Soc. London, Ser. A **356**, 2335–2378.

Campos, L. M. B. C. (**1978a**). ''The spectral broadening of sound by turbulent shear layers. Part 1: The transmission of sound through turbulent shear layers,'' J. Fluid Mech. **89**, 723–749.

Forsyth, A. R.. (**1902**). Theory of differential functions (Cambridge University Press).

Fuller, C. R. (**1984a**). ''Propagation and radiation of sound from flanged circular ducts with circumferentially varying wall admittances. I. Semi-infinite ducts,'' J. Sound Vib. **93**(3), 321–340.

Fuller, C. R. (**1984b**). ''Propagation and radiation of sound from flanged circular ducts with circumferentially varying wall admittances. II. Finite ducts with sources,'' J. Sound Vib. **93**(3), 341–351.

Hocter, S. T. (**1999**). ''Exact and approximate directivity patterns of the sound radiated from cylindrical duct,'' J. Sound Vib. **227**(2), 397–407.

Horowitz, S. J., Sigman, R. K., and Zinn, B. T. (**1982**). ''An iterative method for predicting turbofan inlet acoustics,'' AIAA J. **20**(12), 1693–1699.

Howe, M. S. (**1983**). ''The attenuation of sound in a randomly lined duct,'' J. Sound Vib. **87**(1), 83–103.

Howe, M. S. (**1995**). ''The damping of sound by turbulent wall shear layers,'' J. Acoust. Soc. Am. **98**, 1723–1730.

Joseph, P., and Morfey, C. L. (**1999**). ''Multimode radiation from an unflanged, semi-infinite circular duct,'' J. Acoust. Soc. Am. **105**(5), 2590–2600.
Koch, W. (**1977a**). ''Attenuation of sound in multi-element acoustically lined rectangular ducts in the absence of mean flow,'' J. Sound Vib. **52**(4), 459–496.
Koch, W. (**1977b**). ''Radiation of sound from a two-dimensional acoustically lined duct,'' J. Sound Vib. **55**(2), 255–274.
Koch, W., and Möhring, W. (**1983**). ''Eigensolutions for liners in uniform mean flow ducts,'' AIAA J. **21**(2), 200–213.
Martinez, R. (**1988**). ''Diffracting open-ended pipe treated as a lifting surface,'' AIAA J. **26**(4), 396–404.
Meyer, H. D. (**1996**). ''Effect of inlet reflections on fan noise radiation,'' AIAA J. **34**(9), 1771–1777.
Munt, R. M. (**1977**). ''Acoustic radiation from a semi-infinite circular duct in an uniform subsonic mean flow,'' J. Fluid Mech. **83**, 609–640.
Nelson, P. A., and Morfey, C. L. (**1981**). ''Aerodynamic sound production in low speed flow ducts,'' J. Sound Vib. **79**(2), 263–289.
Pierce, A. D. (**1981**). *Acoustics. An Introduction to Its Physical Principles* (McGraw-Hill, New York).
Plumblee, H. E., and Doak, P. E. (**1979**). ''Duct noise radiation through a jet flow,'' J. Sound Vib. **65**(4), 453–491.
Rawlins, A. D. (**1978**). ''Radiation of sound from an unflanged rigid cylindrical duct with an acoustically absorbing internal surface,'' Proc. R. Soc. London, Ser. A **361**, 65–91.
Rayleigh, J. W. S. (**1945**). *Theory of Sound* (Dover, New York).
Regan, B., and Eaton, J. (**1999**). ''Modeling the influence of acoustic liner nonuniformities on duct modes,'' J. Sound Vib. **219**(5), 859–879.
Rienstra, S. W. (**1984**). ''Acoustic radiation from a semi-infinite duct in a uniform subsonic mean flow,'' J. Sound Vib. **94**(2), 267–288.
Rienstra, S. W. (**1985**). ''Contributions to the theory of sound propagation in ducts with bulk-reacting lining,'' J. Acoust. Soc. Am. **77**(5), 1681–1685.
Savkar, S. D. (**1975**). ''Radiation of cylindrical duct acoustic modes with flow mismatch,'' J. Sound Vib. **42**, 363–386.
Snakowska, A., and Idczak, H. (**1997**). ''Prediction of multitone sound radiation from a circular duct,'' Acustica **83**, 955–962.
Snakovska, A., Idczak, H., and Bogusz, B. (**1996**). ''Modal analysis of the acoustic field radiated from an unflanged cylindrical duct—theory and measurement,'' Acustica **82**, 201–206.
Tam, C. K. W., and Auriault, L. (**1998**). ''The wave modes in ducted swirling flows,'' J. Fluid Mech. **371**, 1–20.
Vaidya, P. G. (**1985**). ''The propagation of sound in ducts lined with circumferentially nonuniform admittance of the form $\eta_0+\eta_q \exp(iq\theta)$,'' J. Sound Vib. **100**(4), 463–475.
Watson, G. N. (**1944**). *Bessel Functions* (Cambridge University Press, Cambridge).
Watson, W. R. (**1984**). ''An acoustic evaluation of circumferentially segmented duct liners,'' AIAA J. **22**(9), 1229–1233.
Whittaker, E. T., and Watson, G. N. (**1927**). *A Course of Modern Analysis* (Cambridge University Press, Cambridge).

Wave interaction in acoustic resonators with and without hysteresis

Lev A. Ostrovsky[a)]
Zel Technologies/University of Colorado, 325 Broadway, Boulder, Colorado, 80305

(Received 21 May 2004; revised 23 September 2004; accepted 3 October 2004)

Nonlinear interaction of counterpropagating waves in solids is considered by using a general approach taking into account only the cumulative (resonant) nonlinear perturbations giving a nonzero contribution over the period and, consecutively, potentially able to significantly modify the linear solution. Different stress-strain relations are addressed, including those with hysteresis which serve as basic models for the recent acoustic experiments with rock and metals. An important case of the interaction of counterpropagating waves with close amplitudes in a high-Q resonator (bar) with hysteresis is specially addressed and compared with the case of a ring resonator. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1823332]

PACS numbers: 43.25.Gf, 43.25.Dc [MFH] Pages: 3348–3353

I. INTRODUCTION

This paper deals with one of the difficult problems of modern nonlinear acoustics: nonlinear resonances in solid bars. It is common knowledge that many real materials, including rock and metals, are characterized by a hysteresis in the dependence between tensors of stress, σ_{ik}, and strain, ϵ_{ik}. Acoustic wave propagation in such media has many peculiarities. Due to important applications such as in the nondestructive evaluation, these effects became a hot topic in the recent years (see Refs. 1 and 2, and references therein).

Although most informative experimental results have been obtained in resonant bars, few theoretical models are yet available to describe and calculate waves interacting in such bars, except for the classical case of quadratic nonlinearity.[3,4] For the latter, in the main approximation, the counterpropagating waves do not interact in a cumulative way. As regards media with hysteresis, a relatively detailed consideration has been made for traveling waves[5,6] that allowed to construct a model of a ring resonator in which a wave propagates in one direction.[2,7] Interaction of oppositely propagating waves in a hysteretic model was considered only for quasiharmonic waves, and only for some rather specific cases.[8–10]

Here a simple but rather general approach is suggested based on accounting of only resonance nonlinear perturbations which cumulate from period to period, whereas the nonresonant terms in the equations provide only small perturbations to the linear solution and can be neglected in the first approximation. Note that this approach is typical of all asymptotic perturbation methods.

II. GENERAL APPROACH

For simplicity, one-dimensional processes and the corresponding scalar values, $\sigma=\sigma_{11}$ and $\epsilon=\epsilon_{11}$, are considered, which is adequate to the majority of experiments performed in resonance bars.

We begin with the nonlinear wave equation

$$\rho\frac{\partial^2\epsilon}{\partial t^2}=\frac{\partial^2\sigma(\epsilon,\epsilon_t)}{\partial x^2}, \tag{1}$$

(ρ is the material density, $\epsilon_t\equiv\partial\epsilon/\partial t$). In media with dominating physical nonlinearity, the kinematic nonlinear terms responsible for the difference between Eulerian and Lagrangian forms of the equation can be neglected;[2] note that all nonlinear terms are still much smaller than linear ones, which is practically always true in solid acoustics without plasticity and destruction effects. We represent the stress-strain relation as

$$\sigma=M\epsilon+\sigma_N(\epsilon,\epsilon_t),$$

where the elastic modulus M is constant, and $\sigma_N=O(\epsilon^2)$ is a small, nonlinear part that may or may not include hysteresis. Hence, from (1):

$$\frac{\partial^2\epsilon}{\partial t^2}-c^2\frac{\partial^2\epsilon}{\partial x^2}=-(1/\rho)\frac{\partial^2\sigma_N}{\partial x^2}. \tag{2}$$

Here $c=\sqrt{M/\rho}$ is the linear wave velocity.

For studying the opposite waves' interaction, we pass from x and t to "traveling" variables, $\eta_+=x-ct$ and $\eta_-=x+ct$. In these variables, Eq. (2) reads

$$2M\frac{\partial^2\epsilon}{\partial\eta_+\partial\eta_-}=\left(\frac{\partial^2}{\partial\eta_+^2}+\frac{2\partial^2}{\partial\eta_+\partial\eta_-}+\frac{\partial^2}{\partial\eta_-^2}\right)\sigma_N, \tag{3}$$

or, after integration,

$$\epsilon=\epsilon_{0+}(\eta_+)+\epsilon_{0-}(\eta_-)+\epsilon_N, \tag{4}$$

where

$$2M\epsilon_N=\int\frac{\partial\sigma_N}{\partial\eta_+}d\eta_-+\int\frac{\partial\sigma_N}{\partial\eta_-}d\eta_++2\sigma_N. \tag{5}$$

Note that the earlier transformations are still exact, and the integral form is more adequate for dealing with nonanalytic functions characteristic of models with hysteresis considered later; however, in each case we shall finally return to a more common differential form of the equations. In the absence of nonlinearity Eq. (3) has an evident general solu-

a)Electronic address: lev.a.ostrovsky@noaa.gov

tion, $\epsilon_0=\epsilon_{0+}(\eta_+)+\epsilon_{0-}(\eta_-)$. The nonlinear perturbation, ϵ_N, is supposed small. However, due to the cumulative character of wave distortion (such as in the simple wave tending to break), the contribution of ϵ_N tends, in general, to secularly increase. Thus, according to the standard asymptotic theory, we represent ϵ as a superposition of two slowly distorting, interacting waves

$$\epsilon=\epsilon_+(\eta_+,\eta_-)+\epsilon_-(\eta_-,\eta_+)+g(\eta_+,\eta_-), \tag{6}$$

where ϵ_+ only slowly depends on η_- (so that $\partial\epsilon/\partial\eta_- \ll \partial\epsilon/\partial\eta_+$), and vice versa for ϵ_-. The interaction function g is small, of order $O(\sigma_N)$, and it disappears in each of traveling waves in the absence of the other. Note also that in literature the presentation in such form as, e.g., $\epsilon_+(\eta_+,\tau)$ where τ is a slowly varying, is more common, but there is no real difference between these presentations (see later).

To be more specific, we suppose that the process is periodic in space with a wavelength Λ, and the period-average of each term in (6) is zero or at least is of the order of $o(\sigma_N)$. This is compatible with all nonlinear models of stress-strain dependence considered later and it is sufficient for practically all realistic situations where the vibrations are excited by an oscillating transducer.

Substitute Eq. (6) into Eq. (5). As mentioned, all nonlinear terms are locally small, and in the main order, one should take into account only the resonance ones the resulting effect of which is not small. Consequently, in the first approximation the terms g in Eq. (6) and σ_N in Eq. (5) can be neglected, whereas in the integrands of Eq. (5) only terms with nonzero period average effectively contribute into the solution. Note also that upon integrating over the period, slow dependencies can be ''frozen.''

III. NONHYSTERETIC MODELS

Consider first two simple, nonhysteretic stress-strain dependencies, starting with the quadratic one

$$\sigma_N=-a\epsilon^2, \quad a=\text{const}. \tag{7}$$

Substituting (7) and (6) into (5), we have

$$\epsilon_N=\epsilon_{N+}+\epsilon_{N-} = -\frac{a}{2M}\cdot\left\{\begin{array}{l}\displaystyle\int\left(\frac{\partial\epsilon_+^2}{\partial\eta_+}+2\frac{\partial\epsilon_+}{\partial\eta_+}\epsilon_-+\frac{\partial\epsilon_-^2}{\partial\eta_-}\right)d\eta_- \\ \displaystyle+\int\left(\frac{\partial\epsilon_-^2}{\partial\eta_-}+2\frac{\partial\epsilon_-}{\partial\eta_-}\epsilon_++\frac{\partial\epsilon_+^2}{\partial\eta_+}\right)d\eta_+\end{array}\right. . \tag{8}$$

As mentioned, we shall neglect all nonresonant terms in which the corresponding integrals over the spatial wave period Λ disappear. Evidently, the last term in each of the earlier integrals is zero over the period, and hence, these terms are noncumulative. The same is true for the second terms if, as already mentioned, $\int_{\eta_-}^{\eta_-+\Lambda}\epsilon_\pm d\eta_\pm=0$. (The respective factors, $\epsilon_\mp$, are slowly varying and can be taken constant over the period of $\eta_\pm$.) The first terms in the integrands contribute to ϵ_+ and ϵ_-, respectively. After differentiating we have for each wave (evidently, $\partial\epsilon_\pm/\partial\eta_\mp=\partial\epsilon_{N\pm}/\partial\eta_\mp$):

$$\frac{\partial\epsilon_\pm}{\partial\eta_\mp}=-\frac{a}{2\rho c^2}\frac{\partial\epsilon_\pm^2}{\partial\eta_\pm}. \tag{9}$$

Thus, the opposite waves in a quadratic medium propagate independently, and each satisfies the known equation of a simple wave (the nonviscous Burgers equation). This result was earlier obtained and used by the author[3] to calculate nonlinear waves in an acoustic resonator (see also Ref. 4).

Note that in the original variables, we have $\partial/\partial\eta_-=\partial/\partial x+(1/c)\partial/\partial t$, and $\partial/\partial\eta_+=\partial/\partial x-(1/c)\partial/\partial t\approx 2\partial/\partial x$ for the ϵ_+ wave (and correspondingly for ϵ_-). Then, in variables $\tau=t\mp x/c$, $x'=x$, with a slow dependence on τ, it transforms Eq. (9) to the standard form

$$\frac{\partial\epsilon_\pm}{\partial x'}=\frac{a}{\rho c^3}\frac{\partial\epsilon_\pm^2}{\partial\tau_\pm}. \tag{10}$$

The similar is true for all subsequent models.

Now consider cubic nonlinearity

$$\sigma_N=b\epsilon^3, \quad b=\text{const}. \tag{11}$$

Representing again $\epsilon=\epsilon_++\epsilon_-$, we have from Eq. (5), for, e.g., ϵ_+:

$$\epsilon_{N+}=\frac{b}{2M}\cdot\left(\begin{array}{c}\displaystyle\int\frac{\partial\epsilon_+^3}{\partial\eta_+}d\eta_-+3\frac{\partial\epsilon_+}{\partial\eta_+}\int\epsilon_-^2 d\eta_- \\ \displaystyle+3\frac{\partial\epsilon_+^2}{\partial\eta_+}\int\epsilon_- d\eta_-\end{array}\right). \tag{12}$$

The last term in (12) gives zero over the period and can be neglected. Differentiating the rest, we have

$$\frac{\partial\epsilon_+}{\partial\eta_-}=\frac{b}{2M}\left(\frac{\partial\epsilon_+^3}{\partial\eta_+}+3\langle\epsilon_-^2\rangle\frac{\partial\epsilon_+}{\partial\eta_+}\right). \tag{13}$$

Similarly for ϵ_-:

$$\frac{\partial\epsilon_-}{\partial\eta_+}=\frac{-b}{2M}\left(\frac{\partial\epsilon_-^3}{\partial\eta_-}+3\langle\epsilon_+^2\rangle\frac{\partial\epsilon_-}{\partial\eta_-}\right). \tag{14}$$

Here $\langle\epsilon_\pm^2\rangle=(1/\Lambda)\int_{\eta_\pm}^{\eta_\pm+\Lambda}\epsilon_\pm^2 d\eta_\pm$ is the average value.

For a traveling wave, each of Eqs. (13) and (14) is equivalent to the modified Burgers equation without losses which can be applicable for transverse waves in, e.g., a string. Each wave propagates with a local velocity of $c[1-(3b/M)(\epsilon_\pm^2+\langle\epsilon_\mp^2\rangle)]$; the last term in these parentheses defines a change in the average wave velocity and the corresponding nonlinear frequency shift in a resonator, whereas the previous term is responsible for the nonlinear profile distortion of the simple wave. In general, these two equations must be solved as a system. In a symmetric case, $\langle\epsilon_-^2\rangle=\langle\epsilon_+^2\rangle$, that is characteristic of a high-quality resonator, these two equations are identical.

IV. HYSTERESIS

More complicated is the problem of wave interaction in media with a hysteretic stress-strain relation. Consider first the most frequently used stress-strain dependence given by (Fig. 1):

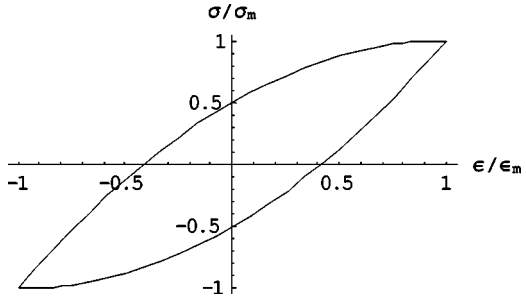


FIG. 1. The hysteretic stress-strain dependence of type (15).

$$-\sigma_N = a\epsilon_m\epsilon + S\gamma(\epsilon^2-\epsilon_m^2); \quad a,\gamma>0. \tag{15}$$

Here $S=\mathrm{Sign}(\partial\epsilon/\partial t)$.

By definition, $\epsilon_m(x)>0$ is the peak value of time-varying strain at a given point x. In a progressive wave, the dependence of ϵ_m on x is slow as compared with the wave length and is due to the wave amplitude damping at singularities in the stress-strain function (15); this process was considered, in, e.g., Refs. 5 and 6. The slow variation of ϵ_m does not affect integration over the period. However, for the case of two opposite waves considered here, strong variations of ϵ_m along x can occur at each spatial period (such as in a quasiharmonic standing wave), and they must be taken into account in the main order when the wave interaction is considered.

Letting $\epsilon_m=\epsilon_m[x=(\eta_+ + \eta_-)/2]$, we have from (5) and (15):

$$2M\epsilon_N = \left\{ \begin{array}{c} a\int \frac{\partial}{\partial\eta_+}(\epsilon_m\epsilon_+ + \epsilon_m\epsilon_-)d\eta_- + a\int \frac{\partial}{\partial\eta_-}(\epsilon_m\epsilon_+ + \epsilon_m\epsilon_-)d\eta_+ \\ +S\gamma\int\left(\frac{\partial\epsilon_+^2}{\partial\eta_+}+2\frac{\partial\epsilon_+}{\partial\eta_+}\epsilon_- + \frac{\partial\epsilon_-^2}{\partial\eta_-}\right)d\eta_- + S\gamma\int\left(\frac{\partial\epsilon_-^2}{\partial\eta_-}+2\frac{\partial\epsilon_-}{\partial\eta_-}\epsilon_+ + \frac{\partial\epsilon_+^2}{\partial\eta_+}\right)d\eta_+ \\ +S\gamma\left(\int\frac{\partial\epsilon_m^2}{\partial\eta_+}d\eta_- + \int\frac{\partial\epsilon_m^2}{\partial\eta_-}d\eta_+\right. \end{array} \right). \tag{16}$$

Evidently, $\partial\epsilon_m/\partial\eta_+ = \partial\epsilon_m/\partial\eta_-$.

For a traveling wave when $\epsilon_-=0$ and ϵ_m is slowly varying, only the first terms in the first and second lines' integrands of the right-hand side of (16) give a cumulative contribution for ϵ_+, and their differentiating over η_- yields

$$2M\frac{\partial\epsilon_+}{\partial\eta_-} = a\epsilon_m\frac{\partial\epsilon_+}{\partial\eta_+} + S\gamma\frac{\partial\epsilon_+^2}{\partial\eta_+}, \tag{17}$$

and similarly for ϵ_-. In the variables $\tau=-\eta_+/c$ and $x=x'$, Eq. (17) corresponds to Eq. (6) of Ref. 6. Its solution is known: a primarily sinusoidal wave transforms to a triangular form asymptotically decreasing in magnitude as x^{-1} (Fig. 2). The term with ϵ_m (here, in general, ϵ_m is slowly decreasing) is responsible for the wave velocity variation.

Let us return to the interaction of opposite waves in the framework of Eq. (16). Considering that ϵ_+ only slowly depends on η_-, this dependence can be neglected at a period, as earlier. The same is true for $\epsilon_-(\eta_+)$. However, the function $\epsilon_m(\eta_+ + \eta_-)$ is now determined by the superposition of both waves and in a general case it can have a rather complicated form. As earlier, the integrable terms can be neglected because they give zero integral at the period. In particular, the integrals of type $\int\epsilon_+(\partial\epsilon_m/\partial\eta_+)d\eta_- = \int\epsilon_+(\partial\epsilon_m/\partial\eta_-)d\eta_-$ over the period are zero. As a result, we obtain

$$2M\epsilon_N = \left\{ \begin{array}{c} a\left(\int\frac{\partial\epsilon_+}{\partial\eta_+}\epsilon_m d\eta_- + \int\frac{\partial\epsilon_-}{\partial\eta_-}\epsilon_m d\eta_+ + \int\frac{\partial\epsilon_m}{\partial\eta_+}\epsilon_- d\eta_- + \int\epsilon_+\frac{\partial\epsilon_m}{\partial\eta_-}d\eta_+\right) \\ +S\gamma\left(\int\frac{\partial\epsilon_+^2}{\partial\eta_+}d\eta_- + \int\frac{\partial\epsilon_-^2}{\partial\eta_-}d\eta_+ + 2\int\frac{\partial\epsilon_+}{\partial\eta_+}\epsilon_- d\eta_- + 2\int\frac{\partial\epsilon_-}{\partial\eta_-}\epsilon_+ d\eta_+\right. \end{array} \right). \tag{18}$$

For the description of resonators (bars), of special interest is again the case when at a given interval of $x=(0,L)$, the opposite waves have almost identical amplitudes and profiles at each moment (i.e., wave damping at the distance L is supposed small). In this case, without loss of generality, one can let the function $\epsilon_-(\eta_-)$ the same as $\epsilon_+(\eta_+)$ (just by a proper positioning of the origin, $x=0$).

Let us consider the third term, $\int\epsilon_-(\partial\epsilon_m/\partial\eta_+)d\eta_- = \int\epsilon_-(\partial\epsilon_m/\partial\eta_-)d\eta_-$, in Eq. (18). In our case, ϵ_- is a periodic function of η_- with a period Λ, and its Fourier expansion contains harmonics proportional to $\exp[(\pm 2n\pi i/\Lambda)\eta_-]$, where $n=1,2,\ldots$. The function $\epsilon_m(x) = Max\{|\epsilon_+ + \epsilon_-|\}$ is an even, positive function varying from 0 to $2A$, where A is the amplitude of each wave (*Max*

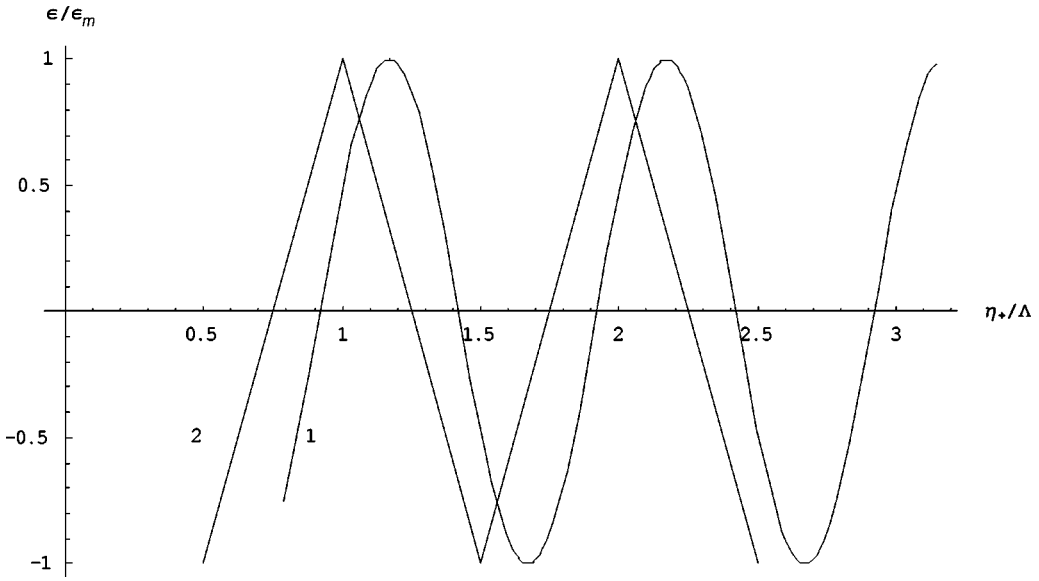


FIG. 2. Scheme of evolution of an initially sinusoidal traveling wave in a medium with the stress-strain dependence from Fig. 1. (1) $t=0$ and (2) $t>0$.

is the time maximum depending on x). Evidently, its period is $\Lambda/2$, and its Fourier series, as well as the series of its derivative, contains terms with $\exp[(\pm 4m\pi i/\Lambda)(\eta_- + \eta_+)]$, where $m=0,1,\ldots$. Hence, the product $\epsilon_- \partial\epsilon_m/\partial\eta_-$ can contain either terms oscillating in η_- that do not give a cumulative effect, or the terms of type $\exp[(\pm 4n\pi i/\Lambda)\eta_+]$ when $n=\pm 2m$. However, in the class of symmetric functions $\epsilon_-(\eta_-)$ with zero average, such as those shown in Fig. 3, there is no even harmonics in the wave, and the latter components are absent (cf. Eq. (12) of Ref. 6 where, in our case, $\beta_1=\beta_2$). This means that the corresponding term in (18) is not cumulative and can be neglected.

At this stage, differentiating (18) gives for ϵ_+:

$$2M\frac{\partial\epsilon_+}{\partial\eta_-}=(a\epsilon_m+2S\gamma\epsilon_-)\frac{\partial\epsilon_+}{\partial\eta_+}+S\gamma\frac{\partial\epsilon_+^2}{\partial\eta_+}. \tag{19}$$

The terms in parentheses define a perturbation of wave velocity. In general, they describe a "jerky" motion of each point of the wave profile (with a fixed η_+). Representation $\epsilon_m(x)=\langle\epsilon_m\rangle+\tilde{\epsilon}_m(x)$ shows that at a fixed η_+, only the average part, $\langle\epsilon_m\rangle=(1/\Lambda)\int_{\eta_-}^{\eta_-+\Lambda}|\epsilon_+ + \epsilon_-|_{\max}dx$, generates a systematic velocity shift in (19). In the hysteretic term, the factor S changes its sign together with the time derivative, $\epsilon_t=c[(\epsilon_-)_{\eta_-}-(\epsilon_+)_{\eta_+}]$. At a fixed η_+ the last term in ϵ_t is fixed (slow dependencies are unimportant at a period), while the first term oscillates. The integral $S\int\epsilon_- d\eta_-$ gives zero at a period, and the corresponding term in (19) can be neglected. Finally, we have

$$2M\frac{\partial\epsilon_+}{\partial\eta_-}=a\langle\epsilon_m\rangle\frac{\partial\epsilon_+}{\partial\eta_+}+2S\gamma\epsilon_+\frac{\partial\epsilon_+}{\partial\eta_+}, \tag{20}$$

and a similar equation for ϵ_-.

This equation looks similar to Eq. (17) written for a traveling wave. However, their solutions differ very significantly. First, instead of ϵ_m coinciding with the peak amplitude of the wave, here we have $\langle\epsilon_m\rangle$ depending on both waves. This quantity depends on the wave profile. If each wave is close to a sinusoid with a peak amplitude A, then $\epsilon_m=2A|\sin kx|$, and $\langle\epsilon_m\rangle=4A/\pi$. Thus, in terms of a traveling wave amplitude, the nonlinear velocity shift and the respective frequency shift in a resonator is somewhat stronger than in a ring resonator of length $2L$. In case of strongly distorted waves of a triangular type with a peak value of A (curve 2 in Fig. 2), the function $\epsilon_m(x)$ is a sawtooth one (Fig. 3). It yields $\langle\epsilon_m\rangle=A$, and no difference with the case of a traveling wave exists in this aspect.

More radical is the difference in the effect of the last, hysteretic term. It describes a nonlinear, cumulative profile distortion but the relative velocity of each profile point changes its sign with $\epsilon_t=c[(\epsilon_-)_{\eta_-}-(\epsilon_+)_{\eta_+}]$. As before the last term here is fixed at a fixed profile point, and the sign variation is related to $(\epsilon_-)_{\eta_-}$ as it shows up at that point. Note first that for a sawtooth wave shown in Fig. 2, the value of $(\epsilon_+)_{\eta_+}$ is the same at the entire wave front, so that the opposite wave of the same (or smaller) steepness cannot change its sign: the entire process is the same as in the traveling wave. On the contrary, in the wave close to a sinusoid (a more realistic case at present) the opposite wave affects the propagation. As before, each profile point propagates with "jerks," except for the points where $\epsilon_+=0$ (and $(\epsilon_+)_{\eta_+}$ is maximal) which does not have nonlinear velocity

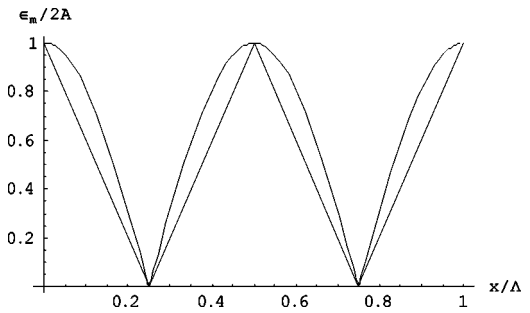


FIG. 3. The dependence $\epsilon_m(x)$ for two counterpropagating waves of sinusoidal and triangular profiles (see Fig. 2) and equal amplitudes.

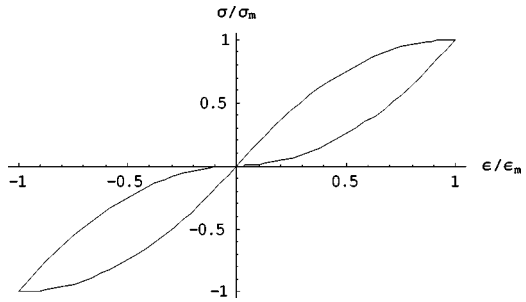


FIG. 4. The hysteretic stress-strain model of type (21).

shift at all, and the change of wave steepness in this point is not affected by the opposite wave. The vicinities of wave crests and troughs, with small $(\epsilon_+)_{\eta_+}$, mostly move back and forth due to the action of oscillating $(\epsilon_-)_{\eta_-}$, and these and other points are progressing slower than in a traveling wave because ϵ_t changes its sign during the period. Thus, the wave distortion, remaining qualitatively the same as in a traveling wave, occurs slower. Correspondingly, harmonic generation is less intensive than in a traveling wave or in a ring resonator considered in Refs. 6 and 7.

Consider briefly another basic model of the hysteretic stress-strain dependence[6] (Fig. 4):

$$\sigma_N = -Mb\begin{cases} \epsilon^2 (\epsilon>0, \epsilon_t>0) \\ -\epsilon^2 + 2\epsilon_m \epsilon (\epsilon>0, \epsilon_t<0) \\ -\epsilon^2 (\epsilon<0, \epsilon_t<0) \\ \epsilon^2 + 2\epsilon_m \epsilon (\epsilon<0, \epsilon_t>0) \end{cases}. \tag{21}$$

The evolution of a traveling wave for this case was also described in Ref. 6. The wave profile is more complex: it contains both cusps as in Fig. 2 and shock fronts [see Fig. 2(b) of Ref. 6]; however, it retains the same symmetries as in the previous case. Now the integrals in Eq. (5) should be taken at four intervals during the period, corresponding to the four branches in (21). After a simple but somewhat cumbersome consideration it can be shown that the counterpropagating waves of similar profiles and equal amplitudes satisfy an equation similar to (20). Hence, each of the oppositely propagating waves in a resonator again acquires an additional constant velocity shift proportional to $\langle \epsilon_m \rangle$, whereas its profile distortions occur slower.

V. RESONATORS

To apply this theory to resonance longitudinal modes in a bar, one must consider reflections of a traveling wave from the bar ends. If the ends are free (a typical situation), stress (pressure) and strain invert their phases upon reflection, whereas the particle velocity is reflected in-phase. Consider the earlier cases from this viewpoint.

Quadratic, nonhysteretic nonlinearity, Eq. (7). As σ and ϵ change the phase by π, the compressions and dilations in the wave are transposed after each reflection. Hence, the nonlinear steepening is not cumulative; in particular, the second harmonic is not generated in a resonance manner. If the wave distortion at the resonator's length L is small, it remains small at all times.

Cubic, nonhysteretic nonlinearity, Eq. (11). In this case, the nonlinear part of the simple wave velocity is proportional to ϵ^2, and change of sign does not affect distortion, which continues to cumulate after each reflection, until a steady-state regime is established. Hence, the cubic effects, such as third harmonic generation and nonlinear resonance frequency shift, are equally well expressed in resonators with hard and free ends.

Hysteretic nonlinearity, Eqs. (15) and (21). As ϵ changes its sign upon reflection, the same is true for ϵ_t, and the wave passes from the ascending branch of the stress-strain curve to the descending part, and vice versa. As a result, the nonlinear distortion continues: although local nonlinearity is quadratic, it will cumulate like in cubic medium.

Note that in the latter three models, only odd harmonics of a primarily sinusoidal field exist, which may explain the fact that in many bar experiments, the observed level of third harmonic exceeds that for the second harmonic, in spite of the natural domination of quadratic terms in the expansion of the material equation of state, $\sigma(\epsilon)$.

Let us briefly compare oscillations in a hysteretic ring resonator[7] and a resonance bar considered here. From the earlier consideration it follows that if the wave has reached the sawtooth stage, there is no difference between these two cases, and all results obtained in Refs. 6 and 7 are applicable to the bar provided the traveling wave amplitude, i.e., half of peak amplitude in the bar, A, is the same as in a ring of a length $2L$. For a quasisinusoidal wave, the value of $\langle \epsilon_m \rangle$ in Eq. (20) differs from the amplitude ϵ_m in Eq. (17) by a factor of $4/\pi$. In the latter case, the relative resonance frequency shift is $\Delta\omega/\omega = \Delta c/c = a\epsilon_{\max}/2\pi M$ (Δc is the nonlinear variation of sound velocity). This agrees with most of resonant bar experiments where the measured resonance frequency shift is proportional to the wave amplitude in the bar.

Detailed analysis of resonance modes in a bar is beyond the scope of this paper. Note, however, that, similarly to the ring resonator, the ratio between the nonlinear frequency shift and losses due to hysteresis, is expected to be amplitude-independent in a quasisinusoidal regime. Also, as mentioned, harmonic generation should be partially suppressed by the interaction of opposite waves. It is not clear yet whether this conclusion agrees with experimental data but some indicators for this do exist. In the framework of the model (15), the ratio of the third harmonic amplitude to that of the main harmonic, A_3/A_1, in a quasisinusoidal traveling wave (and, consequently, in a ring resonator) would be three times larger that the relative sound velocity shift, that is equal to the relative resonance frequency shift, $\Delta f/f$ (cf. Eq. (12) of Ref. 6). In experiments by Johnson *et al.*, at the strain amplitude of, e.g., 10^{-6}, the value of $\Delta f/f$ reaches 10^{-1} for Berea sandstone (see Fig. 6 of Ref. 2). At the same time, from Fig. 10(a) of Ref. 2, the value of A_3/A_1 is of order 5×10^{-3} for Berea sandstone, that is significantly smaller than the frequency shift, which possibly agrees with the above conclusions. This question needs further consideration.

VI. CONCLUSION

The approach developed earlier seems to provide a relatively simple and universal description of counterpropagat-

ing waves' interaction, including resonance processes in bars where the amplitudes of these waves are typically close to each other. Most specific is the hysteretic case in which the interaction of opposite quasiharmonic waves preserves the same order of nonlinear frequency shift as in a traveling wave but tends to decrease nonlinear distortions and harmonic generation at a given strain amplitude (as opposed to a nonhysteretic medium with cubic nonlinearity), whereas at the sawtooth stage, both effects are similar to those in a traveling wave. A possible explanation of this in terms of a standing wave representation is that the diameter of hysteretic loop decreases with the distance from the oscillation peak and, in addition, only a half of the ''doubled'' hysteretic loop at each point can be attributed to a wave traveling in a given direction. Still, a more detailed consideration is needed, taking into account asymmetry of hysteretic loops, linear losses, etc. This work is planned to be continued.

ACKNOWLEDGMENT

The author is grateful to Dr. V. E. Gusev for valuable discussions.

[1] R. A. Guyer and P. A. Johnson, ''Nonlinear mesoscopic elasticity: Evidence for a new class of materials,'' Phys. Today **52**, 30–35 (1999).
[2] L. A. Ostrovsky and P. A. Johnson, ''Dynamic nonlinear elasticity in geomaterials,'' Riv. Nuovo Cimento **24**, 1–46 (2001).
[3] L. A. Ostrovsky, ''Discontinuous oscillations in an acoustic resonator,'' Sov. Phys. Acoust. **20**, 140–142 (1974).
[4] O. V. Rudenko, C. M. Hedberg, and B. O. Enflo, ''Nonlinear standing waves in a layer excited by the periodic motion of its boundary,'' Acoust. Phys. **47**, 452–460 (2001).
[5] V. Gusev and V. Aleshin, ''Strain wave evolution equation for nonlinear propagation in materials with mesoscopic elements,'' J. Acoust. Soc. Am. **112**, 2666–2679 (2002).
[6] V. E. Nazarov, A. V. Radostin, L. A. Ostrovsky, and I. A. Soustova, ''Wave processes in media with hysteretic nonlinearity, Part 1,'' Acoust. Phys. **49**, 344–353 (2003).
[7] V. E. Nazarov, A. V. Radostin, L. A. Ostrovsky, and I. A. Soustova, ''Wave processes in media with hysteretic nonlinearity, Part 2,'' Acoust. Phys. **49**, 529–534 (2003).
[8] V. E. Nazarov, L. A. Ostrovskii, I. A. Soustova, and A. M. Sutin, ''Anomalous acoustic nonlinearity in metals,'' Sov. Phys. Acoust. **34**, 285–289 (1988).
[9] V. Gusev, H. Baillet, P. Lotton, and M. Bruneau, ''Interaction of counterpropagating acoustic waves in media with nonlinear dissipation and hysteretic media,'' Wave Motion **29**, 211–221 (1999).
[10] V. Gusev, ''Parametric attenuation and amplification of acoustic waves in the media with hysteretic quadratic nonlinearity,'' Phys. Lett. A **271**, 100–109 (2000).

Nonlinear acoustic techniques for landmine detection[a)]

Murray S. Korman[b)]
U.S. Army CERDEC, RDECOM, Night Vision and Electronics Sensors Directorate, 10221 Burbeck Road, Fort Belvoir, Virginia 22060 and Department of Physics, U.S. Naval Academy, Annapolis, Maryland 21402

James M. Sabatier
U.S. Army CERDEC, RDECOM, Night Vision and Electronics Sensors Directorate, 10221 Burbeck Road, Fort Belvoir, Virginia 22060 and National Center for Physical Acoustics, University of Mississippi, University, MS 38677

(Received 24 July 2003; revised 14 July 2004; accepted 20 August 2004)

Measurements of the top surface vibration of a buried (inert) VS 2.2 anti-tank plastic landmine reveal significant resonances in the frequency range between 80 and 650 Hz. Resonances from measurements of the normal component of the acoustically induced soil surface particle velocity (due to sufficient acoustic-to-seismic coupling) have been used in detection schemes. Since the interface between the top plate and the soil responds nonlinearly to pressure fluctuations, characteristics of landmines, the soil, and the interface are rich in nonlinear physics and allow for a method of buried landmine detection not previously exploited. Tuning curve experiments (revealing ''softening'' and a back-bone curve linear in particle velocity amplitude versus frequency) help characterize the nonlinear resonant behavior of the soil-landmine oscillator. The results appear to exhibit the characteristics of nonlinear mesoscopic elastic behavior, which is explored. When two primary waves f_1 and f_2 drive the soil over the mine near resonance, a rich spectrum of nonlinearly generated tones is measured with a geophone on the surface over the buried landmine in agreement with Donskoy [SPIE Proc. **3392**, 221–217 (1998); **3710**, 239–246 (1999)]. In profiling, particular nonlinear tonals can improve the contrast ratio compared to using either primary tone in the spectrum. [DOI: 10.1121/1.1814119]

PACS numbers: 43.25.Ts, 43.40.Fz, 43.40.Ga [MFH] Pages: 3354–3369

I. INTRODUCTION

A. Historical background

Acoustic-to-seismic coupling[1,2] has proven to be an extremely accurate technology for locating buried landmines. Focusing mainly on linear acoustic techniques, airborne sound couples into the ground, interacts with the buried mine, and causes increased vibration of the ground above the mine—at the surface. Recently, Donskoy has suggested a nonlinear vibro-acoustic technique[3–5] involving two primary frequencies which couple into the ground and interact nonlinearly with the soil over the top surface of the mine generating nonlinear frequency components that can also be detected at the surface. The nonlinear mechanism involves the soil and the relatively compliant mine to be in contact during the compressional phase of the wave but separating during the tensile phase. Therefore false alarms should be eliminated for relatively noncompliant buried objects. Linear and nonlinear tests have been compared at Army test lanes indicating that nonlinear spatial profiles over buried landmines have more contrast sensitivity than linear measurements.[6] To further understand the nonlinear mechanisms involved, a set of nonlinear tuning curves of the resonant response of the top plate versus amplitude—for several buried landmines—are presented. In one case an inert VS 2.2 anti-tank mine buried 3.6 cm deep (in loess soil) had its peak frequency shift from 90 to 80 Hz as the amplitude of the acceleration increased by a factor of 4. Analogies are made to experimental nonlinear tuning curves for rocks (like sandstone) and for soil alone where the theoretical mechanisms have been established.[7]

1. On "classical" and "nonclassical" nonlinearity

If the excitation of a system is due to a harmonic force that contains only a single ''primary'' frequency component, and if the system's oscillation response contains frequency components other than the primary (i.e., harmonics, subharmonics), then we say that the system exhibits nonlinear behavior. When two primary components (labeled by frequencies f_1 and f_2) excite a system, then nonlinear behavior may be observed in the study of the measured amplitudes of the so-called combination frequencies $|mf_1 \pm nf_2|$ where m and n are integers. To lowest order the combination frequencies are the sum and difference frequencies, f_1+f_2 and f_1-f_2, respectively.

The nonlinear elastic behavior of earth materials (including rock, soils, or granular material) is an extremely rich topic[8] (see the review article by L. A. Ostrovsky and P. A.

[a)]Work presented at the 146th Meeting of the Acoustical Society of America, Austin, TX, 10–14 November 2003. Portions of this work have been presented at SPIE' AeroSense Conference, 21–25 April 2003, Orlando, FL, *Proc. SPIE 17th*, Conference on Detection and Remediation Technologies for Mines and Minelike Targets VIII, edited by R. S. Harmon, J. H. Holloway, and J. T. Broach, ''Nonlinear tuning curve vibration response of a buried landmine,'' **5089**, paper no. 57 (2003); also at the 16th International Symposium on Nonlinear Acoustics, Moscow, Russia 19–23 August 2002, session on Devices and Industrial Applications of Nonlinear Acoustics.
[b)]Electronic mail: korman@usna.edu

Johnson[9]), and some distinction must be made for characterizing the *structural nonlinear elasticity* behavior (also called *mesoscopic/nanoscale elasticity*) of this class of materials compared to the *atomic elastic* behavior such as conventional fluids and monocrystalline solids.[10] The nonlinearity of atomic elastic materials (due to the anharmonicity of the lattice on an atomic or molecular level) is small because the intermolecular forces are large. However, for the geomaterials, the cracks, contacts, and dislocations of the so-called *bond system* forces (formed by the hard matrix of grains and crystals—exhibiting ''classical'' nonlinearity) are relatively small. These geomaterials are said to exhibit *slow dynamic effects*, involving hysteresis and relaxation phenomenon.

2. Objectives

This paper involves a study of nonlinear acoustic mine detection and demonstrates that some of the nonlinear oscillatory experiments reported on here exhibit the same characteristics as mesoscopic nonlinear behavior. Is it possible that the mechanisms that result in mesoscopic behavior are also present in the landmine detection problem? Here, an inert VS 2.2 anti-tank mine[11] is used in the investigation.

Six experiments are reported. In experiment 1, nonlinear tuning curves are measured involving a set of rms acceleration measurements of the top plate of the buried mine versus frequency, using a small accelerometer. A family of tuning curves is obtained corresponding to a set of airborne constant amplitude swept sinusoidal acoustic pressure drive levels that incrementally increase from the previous curve. Here the mine, which rests on a smooth thick concrete slab (located outdoors at the Army test site), is buried in finely sifted, very dry, loess soil (a clay–silt mix).

In experiment 2, the apparatus is taken to the gravel lane 14 at the Army test site, to perform acoustic experiments on another VS 2.2 landmine (which has been buried for several years). In this case nonlinear tuning curves are measured involving a set of rms particle velocity measurements of the ground (above the center of the mine) versus frequency using the swept sinusoid technique described in experiment 1.

Experiment 3 uses the setup in experiment 2 with a few modifications to perform a two-tone nonlinear test on the buried VS 2.2 mine in the gravel lane 14. Here, two subwoofers are individually driven to produce individual acoustic pressure amplitudes at the so-called primary components which are designated by the frequencies f_1 and f_2, respectively. The geophone used in experiment 2 is also used in experiment 3 to detect the nonlinearly generated harmonics and combination frequencies that are produced by the interaction of the primaries which couple into the ground and interact nonlinearly with the soil and the buried landmine. The geophone is repositioned across the soil surface so that the rms particle velocity profile corresponding to each tonal in the received spectrum can be measured.

The nonlinear experiments described in experiment 3 are repeated in experiment 4, where the VS 2.2 mine (used in experiment 1) is buried in finely sifted, very dry, loess soil. Here, a layer of soil is sandwiched between the concrete slab and the bottom of the mine.

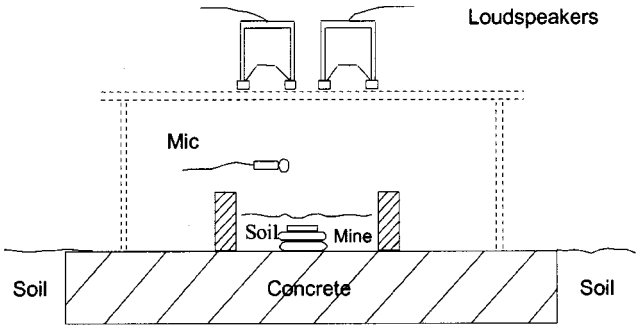


FIG. 1. Experimental setup of the inert VS 2.2 landmine for measuring the mine's top plate acceleration versus frequency for a variety of soil layer depths. Concrete slab is located outside at the Army test site (7 August 2002).

In experiment 5, an apparatus called the soil-plate oscillator is used to model the buried landmine as a thin clamped circular elastic plate (mounted within a short rigid tube) with a layer of finely sifted loess soil placed over the plate. A geophone that rests on the surface of the soil is used to measure nonlinear tuning curves in the manner described above. From these tuning curves, first measured with increasing drive amplitudes, then decreasing drive amplitudes, one can measure effects that appear to exhibit mesoscopic nonlinear elastic behavior including slow dynamics and hysteresis.

The final experiment 6 is a nonlinear two-tone test study involving another VS 2.2 landmine buried in sifted, very dry, loess soil at the National Center for Physical Acoustics field test site. Here, the nonlinearly generated difference frequency component mean-squared velocity profile across the mine is found to be more narrow compared to the profile results when the experiment is repeated (using linear techniques) where a single primary frequency is chosen to be located at the difference frequency. Whereas in experiments 2–4 the axes of the subwoofers are perpendicular to the ground plane, in experiment 6, the axes of the subwoofers are nearly horizontal to the ground in a ''forward looking'' arrangement.

An additional experiment involving a soil-mass oscillator has been placed in the Appendix. Here, the effective longitudinal sound speed is measured and an attempt is made to quantify the soil nonlinearity using resonant tuning curve behavior. The relative frequency shift with vibration amplitude is then compared with the buried landmine and with the soil-plate oscillator tuning curve results.

We now present these experiments and attempt to show a strong connection to the mesoscopic nonlinear phenomenon we have outlined above.

II. EXPERIMENTAL RESULTS

A. Experiment 1: Resonant behavior of a VS 2.2 landmine in sifted soil

Figure 1 shows the experimental setup of the inert VS 2.2 landmine that is placed on a smooth concrete slab. Not shown in the figure is a small accelerometer that is glued onto the center of the mine's top plate. The acceleration response and the acoustic pressure response (above the center of the top plate or the soil over it) versus acoustic excitation

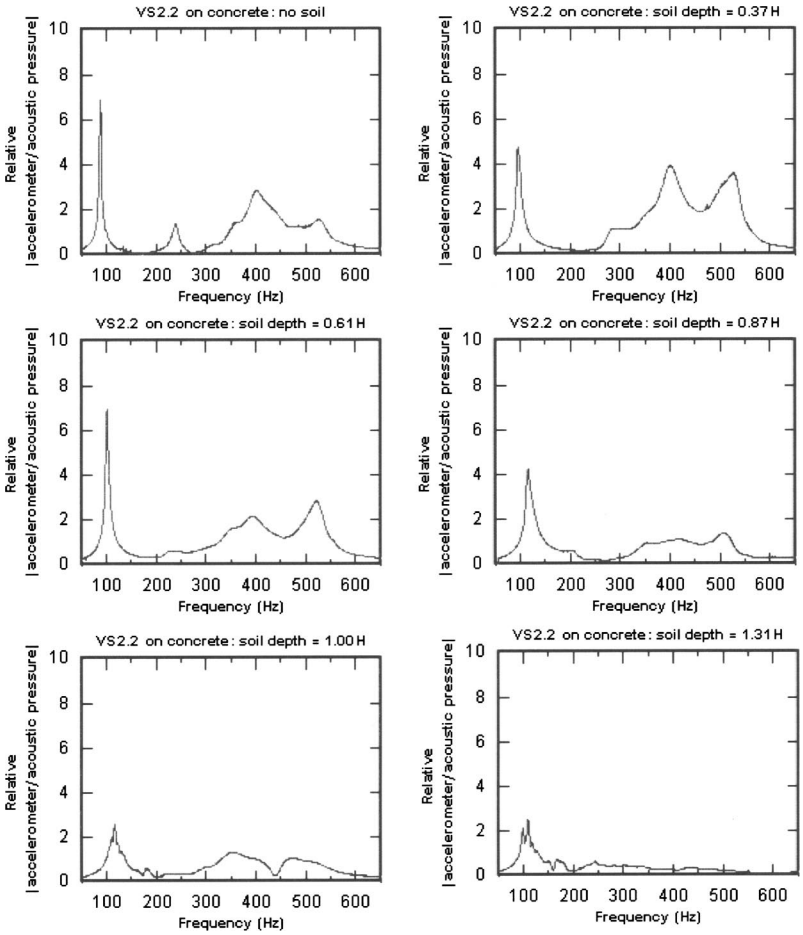


FIG. 2. The response versus frequency of the center of the top plate of the VS 2.2 resting on the concrete, with no soil, with soil layers 0.37 H, 0.61 H, 0.87 H deep, with a soil layer that is flush with the height of the mine 1.00 H and 1.31 H deep. The mine's height is H=11.5 cm. Multiply the relative vertical scale by 0.0167 $(m/s^2)/Pa$ for actual values.

frequency are recorded using an Agilent 35670A dynamic spectrum analyzer to find the location of its resonances. Here, an amplified swept sinusoidal drive signal is applied to both the subwoofers using a tee connector. The woofers (30.5 cm diam) are located 80 cm above the concrete on a light parallel bar plastic frame that is 1.5 m long and 30 cm wide. The soil is supported in a concrete rectangle, with a 61 cm by 61 cm inside dimension, that is 23 cm high and 7.6 cm thick. The experiment is repeated several times using homogeneous fine sifted loess soil to bury the mine at various depths in order to keep track of how the resonant frequency changes with soil mass loading.

Vibration studies of the VS 2.2 landmine placed on the concrete surface, first with no soil and then with different thickness layers of loess soil burying the mine, are studied. Figure 2 shows the relative response of the |accelerometer's complex voltage signal/microphone's complex voltage signal| versus frequency between 50 and 650 Hz for several cases. In each case the amplitude acoustic drive level is set very low. It is observed that for the lowest frequency mode, the frequency increases with mass loading up to the point where the soil is flush with the top plate of the mine, while further increase in soil depth exhibits a decrease in frequency.

Next, a set of tuning curves is obtained for the buried landmine for the case where the soil layer is 3.6 cm above the top of the mine. The results are shown in Fig. 3(a) in the frequency range of interest which covers the lowest resonant peaks. Nonlinear tuning curves are generated by sinusoidally sweeping the subwoofers between 50 and 150 Hz at different drive amplitudes. Notice that as the acoustic drive amplitude increases, the peak frequencies become shifted to lower frequencies, suggesting that the nonlinear interaction is getting ''softer.'' A plot of the peak resonant accelerometer signal voltage versus the corresponding resonant frequency—for each peak—exhibits linear behavior. See Fig. 3(b). The results shown in Fig. 3 involve acoustic measurements performed in finely sifted loess soil. The linear behavior of the so-called backbone curves in Fig. 3(b) are characteristic of geophysical materials which exhibit *mesoscopic/nanoscale elasticity* undergoing vibration near a resonance. (See Ref. 9, pp. 8–9.)

These results are now compared to experiments where a VS 2.2 has been buried over a long period of time in a gravel lane at the Army test site.

B. Experiment 2: Resonant behavior of a VS 2.2 landmine in weathered gravel

Experiments involving a VS 2.2 landmine are now performed in the gravel lane 14 at the Army test site, on 7 October 2002. The mine has been buried at 1 in. and has been in the ground for several years. The experimental setup is shown in Fig. 4.

A geophone (Input/Output, Inc. model SM-11) located at the center of the mine on the surface and a microphone

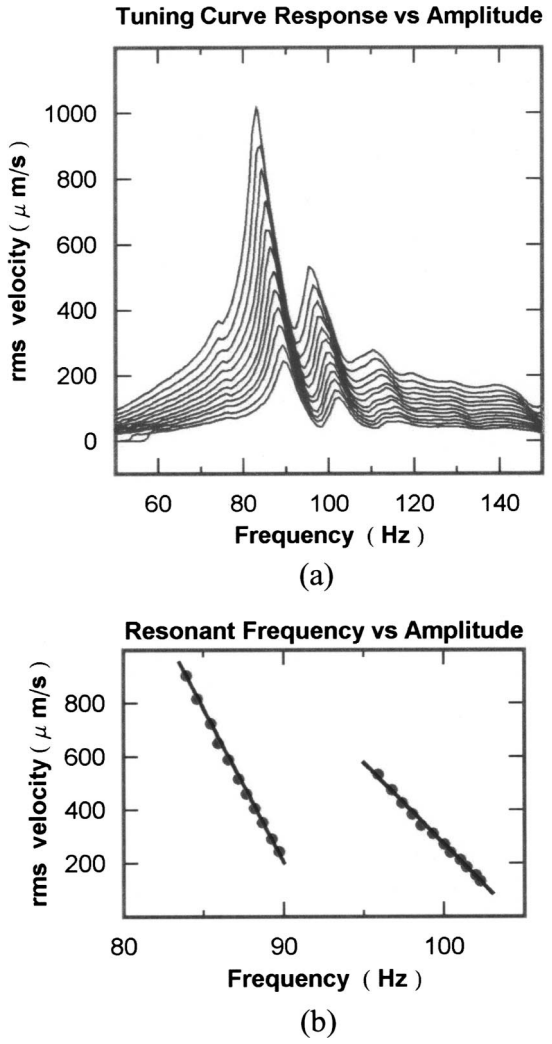


FIG. 3. (a) The relative velocity response versus frequency of the center of the top plate of the VS 2.2 resting on the concrete with a soil layer that is 3.6 cm above the height of the mine. The family of nonlinear tuning curves suggests that for increasing acoustic drive the nonlinear interaction is getting ''softer.'' (b) The ''back-bone'' curve near resonance of the peak velocity versus the corresponding frequency is linear for two closely separated peaks in (a).

(B&K $\frac{1}{2}$-in.-diam condenser type) located at 1 in. above the soil at the center of the mine are used to measure the rms geophone/microphone response function versus frequency. Here, the subwoofer loudspeakers (connected by a tee) are driven by an amplified swept sinusoidal tone in the frequency range of 50 to 400 Hz. The airborne sound coupling

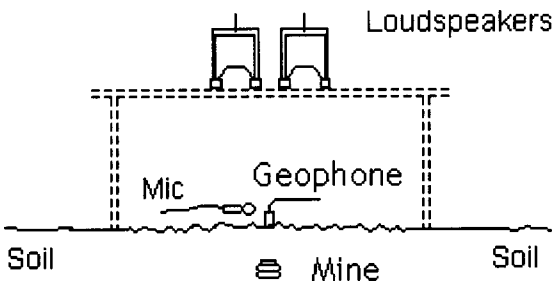


FIG. 4. Experimental setup of the inert VS 2.2 landmine buried at 1 in. deep in gravel lane 14 at the Army test site (7 October 2002) for measuring the particle velocity versus frequency across the mine.

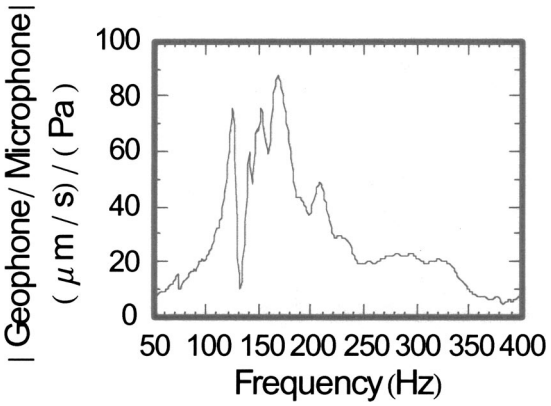


FIG. 5. Geophone/microphone response function for the VS 2.2 buried at 1 in. in lane 14 at the Army test site (7 October 2002).

into the ground shows a strong resonant response near 170 Hz. The results are shown in Fig. 5.

From the geophone/microphone response, one can now choose a suitable frequency band (in this case 120 to 250 Hz) to generate a family of tuning curves—near resonance—for studying the acoustic behavior of the buried VS 2.2 landmine as a nonlinear oscillating system. The tuning curves shown in Fig. 6 are a measure of the geophone rms velocity versus frequency for a series of increasing acoustic drive levels. Figure 7 shows that the bending of the tuning curves towards resonant frequencies that are progressively lower yields a linear relationship between the peak rms particle velocity versus the corresponding resonant frequency. This relationship is characteristic of tuning curve bending in experiments that are performed on rocks and granular material. See pp. 8–9 of the review article by L. A. Ostrovsky and P. A. Johnson[9] for more details.

C. Experiment 3: Nonlinear two-tone results for a VS 2.2 landmine in weathered gravel

1. Background

Donskoy's nonlinear vibro-acoustic technique for buried landmine detection[3,4] involved insonifying the soil above the

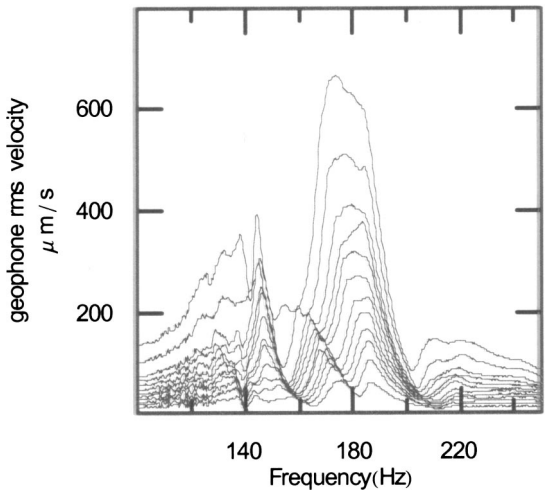


FIG. 6. Tuning curves are shown of the geophone's response versus frequency for increasing levels of acoustic drive amplitude for the VS 2.2 buried in gravel lane 14 at the Army test site (8 October 2002).

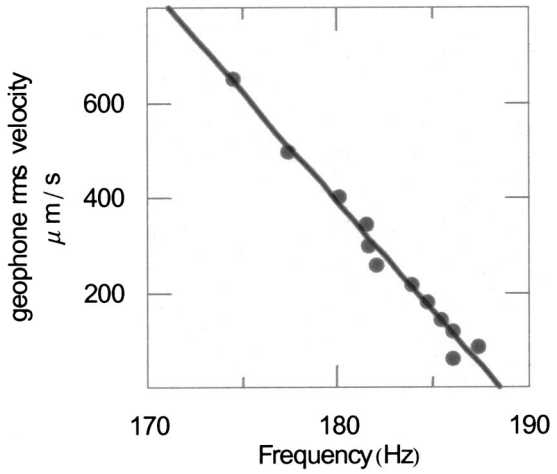


FIG. 7. The geophone's peak resonant response is plotted versus the associated frequency of the peak. The linear behavior is typical of the slow dynamic nonlinear effects observed in rock samples.

mine with two acoustic pressure discrete tones (designated f_1 and f_2) and measuring the amplitudes of the nonlinearly generated combination frequencies with a contact sensor placed on the soil surface above the mine and well off the mine's center. In the absence of the mine or when the mine was replaced by buried objects that have a compressional stiffness K much greater than the mine's stiffness K_m, the combination frequencies greatly diminished. According to Donskoy's lumped element model, it is necessary to have the mine's stiffness be much lower than the stiffness K_s of the soil layer above the mine for the nonlinear coupling to be significant. See Figs. 8(a) and (b).

Donskoy's model reveals the essential points of a more complex nonlinear mechanism where the grains of a soil's surface are to be in contact with an elastic boundary in the compressional phase and perhaps in the tensile phase some lose contact at different locations across the surface; i.e., the opening and closing of soil cracks. All surfaces (of the mine case) may need to be considered in regards to the soil contact problem.

Sabatier *et al.*[6] applied this nonlinear technique to buried landmine detection using a computer controlled laser Doppler vibrometer (LDV) to scan the soil surface over a relatively large area and to systematically study the effects of placing the two primary acoustic tonals (f_1 and f_2) close together and on either side of a strong resonance from the geophone/microphone response function. Sabatier's measurements of each individual nonlinearly generated surface particle velocity component versus scan position for a VS 2.2 buried in a gravel lane demonstrated the usefulness of using nonlinear acoustic detection schemes versus linear acoustic schemes. However, the dynamic range of the LDV can sometimes be a limiting factor in detecting the full contrast ratio involving particle velocity detection on and off the buried landmine in some nonlinear detection schemes.

Concurrently, Donskoy *et al.*[5] also reported LDV field test measurements (on and off the mine) for sum and difference frequencies from their two-tone test excitation. Their test included the nonlinear detection of buried plastic antitank mines (M19, VS 1.6) and plastic antipersonnel mines

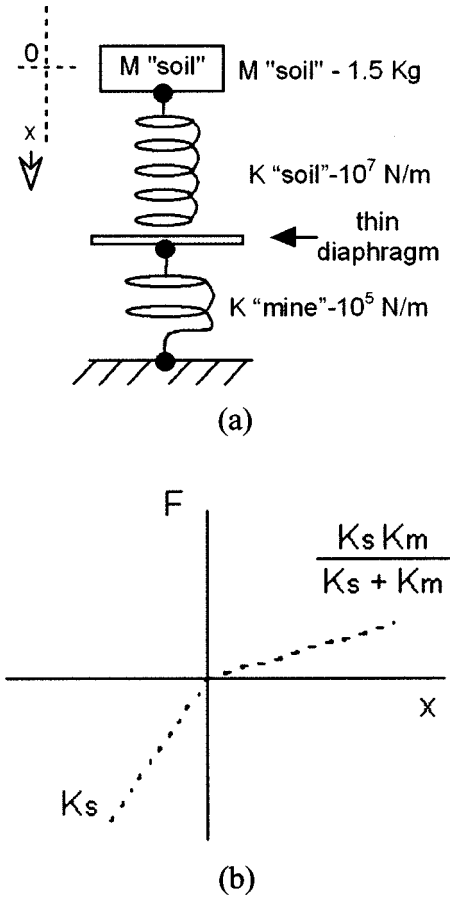


FIG. 8. (a) Donskoy's one-dimensional model with typical parameters. Shear stiffness and damping elements are not shown. (b) The effective force F versus displacement x shows bi-modular behavior. The springs stay together during the compression phase and separate under the tensile phase. In compression, the effective spring constant is the series equivalent of K_s and K_m.

(M14,VS50). They used a buried shaker for one primary frequency (seismic excitation), f_1, and a loudspeaker for the other primary frequency (acoustic-to-seismic excitation), f_2. Both frequencies are swept in increments given by $f_1 = f_{01} + 5n$ Hz and $f_2 = f_{02} - 5n$ Hz, where f_{01} and f_{02} are the initial values of the primaries and the step number $n = 1,2,...,23$, such that the sum frequency remains fixed at $f_{01} + f_{02}$. Donskoy's measurements showed the significance of finding optimal resonant conditions in detecting M19 and VS50 mines over a set of burial depths.

Further, in a laboratory experiment, involving a mine stimulant buried in frozen wet sand, Donskoy's group used a scanning LDV to demonstrate that the nonlinear technique could profile the object, whereas the linear approach was found to be successful only at normal temperatures, due to a reduction in the on/off mine impedance contrast.

2. Results

The nonlinear experiments presented below are motivated in part by extending Donskoy's and Sabatier's work on acoustic landmine detection and in part by the evidence (from the previous experiments reported here) that the non-

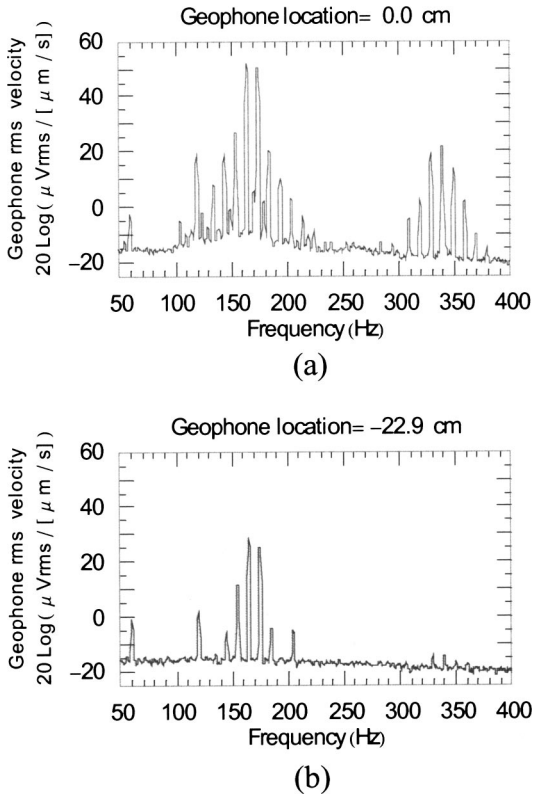


FIG. 9. The primary acoustic components $f_1=165$ Hz and $f_2=175$ Hz (with incident amplitudes $p_{1\mathrm{rms}}=3.5$ Pa, $p_{2\mathrm{rm}}=3.5$ Pa) drive the buried VS 2.2 landmine. Measurements show the nonlinearly generated components that are characteristically separated by 10-Hz intervals across the spectrum. The geophone is located over the center of the mine in (a) and well off the mine in (b).

linearity involves features that may involve mesoscopic elasticity, and feature slow dynamical (relaxation) response and hysteresis effects.

Figures 9(a) and (b) show the results of a nonlinear two-tone test, where the primary acoustic frequency components are $f_1=165$ Hz and $f_2=175$ Hz. The experimental setup shown in Fig. 4 remains the same, except that the individual loud speakers are driven separately with amplified tones at 165 and 175 Hz, respectively. The microphone (located at $x=0.0$ and 2.5 cm above the soil) generated a signal spectrum that was recorded with the spectrum analyzer. The sound pressure levels of interest are $\mathrm{SPL}(f1)=105$ dB, $\mathrm{SPL}(f2)=105$ dB due to the primary wave components and $\mathrm{SPL}(2f1)=68$ dB, $\mathrm{SPL}(2f2)=66$ dB and $\mathrm{SPL}(f1+f2)=60$ dB re 20 μPa indicating that there is some distortion at the source (and some in the microphone as well). The ambient SPL was typically 45 dB.

In Fig. 9(a), the geophone is placed on the soil surface at the location $x=0.0$ cm (on the mine) marking the axis through the center of the buried VS 2.2. In Fig. 9(b), the geophone is placed on the surface at the location $x=-22.9$ cm (off the mine), keeping all other conditions the same. The results presented in Fig. 9 and later on involve an average of 40 FFTs.

It is estimated that the contrast ratio is about 20 dB for the geophone rms velocity when one is looking at either primary component f_1 or f_2. However, nonlinearly generated components near the sum frequency component f_1+f_2 have a contrast ratio of about 40 dB. The nonlinear two-tone experimental results from Fig. 9 are recorded along with

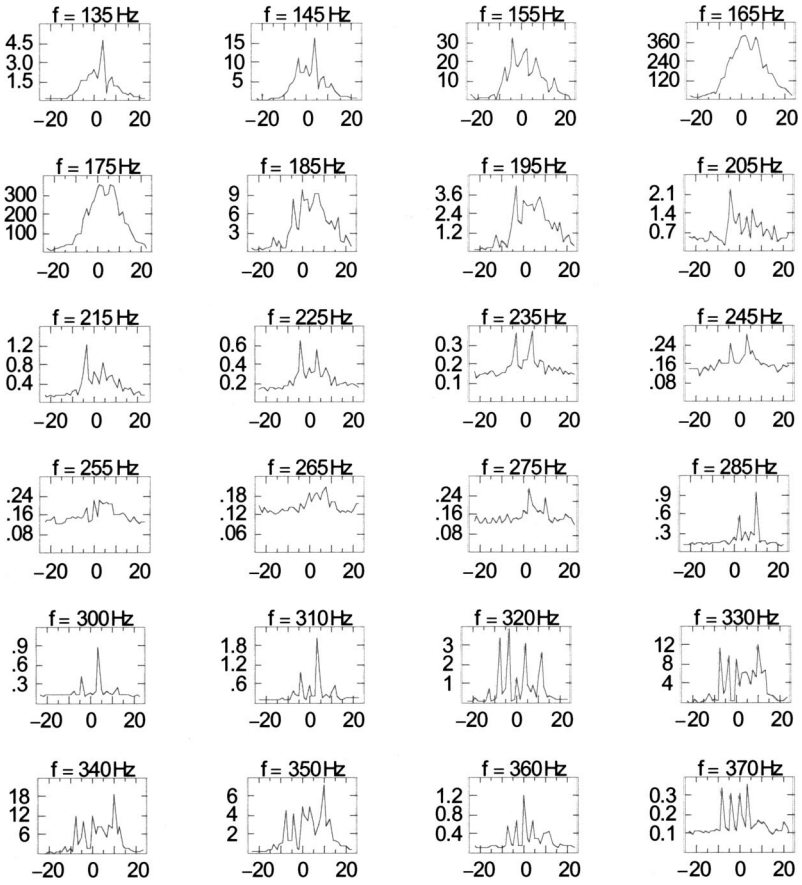


FIG. 10. Geophone rms velocity versus scan position given for individual tonals. The primary frequency components are $f_1=165$ Hz and $f_2=175$ Hz, respectfully. The horizontal scale is in cm and the vertical scale μm/s. The VS 2.2 landmine is buried in gravel.

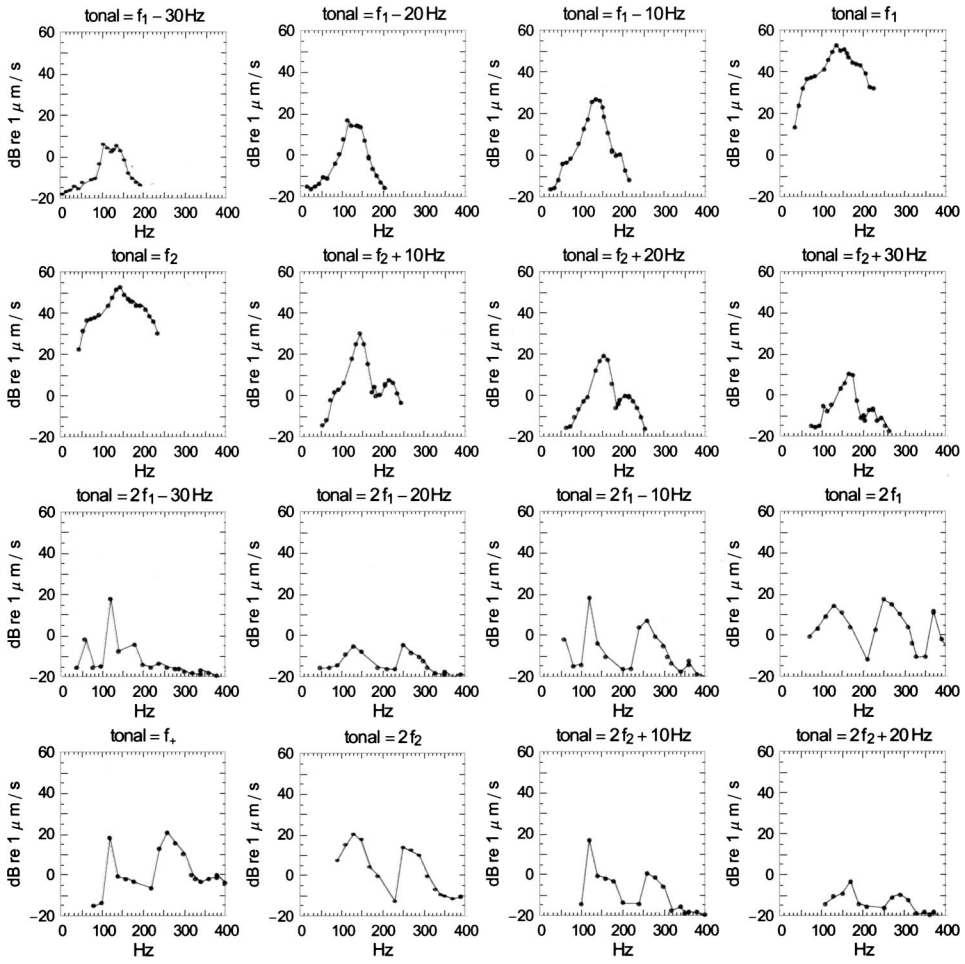


FIG. 11. Results for the ''swept'' nonlinear two tone test experiment involves plotting (for a specific tonal location relative to the two primaries f_1 and f_2) the geophone rms velocity response versus frequency. The two-tone primary excitation frequencies are labeled $f_1[n]$ and $f_2[n]$ where $f_1[n]=(35+10n)$ Hz and $f_2[n]=(45+10n)$ Hz where $n=1,...,15$. First let $f_1[1]=35$ Hz, $f_2[1]=45$ Hz. Then increment by 10 Hz to $f_1[2]=45$ Hz, $f_2[2]=55$ Hz, up to $f_1[15]=225$ Hz and $f_2[15]=235$ Hz. For each of the 15 two-tone tests there will be two primaries and a set of combination tones. When the results are collected for all the sum frequency tonals, designated by tonal$=f_+$, one can observe two peaks in the geophone response versus frequency in the graph located in row 4, column 1. When the results are collected for all the tonals that are 10 Hz below the first primary, designated by tonal$=f_1-10$ Hz, there is a single peak located at 130 Hz, shown in the graph located in row 1, column 3. The graphs at locations row 1, column 4, and row 2, column 1, represent the geophone response versus the primary tonals f_1 and f_2, respectively.

other trials for a set of geophone locations going across the mine. Some of these nonlinearly generated rms velocity profiles are shown in Fig. 10 along with the profiles of the primary components f_1 and f_2.

A two-tone test involving a linear vibrating surface was used to establish that any nonlinearity generated by the geophone and supporting spectrum analyzer was greater than 65 dB down from the primaries. Therefore a 0.2 μm/s threshold is placed on the detection of any nonlinearly generated tone in Fig. 10.

The continuous wave experiments for buried landmine detection involving the two primary frequency components f_1 and f_2 produced side bands of first order, namely a sum frequency component f_1+f_2 and second harmonic generation components $2f_1$ and $2f_2$. (The difference frequency f_2-f_2 was too low to observe in this experiment since the geophone's response is very low for frequencies much less that 100 Hz.) The experiment also produced side bands of second order f_2+2f_1, f_2-2f_1 with significant amplitude, and higher orders as well, indicating strong evidence for nonclassical nonlinearity.[8]

[In Ref. 9 nonclassical nonlinearity is reported for experiments in Berea sandstone where two-tone excitation produced a rich spectrum which included first and second order sidebands. (See pp. 14–15.) Further, in the experimental technique called nonlinear wave modulation spectroscopy (NWMS), an object (such as an automobile engine bearing cap) struck by an impulse while being simultaneously driven by a pure tone continuous wave shows a linear superposition in its spectrum of resonant frequencies along with the pure tone—without a crack. When the part is damaged due to a crack, the spectrum includes a significant nonlinear response characterized by sidebands. (See pp. 40–42.)]

3. Nonlinear two-tone sweep experiment

The geophone/microphone response function (Fig. 5) for the VS 2.2 buried at 1 in. in lane 14 at the Army test site was used to select the two primary frequency components $f_1=165$ Hz and $f_2=175$ Hz, for the two-tone nonlinear experiment, which were chosen to be 10 Hz apart, near the resonant peak of 170 Hz. The following questioned is posed. Given conditions where the resonant frequencies are not known, can one perform a series of two-tone nonlinear experiments where the primary components are $f_1=f$ and $f_2=f+10$ Hz and f ranges in steps from frequencies well below to well above the expected resonant peaks, and discriminate landmines from false alarms?

The ''swept'' nonlinear two-tone test experiment involves plotting (for a specific tonal) the geophone rms velocity response versus frequency. The two-tone primary excitation frequencies are labeled f_1 and f_2 where f_2 is always set at f_1+10 Hz. The data set, shown in Fig. 11, involves utilizing a set of Fourier spectra where first $f_1=35$ Hz, $f_2=45$ Hz, then increment by 10 Hz to $f_1=45$ Hz, $f_2=55$ Hz,..., up to $f_1=225$ Hz and $f_2=235$ Hz. Whereas the

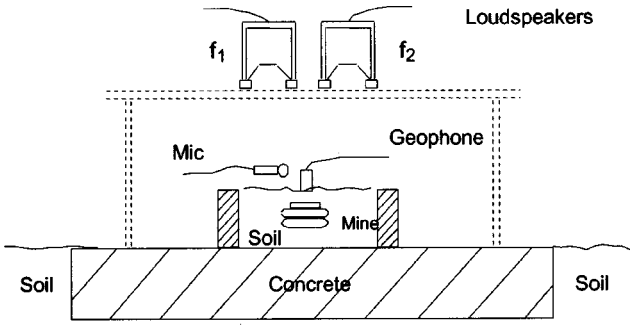


FIG. 12. Experimental setup of the inert VS 2.2 (buried at 3.6 cm below the surface and 7.9 cm above the concrete in sifted loess) for the nonlinear two-tone acoustic test (3 October 2002).

geophone response versus frequency labeled by f_1 or f_2 represent the ''conventional'' linear response of the soil above the center of the mine, the tonal labeled by f_+ represents the sum frequency component's tonal response versus frequency, as the primaries are incremented in steps of 10 Hz. The experiments where done on 8 October 2002 with the VS 2.2 landmine buried in gravel lane 14 at the Army test site.

For example, if one chose to detect the mine with the sum frequency component, f_+, then the peak geophone response occurs at the frequencies 120 and 260 Hz, respectively. Therefore, primary frequency component sets $\{f_1 = 55\,\text{Hz}$ and $f_2 = 65\,\text{Hz}\}$ and $\{f_1 = 125\,\text{Hz}$ and $f_2 = 135\,\text{Hz}\}$ respectively achieve these results. Most of the nonlinearly generated tonals appear to have two similar distinct peaks in their response versus frequency data set.

D. Experiment 4: Laboratory nonlinear mine detection experiment

The VS 2.2 buried at the Army test site in lane 14, a gravel lane, has been exposed to natural weathering for several years. In comparison we now explore a nonlinear acoustic landmine detection experiment where a similar VS 2.2 (the one used in experiment 1) is buried in finely sifted homogeneous loess soil. In particular, the experimental results presented in Fig. 10 (involving the profiles of the geophone response across the gravel surface for individual tonals) will be our base for comparison.

The nonlinear mine detection experimental setup shown in Fig. 12 is similar to Fig. 1 except that the landmine is buried at 3.6 cm below the surface and is 7.9 cm above the concrete slab. Figures 13(a) and (b) show the results of a nonlinear two-tone test, where the primary acoustic frequency components are $f_1 = 75\,\text{Hz}$ and $f_2 = 85\,\text{Hz}$. These primary components are selected from the geophone/microphone response function which shows a strong resonant peak at 80 Hz. The experimental technique shown in Fig. 4 remains the same, except that the individual loud speakers are driven separately with amplified tones at 75 and 85 Hz, respectively. The microphone (located at $x = 0.0$ and 2.5 cm above the soil) generated a signal spectrum that was recorded with the spectrum analyzer. The sound pressure levels of interest are $\text{SPL}(f1) = 110\,\text{dB}$, $\text{SPL}(f2) = 110\,\text{dB}$ due to the primary wave components and $\text{SPL}(2f1) = 70\,\text{dB}$, $\text{SPL}(2f2) = 65\,\text{dB}$ and $\text{SPL}(f1+f2) = 53\,\text{dB}$ re 20 μPa indicating that there is some distortion at the source (and some in the microphone as well). In Fig. 13(a), the geophone is placed on the soil surface at the location $x = 0.0\,\text{cm}$ (on the mine) marking the axis through the center of the buried VS 2.2. In Fig. 13(b), the geophone is placed on the surface at the location $x = -22.9\,\text{cm}$ (off the mine), keeping all other conditions the same.

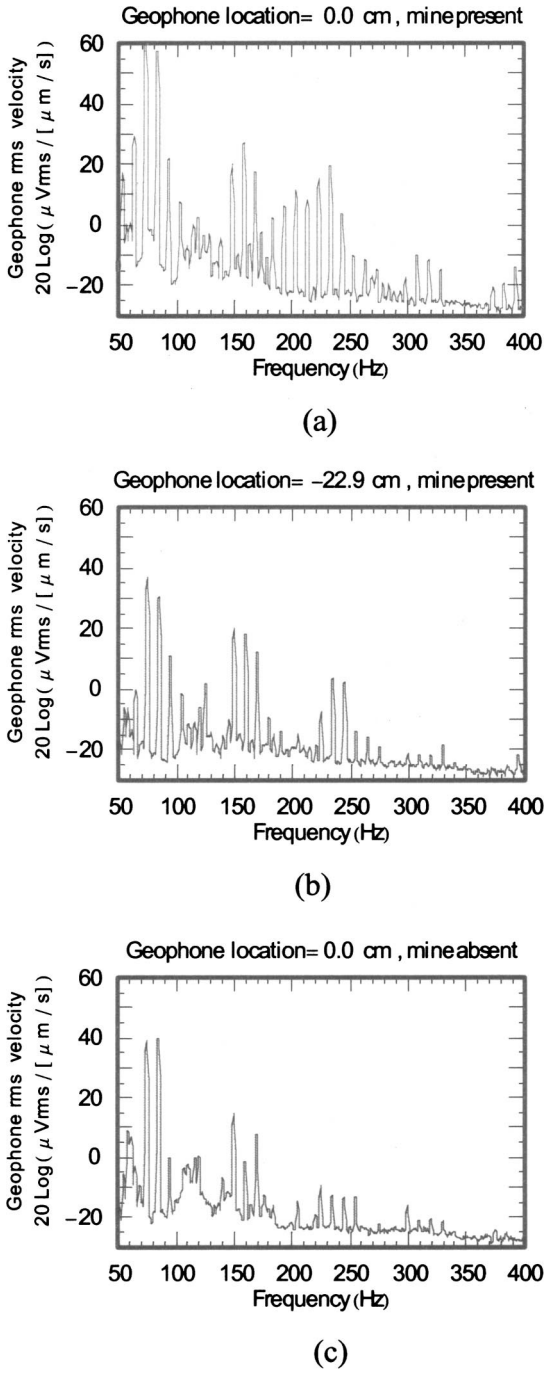


FIG. 13. The primary acoustic components $f_1 = 75\,\text{Hz}$ and $f_2 = 85\,\text{Hz}$ (with incident amplitudes $p_{1\,\text{rms}} = 6.3\,\text{Pa}$, $p_{2\,\text{rms}} = 6.3\,\text{Pa}$) drive the buried VS 2.2 landmine. Measurements show the nonlinearly generated components that are characteristically separated by 10-Hz intervals across the spectrum. The geophone is located over the center of the mine in (a) and well off the mine in (b). In (c) the mine has been removed, the soil has been raked and filled and leveled to the preexisting height, and the geophone placed over the center.

It is estimated that the contrast ratio is about 22 dB for

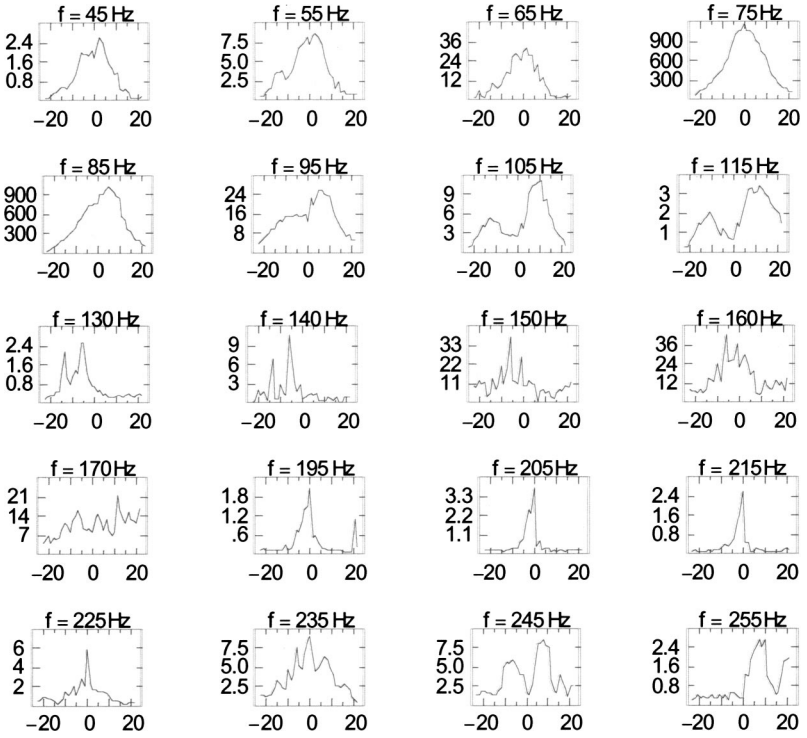


FIG. 14. Geophone rms velocity versus scan position given for individual tonals. The primary frequency components are $f_1 = 75$ Hz and $f_2 = 85$ Hz, respectfully. The horizontal scale is in cm and the vertical scale is μm/s. The VS 2.2 landmine is buried in sifted loess soil.

the geophone rms velocity when one is looking at either primary component f_1 or f_2. However, nonlinearly generated components near the sum frequency component f_1+f_2 have a contrast ratio of about 10 dB. At the end of all the detection experiments the mine is removed and the soil raked, filled in, and leveled. From Fig. 13(c) the contrast ratio improves to 25 dB at the sum frequency component and about 30 dB at the $2f_1+f_2=235$ Hz component.

The nonlinear two-tone experimental results from Fig. 13 are recorded along with other trials for a set of geophone locations going across the mine. Some of these nonlinearly generated rms velocity profiles are shown in Fig. 14 along with the profiles of the primary components f_1 and f_2. (Place a 0.6 μm/s threshold on the detection of any nonlinearly generated tone in Fig. 14 due to geophone nonlinearity.)

The nonlinear two-tone test results for the VS 2.2 buried in finely sifted loess soil demonstrate that nonlinear acoustic mine detection techniques also work for a loosely packed granular medium. However, detailed comparisons between the nonlinearly generated velocity profiles for the individual tonals shown in Figs. 10 and 14 present challenges that probably go beyond understanding the differences between the acoustics of weathered gravel versus finely sifted loess soil. Whereas the results of Fig. 10 involve primaries $f_1 = 165$ Hz and $f_2 = 175$ Hz, and in Fig. 14 $f_1 = 75$ Hz and $f_2 = 85$ Hz, the mass loaded VS 2.2 mine case and soil interface probably involve different ''primary'' modal shapes. In spite of these differences, performing experiments in sifted loess soil, a weathered soil (or another granular material that is made homogeneous) in a limited size concrete rectangular vessel should prove useful in modeling the vibro-acoustic nonlinear interaction. Referring to Fig. 14, several of the profiles (for example $f = 150$, 160, 170 Hz) do not fall off into the noise floor when the geophone is well off the mine. So, the effects of the rectangular boundary on this experiment might prove to be a significant study. The soil-plate oscillator apparatus described in the next section is an attempt to model the buried landmine using a simplified approach.

E. Experiment 5: The soil-plate oscillator

It is necessary to try to model the interaction of a buried landmine with the soil for the purpose of understanding the physics of this complicated nonlinear oscillating system at a fundamental level. The normal modes of a landmine case are further complicated by the soil loading, the moisture in the soil, the layering of the natural soil in the ground as well as the inhomogeneous soil content, vegetation, and weathering. Figure 15 shows a laboratory apparatus called the soil-plate oscillator which is designed to model the oscillating top plate of a buried landmine. It is filled with a layer of finely sifted,

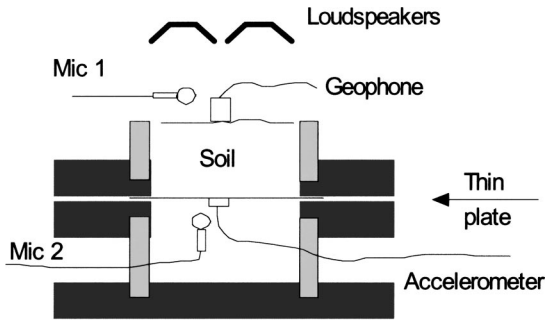


FIG. 15. Cross-sectional diagram of the soil-plate oscillator used for modeling the oscillating top plate of a landmine that is mass loaded with sifted loess soil. The oscillator is setup for the swept sinusoidal acoustic excitation experiment when the speakers are connected by a tee and for a nonlinear two-tone test when they are individually driven at the primary frequencies f_1 and f_2.

dried, crushed and resifted Mississippi loess soil (a clay–silt mix). The density of the loess soil is 1.22 g/cm^3. Both loudspeaker enclosures have a speaker diameter of 8 cm.

The soil-plate oscillator has a thin elastic polycarbonate circular plate that is clamped around the edges by two circular aluminum flanges. The upper flange (thickness$=2L$, i.d.$=8L$, o.d.$=16L$) is fitted with an aluminum tube (thickness$=L$, i.d.$=8L$, length$=2.5L$). A column of soil up to $3.5L$, where $L=2.54$ cm, can be placed in the tube above the circular plate. The polycarbonate plate is 2.78 ± 0.04 mm thick, has a physical diameter of $8.5L$ and is clamped at $8L$, the inside diameter. The density of the plate is 1.23 g/cm^3. The mass of the clamped portion of the plate is $M_{\text{plate}}=111$ g. A small accelerometer is glued to the underside of the plate at the center. The apparatus is placed in the Michelson Hall lab B-2 anechoic chamber facility at the Naval Academy. The temperature was recorded to be 23 °C. Experiments performed with the thin clamped plate (oscillating with no soil loading) show that the frequencies of the lowest normal modes agree with the theory for a thin plate clamped at its edges. The frequency of the fundamental mode, $f_{01}=170.3$ Hz, is obtained using the accelerometer signal versus frequency tuning curve that is measured on the spectrum analyzer operating in the swept sinusoidal mode between 50 and 850 Hz. The $f_{11}=338.5$ Hz and the $f_{20}=677.6$ Hz modes were also observed.

1. Effects of mass loading and tuning curve response

When soil is loaded on the elastic plate, the frequency of the fundamental mode decreases until the frequency reaches a minimum of around 70 Hz. This occurs for a soil mass M_{soil} of about 900 g (or a soil layer 2.3 cm in height). A further increase in mass loading results in an increase in the fundamental mode that is shown in Fig. 16(a). The f_{11} peak reduced to 207.6 Hz when 255 g of soil were added. This resonant peak (from the spectrum analyzer trace) flattened out so that the peak could not be distinguished when further mass was added, while the "f_{21}" mode is not observed at all. Figure 16(b) shows the results of mass loading for the f_{20} (second radial mode) peak. The "f_{20}" radial mode has decreased from 694.2 Hz to about 246 Hz when 1013 g of soil are added. Results depend on the temperature.

Initially it was thought that the soil loading must be deforming the shape of the elastic plate so that the plate's increased stiffness with mass loading dominates the effect of adding elastic material mass to the oscillator. The theory of the fundamental oscillation of a clamped elastic plate under a static load can be modeled as a Duffing oscillator, with a damping term proportional to the velocity, a linear restoring force term, a nonlinear cubic restoring force term, and a static as well as a dynamic forcing term. (See the work done by Farnsworth[12] and Kung.[13]) [*Note added in proof:* Donskoy's work (presented at the 147th Meeting of the ASA, New York, May 2004) suggests that in some cases the effects of the flexural deformation of the soil column involving shear stiffness are more important than the bulk stiffness in predicting the resonant frequency versus depth in buried landmine studies. In response, the authors replaced the soil loading by water and observed a continuous decrease in frequency closely following $f\propto((M_{water}/M_{plate})+1)^{-1/2}$ for $0<M_{water}/M_{plate}<25$. For this oscillator, the plate's in-

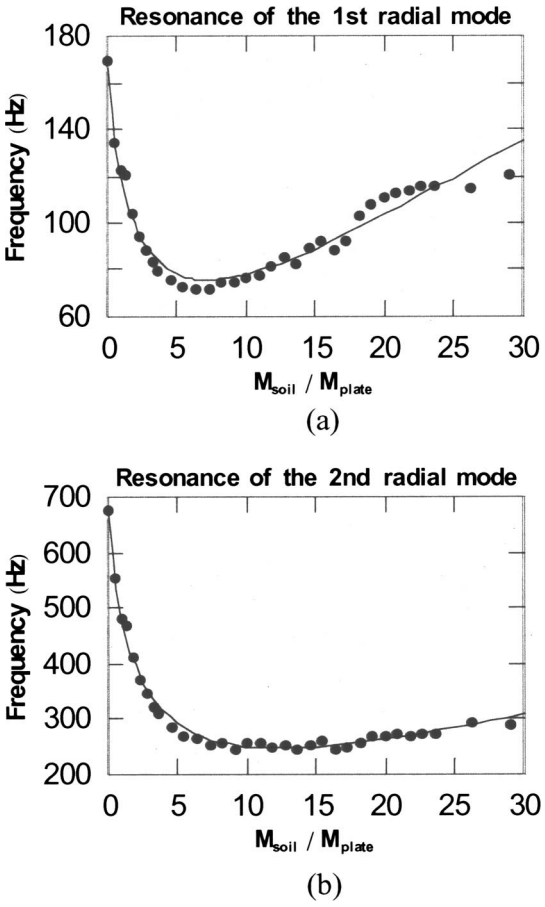


FIG. 16. (a) The fundamental resonant frequency f (first radial mode) of the soil-plate oscillator versus $z=M_{\text{soil}}/M_{\text{plate}}$. The solid line is a semi-empirical fit to the function $f=(A/(z+1)+Bz+Cz^2)^{1/2}$, where $A=27\,800$, $B=241$, and $C=11.4$. (b) The resonant frequency f (second radial mode) of the soil-plate oscillator versus $z=M_{\text{soil}}/M_{\text{plate}}$. The solid line is a semi-empirical fit to the function $f=(A/(z+1)+Bz+Cz^2)^{1/2}$, where $A=449\,000$, $B=1910$, and $C=25.6$.

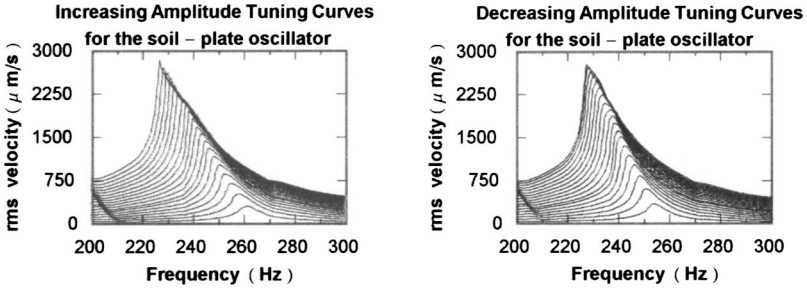


FIG. 17. Soil-plate oscillator nonlinear tuning curves. The data sets are for increasing (a) then decreasing (b) amplitudes.

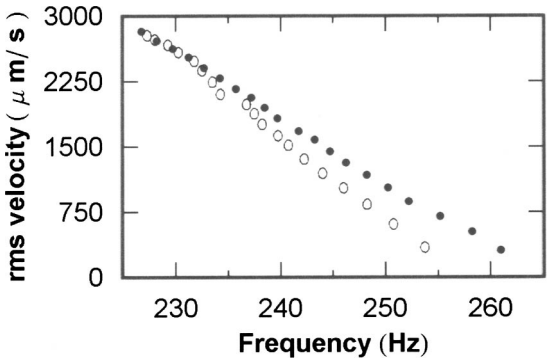


FIG. 18. Soil-plate oscillator tuning curves exhibit hysteresis in the resonant frequency versus corresponding peak geophone signal. Solid circles: increasing pressure amplitude. Open circles: decreasing pressure amplitude.

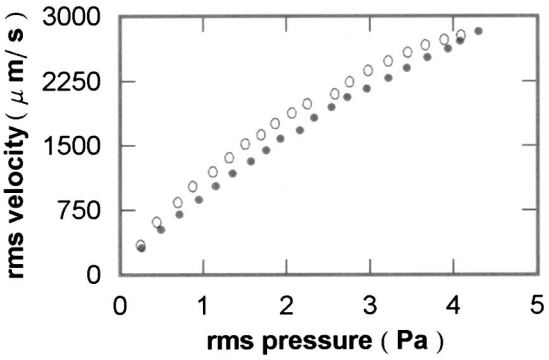


FIG. 19. Soil-plate oscillator tuning curves exhibit hysteresis in the stress (microphone signal) versus strain (geophone signal). Solid circles: increasing pressure amplitude. Open circles: decreasing pressure amplitude.

creased stiffness effects will require a relatively thicker soil column to become more significant.] The experimental data and semi-empirical fit (of the frequency versus mass loading shown in Fig. 16) are for the condition of an infinitesmal amplitude dynamic sinusoidal force.

Tuning curves for the soil-plate oscillator are now obtained in the frequency range between 200 and 300 Hz corresponding to this "f_{20}" radial mode. The soil surface is insonified by an acoustic pressure that is obtained by driving the loud speaker by a swept sinusoidal signal. The geophone (Mark Products model L40A2-100H7510-K6) placed at the center of the soil surface is used to measure the rms particle velocity versus frequency. A miniature electret microphone placed very close to the geophone is used to monitor the incident pressure versus frequency. In Fig. 17(a), the rms velocity versus frequency tuning curves are measured for a series of pressure amplitudes that incrementally increase. The tuning curve data set is collected uniformly in time over a period of 33 min. Then immediately after, under the same experimental conditions, a second set of tuning curves is measured for a series of pressure amplitudes that incrementally decrease starting from the highest amplitude of the previous data set. See Fig. 17(b). The geophone sensitivity calibration is 25.5 V/(m/s) from 200 to 300 Hz. There are 900 g of sifted loess soil resting on the plate for the results in Figs. 17–21.

Figure 18 shows a plot of the peak rms geophone signal versus the corresponding resonant frequency for increasing pressure amplitude tuning curves, along with a second plot for the decreasing pressure amplitude tuning curves. The effects of hysteresis are present. In Fig. 19, the geophone signal (corresponding to the strain) versus the microphone signal (corresponding to the stress) shows nonlinear behavior as well as hysteresis effects.

In Figs. 18 and 19, the geophone calibration sensitivity is 25.5 V/(m/s) and the microphone calibration sensitivity is 7.5 mV/Pa. The results in Fig. 19 are similar to the hysteretic nonlinear behavior of rock elasticity with discrete memory. The experimental results for rock have been predicted by the Priesach-Mayergoyz space (PM space) model. The theory is

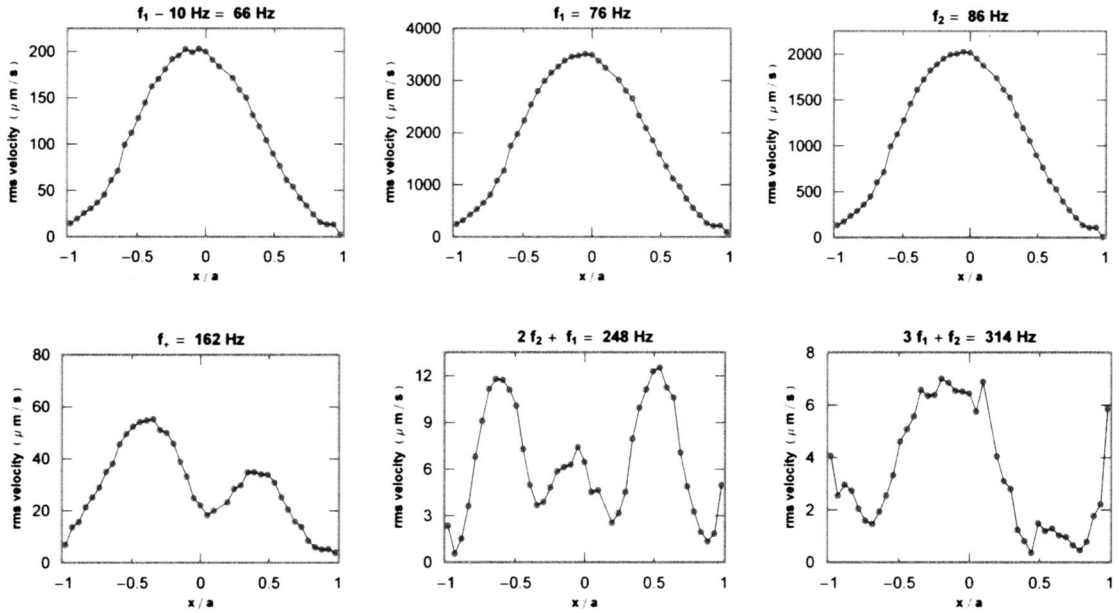


FIG. 20. The nonlinear two-tone test using the soil plate oscillator. LDV rms velocity versus scan position for the primary components $f_1=76$ Hz, $f_2=86$ Hz and the sum frequency component $f_1+f_2=162$ Hz, respectfully. The horizontal scale x/a is the position normalized to the radius a of the clamped plate and the vertical scale is in μm/s.

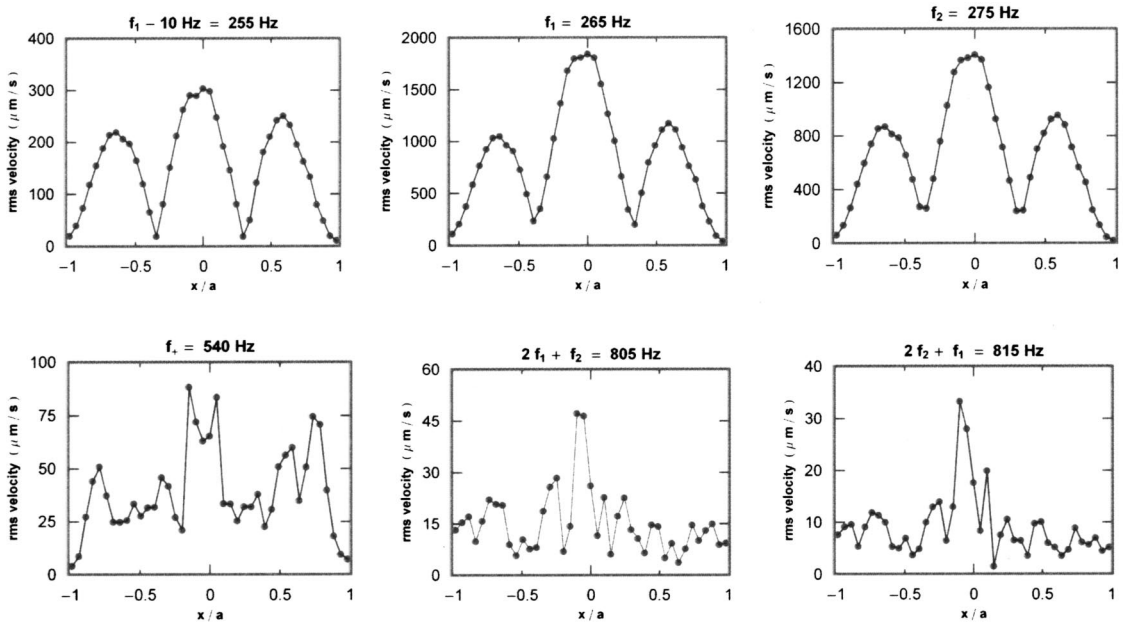


FIG. 21. The nonlinear two-tone test using the soil-plate oscillator. LDV rms velocity versus scan position for the primary components $f_1=265$ Hz, $f_2=275$ Hz and the sum frequency component $f_1+f_2=540$ Hz, respectfully. The horizontal scale x/a is the position normalized to the radius a of the clamped plate and the vertical scale is in μm/s.

summarized in the review article by Ostrovsky and Johnson.[9] (See pp. 22–26.)

2. Two-tone nonlinear test using the soil-plate oscillator

The soil-plate oscillator has been shown to exhibit nonlinear behavior from the tuning curve experiments that to some degree model the behavior of the VS 2.2 that was either buried in fine sifted homogeneous loess soil (Fig. 3) or in the gravel lane 14 at the Army test site (Figs. 6 and 7). Figure 15 shows the experiment for the two-tone nonlinear test using the soil-plate oscillator. The loudspeaker enclosures are metal with perforated grills for shielding.

When the soil-plate oscillator is driven with airborne sound at two closely spaced primary frequencies, the geophone's signal is rich in nonlinearly generated tonals, similar to the experimental Fourier spectrum shown in Fig. 13(a). Here, the primary frequency components $f_1=76$ Hz and $f_2=86$ Hz are chosen around the fundamental resonance "f_{01}" of the oscillator, which was 81 Hz for the experimental conditions. Six profiles of the normal component of the soil surface response versus scan position across the diameter of the soil plane are shown in Fig. 20 for the primary components and several nonlinearly generated components including the sum frequency component $f_+=162$ Hz. The 3.2 cm diameter of the geophone makes it difficult to obtain data points close to the inner cylindrical wall of the apparatus. Therefore, measurements were taken using a Polytec 100 laser Doppler vibrometer (LDV) in the NCPA anechoic chamber facility. Referring to Fig. 15, the LDV profile runs from "out of the page to "into" the page.

Earlier it was mentioned that the "f_{11}" mode exhibited a very broad peak when the polycarbonate clamped plate is loaded with soil. This peak is estimated to be near 150 Hz. Moving now to the more sharply peaked "f_{20}" radial mode (where the peak is estimated to be near 270 Hz), results in Fig. 21 are shown where the nonlinear two-tone test is performed where the primary components are given by $f_1=265$ Hz and $f_2=275$ Hz.

In Fig. 20, the profiles at the primary components $f_1=76$ Hz and $f_2=86$ Hz and the nonlinearly generated f_1-10 Hz$=66$ Hz component closely resemble the fundamental mode shape of a thin plate or membrane. The $f_+=162$ Hz sum frequency component profile has some resemblance to the common f_{11} mode of a timpani drum, while the $2f_2+f_1=248$ Hz component has a mode structure resembling the f_{02} radial mode—with diminished response near the center. The $3f_1+f_2=314$ Hz component is not symmetric. It might be described using a combination of modes with more emphasis placed on the fundamental.

In Fig. 21, the profiles at the primary components $f_1=265$ Hz and $f_2=275$ Hz and the nonlinearly generated f_1-10 Hz$=255$ Hz component closely resemble the mode shape at the second radial mode frequency, f_{02}. The $f_+=540$ Hz sum frequency component exhibits a deteriation in the f_{02} mode along with a small but distinguishable notch at the center. The $2f_1+f_2=805$ Hz and $2f_2+f_1=815$ Hz profiles are more narrow at the center and exhibit very complicated structure away from the center. This structure cannot be explained at this time.

F. Experiment 6: Nonlinear mine detection profiles at the difference frequency

Nonlinear acoustics experiments are performed at the National Center for Physical Acoustics' mine lane field test site that is located in Oxford, MS about 2 miles from the University of Mississippi. The mine lane is made up of a natural Mississippi loess soil that is comprised of a clay–silt

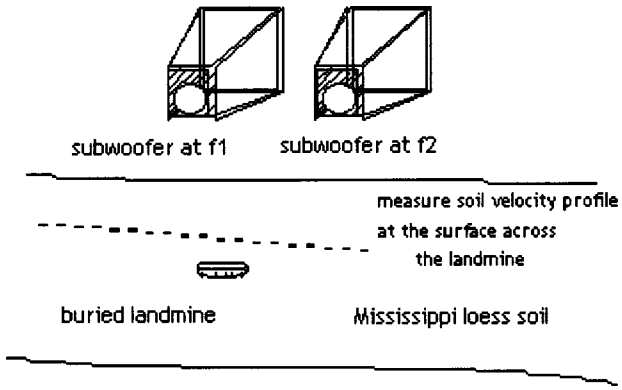


FIG. 22. Experimental setup: Nonlinear acoustic detection of buried land mines. A geophone measures the velocity profile at the f_2-f_1 difference frequency.

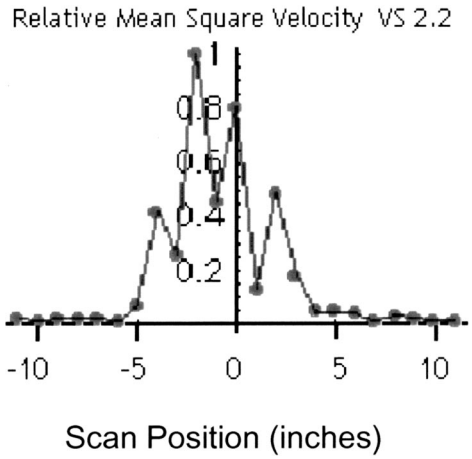


FIG. 23. Mean squared velocity profile at the 115-Hz difference frequency. Horizontal scale is in inches. The primary frequencies are $f_1=800$ Hz and $f_2=915$ Hz. The peak corresponds to 0.2 μm/s. The VS 2.2 is buried in naturally weathered loess soil.

mixture that is free from rocks or even small pebbles. The lane is kept free from grass or other vegetation growth. Figure 22 shows the experimental setup.

Individual loudspeakers (Peavey Impulse 200 Subwoofers) are decoupled from the soil and are placed at a height of 22 cm above the soil's surface. The speakers are separated by 100 cm and centered 186 cm from the center spot of the buried land mine. A VS 2.2 inert anti-tank mine was buried months before the test at a depth of 9 cm below the soil's surface. A model L-10A (100 Hz) geophone (manufactured by Mark Products) was used to measure the velocity profile at the surface across the landmine.

The nonlinear acoustic detection of the buried land mines required a preliminary linear test. First the geophone is placed at the center of the landmine (on the soil's surface) and a swept sine wave is used to drive the speakers which are connected in parallel. A strong linear resonance at 115 Hz was found for the buried VS 2.2. A trial and error study determined that using primary frequencies of $f_1=915$ Hz and $f_2=800$ Hz yielded the strongest difference frequency f_1-f_2 component. Nonlinear experiments are performed with drive voltages of 11.3 and 10.8 Vrms on the speakers labeled S_1 and S_2, respectively, which generate separate primary tones at f_1 and f_2. The geophone had a linear gain of 500 (Tektronix 502 preamp) and was filtered in a 50-Hz bandwidth about the nonlinearly generated difference frequency component using a Stanford Research (model number 650) elliptical bandpass filter.

There is always the question in nonlinear experiments that the receiving electronics is generating the nonlinearities. A field-test check using two tones $f_1=945$ Hz and $f_2=800$ Hz (that excited the geophone suspended in the nearfield of the two speakers) demonstrated that there are no nonlinearites being generated (at the difference frequency) in the electronics at least above the noise floor of the spectrum analyzer.

Figure 23 shows the relative mean square velocity profile on the surface detected at the difference frequency across the center of the buried mine for the VS 2.2.

Here the geophone (whose diameter is 1 in.) is positioned at 1-in. intervals to generate the profile. Clearly, the nonlinear technique is able to detect the buried land mine but one could not determine to what extent the profile is due to the nonlinear effects of the mine-soil interface, or the nonlinearites manifested by inhomogeneities of the soil alone.

These experiments, performed on 20 June 2001, were modified by planning similar experiments where the natural loess soil was excavated forming a 200-cm-diam, 70-cm-deep cavity and painstakingly refilled with very fine sifted—homogeneous and very dry—Mississippi loess soil. The VS 2.2 mine was then buried (at a depth of 3.8 cm) in this pool of fine loess soil on 30 September 2001 and the nonlinear experiments were continued.

Using the setup described earlier, it was determined, from a swept sine wave driving both speakers, that a strong linear resonance occurred at 92 Hz. It was suggested that a linear profile across the mine be taken, for later comparison with the corresponding profile. However, in order to disturb the soil's surface as little as possible, the velocity profile was determined every 4 in. over a 36-in. scan. Next, speakers S_1 and S_2 were driven at frequencies $f_1=300$ Hz and $f_2=392$ Hz and the geophone is positioned at 1-in. intervals over the 36-in. scan. Here the gain on the geophone is 1000 and the bandpass filter is set to pass from 80 to 100 Hz. Each data point corresponds to an average of 100 spectra that are measured on the Agilent spectrum analyzer which was set at a resolution of 2 Hz using a span of 80 to 100 Hz.

From the results of Fig. 24 one can see that the relative mean square velocity profile from the geophone's response is more sharply peaked over the mine and displays more relative sensitivity for the nonlinearly generated difference frequency component of 92 Hz, compared to the linear response at 92 Hz. A detailed velocity profile over the VS 2.2 mine using the nonlinearly generated 92 Hz difference component is shown in Fig. 25.

One might conclude that some of the fluctuating nonlinearity has been eliminated by using a more homogeneous soil. Here, the mean-squared velocity contrast ratio between ''on'' the mine, and well ''off'' is ~50 or 17 dB.

III. DISCUSSION AND CONCLUSIONS

The outcomes of experiments 1–5 clearly exhibit nonlinear behavior that is characteristic of the mesoscopic non-

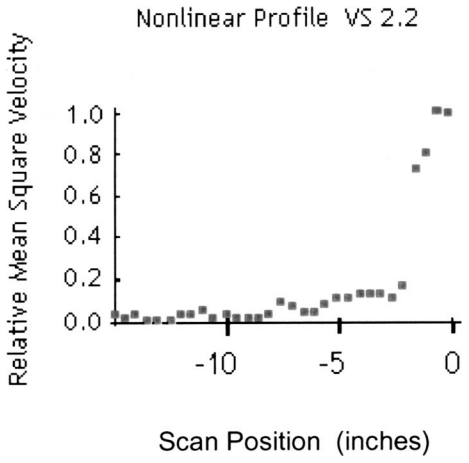


FIG. 25. Using primary frequencies $f_1=300$ Hz and $f_2=392$ Hz the nonlinearly detected profile is obtained at the 92-Hz difference frequency—for the VS 2.2 mine—at finer increments in a second experimental trial. The horizontal scale is 1 in. The peak corresponds to 6 μm/s.

linear behavior that is used to describe rocks and granular material. That is, the amplitudes of the lowest order and higher order nonlinearly generated components are relatively large for low drive level excitation. When tuning curve behavior is measured in experiments 1, 2, and 5, it shows a linear relationship between the resonant amplitude of oscillation versus the corresponding resonant frequency. In these experiments (1,2,5) the tuning curve bending indicates a softer spring constant at higher amplitudes. In the two-tone test experiments reported here, the rich spectrum of nonlinearly generated tones is in agreement with Donskoy,[3,4] whose work provided much of the motivation for the authors' research.

The nonlinearly generated profiles shown in Figs. 10 and 14 and their trends when compared to neighboring tonals present a formidable challenge to the theoreticians. Here, the normal modes of the mine case and their connection with the

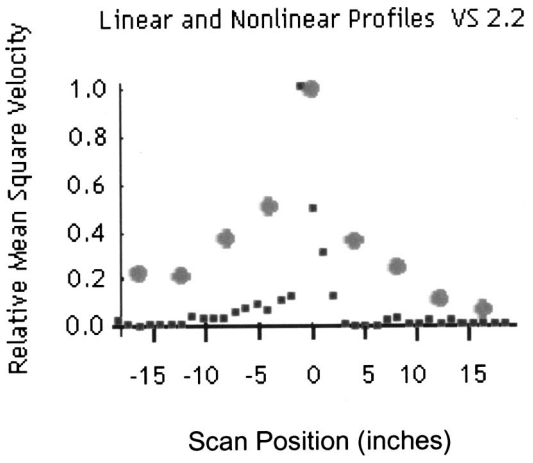


FIG. 24. Using primary frequencies $f_1=300$ Hz and $f_2=392$ Hz the nonlinearly detected profile (small data points) is obtained at the 92-Hz difference frequency and compared with the linear detection profile (large data points) at a primary of 92 Hz—for the VS 2.2 mine. The horizontal scale is 1 in. The linear and nonlinear peaks correspond to 6000 and 9 μm/s, respectively. The primary sound pressure levels just over the soil above the mine are SPL(f_1)=113 dB and SPL(f_2)=114 dB *re* 20 μPa. The VS 2.2 is buried in finely sifted loess soil.

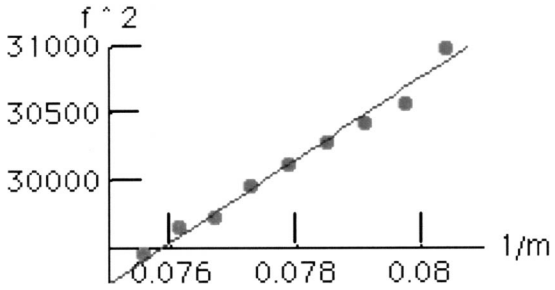


FIG. 26. Soil-mass oscillator results. Horizontal scale: 1/Kg. Vertical scale: Hz^2. Slope=306 700 Kg Hz^2.

nonlinearity at the interface are specific issues that need to be addressed. Further studies using the soil-plate oscillator model might be useful in learning more about the nonlinear dynamics at the interface and predicting the particle velocity profiles (at different tonals) at the soil's surface. A theoretical model of this oscillator should consider both the nonlinearity of the plate due to soil loading as well as the nonlinear soil-plate interface effects that have been suggested by Donskoy.

In experiment 6, the nonlinear profile (at the difference frequency) is shown to be considerably narrower when compared to the linear profile obtained at the same frequency. There is a considerable advantage to using high-frequency (more directive) primary sources since laser Doppler velocimeter measurements of ground vibration require a significant decoupling from the incident sound field. Although no theory has been developed for these nearfield profiles, there might be some comparisons to the nonlinear interaction of sound beams in the presence of an obstacle or bubbly media. (See Ref. 10, pp. 318–331.) Recently, a linear model for landmine detection in an effective fluid (involving acoustic-to-seismic coupling) has been proposed and offers a theoretical prediction of the particle velocity profile at the fluid surface.[14]

In the experimental study of the nonlinear acoustic detection of buried landmines, the characteristics of a family of tuning curves reveals nonlinear behavior that is similar to the nonlinear mesoscopic elastic behavior that describes rocks or granular material. The nonlinearity of soil, the mine case under static and dynamic stress, and the interaction of the soil with the surface of the mine case are all participants in this rich nonlinear mine detection problem. The nonlinear mine detection problem is a worthwhile venture because the nonlinear approach eliminates some types of false alarms and the contrast ratio on and off the mine is improved; however its study and effective application involve using the diagnostic tools of linear acoustic mine detection—a proven technology.[15]

ACKNOWLEDGMENTS

This work is supported by the United States Army Communications-Electronics Command Night Vision and Electronic Sensors Directorate. The authors wish to express their gratitude to Kelly Sherbondy for all his support at Night Vision while they were on sabbatical in 2002. MSK wishes to acknowledge support from the National Center for Physical Acoustics (University of Mississippi), the Physics Department at the U.S. Naval Academy, and the Naval Acad-

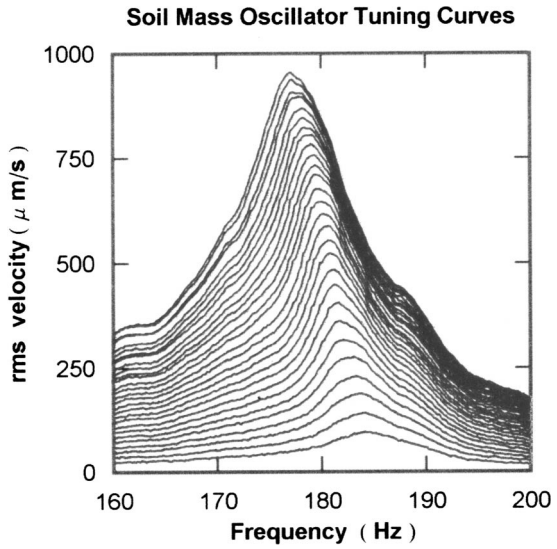


FIG. 27. Soil-mass oscillator nonlinear tuning curves. Data set for increasing levels of acoustic drive amplitude.

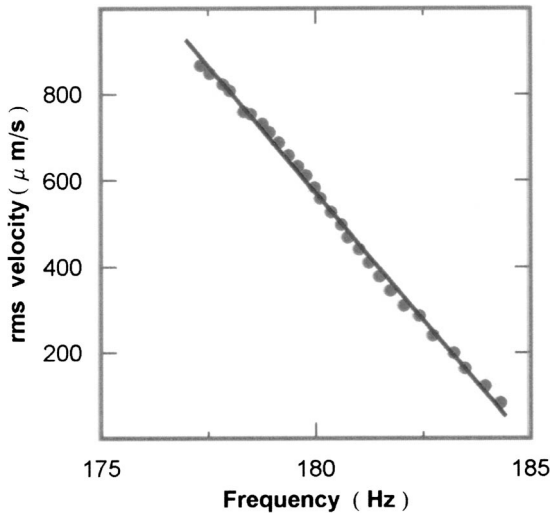


FIG. 28. The peak velocity response is plotted versus the associated frequency of the peak. Solid line: linear fit of the data points.

emy Research Council. Thanks go to Eric Friemark who was involved in the massive soil preparation project at NCPA and to Doug Fenneman and Waini Karen Tai who helped with some of the experiments and data taking and to Roger Waxler, who read the manuscript and provided guidance. Special thanks go to Demitri Donskoy for many discussions on nonlinear acoustic mine detection and to Paul Johnson for discussions on mesoscopic nonlinearity.

APPENDIX: THE SOIL-MASS OSCILLATOR

1. Estimating some elastic properties of a soil sample

An apparatus called the soil-mass oscillator was used to determine the effective longitudinal sound speed in a very finely sifted, dried, crushed and rolled and resifted sample of Mississippi loess soil. Here a rigid rectangular frame of inside dimensions $L_x = 33.0\,\text{cm}$, $L_y = 26.7\,\text{cm}$ is filled to a depth of $L = 8.6\,\text{cm}$ with soil. A concrete plate of mass $m_p = 9.272\,\text{Kg}$ rests on the surface of the soil. The density of the soil is $\rho = 1220\ \text{Kg/m}^3$ and the total soil mass is $m_s = 9.2\,\text{Kg}$. In order to perform the experiment, a pair of subwoofer drivers, located 2 m above the buried concrete slab (roughly 1 m^3), are driven with a swept sinusoidal tone through the fundamental resonance of the system. A spectrum analyzer records the response of an accelerometer that is glued to the surface of the concrete plate. A family of resonance curves is obtained at a very low acoustic drive pressure from the subwoofer, by successively adding 100-g weights to the top of the plate, up to 900 g. At low frequencies (long wavelengths) the effective mass of the soil contributes a value of $m_s/3$. Therefore the total mass of the oscillator is $m = m_p + (m_s/3) + m_a$, where m_a is the added mass.

Figure 26 shows a plot of the square of the peak (resonant) frequency versus the inverse of the total mass of the oscillator. For small damping the spring constant of the soil is given using the relation $\omega = \sqrt{k/m}$, where $\omega = 2\pi f$ and k is the spring constant in N/m. It can be shown that for long wavelengths that the spring constant k can be expressed in the form $k = \rho c^2 S/L$ where S is the cross-sectional area of

TABLE I. Linear fit of the back-bone curve from a set of tuning curves. Results are for a buried VS 2.2 landmine, the soil plate oscillator, and the soil mass oscillator.

Back-bone curve fit function $v = b + m \times f$ f: frequency in Hz	Experimental setup	Comments	Temperature range (°C)
10 580 μm/s−(115 μm) f	VS 2.2 buried in sifted loess soil at 3.6 cm deep	first of two peaks closely spaced	29.5–30.3
6360 μm/s−(60.9 μm) f	VS 2.2 buried in sifted loess soil at 3.6 cm deep	second of two peaks closely spaced	30.5–28.6
8660 μm/s−(60.0 μm) f	VS 2.2 buried in weathered gravel at 2.5 cm deep	one dominant peak	~22
19 880 μm/s−(75.2 μm) f	soil plate oscillator sifted loess soil is 2.3 cm deep	increasing amplitude tuning curves	18.5
24 480 μm/s−(95.2 μm) f	soil plate oscillator sifted loess soil is 2.3 cm deep	decreasing amplitude tuning curves	18.5
21 710 μm/s−(117.5 μm) f	soil mass oscillator sifted loess soil is 9 cm thick	increasing amplitude tuning curves	22.4–22.4
19 900 μm/s−(107.9 μm) f	soil mass oscillator sifted loess soil is 9 cm thick	increasing amplitude tuning curves	36.1–31.8

the soil and L is the soil's thickness. The spring constant is determined to be $k=1.2\times10^7$ N/m and the effective longitudinal sound speed in the soil is $c=98$ m/s.

If the damping force is given by $-rv$, where r is the effective damping constant and v is the velocity of the concrete plate, then one can estimate the damping from $r=m\omega/\mathrm{Q}$. For zero added mass, the resonant frequency $f=261$ Hz and quality factor Q=31.6 were used to determine $r=640$ Kg/s.

2. Nonlinear tuning curves

When the soil-mass oscillator is driven at sufficiently high acoustic pressure amplitudes (while the total mass m remains constant) the tuning curves shift their peak to lower frequencies. This suggests that the soil driven nonlinearly is acting as a soft spring. Figure 27 shows a family of these tuning curves.

If the equation of motion of this oscillator contains weak nonlinear terms in the spring's restoring force, such as a quadratic or a cubic term, then the tuning curve near resonance would show the effects of bending. That is, the locus of points representing the peak amplitude response (taken from the horizontal tangent) versus resonant frequency would bend, but the functional dependence would not be a linear relationship.[16]

For comparison, the nonlinear restoring force is first modeled to have a quadratic nonlinearity which is given by $f_{\mathrm{spring}}=-k_1x-k_2x^2$. The damping force is $f_{\mathrm{damping}}=-k_4dx/dt$ and the externally applied force is $f_{\mathrm{external}}=F_c\cos(\omega t)-F_s\sin(\omega t)$. Using Newton's second law the equation of motion is $f_e+f_s+f_d=md^2x/dt^2$. Using some abbreviations this can be written in the form $d^2x/dt^2+cdx/dt+\alpha x+\beta x^2=H\cos(\omega t)-G\sin(\omega t)$. A trial solution of the form $x=a\cos(\omega t)+b\cos(2\omega t)+d$ can be used to obtain information about the shape of the nonlinear tuning curve. The solution is just given, $S^2(\omega,a)+a^2c^2\omega^2=H^2+G^2$, where S is defined as $S(\omega,a)=a(\alpha-\omega^2)-5a^3\beta^2/6\alpha$. One can obtain horizontal tangents to the frequency response curves from $da/d\omega=0$. Using implicit differentiation one can show that the result for the back-bone curve is $-5a^2\beta^2/6\alpha=\omega^2-\omega_0^2$, where $\omega_0=\alpha-c^2/2$ is the infinitesimal amplitude resonant angular frequency.

In the case where the restoring force is modeled to have a cubic nonlinearity $f_{\mathrm{spring}}=-k_1x-k_3x^3$ the equation of motion becomes $d^2x/dt^2+cdx/dt+\alpha x+\gamma x^3=H\cos(\omega t)-G\sin(\omega t)$, which is the Duffing equation. A trial solution of the form $x=a\cos(\omega t)$ can be used to obtain information about the tuning curve shape. Here, the back-bone curve result is given by $-3a^2\gamma/4=\omega^2-\omega_0^2$.

It is difficult to know from the back-bone curve shown in Fig. 28, what portion of its behavior is due to the conventional (or classical) nonlinearities that are often modeled with higher order terms, and how much is due to the possibility of nonlinear mesoscopic elasticity. However, the back-bone curve is fairly linear, which suggests that a comparison could be made with the other tuning curve measurements presented in this paper.

This comparison is presented in Table I, which shows that the slopes of all the tuning curve experiments have the same order of magnitude.

In regard to nonlinear acoustic mine detection (using tuning curves near a resonance) the ''on the mine''/''off the mine'' contrast ratio might be enhanced for high quality factor, Q, resonance conditions on the mine and low Q factor, far from resonant conditions of ground soil, off the mine. This might explain in part the success of driving the soil-mine oscillator with two tones placed close together on either side of a resonance—in nonlinear acoustic landmine detection.

[1] J. M. Sabatier and N. Xiang, ''Laser-Doppler based acoustic-to-seismic detection of buried mines,'' in *Detection and Remediation Technologies for Mines and Minelike Targets IV*, edited by A. C. Dubey, J. F. Harvey, J. T. Broach, and R. E. Dugan, SPIE Proc. **3710**, 215–222 (1999).

[2] J. M. Sabatier and N. Xiang, ''An investigation of a system that uses acoustic-to-seismic coupling to detect buried anti-tank landmines,'' IEEE Trans. Geosci. Remote Sens. **39**, 1146–1154 (2001).

[3] D. M. Donskoy, ''Nonlinear vibro-acoustic technique for land mine detection,'' in *Detection and Remediation Technologies for Mines and Minelike Targets III*, edited by A. C. Dubey, J. F. Harvey, and J. T. Broach, SPIE Proc. **3392**, 211–217 (1998); ''Detection and discrimination of nonmetallic land mines,'' in *Detection and Remediation Technologies for Mines and Minelike Targets IV*, edited by A. C. Dubey, J. F. Harvey, J. T. Broach, and R. E. Dugan, SPIE Proc. **3710**, 239–246 (1999).

[4] D. M. Donskoy, A. Ekimov, N. Sedunov, and M. Tsionskiy, ''Nonlinear seismo-acoustic land mine detection and discrimination,'' J. Acoust. Soc. Am. **111**, 2705–2714 (2002).

[5] D. M. Donskoy, A. Ekimov, N. Sedunov, and M. Tsionskiy, ''Nonlinear seismo-acoustic land mine detection: field test,'' in *Detection and Remediation Technologies for Mines and Minelike Targets VI*, edited by J. T. Broach, R. S. Harmon, and G. T. Dobeck, SPIE Proc. **4742**, 685–694 (2002).

[6] J. M. Sabatier, M. S. Korman, and N. Xing, ''Linear and nonlinear acoustic velocity profiles over buried landmines,'' in *Detection and Remediation Technologies for Mines and Minelike Targets VII*, edited by J. T. Broach, R. S. Harmon, and G. T. Dobeck, SPIE Proc. **4742**, 695–700 (2002).

[7] P. Johnson, B. Zinszner, and P. Rasolofosaon, ''Resonance and elastic phenomena in rock,'' J. Geophys. Res. **101(B5)**, 11553–11564 (1996).

[8] R. A. Guyer and P. A. Johnson, ''The astonishing case of mesoscopic elastic nonlinearity,'' Phys. Today **52**, 30–35 (1999).

[9] L. A. Ostrovsky and P. A. Johnson, ''Dynamic nonlinear elasticity in geomaterials,'' Riv. Nuovo Cimento **24(7)**, 1–46 (2001).

[10] R. T. Beyer, *Nonlinear Acoustics* (Naval Ship Systems Command, Department of the Navy, 1974), p. 99.

[11] *Jane's Mines and Mine Clearance, 6th ed., 2001–2002*, edited by C. King (Jane's Info. Group, Alexandria, VA.), pp. 183–184.

[12] C. E. Farnsworth and R. M. Evan-Iwanowski, ''Resonance response of nonlinear circular plates subjected to uniform static load,'' Trans. ASME, J. Appl. Mech. **37(4)**, 1043–1049 (1970).

[13] G. C. Kung and Y.-H. Pao, ''Nonlinear flexural vibrations of a clamped circular plate,'' Trans. ASME, J. Appl. Mech. **39(3)**, 1050–1054 (1972).

[14] D. Velea, R. Waxler, and J. M. Sabatier, ''An effective fluid model for landmine detection using acoustic to seismic coupling,'' J. Acoust. Soc. Am. **115**, 1993–2002 (2004).

[15] N. Xiang and J. M. Sabatier, ''An experimental study on antipersonnel landmine detection using acoustic-to-seismic coupling,'' J. Acoust. Soc. Am. **113**, 1333–1341 (2003).

[16] J. J. Stoker, *Nonlinear Vibrations* (Wiley, New York, 1950).

Multiple-order derivatives of a waveguide acoustic field with respect to sound speed, density, and frequency[a)]

Aaron Thode[b)] and Katherine Kim
Marine Physical Laboratory, Scripps Institution of Oceanography, San Diego, California 92093-0205

(Received 22 April 2003; revised 16 March 2004; accepted 19 April 2004)

An adjoint perturbative method is used to derive expressions for the first- through third-order derivatives of a pressure field with respect to sound speed, density, and frequency, for the restricted case of a laterally homogenous waveguide in which environmental parameters are only a function of depth. By using a normal-mode Green's function, the three-dimensional spatial correlation required by the standard acoustic adjoint equation can be reduced to a set of one-dimensional depth integrals. The resulting expressions for the first-order derivative are similar to those obtained by previous perturbative approaches based on the depth-separated wave equation, but the approach followed here permits straightforward extension to higher-order derivatives. Explicit evaluations of the expressions for a representative shallow-water waveguide model are in excellent agreement with numerical finite-difference computations. An analysis of the expressions as a function of source-receiver range finds the contributions to the mode amplitude derivatives to be non-negligible at ranges less than a few modal interference lengths, for parameters associated with the ocean bottom. Therefore, linear perturbative inversion methods that perturb only horizontal wavenumbers and not mode amplitudes should either be used with caution or modified to incorporate the expressions presented here. [DOI: 10.1121/1.1760107]

PACS numbers: 43.30.Pc, 43.30.Bp [WLS] Pages: 3370–3383

I. INTRODUCTION

The nonlinear inversion of ocean acoustic data for environmental parameters often requires estimates of the first- and higher-order derivatives of a modeled acoustic pressure field with respect to various environmental parameters of interest, as well as with respect to frequency. These ''pressure derivatives'' are fundamental theoretical tools in ocean acoustic parameter estimation theory. The first-order pressure derivatives are required to determine the theoretical minimum variance of an environmental parameter estimate, via the Cramer–Rao lower bound (CRLB).[1] Recent work has also demonstrated that if the second- and third-order pressure derivatives can be obtained, then the corresponding limiting bias can be computed, along with an estimate of the number of independent data samples required for a maximum-likelihood estimate to attain the CRLB.[2,3]

Pressure derivatives also arise in applied ocean acoustic work whenever linear relationships between small environmental fluctuations and resulting changes in the acoustic signal are sought. Some applications include the effect of small sound-speed fluctuations on the acoustic signal variance,[4] and inverse perturbative methods for directly extracting environmental parameter estimates from small deviations between a measured and modeled signal.[5–7] In addition, nonlinear inversion procedures, including gradient-search and other local optimization methods,[8] can also exploit information from first- and higher-order pressure derivatives to speed convergence to a local minimum.[9] The derivative of pressure with respect to frequency is also a matter of interest in studies of full-field ocean tomography[10] and the acoustic invariant.[11,12]

The evaluation of these derivatives can be problematic, even for the restricted case of a laterally homogeneous waveguide, where the environmental variation is assumed to be only depth-dependent, and the acoustic field propagating from source depth z_s to receiver depth z_i with frequency ω can be conveniently expressed as a weighted sum of normal modes:[13]

$$p(z_i,z_s,r)=\frac{ie^{-i\pi/4}S(\omega)}{\rho(z_s)\sqrt{8\pi r}}\sum_f U_f(z_i)U_f(z_s)\frac{e^{ik_{rf}r}}{\sqrt{k_{rf}}}. \quad (1)$$

Here k_{rf} is the horizontal wavenumber of mode U_f, $S(\omega)$ is a source strength, ρ is the density of the medium at the source, and $k_{rf}r$ is assumed much greater than one.

One approach for obtaining the pressure derivative of Eq. (1) is to perturb the environment by a small amount, compute a new set of modes, and estimate the pressure derivative via finite differences. However, issues of computational time and numerical stability arise. For example, finite-difference estimates for M parameters require $M+1$ normal mode evaluations to obtain the M first-order pressure derivatives, and approximately $M^2/2$ and $M^3/6$ separate normal mode evaluations to obtain all possible second- and third-order derivatives, respectively. Even a laterally homogenous waveguide model typically has at least six environmental parameters, and thus computational time can be a significant issue, even for only first-order derivatives.

[a)]Portions of this work have been presented at the 143rd (Cancun) Meeting of the Acoustical Society of America.
[b)]Electronic mail: athode@ucsd.edu

A more serious problem is the selection of appropriate step and grid sizes for the finite difference estimates. Not only are additional runs required for a convergence test of these parameters, but the appropriate step size changes with source/receiver horizontal separation. For example, taking the derivative of Eq. (1) with respect to an environmental parameter ε yields terms such as $ir(\partial k_{rf}/\partial\varepsilon)$, indicating that the pressure derivative grows linearly with range, and thus the appropriate finite-difference step size should shrink with range. Thus optimizing finite difference step sizes over a wide set of ranges is not straightforward.

An alternative method for estimating first-order pressure derivatives for a laterally homogenous waveguide involves the application of perturbation theory to the depth-separated component of the wave equation, usually to obtain the derivative of a waveguide horizontal number with respect to the environmental parameter of interest.[5,6] For example, for a sound speed perturbation one obtains

$$\frac{\partial k_{rf}}{\partial a} \equiv -\frac{\omega^2}{k_{rf}} \int_0^\infty \frac{\partial c(z)/\partial a}{c(z)^3 \rho(z)} |U_f(z)|^2 \, dz, \tag{2a}$$

where $c(z)$ is the depth-dependent sound speed profile that is a function of environmental parameter a. These ''wavenumber derivatives'' are then applied to Eq. (1) via the chain rule:

$$\frac{\partial p(z_i, z_s, r)}{\partial a} = -\sqrt{\frac{r}{8\pi}} \frac{e^{-i\pi/4} S(\omega)}{\rho(z_s)} \times \sum_f \left(\frac{\partial k_{rf}}{\partial a} \right) U_f(z_i) U_f(z_s) \frac{e^{ik_{rf} r}}{\sqrt{k_{rf}}}. \tag{2b}$$

These expressions have also been extended by Ref. 14 to incorporate the derivative of the normal mode amplitudes U_f with respect to the environmental parameter, defined here as the ''mode amplitude derivative.'' Equation (2) requires only one normal mode computation, regardless of the number of environmental parameters desired, and the vertical integration can be computed numerically without the convergence and range-dependent issues faced by direct computation of finite differences. However, extending these results to multiple-order derivatives quickly becomes cumbersome, particularly if the modal amplitude derivatives are retained.

In this paper, a third approach for computing multiple-order pressure derivatives is derived using an acoustic adjoint perturbative approach that is well established in the geophysical literature and has recently been extended to range-dependent ocean environments.[15,16] The resulting expressions compactly incorporate the effects of both modal amplitude and wavenumber changes on the pressure field for the first-through third-order derivatives.

After providing some preliminary background for the acoustic adjoint equation in Sec. II A, Sec. II B derives normal-mode expressions for the first- and second-order pressure derivatives with respect to sound speed, density, and frequency. Expressions for the third-order derivative with respect to a single parameter are also given. Section III discusses how the first-order expressions reduce to Eq. (2) when the mode amplitude derivatives are neglected, but illustrates situations where this assumption is *not* valid, and Eq. (2) will not be accurate. Section III concludes by comparing the adjoint pressure derivatives with finite-difference estimates for a representative shallow water ocean waveguide model.

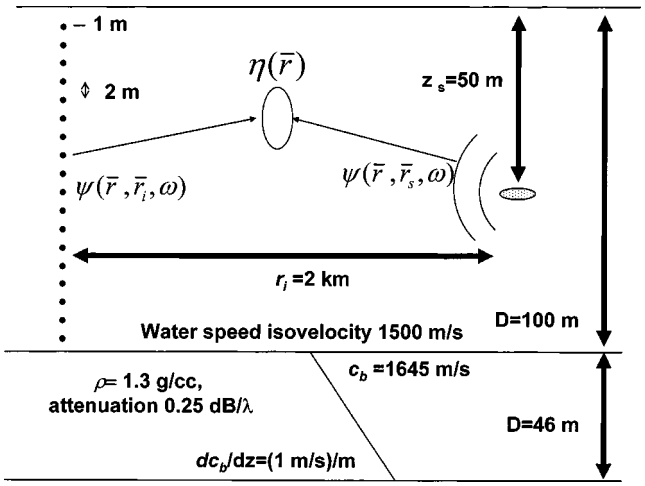


FIG. 1. Waveguide model used to illustrate adjoint formulas. A 48-element vertical array, with 2-m spacing and first element at 1-m depth, is positioned in a 100-m isovelocity waveguide with bottom speed of 1645 m/s, subsurface compressional speed gradient of 1 m/s per m, bottom density 1.3 g/cc, and bottom attenuation of 0.25 dB/wavelength. The acoustic source depth is 50 m. The bottom density is uniform in both the sediment and half-space, which begins 46 m below the sediment/water interface.

II. THEORY

A. Preliminaries

1. Acoustic adjoint equation for environmental pressure derivative

This section outlines the standard derivation of the acoustic adjoint equation, in preparation for the new applications in Sec. II B. More extensive derivations are provided in Refs. 17–20.

Let $g(\overline{r,r_s},\omega)$ define a Green's function for a waveguide acoustic field generated at position $\overline{r_s}$ and measured at location $\overline{r}$ (Fig. 1) at angular frequency ω. The propagation of this field is governed by the density-dependent Helmholtz equation:

$$\rho_0(\overline{r})\nabla\cdot\left(\frac{1}{\rho_0(\overline{r})}\nabla g(\overline{r,r_s},\omega)\right) + k_{ref}^2 \eta_0(\overline{r}) g(\overline{r,r_s},\omega)$$
$$= -\delta(\overline{r}-\overline{r_s}) = \nabla^2 g(\overline{r,r_s},\omega) + k_{ref}^2 \eta_0(\overline{r}) g(\overline{r,r_s},\omega)$$
$$- \frac{1}{\rho_0(\overline{r})}\nabla\rho_0(\overline{r})\cdot\nabla g(\overline{r,r_s},\omega) \equiv L_0 g(\vec{r},\overline{r_s},\omega). \tag{3}$$

Here k_{ref} is a spatially invariant medium wavenumber, while $\eta_0(\overline{r},\omega) \equiv k_0(\vec{r})^2/k_{ref}^2 = c_{ref}^2/c_0(\vec{r})^2$ describes a spatially dependent refractive index squared that defines the propagation environment described by a spatially dependent sound speed profile $c_0(\vec{r})$ and density $\rho_0(\vec{r})$. The linear operator L_0 provides a shorthand way of writing this differential equation. From this point the explicit dependence of g on frequency is dropped, and the sound speed and density are assumed to vary only with depth.

Imagine that the propagation environment is perturbed such that the new propagation environment can be expressed

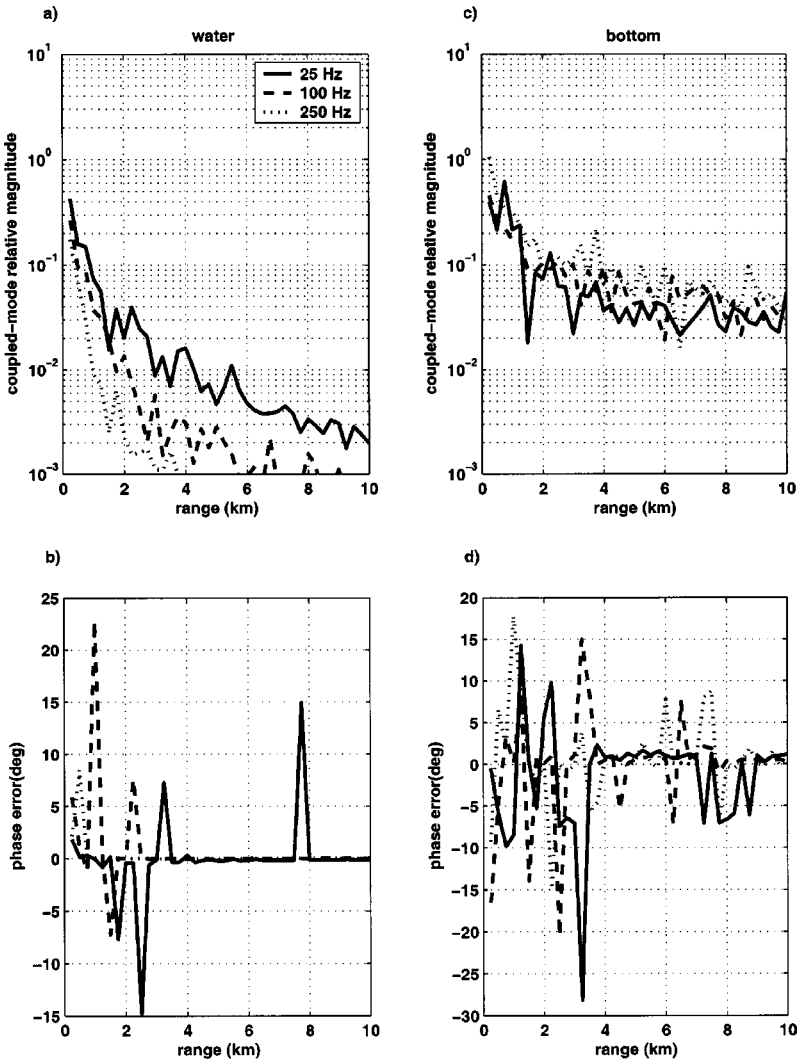


FIG. 2. The relative contributions of mode amplitude derivative terms to the environmental pressure derivative at 25, 100, and 250 Hz, as a function of source-receiver separation. The top row shows the ratio of the magnitude of the sum of the off-diagonal terms in Eq. (18a) to the magnitude of the sum of all terms in Eq. (18a), for derivatives with respect to (a) uniform water column sound speed and (c) bottom interface speed. The bottom row shows the difference in phase between Eqs. (18a) and (2) for derivatives with respect to (b) uniform water column sound speed and (d) bottom interface speed. All quantities have been averaged across all array receiver depths to produce a depth-averaged result. As expected, as the horizontal separation increases, the relative influence of the modal amplitude derivatives decrease.

as a linear sum of a set of basis functions; i.e., $c(z)=c_0(z)+\Sigma_j a_j c_j(z)$, and $\rho(z)=\rho_0(z)+\Sigma_j b_j \rho_j(z)$, with the coefficients a_j and b_j defined here as environmental parameters, and expressions like $c_{a_j} \equiv \partial c/\partial a_j|_{a_j=0}=c_j(z)$, etc., defined as "first-order environmental derivatives." Most physical processes in the ocean can be parametrized in this form.

Given these expressions, asymptotic expansions can be written for the square refractive index and density, based on the set of *infinitesimal* nondimensional environmental parameter magnitudes α_j and β_j. The parameter α_j has the form a_j/c_{ref} with the differential terms a_j having respective units of compressional sound speed (m/s), while β_j has the form $b_j/\max[\rho_0(z)]$, with the density parameter b_j in terms of (g/cc). These normalizations ensure that all α and β are dimensionless and much less than order one. The expansion for a single environmental parameter set α_1 and β_1 thus has the following form:

$$\eta(z) \equiv \eta_0(z)+\alpha_1 \eta_{\alpha_1}(z)+\frac{\alpha_1^2}{2}\eta_{\alpha_1\alpha_1}(z)+\frac{\alpha_1^3}{6}\eta_{\alpha_1\alpha_1\alpha_1}(z)\cdots,$$
$$\rho(z) \equiv \rho_0(z)+\beta_1 \rho_{\beta_1}(z). \tag{4}$$

The addition of other parameters α_2, α_3 etc. would yield additional cross-terms and "mixed derivatives" in Eq. (4). Despite the linear relationship between the perturbations and sound speed, higher-order environmental derivatives like $\eta_{\alpha_1\alpha_1}$ must be retained for the square refractive index, because of the nonlinear relationship between it and the compressional sound speed. For example, $\eta_{\alpha_1}(z)=-2\lfloor c_{ref}^2/c(z)^3 \rfloor c_{\alpha_1}(z)$ and $\eta_{\alpha_1\alpha_1}(z)=6\lfloor c_{ref}^2/c(z)^4 \rfloor c_{\alpha_1}(z)^2 \neq 0$. Note that an environmental derivative for attenuation can be obtained by adding an imaginary term to the refractive index via standard methods.[13]

The Green's function $g(\overline{r,r_s})$ of the perturbed environment can be expressed as a similar expansion, illustrated here for only a single parameter set:

$$g(\overline{r,r_s}) \equiv g_0(\overline{r,r_s})+\alpha_1 g_{\alpha_1}+\beta_1 g_{\beta_1}+\alpha_1\beta_1 g_{\alpha_1\beta_1}+\frac{\alpha_1^2}{2}g_{\alpha_1\alpha_1}+\frac{\beta_1^2}{2}g_{\beta_1\beta_1}+\cdots. \tag{5}$$

Here the order-one terms $g_{\alpha_1} \equiv \partial g/\partial \alpha_1|_{\alpha_1=0}$ and $g_{\alpha_1\beta_1} \equiv \partial^2 g/\partial\alpha_1\partial\beta_1|_{\alpha_1,\beta_1=0}$ are the derivatives of the Green's function with respect to the nondimensional environmental perturbations. Once relationships between these terms and

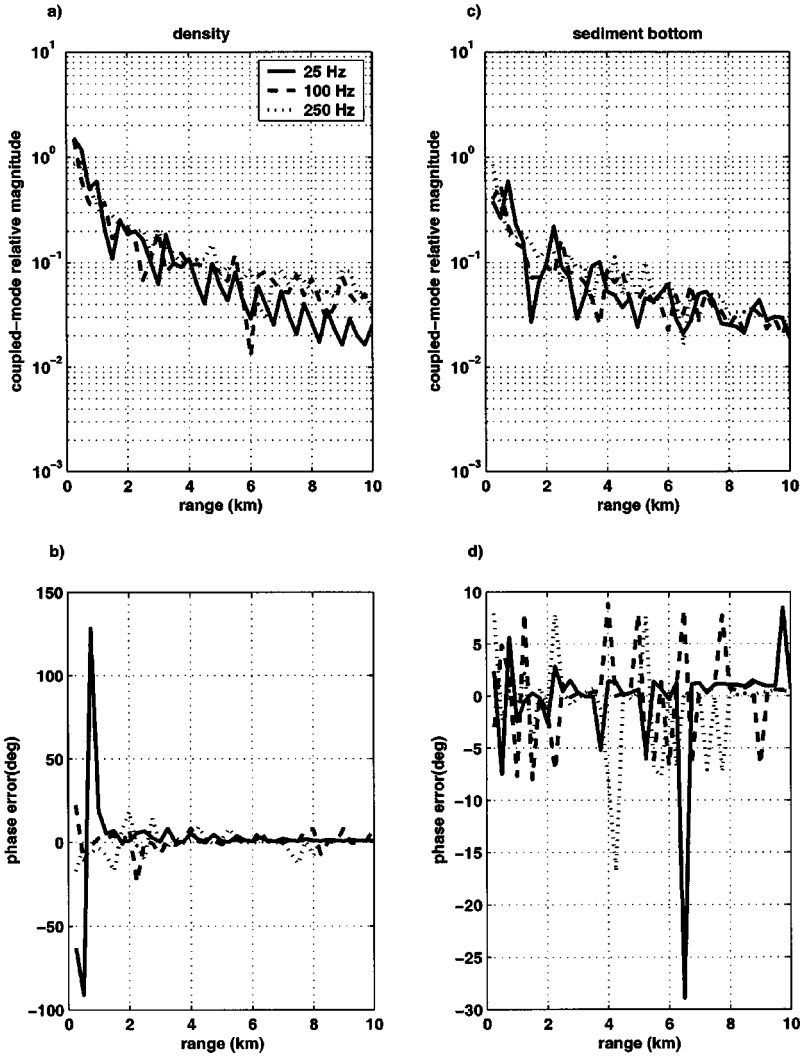


FIG. 3. Same as Fig. 2, but with (a) and (b) with respect to bottom density and (c) and (d) with respect to subsurface compressional speed gradient.

the environmental derivatives have been established, the Green's function derivatives can be expressed in terms of dimensional units.

By inserting Eqs. (4) and (5) into (3) and rearranging, an *exact* equation is obtained for the derivative of the Green's function with respect to an environmental parameter:

$$\left[\nabla^2 + k_{ref}^2 \eta_0(\overline{r}) - \frac{1}{\rho_0(\overline{r})}\nabla\rho_0(\overline{r})\cdot\nabla\right] g_\varepsilon = -\left[k_{ref}^2 \delta_{\alpha_j \varepsilon}\eta_{\alpha_j}(\overline{r}) + \delta_{\epsilon\beta_j}\left(\frac{\rho_{\beta_j}}{\rho_0^2}\nabla\rho_0 - \frac{1}{\rho_0}\nabla\rho_{\beta_j}\right)\cdot\nabla\right] \times g_0 \equiv L_\varepsilon g_0 . \tag{6}$$

The symbol $\delta_{\varepsilon\nu}$ is the Kronecker delta, with the dummy parameters ε and ν representing any parameter α_j or β_j, thus covering the possibility of mixed parameter derivatives. A general way for solving Eq. (6) for g_ε is to use an "adjoint" Green's function $g(\overline{r,r_i})$, for a source excitation at a receiver location r_i:

$$L_0 g_0(\overline{r,r_i}) = -\delta(\overline{r}-\overline{r_i}). \tag{7}$$

(The minus sign in front of the delta function is a matter of convention.) Multiplying Eq. (7) by $g_\varepsilon(\overline{r,r_s})/\rho_0(\overline{r})$ and Eq. (6) by $g_0(\overline{r,r_i})/\rho_0(\overline{r})$, subtracting the results from each other, and then integrating each term over the entire spatial domain produces a solution for $g_\varepsilon(\overline{r_i},\overline{r_s})$:

$$\frac{g_\varepsilon(\overline{r_i},\overline{r_s})}{\rho_0(\overline{r_i})} = -\int_V\int\int \frac{g_0(\overline{r,r_i})}{\rho_0(\overline{r})} L_\varepsilon(\overline{r}) g_0(\overline{r,r_s})\,\mathbf{d^3r} + \int_V\int\int \left(\frac{1}{\rho_0(\overline{r})}\right)[g_0(\overline{r,r_i})L_0 g_\varepsilon(\overline{r,r_s}) - g_\varepsilon(\overline{r,r_s})L_0 g_0(\overline{r,r_i})]\,\mathbf{d^3r}. \tag{8}$$

As discussed in Refs. 17–19 among others, the second volume integral can be converted into a surface integral, which vanishes as the domain of integration is extended to infinity by invoking the Sommerfeld radiation condition.

All pressure derivatives from Sec. II B onward will be written with respect to dimensional units. Thus, the final expression for the derivative of acoustic pressure p with respect to a dimensional sound speed parameter a is

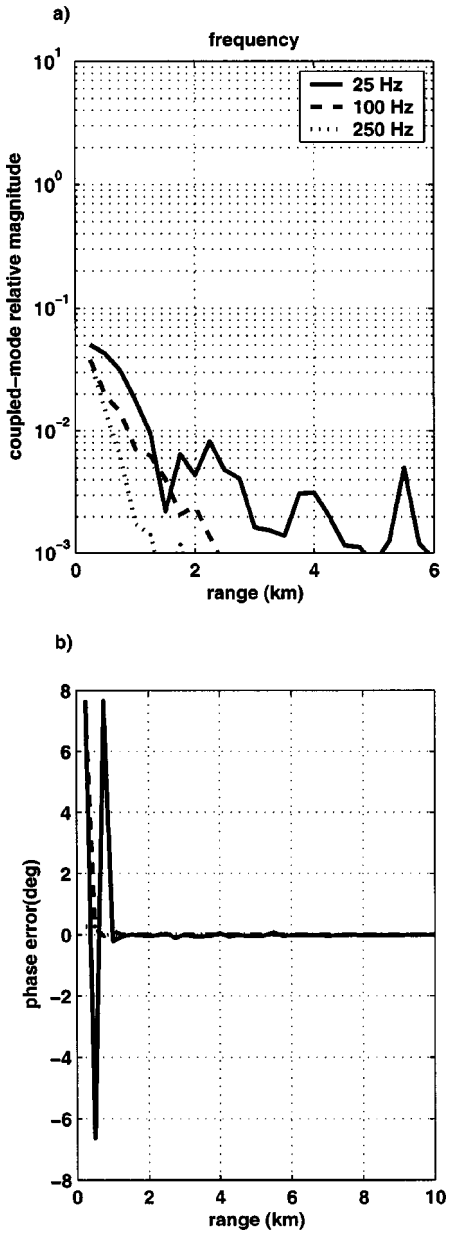


FIG. 4. Same as Fig. 3, but with respect to acoustic frequency.

$$p_a(\overline{r_i},\overline{r_s}) = \rho_0(\overline{r_i})\left(\frac{S}{S_0}\right)S_0 \times \int\int\int_V \left[k_{ref}^2 \eta_a(\overline{r})\frac{g_0(\overline{r,r_i})}{\rho_0(\overline{r})} g_0(\overline{r,r_s})\right]\mathbf{d^3r}, \tag{9a}$$

or for a dimensional density parameter b,

$$p_b(\overline{r_i},\overline{r_s}) = \rho_0(\overline{r_i})\left(\frac{S}{S_0}\right)S_0 \times \int\int\int_V [g_0(\overline{r,r_i})\overline{V}_b(\overline{r,\rho})\cdot\nabla g_0(\overline{r,r_s})]\,\mathbf{d^3r}, \tag{9b}$$

where $\overline{V}_b(\overline{r,\rho}) \equiv (\rho_b(\overline{r})/\rho_0^3(\overline{r}))\nabla\rho_0(\overline{r}) - 1/\rho_0^2(\overline{r})\nabla\rho_b(\overline{r})$, which requires both the density environmental derivative and the spatial gradient of the environmental derivative.[6,21,22] To convert the Green's function derivative into a pressure derivative, a reference source strength $S_0 = \rho_0\omega^2 V_0$ and a true source strength S (both in units of Pa-m) have been added, with V_0 being the volume injection sufficient to generate a pressure level of 1 Pa at 1 m range (120 dB *re* 1 μPa @1 m source level). This choice of V_0 sets $S_0 = 1$ Pa-m.

Equations (9a) and (9b), subsequently labeled the "acoustic adjoint equations," are well established in the geophysical literature (e.g., Refs. 19, 20 and 23). Equation (9) is an *exact* expression for the pressure derivative, even though it was derived using a perturbational approach, and its appearance is very similar to that of the Born approximation.[24] Without further simplification the spatial integral of Eq. (9) faces the same issues with computational speed and convergence as finite-difference approaches; indeed, in many geophysical applications this integration must be performed numerically, e.g., Refs. 23 and 25. However, it will be shown below that the particular geometry of a laterally homogenous waveguide permits this three-dimensional integration to be reduced to a set of one-dimensional depth integrals, reducing the computational load normally associated with the adjoint equation.

2. Higher-order environmental pressure derivatives

Having solved for the first-order pressure derivative, the second-order derivatives $g_{\varepsilon\nu}$ may be obtained[25] by permitting the nondimensional magnitudes of ε and ν to increase to the point that the higher-order terms in Eqs. (4) and (5) are no longer negligible, yielding a new equation:

$$\begin{aligned} L_0 g_{\varepsilon\nu}(\overline{r}) = &-2k_{ref}^2\delta_{\alpha\varepsilon}\eta_\alpha(\overline{r})g_\nu(\overline{r,r_s}) \\ &-k_{ref}^2\delta_{\alpha\varepsilon}\delta_{\alpha\nu}\eta_{\alpha\alpha}(\overline{r})g_0(\overline{r,r_s}) \\ &-2\delta_{\beta\varepsilon}\rho_0(\overline{r})V_\beta(\overline{r})\cdot\nabla g_\nu(\overline{r,r_s}) \\ &+2\delta_{\beta\varepsilon}\delta_{\beta\nu}\rho_\beta(\overline{r})V_\beta(\overline{r})\cdot\nabla g_0(\overline{r,r_s}). \end{aligned} \tag{10}$$

Using a series of steps nearly identical to Eqs.(7) and (8), the second-order environmental pressure derivatives can be computed with respect to dimensional sound-speed parameters a_1 and a_2:

$$p_{a_1a_2}(\overline{r_i},\overline{r_s}) = \rho_0(\overline{r_i})\left(\frac{S}{S_0}\right)S_0k_{ref}^2\int\int_{\overline{r}}\int g_0(\overline{r,r_i}) \times\left[2\frac{\eta_{a_1}(\overline{r})}{\rho_0(\overline{r})}g_{a_2}(\overline{r,r_s}) + \frac{\eta_{a_1a_2}(\overline{r})}{\rho_0(\overline{r})}g_0(\overline{r,r_s})\right]\mathbf{d^3r}, \tag{11a}$$

or with respect to dimensional density parameters b_1 and b_2:

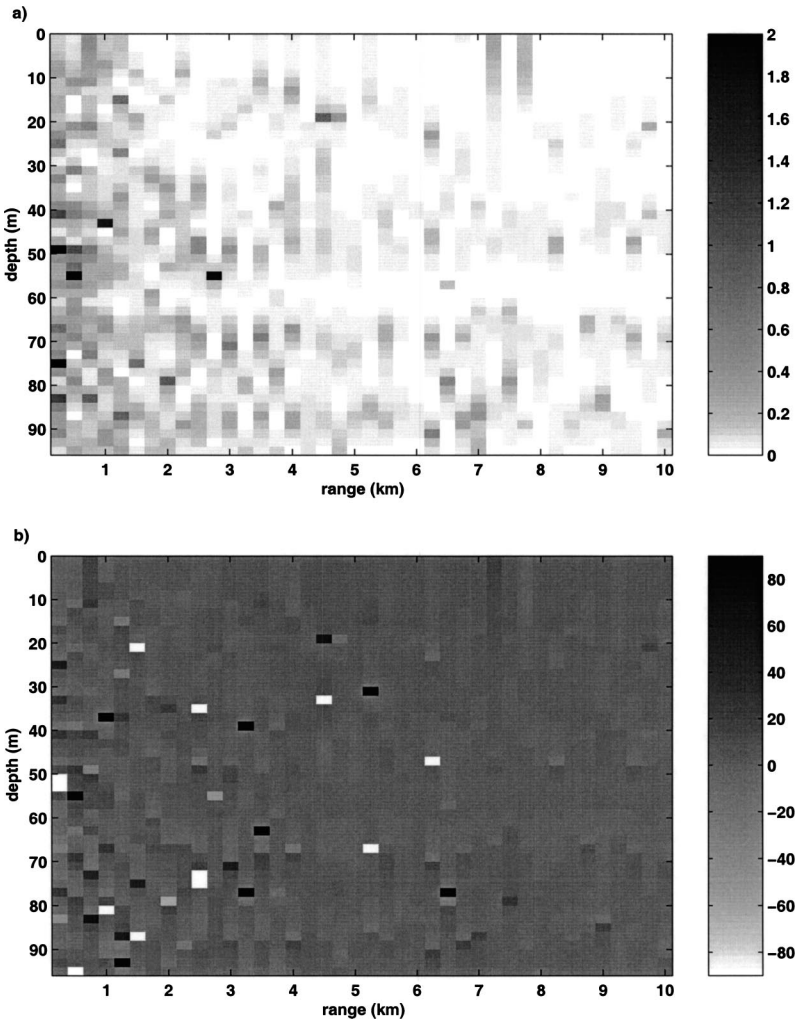


FIG. 5. The relative contribution of modal amplitude derivative terms to the environmental pressure derivative for a 100-Hz signal, as a function of source-receiver separation and receiver depth. Subplot (a) shows the ratio of the magnitude of the sum of the off-diagonal terms in Eq. (18a) to the magnitude of the sum of all terms in Eq. (18a), for derivatives with respect to bottom speed. Subplot (b) shows the difference in phase between Eqs. (18a) and (2), for derivatives with respect to bottom speed.

$$p_{b_1 b_2}(\overline{r_i},\overline{r_s}) = 2\rho_0(\overline{r_i})\left(\frac{S}{S_0}\right)S_0\int\int_{\overline{r}}\int g_0(\overline{r,r_i}) \times\left[V_{b_1}(\overline{r})\cdot\nabla g_{b_2}(\overline{r,r_s}) - \left(\frac{\rho_{b_1}(\overline{r})}{\rho_0(\overline{r})}\right)V_{b_2}(\overline{r})\cdot\nabla g_0(\overline{r,r_s})\right] \times \mathbf{d}^3\mathbf{r}, \tag{11b}$$

as well as "mixed" derivatives with respect to density and sound speed parameters:

$$p_{ab}(\overline{r_i},\overline{r_s}) = 2\rho_0(\overline{r_i})\left(\frac{S}{S_0}\right)S_0\int\int_{\overline{r}}\int g_0(\overline{r,r_i}) \times\left[k_{ref}^2\frac{\eta_a(\overline{r})}{\rho_0(\overline{r})}g_b(\overline{r,r_s}) + V_b(\overline{r})\cdot\nabla g_a(\overline{r,r_s})\right]\mathbf{d}^3\mathbf{r}. \tag{11c}$$

As was argued in the previous section, Eq. (11) is *exact*, and not an approximation, despite the appearance of terms similar to the first and second Born approximations.

The procedure for obtaining third- and higher-order derivatives proceeds much the same way; here only the "pure" third-order derivative with respect a single sound speed parameter a is explicitly presented:

$$p_{aaa}(\overline{r_i},\overline{r_s}) = \rho(\overline{r_i})\left(\frac{S}{S_0}\right)S_0k_{ref}^2\int\int_{\overline{r}}\int g_0(\overline{r,r_i}) \times\left[\begin{array}{c} 3\dfrac{\eta_{aa}(\overline{r})}{\rho_0(\overline{r})}g_a(\overline{r,r_s}) + \dfrac{\eta_{aaa}(\overline{r})}{\rho_0(\overline{r})}g_0(\overline{r,r_s}) + \cdots \\ 3\dfrac{\eta_a(\overline{r})}{\rho_0(\overline{r})}g_{aa}(\overline{r,r_s}) \end{array}\right]\mathbf{d}^3\mathbf{r}. \tag{12}$$

3. Derivatives with respect to frequency

The medium wavenumber $k_{ref}=\omega/c_{ref}(z)$ visible in Eq. (3) involves a ratio between frequency and the sound speed profile. Therefore, a perturbation in acoustic frequency $\delta\omega$ produces a change in the medium wavenumber identical to that produced by a sound speed perturbation $\delta c(z) = -c(z)d\omega/\omega$ at the unperturbed frequency. Thus any expression for an environmental pressure derivative can be re-

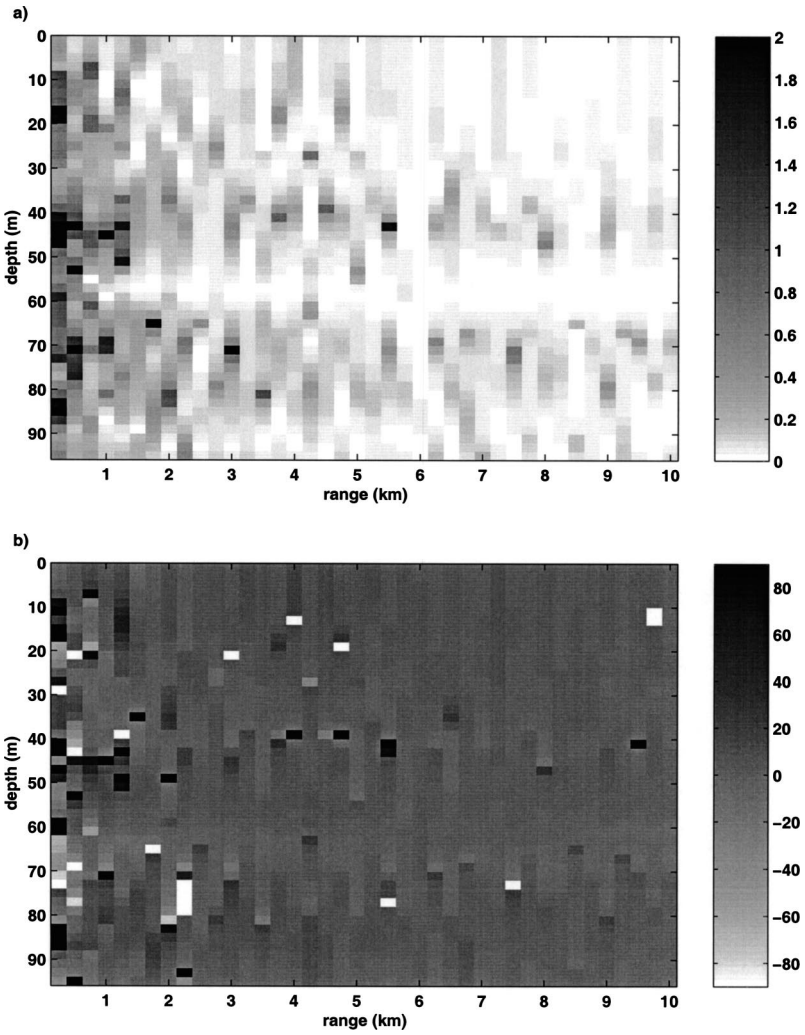


FIG. 6. Same as Fig. 5, but all derivatives are with respect to bottom density.

interpreted as a derivative with respect to frequency. The most direct approach for obtaining the frequency derivative is to extend the definition of η in Sec. II A 1 to $\eta(\overline{r},\omega) \equiv k(\overline{r})^2/k_{ref}^2 = \omega^2 c_{ref}^2/\omega_0^2 c(\overline{r})^2$, with ω_0 defined as the original "background" frequency for which the original Green's function in Eq. (3) is computed. Thus, if the environmental parameter a is reinterpreted as frequency ω, then the extended definition of η produces terms like $\eta_\omega(\overline{r},\omega_0) = (2\omega/\omega_0^2)[c_{ref}^2/c(\overline{r})^2]|_{\omega=\omega_0} = 2\eta(\overline{r},\omega_0)/\omega_0$ and $\eta_{\omega\omega}(\overline{r},\omega_0) = (2/\omega_0^2)[c_{ref}^2/c(\overline{r})^2]|_{\omega=\omega_0} = 2\eta(\overline{r},\omega_0)/\omega_0^2$. These terms can now be substituted directly into Eqs. (9), (11), and (12). In the next section all environmental derivative expressions will also contain definitions that convert them into frequency derivatives.

B. Normal-mode formulation of acoustic adjoint equation

1. First-order derivative

In this section the standard acoustic adjoint equations in Eqs. (9a) and (9b) are used to derive normal-mode expressions for the environmental pressure derivatives in an ocean acoustic waveguide, which have been previously derived by perturbing the depth-dependent component of the wave equation.[14,22] The inversion parameters of interest will be assumed to be depth dependent only, e.g., $\eta_0(\overline{r}) = \eta_0(z)$ and $\nabla\rho_0(\overline{r}) = \partial\rho_0(z)/\partial z$. In a related paper this latter restriction is relaxed.[26]

Neglecting the density term momentarily, Eq. (9a) can be written as

$$\frac{\partial p(\overline{r_i},\overline{r_s})}{\partial a} = k_{ref}^2 \rho(\overline{r_i}) \left(\frac{S}{S_0}\right) S_0 \int dz' \frac{\eta_a(z')}{\rho(z')} \times \left[\int \int \mathbf{d^2 r'}\, g(|\overline{r'} - \overline{r_i}|, z', z_i) g(\overline{r'}, z', z_s) \right]. \tag{13}$$

Here the acoustic source's horizontal location has been set to the horizontal origin ($\overline{r_s} = 0$), z_i, r_i are the receiver ranges and depths, and z_s is the source depth. The notation $g(|\overline{r} - \overline{r_s}|, z, z_s)$ for the Green's function describes the propagation of an acoustic field from point $(z_s, \overline{r_s})$ to a point at depth z and horizontal range separation $|\overline{r} - \overline{r_s}|$. From Eq. (13) onward the zero subscript for the unperturbed Green's functions and density are dropped, with the understanding that the Green's function and density terms in subsequent discussion refer to the *unperturbed* functions.

This equation can be solved in a manner reminiscent to normal-mode formulations for ambient ocean noise.[27] First,

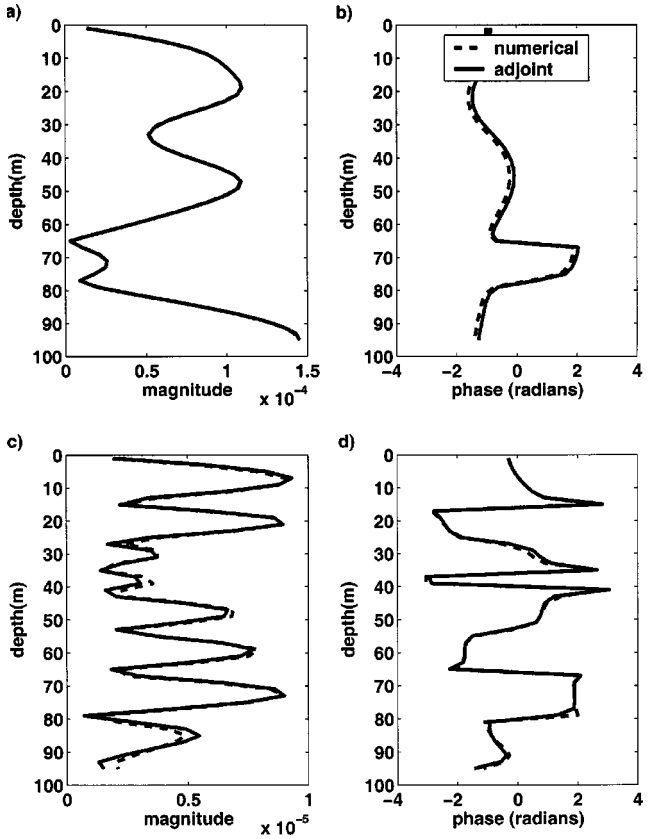


FIG. 7. Adjoint (solid line) and numerical (dashed line) first-order pressure derivative estimates, as a function of receiver depth, for a 100 Hz signal with 120 dB *re* 1 μPa@ 1 m source level propagating 2 km through Fig. 1 environment: (a) magnitude and (b) phase of derivative with respect to uniform water column sound speed, and (c) magnitude and (d) phase of derivative with respect to bottom interface speed. Numerical parameter step sizes are 0.1 m/s, mesh grid size is 500 in water column.

the waveguide Green's functions are decomposed into their spatial Fourier components:

$$g(\overline{r}',z',z_s)=\frac{1}{2\pi}\left[\int\int \mathbf{d^2k}\, g(k_x,k_y,z',z_s)e^{i(k_xx'+k_yy')}\right],$$
$$g(|\overline{r}'-\overline{r_i}|,z',z_i)=\frac{1}{2\pi}\left[\int\int \mathbf{d^2k'}\, g(k_x',k_y',z',z_i)\times e^{i[k_x(x'-x_i)+k_y(y'-y_i)]}\right]. \tag{14}$$

Using Eq. (14) and integral definitions of the delta function, e.g., $\delta(k_x+k_x')=(1/(2\pi))[\int e^{i(k_x+k_x')x'}\,dx']$, the zeroth order Bessel function, $J_0(x)=(1/(2\pi))[\int_0^{2\pi}e^{ix\cos\theta}d\theta]$, Eq. (13) becomes

$$\frac{\partial p(\overline{r_i},\overline{r_s})}{\partial a}=2\pi k_{ref}^2\rho(\overline{r_i})\left(\frac{S}{S_0}\right)S_0\int dz'\frac{\eta_a(z')}{\rho(z')}\times\left[\int_0^\infty dk_r\, k_r g(k_r,z',z_i)g(k_r,z',z_s)\mathbf{J}_0(k_rr_i)\right]. \tag{15}$$

For a laterally homogeneous waveguide, the wavenumber Green's function can be expressed as a sum of normal modes:

$$g(k_r,z,z_s)=\frac{1}{2\pi\rho(z_s)}\sum_f\frac{U_f(z)U_f(z_s)}{k_r^2-k_{rf}^2}, \tag{16}$$

where k_{rf} is the horizontal wavenumber of mode U_f. Substituting Eq. (16) into (15) yields integrals of the form

$$I_{fg}(r_i)\equiv\int_{-\infty}^{\infty}\frac{k_rH_0^{(1)}(k_rr_i)dk_r}{(k_r^2-k_{rf}^2)(k_r^2-k_{rg}^2)}, \tag{17}$$

which can be solved via a contour integration in a manner reminiscent of Ref. 27. When $f\neq g$, the residues are simple poles; when $f=g$, the residues are second-order poles. The final result reduces the acoustic adjoint equation of Eq. (9a) from a three-dimensional spatial integration to a sum of one-dimensional depth integrals:

$$\frac{\partial p(\overline{r_i},\overline{r_s})}{\partial a}=\left(\frac{S}{S_0}\right)S_0\frac{i}{4\rho(z_s)}\sum_{f,g}Z_{fg}(\eta_a)\times R_{fg}(r_i)U_f(z_i)U_g(z_s), \tag{18a}$$

with

$$Z_{fg}(\chi)\equiv k_{ref}^2\int_0^\infty\frac{\chi(z')}{\rho(z')}U_f(z')U_g(z')\,dz', \tag{18b}$$

where $\chi(z)$ is an arbitrary depth-dependent function, and

$$R_{fg}(r_i)=\begin{cases}\dfrac{H_0^{(1)}(k_{rf}r_i)-H_0^{(1)}(k_{rg}r_i)}{k_{rf}^2-k_{rg}^2}, & f\neq g,\\ \dfrac{-r_iH_1^{(1)}(k_{rf}r_i)}{2k_{rf}}\xrightarrow[k_{rf}r_i\gg1]{}\dfrac{ir_iH_0^{(1)}(k_{rf}r_i)}{2k_{rf}}, & f=g\end{cases}. \tag{18c}$$

Here $H_i^{(1)}$ is the ith-order outgoing Hankel function of the first kind.

The first-order derivative of pressure with respect to frequency is the same as above, except in Eq. (18a) $Z_{fg}(\eta_a)$ is replaced by $Z_{fg}(2\eta/\omega)$. The equivalent expression for density in Eq. (9b) can be derived in exactly the same manner, obtaining a result identical in form to Eq. (18a), but with a slightly altered form for the depth integral, Eq. (18b):

$$\overset{\rho}{Z}_{fg}(V_b)\equiv\int_0^\infty U_f(z')V_b(z',\rho)\frac{\partial U_g(z')}{\partial z'}dz'. \tag{18d}$$

For the idealized case of constant densities ρ_{water} and ρ_{sed} for the water and bottom sediment, respectively, with a discontinuity at depth D, the fact that the derivative of a step discontinuity can be treated as a delta function leads to the expression

$$\overset{\rho}{Z}_{fg}=U_f(D)\left(\frac{\Delta\rho}{\overline{\rho}^3}-\frac{1}{\overline{\rho}^2}\right)\frac{\partial U_g(D)}{\partial z}, \tag{18e}$$

where $\overline{\rho}=(\rho_{sed}+\rho_{water})/2$ and $\Delta\rho=(\rho_{sed}-\rho_{water})/2$. Thus only the modal depth derivatives at the bottom interface influence the density derivative.

2. Second-order derivatives

The power of the adjoint perturbative approach becomes apparent when progressing to higher-order terms [Eq. (11)], where the second-order pressure derivative has two refractive index terms. The second term can be solved in a manner similar to the previous section; the first term requires some additional steps, beginning by combining Eqs. (11) and (13):

$$\frac{\partial p(\overline{r_i},\overline{r_s})}{\partial a_1 \partial a_2}(term\ \ 1)$$
$$= 2k_{ref}^4 \rho(\overline{r_i}) \left(\frac{S}{S_0}\right) S_0 \int \int_{\overline{r}} \int g(|\overline{r'}-\overline{r_i}|,z',z_i)$$
$$\times \eta_{a_1}(\overline{r'})\, \mathbf{d^3 r'} \int \int_{\overline{r}} \int \mathbf{d^3 r''} \left(\frac{g(|\overline{r''}-\overline{r'}|,z'',z')}{\rho(\overline{r''})}\right)$$
$$\times \eta_{a_2}(\overline{r''}) g(|\overline{r''}-\overline{r_s}|,z'',z_s). \tag{19}$$

Substituting the wavenumber integrals for the Green's function as before yields integrals of the form

$$I_{fgh}(r_i) \equiv \int_{-\infty}^{\infty} \frac{k_r H_0^{(1)}(k_r r_i) dk_r}{(k_r^2-k_{rf}^2)(k_r^2-k_{rg}^2)(k_r^2-k_{rh}^2)},$$

requiring contour integration around first-through third-order poles. When combined with the second term, the full expression for the second-order pressure derivative with respect to refractive index parameters becomes

$$\frac{\partial^2 p(\overline{r_i},\overline{r_s})}{\partial a_1 \partial a_2} = \left(\frac{S}{S_0}\right) S_0 \frac{i}{2\rho(z_s)} \times \left[\sum_{f,g,h} Z_{fg}(\eta_{a_1}) Z_{gh}(\eta_{a_2}) R_{fgh}(r_i) U_f(z_i) U_h(z_s) + \frac{1}{2} \sum_{f,g} Z_{fg}(\eta_{a_1 a_2}) R_{fg}(r_i) U_f(z_i) U_g(z_s) \right], \tag{20a}$$

$$R_{fgh}(r_i) = \begin{cases} \dfrac{H_0^{(1)}(k_{rf}r_i)}{(k_{rf}^2-k_{rg}^2)(k_{rf}^2-k_{rh}^2)} + \dfrac{H_0^{(1)}(k_{rg}r_i)}{(k_{rg}^2-k_{rf}^2)(k_{rg}^2-k_{rh}^2)} + \dfrac{H_0^{(1)}(k_{rh}r_i)}{(k_{rh}^2-k_{rf}^2)(k_{rh}^2-k_{rg}^2)}, & f \neq g \neq h, \\ \dfrac{-r_i H_1^{(1)}(k_{rg}r_i)}{2k_{rg}(k_{rg}^2-k_{rf}^2)} \xrightarrow[k_{rg}r_i \gg 1]{} \dfrac{i r_i H_0^{(1)}(k_{rg}r_i)}{2k_{rg}(k_{rg}^2-k_{rf}^2)}, & f \neq g = h, \\ \dfrac{r_i^2 H_2^{(1)}(k_{rg}r_i)}{8k_{rg}^2} \xrightarrow[k_{rg}r_i \gg 1]{} \dfrac{-r_i^2 H_0^{(1)}(k_{rg}r_i)}{8k_{rg}^2}, & f = g = h \end{cases} . \tag{20b}$$

If frequency is the parameter of interest, then $Z_{fg}(\eta_{a_1 a_2})$ in Eq. (20a) is replaced by $Z_{fg}(2\eta/\omega^2)$, $Z_{fg}(\eta_a)$ is replaced by $Z_{fg}(2\eta/\omega)$, etc. The corresponding solution for the second-order density derivative is

$$\frac{\partial^2 p(\overline{r_i},\overline{r_s})}{\partial b_1 \partial b_2} = \left(\frac{S}{S_0}\right) S_0 \frac{i}{2\rho(z_s)} \left[\sum_{f,g,h} \overset{\rho}{Z}_{fg}(V_{b_1}) \overset{\rho}{Z}_{gh}(V_{b_2}) \times R_{fgh}(r_i) U_f(z_i) U_h(z_s) - \frac{1}{2} \times \sum_{f,g} \overset{\rho}{Z}_{fg}\left(\frac{\rho_{b_1} V_{b_2}}{\rho_0}\right) R_{fg}(r_i) U_f(z_i) U_g(z_s) \right], \tag{21}$$

and for the mixed-order derivative

$$\frac{\partial^2 p(\overline{r_i},\overline{r_s})}{\partial a \partial b} = \left(\frac{S}{S_0}\right) S_0 \frac{i}{\rho(z_s)} \times \left[\sum_{f,g,h} Z_{fg}(\eta_a) \overset{\rho}{Z}_{gh}(V_b) R_{fgh}(r_i) U_f(z_i) U_h(z_s) \right]. \tag{22}$$

3. Third-order derivatives

Equation (12), the adjoint expression for the third-order pressure derivative with respect to a single sound speed parameter, has many terms that are nearly identical to those in previous sections; only the term involving a spatial integral over g_{aa} requires the solution of new contour integrals similar in form to Eq. (17), except that four terms are present in the denominator. Some care must be taken to identify all situations where multiple poles exist. The final expression for the third-order "pure" pressure derivative is

$$\frac{\partial^2 p(\overline{r_i},\overline{r_s})}{\partial a^3}=\left(\frac{S}{S_0}\right)S_0\frac{i}{2\rho(z_s)}\left[\begin{array}{l}3\sum_{f,g,h} Z_{fg}(\eta_{aa})Z_{gh}(\eta_a)R_{fgh}(r_i)U_f(z_i)U_h(z_s)\\ +\frac{1}{2}\sum_{f,g} Z_{fg}(\eta_{aaa})R_{fg}(r_i)U_f(z_i)U_g(z_s)\\ +3\sum_{f,g,h,j} Z_{fg}(\eta_a)Z_{gh}(\eta_a)Z_{hj}(\eta_a)R_{fghj}(r_i)U_f(z_i)U_j(z_s)\end{array}\right] \tag{23a}$$

with the new definition

$$R_{fghj}(r_i)=\begin{cases}\frac{H_0^{(1)}(k_{rf}r_i)}{(k_{rf}^2-k_{rg}^2)(k_{rf}^2-k_{rh}^2)(k_{rf}^2-k_{rp}^2)}+\frac{H_0^{(1)}(k_{rg}r_i)}{(k_{rg}^2-k_{rf}^2)(k_{rg}^2-k_{rh}^2)(k_{rg}^2-k_{rp}^2)}+\frac{H_0^{(1)}(k_{rh}r_i)}{(k_{rh}^2-k_{rg}^2)(k_{rh}^2-k_{rf}^2)(k_{rh}^2-k_{rp}^2)}\\ \quad+\frac{H_0^{(1)}(k_{rj}r_i)}{(k_{rj}^2-k_{rf}^2)(k_{rj}^2-k_{rh}^2)(k_{rj}^2-k_{rf}^2)}, \quad f\neq g\neq h\neq j\\ \frac{H_0^{(1)}(k_{rf}r_i)}{(k_{rf}^2-k_{rg}^2)(k_{rf}^2-k_{rh}^2)^2}+\frac{H_0^{(1)}(k_{rg}r_i)}{(k_{rg}^2-k_{rf}^2)(k_{rg}^2-k_{rh}^2)^2}+\frac{H_0^{(1)}(k_{rh}r_i)}{(k_{rg}^2-k_{rh}^2)(k_{rf}^2-k_{rh}^2)^2}+\frac{H_0^{(1)}(k_{rh}r_i)}{(k_{rf}^2-k_{rh}^2)(k_{rg}^2-k_{rh}^2)^2}\\ \quad+\frac{r_iH_1^{(1)}(k_{rh}r_i)}{k_{rh}(k_{rh}^2-k_{rf}^2)(k_{rg}^2-k_{rh}^2)}, \quad f\neq g\neq h=j,\\ \frac{H_0^{(1)}(k_{rh}r_i)-H_0^{(1)}(k_{rf}r_i)}{(k_{rf}^2-k_{rh}^2)^3}-\frac{r_iH_1^{(1)}(k_{rh}r_i)}{4k_{rh}(k_{rh}^2-k_{rf}^2)^2}-\frac{r_iH_1^{(1)}(k_{rf}r_i)}{4k_{rf}(k_{rh}^2-k_{rf}^2)^2}, \quad f=g\neq h=j\\ \frac{H_0^{(1)}(k_{rf}r_i)-H_0^{(1)}(k_{rg}r_i)}{(k_{rf}^2-k_{rg}^2)^3}+\frac{r_iH_1^{(1)}(k_{rg}r_i)}{2k_{rg}(k_{rf}^2-k_{rg}^2)^2}-\frac{r_i^2H_2^{(1)}(k_{rg}r_i)}{8k_{rg}^2(k_{rf}^2-k_{rg}^2)}, \quad f\neq g=h=j\\ \quad-\frac{r_i^3H_3^{(1)}(k_{rf}r_i)}{48k_{rf}^3}, \quad f=g=h=j\end{cases} \tag{23b}$$

The extension of these expressions to density and mixed-parameter derivatives presents no major conceptual obstacles.

III. DISCUSSION

In this section, the full-field perturbative expressions in Sec. II B are compared with expressions based on perturbations of modal horizontal wavenumber only. These expressions are then compared with derivatives obtained numerically by finite differences, using a representative waveguide propagation model.

A. Ocean waveguide model

The normal modes and associated horizontal wavenumbers for this section are modeled assuming a 100-Hz field generated by a 50-m-deep acoustic source in a 100-m-deep waveguide (Fig. 1). The received field is computed at 48 receiver depths that span a water column with a constant sound speed of $c_{ref}=1500$ m/s. The bottom properties are an interface speed of $c_b=1645$ m/s, subsurface sound speed gradient of 1 (m/s)/m, sediment bed density of 1.3 g/cc, and medium attenuation of 0.25 dB/wavelength. The modes were computed using a MATLAB version of KRAKEN developed by J. Ianniello,[28] using a mesh size of 10 points per meter in the water column.

The pressure derivative with respect to a uniform change in water sound speed uses the refractive index derivative

$$\eta_{c_w}^{water}(z)=\begin{cases}-2\frac{c_{ref}^2}{c(z)^3}\frac{\partial c(z)}{\partial c_w}=-\frac{2}{c_{ref}}, & 0<z<D\\ 0, & z>D\end{cases}, \tag{24}$$

where D is the waveguide depth. The pressure derivative with respect to the bottom interface sound speed (c_b) uses the complementary refractive index derivative

$$\eta_{c_b}^{bottom}(z)=\begin{cases}0, & 0<z<D,\\ -2\frac{c_{ref}^2}{c(z)^3}, & z>D\end{cases}. \tag{25}$$

Finally, the pressure derivatives with respect to the subsurface compressional speed gradient (s) use

$$\eta_s^{gradient}(z)=\begin{cases}0, & 0<z<D,\\ -2\left(\frac{c_{ref}^2}{c(z)^3}\right)\frac{(z-D)}{L_{sed}}, & z>D\end{cases}. \tag{26}$$

Here L_{sed} is the 46-m sediment thickness. Equations (24)–(26) were inserted into Eqs. (18), (20), and (23) for the first- through third-order pressure derivatives, respectively. The pressure derivatives for density were computed using Eqs. (18) and (21). The modal depth derivative for Eq. (18e) was evaluated using a simple finite difference of the mode amplitude.

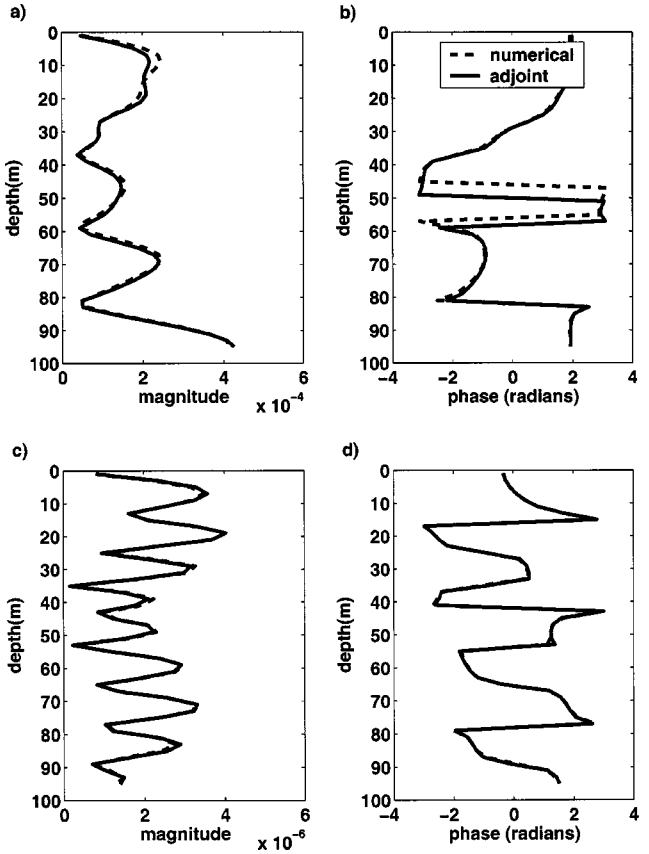


FIG. 8. Same as Fig. 7, but showing first-order pressure derivatives with respect to bottom density and subsurface compressional speed gradient: (a) magnitude and (b) phase of derivative with respect to bottom density, and (c) magnitude and (d) phase of derivative with respect to sediment subsurface compressional speed gradient. Numerical parameter step sizes are 1.25×10^{-3} g/cc and 0.1 m/s per m, respectively.

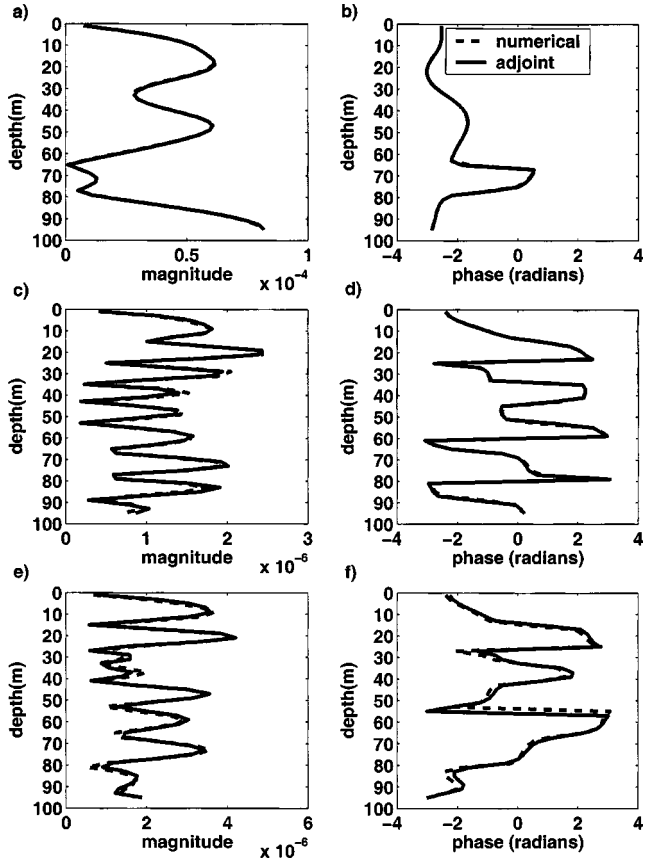


FIG. 9. Adjoint (solid line) and numerical (dashed line) second-order pressure derivative estimates for same situation described in Fig. 7, as a function of receiver depth: (a) magnitude and (b) phase of second-order derivative with respect to uniform water column sound speed, (c) magnitude and (d) phase of second-order derivative with respect to bottom interface speed, and (e) magnitude and (f) phase of the mixed second-order derivative of water and bottom speeds.

B. Relationship with wavenumber perturbation methods

An examination of the matrix R_{fg} in Eq. (18c) suggests that the off-diagonal components are much smaller than the diagonal terms, for horizontal ranges much greater than the modal interference length between two modes. If the off-diagonal terms are neglected, Eq. (18a) reduces to Eq. (2). The diagonal terms in R_{fg} are thus associated with the modal wavenumber derivatives, while the off-diagonal terms are associated with the mode amplitude derivatives, as can be also deduced from Ref. 14.

Equation (18) provides an opportunity to examine under what circumstances mode amplitude derivatives can be safely neglected when relating environmental perturbations to acoustic field changes. Figures 2–4 examine this question for the waveguide environment discussed previously, as a function of source-receiver range and acoustic frequency. The top row of each figure shows the ratio of the magnitude of the sum of the off-diagonal terms in Eq. (18a) to the magnitude of the sum of all the terms in Eq. (18a). The bottom row shows the difference in phase between Eqs. (18a) and (2). All quantities have been averaged across all array receiver depths to produce a depth-averaged ratio. As expected, as the horizontal separation increases the relative influence of the mode amplitude derivatives decrease. For some parameters, such as the water-borne sound speed profile and frequency, the relative magnitudes never exceed 10%. However, parameters associated with the ocean bottom, such as the bottom interface compressional speed, bottom density, and subsurface compressional speed gradient, show significant mode amplitude derivative contributions out to horizontal ranges of 2 to 3 km, which correspond to the modal interference distance of the two lowest-order modes. At certain receiver depths, these contributions can persist out to much further ranges. Figures 5 and 6 show the magnitude ratios and phase error of a 100-Hz signal as a function of both source-receiver range and receiver depth. Magnitude ratios of over 10% that persist to at least 10 km are present at several receiver depths.

These results suggest that the derivatives of normal mode amplitudes with respect to bottom parameters can be non-negligible, particularly when the horizontal range between source and receiver is less than a few modal-interference wavelengths, corresponding to ranges of a few km in water depths on the order of 100 m. Therefore linear inversion methods which assume that environmental perturbations only have first-order effects on horizontal wavenumbers, i.e., use Eq. (2), should either exercise caution at short ranges or use the complete expression in Eq. (18).

The effects of frequency on mode amplitude are con-

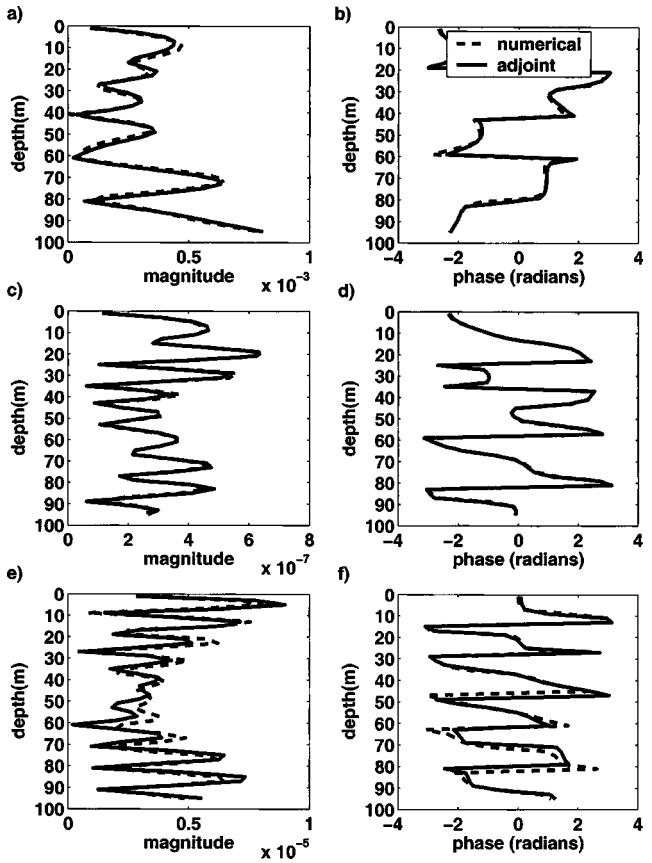


FIG. 10. Additional second-order pressure derivatives for same situation described in Fig. 7: (a) magnitude and (b) phase of second-order derivative with respect to bottom density, (c) magnitude and (d) phase of second-order derivative with respect to subsurface sediment compressional speed gradient, and (e) magnitude and (f) phase of the mixed second-order derivative of density and subsurface gradient.

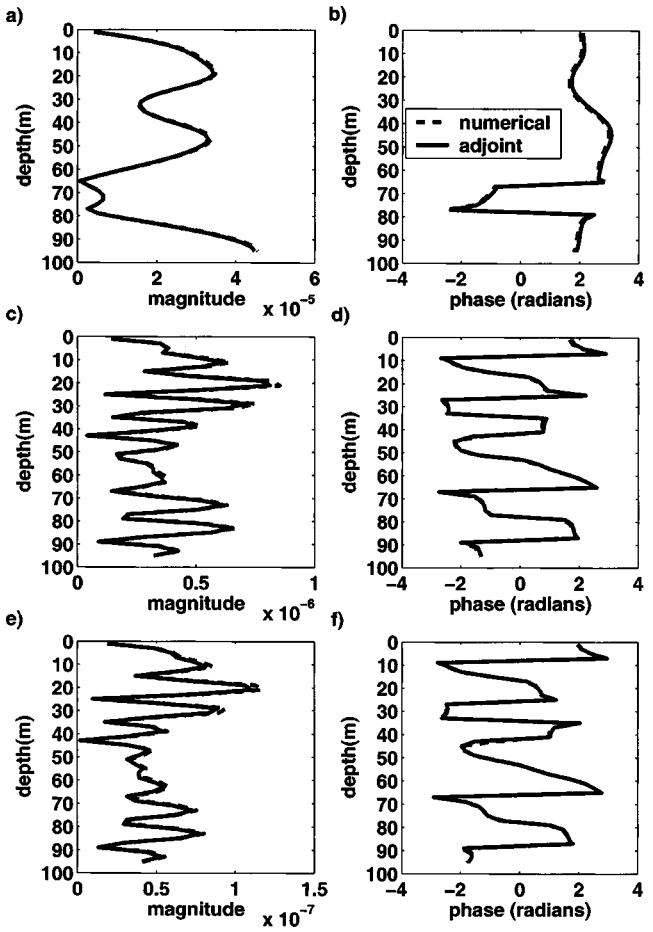


FIG. 11. Adjoint (solid line) and numerical (dashed line) estimates for selected third-order pressure derivatives, for the same situation described in Fig. 7: (a) magnitude and (b) phase of third-order derivative with respect to uniform water column sound speed, (c) magnitude and (d) phase of third-order derivative with respect to bottom interface speed, and (e) magnitude and (f) phase of third-order derivative with respect to sediment subsurface compressional speed gradient.

firmed to be less than 10% at all ranges for this environment, and drop off substantially with at higher frequencies. Thus, the assumption that the pressure derivative with respect to frequency is primarily driven by changes in the horizontal wavenumber, a crucial assumption behind the derivation of the waveguide ''invariant,'' seems accurate for this environment.

This discussion has concerned first-order derivatives only; however, similar conclusions can be drawn about the relative importance of the off-diagonal expressions of R_{fgh} in Eq. (20b) and R_{fghj} in Eq. (23b).

C. Comparison with finite-difference computations

In this section the environmental pressure derivative expressions in Eqs. (18), and (20)–(23) are directly compared with convergence-tested finite-difference estimates, for a 100-Hz field with 120 dB *re* 1 μPa @1m source level propagating through the environment described in Sec. III A. A horizontal separation of 2 km was chosen because the derivatives at this range incorporate significant contributions from all components of R_{fg}. Figures 7 and 8 display the first-order pressure derivatives with respect to water-column sound speed c_w, the interface sound speed c_b, bottom density, and the bottom subsurface sound speed gradient. The magnitudes and phases of the derivatives are plotted for each element in the vertical receiving array. The appropriate step sizes for the finite-difference derivatives were obtained via a convergence test. As discussed in the Introduction, the optimal step sizes vary with range, but at 2-km source range sufficient step sizes were found to be 0.1 m/s for sound speed and 1.25×10^{-3} g/cc for density.

Figures 9 and 10 use Eqs. (20)–(22) to derive the phases and magnitudes of six of the ten possible second-order derivatives that can be formed between the four environmental parameters. All of the ''pure'' derivatives are shown, as well as the mixed derivatives between the water and bottom sound speed, and the bottom density and bottom slope. Figure 11 compares the third-order expressions of Eq. (23) with the numerical estimates for the ''pure'' derivatives with respect to c_w, c_b, and the bottom subsurface sound speed gradient. Finally, Fig. 12 shows the first- and second-order derivatives of the acoustic field with respect to frequency, demonstrating how the second-order derivative is an order of magnitude greater than the first-order derivative. All the first- and second-order adjoint derivatives match the finite-difference estimates well. Simulations with more realistic

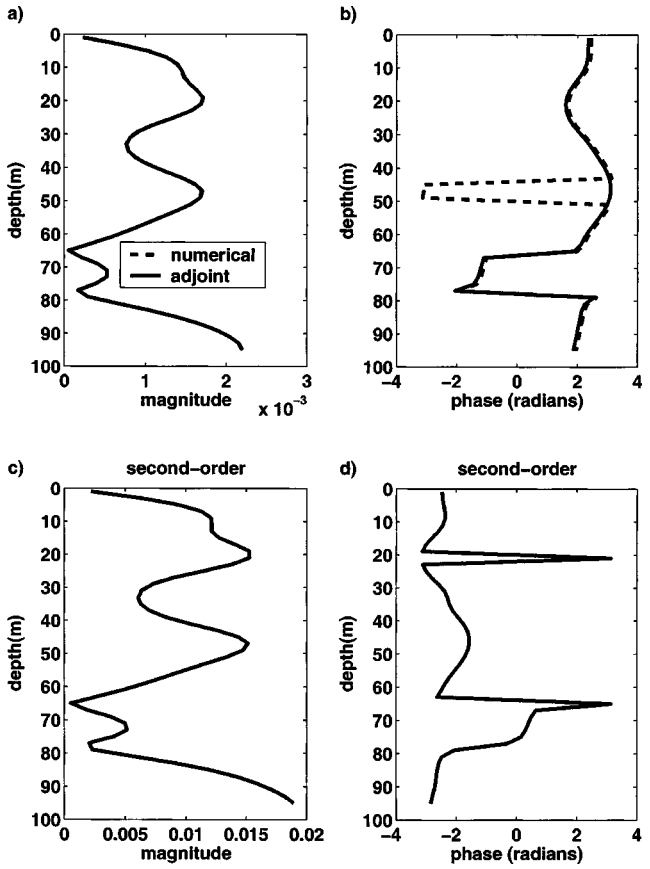


FIG. 12. Pressure derivatives with respect to 100-Hz source frequency for the same situation described by Fig. 7: (a) magnitude and (b) phase of the first-order frequency derivative and (c) magnitude and (d) phase of the second-order derivative. The odd-looking jump in phase for the numeric derivative in (b) is an artifact of the phase wrap-around from $+\pi$ to $-\pi$.

sound-speed profiles, including upward- and downward-refracting profiles, produced similar results.

IV. CONCLUSION

An adjoint perturbative approach has been used to derive the first- through third-order derivatives of pressure with respect to waveguide environmental parameters [Eqs. (18)–(23)], as a function of the normal mode amplitudes and wavenumbers of a laterally homogenous waveguide. These expressions incorporate the effects of the environmental perturbations on the modal horizontal wavenumbers and mode amplitudes, the latter often being neglected in practice, excepting Refs. 14 and 22. Most common environmental parameters, except waveguide depth, can be incorporated into these expressions. Some parameters explored here include water and bottom sound speed, bottom density, subsurface compressional speed gradient, and frequency. The extension of these expressions to bottom attenuation is straightforward.

Explicit evaluations of the expressions for a standard shallow-water waveguide model are in excellent agreement with numerical finite-difference computations. An analysis of the expressions as a function of source-receiver range finds the contribution of the mode amplitude derivatives to be non-negligible at ranges less than a few modal interference wavelengths, for parameters associated with the ocean bottom. However, at certain receiver depths these mode amplitude contributions can persist to at least 10 km range. Thus, linear perturbative inversion methods that perturb only modal wavenumbers and not mode amplitudes should either be used with caution or modified to incorporate the expressions presented here.

The scope of this paper has been restricted to environmental pressure derivatives with respect to parameters that are depth-dependent only. In a related work[26] the normal-mode expressions presented here are extended to include three-dimensional perturbations in a laterally homogenous waveguide.

ACKNOWLEDGMENTS

The authors thank Peter Gerstoft, Paul Hursky, Bruce Cornelle, and Bill Kuperman for helpful discussions on adjoint and perturbation theory, as well as three anonymous reviewers for their comments on this paper. The ONR Uncertainty program (ONR Contract No. N00014-00-F-0395) provided support for the authors.

[1] A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, "Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem," J. Acoust. Soc. Am. **83**, 571–587 (1988).
[2] J. L. Krolik and S. Narasimhan, "Performance bounds on acoustic thermometry of ocean climate in the presence of mesoscale sound-speed variability," J. Acoust. Soc. Am. **99**, 254–265 (1996).
[3] M. Zanolin, I. Ingram, A. M. Thode, and N. C. Makris, "Asymptotic accuracy of geoacoustic inversions," J. Acoust. Soc. Am. **116**(4), 2031–2042 (2003).
[4] J. L. Krolik, "Matched-field minimum variance beamforming in a random ocean channel," J. Acoust. Soc. Am. **92**, 1408–1419 (1992).
[5] E. C. Shang, "Ocean acoustic tomography based on adiabatic mode theory," J. Acoust. Soc. Am. **85**, 1531–1537 (1989).
[6] S. D. Rajan, J. F. Lynch, and G. V. Frisk, "Perturbative inversion methods for obtaining bottom geoacoustic parameters in shallow-water," J. Acoust. Soc. Am. **82**, 998–1017 (1987).
[7] K. M. Becker, S. D. Rajan, and G. V. Frisk, "Results from the geoacoustic inversion techniques workshop using modal inverse methods," IEEE J. Ocean. Eng. **28**, 331–341 (2003).
[8] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization* (Academic, London, 1981).
[9] P. Gerstoft, "Inversion of acoustic data using a combination of genetic algorithms and the gauss-Newton approach," J. Acoust. Soc. Am. **97**, 2181–2190 (1995).
[10] B. J. Sperry, B. E. McDonald, and A. B. Baggeroer, "Consistency and accuracy of perturbative inversion methods for group travel time data," J. Acoust. Soc. Am. **114**, 1851–1860 (2003).
[11] S. D. Chuprov, "Interference structure of a sound field in a layered ocean," in *Akustika okeana. Sovremennoe sostoyanie (Ocean Acoustics, Current State)*, edited by L. M. Brekhovskikh and I. B. Andreevoi (Nauka, Moscow, 1982), pp. 71–91.
[12] G. L. D'Spain and W. A. Kuperman, "Application of waveguide invariants to analysis of spectrograms from shallow water environments that vary in range and azimuth," J. Acoust. Soc. Am. **106**, 2454–2468 (1999).
[13] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt. *Computational Ocean Acoustics* (American Institute of Physics, New York, 1994).
[14] C. T. Tindle, L. M. O'Driscoll, and C. J. Higham, "Coupled mode perturbation theory of range dependence," J. Acoust. Soc. Am. **108**, 76–83 (2000).
[15] P. Hursky, M. B. Porter, B. Cornuelle, W. S. Hodgkiss, and W. A. Kuperman, "Adjoint modeling for acoustic inversion," J. Acoust. Soc. Am. **115**, 607–619 (2003).
[16] J.-C. L. Gac, Y. Stephan, M. Asch, P. Helluy, and J.-P. Hermand, "A variational approach for geoacoustic inversion using adjoint modeling of a PE approximation model with non local impedance conditions," in *Sixth International Conference on Theoretical and Computational Acoustics*, edited by A. Tolstoy, E.-C. Shang, and Y.-C. Teng (World Scientific, Naval Undersea Warfare Center Division, Singapore, 2003).
[17] P. R. McGillivray and D. W. Oldenburg, "Methods for calculation of

Frechet derivatives and sensitivities for the non-linear inverse problem: A comparative study,'' Geophys. Prospect. **38**, 499–524 (1990).
[18] S. J. Norton, ''Iterative inverse scattering algorithms: Methods of computing Frechet derivatives,'' J. Acoust. Soc. Am. **106**, 2653–2660 (1999).
[19] A. Tarantola, ''Inversion of seismic reflection data in the acoustic approximation,'' Geophysics **49**, 1259–1266 (1984).
[20] M. S. Zhdanov, *Geophysical Inverse Theory and Regularization Problems* (Elsevier, Amsterdam, 2002).
[21] P. G. Bergmann, ''The wave equation in a medium with a variable index of refraction,'' J. Acoust. Soc. Am. **17**, 329–333 (1946).
[22] C. J. Higham and C. T. Tindle, ''Coupled perturbed modes over a sloping penetrable bottom,'' J. Acoust. Soc. Am. **114**, 3119–3124 (2003).
[23] R. G. Pratt and M. H. Worthington, ''Inverse theory applied to multi-source cross-hole tomography. Part 1: Acoustic wave-equation method,'' Geophys. Prospect. **38**, 287–310 (1990).
[24] D. N. G. Roy and G. J. Orris, ''A Born scatterer in an acoustical waveguide,'' J. Acoust. Soc. Am. **114**, 626–633 (2003).
[25] R. G. Pratt, C. Shin, and G. J. Hicks, ''Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion,'' J. Geophys. Int. **133**, 341–362 (1998).
[26] A. Thode, ''The derivative of a waveguide acoustic field with respect to a three-dimensional sound speed perturbation,'' J. Acoust. Soc. Am. **115**(6), 2824–2833 (2004).
[27] W. A. Kuperman and F. Ingenito, ''Spatial correlation of surface generated noise in a stratified ocean,'' J. Acoust. Soc. Am. **67**, 1988–1996 (1980).
[28] J. Ianniello, ''A MATLAB version of the KRAKEN normal mode code,'' Naval Undersea Warfare Center Report TM 94-1096 (1994).

Development of an accelerometer-based underwater acoustic intensity sensor

Kang Kim[a)]
The Pennsylvania State University, Graduate Program in Acoustics, 217 Applied Science Building, University Park, Pennsylvania 16802

Thomas B. Gabrielson
The Pennsylvania State University, Applied Research Laboratory, P.O. Box 30, State College, Pennsylvania 16804-0030

Gerald C. Lauchle
The Pennsylvania State University, Graduate Program in Acoustics, 218-B Applied Science Building, University Park, Pennsylvania 16802

(Received 23 March 2004; revised 26 July 2004; accepted 13 August 2004)

An underwater acoustic intensity sensor is described. This sensor derives acoustic intensity from simultaneous, co-located measurement of the acoustic pressure and one component of the acoustic particle acceleration vector. The sensor consists of a pressure transducer in the form of a hollow piezoceramic cylinder and a pair of miniature accelerometers mounted inside the cylinder. Since this sensor derives acoustic intensity from measurement of acoustic pressure and acoustic particle acceleration, it is called a *p-a* intensity probe. The sensor is ballasted to be nearly neutrally buoyant. It is desirable for the accelerometers to measure only the rigid body motion of the assembled probe and for the effective centers of the pressure sensor and accelerometer to be coincident. This is achieved by symmetric disposition of a pair of accelerometers inside the ceramic cylinder. The response of the intensity probe is determined by comparison with a reference hydrophone in a predominantly reactive acoustic field.
[DOI: 10.1121/1.1804632]

PACS numbers: 43.30.Yj, 43.20.Ye, 43.58.Fm [RAS] Pages: 3384–3392

I. INTRODUCTION

The acoustic intensity vector provides important information about acoustic power transport. The direction gives the local direction of power flow and the magnitude is equal to the time-averaged rate of acoustic power transport through a unit area perpendicular to the power-flow direction.

At any time, t, and any point given by the location vector, $\vec{\mathbf{r}}$, in a sound field, the instantaneous acoustic intensity, $\vec{\mathbf{I}}_i$, is the product of acoustic pressure, p, and acoustic particle velocity, $\vec{\mathbf{u}}$:

$$\vec{\mathbf{I}}_i(\vec{\mathbf{r}},t)=p(\vec{\mathbf{r}},t)\cdot\vec{\mathbf{u}}(\vec{\mathbf{r}},t). \tag{1}$$

For a sinusoidal signal, the instantaneous intensity would oscillate about some average value at twice the frequency of the sinusoid. The ordinary acoustic intensity is the time-average value:

$$\vec{\mathbf{I}}(\vec{\mathbf{r}})=\lim_{T\to\infty}\frac{1}{T}\int_0^T\vec{\mathbf{I}}_i(\vec{\mathbf{r}},t)dt. \tag{2}$$

This definition applies to a strictly stationary process. In practice, the nature of the signal and the type of processing determine the time used for the average. For a noise-free periodic waveform, an average over a single period is sufficient. For signals in the inevitable noise, the intensity is normally calculated as an average of some number of time records with finite length.

According to Fahy,[1] acoustic intensity can also be calculated from the cross spectrum between pressure and particle velocity:

$$\vec{\mathbf{S}}_I=\vec{\mathbf{S}}_{pu}(\vec{\mathbf{r}},\omega)=\lim_{T\to\infty}\frac{1}{T}E[p(\vec{\mathbf{r}},\omega,T)\cdot\vec{\mathbf{u}}^*(\vec{\mathbf{r}},\omega,T)], \tag{3}$$

where $\vec{\mathbf{S}}$ is the two-sided spectral density, ω is radian frequency, boldface indicates a complex quantity, the asterisk denotes a complex conjugate, and E is the average over many records. The quantities p and $\mathbf{u}$ are the finite Fourier transforms over the record length, T, of the acoustic pressure and particle velocity, respectively.

For stationary, ergodic processes, Eq. (2) is the cross-correlation function, $\vec{\mathbf{R}}_{pu}$, between p and $\vec{\mathbf{u}}$ for zero time lag. Then, by the Wiener–Khintchine theorem:[2]

$$\vec{\mathbf{I}}(\vec{\mathbf{r}})=\vec{\mathbf{R}}_{pu}(\vec{\mathbf{r}},0)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\vec{\mathbf{S}}_I(\vec{\mathbf{r}},\omega)d\omega. \tag{4}$$

Because the real part of $\vec{\mathbf{S}}_I$ is an even function of ω and the imaginary part is odd, we can write:

$$\vec{\mathbf{I}}(\vec{\mathbf{r}})=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{Re}[\vec{\mathbf{S}}_I(\vec{\mathbf{r}},\omega)]d\omega$$

$$=\frac{1}{\pi}\int_0^{\infty}\mathrm{Re}[\vec{\mathbf{S}}_I(\vec{\mathbf{r}},\omega)]d\omega, \tag{5a}$$

[a)]Present address: University of Michigan, Biomedical Engineering Department, Ann Arbor, MI 48109-2099; electronic mail: kangkim@umich.edu

or, equivalently,

$$\vec{\mathbf{I}}(\vec{\mathbf{r}}) = \frac{1}{2\pi}\int_0^\infty \mathrm{Re}[\vec{\mathbf{G}}_I(\vec{\mathbf{r}},\omega)]\,d\omega, \tag{5b}$$

where $\vec{\mathbf{G}}$ is the single-sided spectral density. (The single-sided density, $\vec{\mathbf{G}}$, is the normal output of signal analyzers; it is twice $\vec{\mathbf{S}}$ but only taken over positive frequencies.) The real part of $\vec{\mathbf{G}}$ determines completely the ordinary acoustic intensity. Often, the real part is called the ''active'' intensity spectrum and the imaginary part the ''reactive'' intensity spectrum. The active intensity is directly related to the ordinary acoustic intensity and describes net acoustic power flow. The reactive intensity is related to oscillatory exchange of acoustic energy.

The single-sided intensity spectrum, $\vec{\mathbf{G}}_I$, measured with a p-a probe is determined from the single-sided cross spectrum, $\vec{\mathbf{G}}_{pa}$, between the pressure signal and the accelerometer output, i.e.,

$$\vec{\mathbf{G}}_I(x,\omega) = \vec{\mathbf{G}}_{pu} = \frac{j}{\omega}\cdot\vec{\mathbf{G}}_{pa} = \frac{j}{\omega}\cdot\frac{2}{T}E[\mathbf{p}(x,\omega,T) \cdot \vec{\mathbf{a}}^*(x,\omega,T)], \tag{6}$$

where $\vec{\mathbf{a}}$ is the finite Fourier transform of the acoustic particle acceleration. The single-axis p-a probe used here measures only one component of particle acceleration so the intensity is determined in the direction, x. Notice that the active intensity is associated with the *imaginary* part of the cross-spectrum, $\vec{\mathbf{G}}_{pa}$, whereas the reactive intensity is associated with the real part.

Determining acoustic power flow has been of interest to researchers for many years.[3–10] Recently, efforts in this area have concentrated on the development of measurement techniques and devices. With the advent of new signal processing techniques and advances in transducer technology, intensity measurement devices have been improved to make them more reliable, accurate, and compact.[11] Much of the previous research on acoustic intensity measurement has been for in-air applications. However, some effort has been given to underwater intensity measurements. Corbett[12] demonstrated in water the methods developed by Seybert and Ross[13] and Chung and Blaser[14] for the measurement of the acoustic properties of materials using a two-microphone technique (the p-p technique). Clark and Watkinson[15] measured the underwater intensity due to a point-excited, periodically ribbed cylinder. DeVilbiss[16] applied the p-p technique to the measurement of the complex reflectivity of thin plates in water, while Segota[17] built a p-p probe with two closely spaced PVDF hydrophones. Van Tol[18] added a third hydrophone between the two used for determining the pressure gradient to obtain the pressure component directly, thus eliminating the need to average two hydrophone signals. Ng[19] was granted a patent for an underwater p-p intensity probe.

While many intensity sensors for underwater use are based on pressure differencing (the p-p concept) there has been interest in measuring acoustic particle velocity directly, using neutrally buoyant inertial sensors.[20] A two-component inertial vector sensor using moving-coil transducers was developed by Bauer and DiMattia,[21] while D'Spain *et al.*[22] describe the Marine Physics Laboratory's freely drifting Swallow Float System as a very low-frequency (0.5–20 Hz) underwater vector sensor that can be used to measure particle velocity and intensity.

Gabrielson *et al.*[23] built and successfully tested a neutrally buoyant inertial sensor consisting of a glass-microballoon-and-epoxy composite cast surrounding a small, commercial geophone. When this type of velocity sensor includes a co-located hydrophone, a p-u type underwater intensity sensor results.[24] This type of sensor is not limited by the finite difference and phase mismatch errors associated with p-p probes.[10] McConnell *et al.*[25] describe an intensity sensor based on two closely spaced velocity sensors. This concept infers acoustic pressure through the use of the continuity equation and measures the acoustic particle velocity directly. This u-u probe is basically a velocity gradient hydrophone that can also be used to measure several components of the acoustic particle velocity tensor field, components that are associated with density fluctuations.[26,27] However, the u-u probe has finite-difference and phase mismatch errors similar to those that affect the p-p probe. Furthermore, the u-u probe has greater sensitivity to nonacoustic motion and vibrations (see Bastyr *et al.*[28]).

The subject of the present paper is the design, development, and evaluation of an underwater pressure-particle acceleration intensity probe: the p-a probe. The one-component probe utilizes a neutrally buoyant body composed of a hollow piezoceramic cylinder to measure pressure and a pair of internally mounted accelerometers to measure the rigid body probe acceleration. (An obvious extension to three components involves three pairs of accelerometers mounted inside a spherical piezoceramic shell.) The rigid body acceleration is the same as the acoustic particle acceleration as long as the probe is neutrally buoyant, and as long as the acoustic wavelength remains large compared to sensor dimensions.[20] The inherent high resonance frequency of an accelerometer allows this type of intensity sensor to be used at higher frequencies than those of geophone-based intensity sensors.[24] However, the self-noise of miniature accelerometers is typically much higher than that of geophones.

II. DESIGN AND CONSTRUCTION OF THE INTENSITY SENSOR

A. Introduction

An acoustically small body immersed in water and free to move responds directly to the acoustic particle motion. If the body is neutrally buoyant then the body's motion is the same as that of the surrounding fluid. If the body is negatively buoyant the amplitude of the motion of the body is reduced compared to that of the fluid; if the body is positively buoyant the amplitude of the body's motion is larger than that of the fluid. Consequently, if an inertial transducer (an accelerometer or geophone) is embedded in the body, a signal is produced that can be related to the acoustic particle motion. If this sensor is coupled with a pressure sensor, acoustic intensity can be determined by Eqs. (1), (3), or (6).

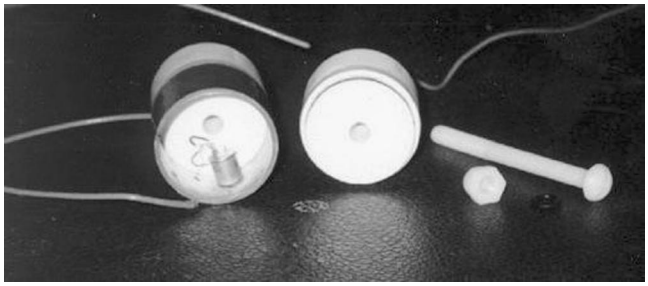


FIG. 1. Partially assembled, single-axis p-a intensity sensor. Only one accelerometer is shown here. Later in the assembly, the other accelerometer is attached to the cylinder wall directly opposite the one shown here.

Because the particle velocity is measured directly, there is no error associated with the subtraction of nearly equal signals common with gradient sensors.

For the single-component intensity sensor described here, a hollow piezoceramic cylinder is employed as the pressure sensor. A pair of Oceana Sensor Technologies piezoceramic miniature accelerometers (A5100-CO2-0/5) is used to measure acoustic particle acceleration. The accelerometers are bonded directly to the inner surface of the pressure sensor with epoxy, as shown for an early prototype in Fig. 1.

The effective centers of the velocity sensor and the pressure sensor of an intensity probe should be co-located. An accelerometer embedded at the acoustic center of a pressure sensor is the configuration considered here. The pair of accelerometers is connected electrically in series as shown in Fig. 2 and arranged symmetrically with respect to the cylinder axis so that the effective center of acceleration sensing is on the cylinder axis, which is also the effective center of the pressure sensor.

This arrangement also cancels the output that results from the radial vibration of the ceramic cylinder. In practice, this is not a significant advantage because the radial vibration that results from the acoustic pressure is important only at very high frequencies.[29]

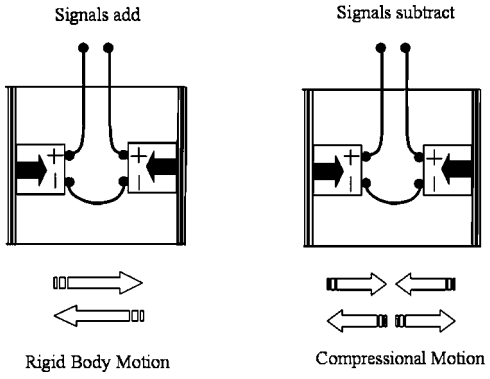


FIG. 2. Schematic connection and mechanical mounting of the accelerometers. The dark arrows inside the accelerometers indicate the directions of the sensing. On the left is the rigid-body motion of the neutrally buoyant intensity sensor, corresponding to the acoustic particle motion of the surrounding fluid. On the right is the radial vibration of the ceramic cylinder due to the same acoustic field. Accelerometers connected in series reduce the response to the radial vibration but increase the response to the rigid-body motion. If the two accelerometers are perfectly matched in sensitivity, the sensor will not respond to the radial vibration but double the response to the rigid-body motion.

TABLE I. The masses and fluid-displacement volumes of the components of the one-component p-a intensity sensor. The sensor is designed to be slightly positively buoyant to compensate for the signal cable mass.

Component	Mass (g)	Volume[a] (cc)
Piezoceramic cylinder	36.9	29.0
Two Accelerometers (with wires)	7.8	—
Two End-caps	23.8	52.1
Two O-rings	0.4	—
Adhesive and sealant	2.0	—
Added mass[b]	9.3	—
Total	80.2	81.1[c]

[a]Fluid displacement volume.
[b]Melted alloy (Cerro Bend: a low melting point alloy of lead, tin, zinc, and bronze manufactured by Cerro Metal Products in Bellefonte, PA) is evenly spaced inside the end caps.
[c]The density of water at 20 °C is 0.998 g/cc.

B. Neutral buoyancy

If the sensor body is neutrally buoyant then the sensor body acceleration is equal to the fluid particle acceleration. While neutral buoyancy is not a requirement for successful operation, it does make interpretation of the measurements simpler. Both the hollow piezoceramic cylinder and the miniature piezoceramic accelerometers are negatively buoyant. To achieve neutral buoyancy, two syntactic foam end caps with density 0.38 g/cm^3 are fitted into the ends of the hollow cylinder. The end-cap volume is determined so that the overall density of the sensor is slightly less than 1 g/cm^3 (see Table I). A low-melting point metal alloy is added to fine tune the buoyancy.

The outer diameter of the piezoceramic cylinder is 3.81 cm and the length is 2.54 cm. The diameter of the end-cap matches that of the cylinder at each end and the length exposed to water at each end is 2.29 cm. Thus, the finished intensity sensor has an outside diameter of 3.81 cm and it is 7.11 cm long. The mass of water displaced by the intensity sensor is 80.9 g (at 20 °C) and the mass of the sensor is 80.2 g. The slight positive buoyancy helps compensate for the additional mass loading of the signal cable.

C. Suspension system

Inertial sensors are sensitive to any motion. Besides the desired motion that results from acoustic waves, structure-borne vibration, and flow-turbulence forces can produce undesired output.[23,30–34] Consequently, the intensity probe suspension system should be designed in a way that isolates the sensors from all nonacoustic vibration. The referenced studies have detailed descriptions for the design of inertial-sensor suspension systems. It is important that the suspension system (a) have a natural frequency well below the intended range of acoustic sensing, (b) fix the average position and orientation of the sensor body, (c) permit movement of the sensor body in response to the acoustic field, (d) isolate the

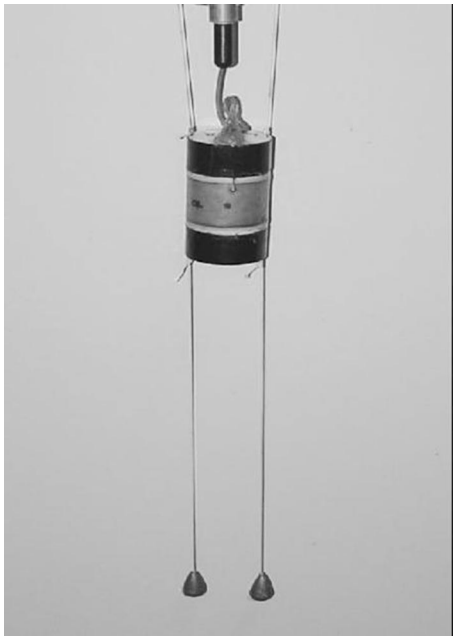


FIG. 3. The two-line suspension system for testing the *p-a* probe. The accelerometer response axis is perpendicular to the plane of the photograph. The electrical cable that extends from the top of the probe is coiled loosely to minimize its influence on the probe motion.

sensor from structure-borne vibration, (e) not distort the response of the sensor in either magnitude, phase, or apparent sound incidence angle, and (f) withstand operational shock.

The present intensity sensor is intended to measure one component of the acoustic vector field, so the acoustically induced motion of the sensor body must be unimpeded in this direction. A taut-line suspension (shown in Fig. 3) was used here. Each line of the two-line suspension has its upper end fixed to a positioner above the water surface. Each line is attached to the cylinder at the top and bottom and the lower ends are attached to 13.5 g weights. The lines are attached to the cylinder so that the accelerometer sense axis is perpendicular to the plane of attachment. This allows free motion in the sense direction but constrains rotation.

D. Pressure sensor

A hollow, piezoceramic cylinder of length $l=2.54\,\text{cm}$, outer diameter $d_o=3.81\,\text{cm}$, and thickness $t=1.5\,\text{mm}$ is used as the pressure sensor. The material is Channel Industries type C5500 (Channel Industries, Santa Barbara, CA; www.channelindustries.com). The cylinder is electrically poled in the radial direction. The primary strain for the breathing mode of the cylinder is tangential; therefore, the predominant piezoelectric coupling is 3-1. Some physical and electrical properties of the cylinder are given in Tables II and III.

TABLE III. Primary relations between the physical and electrical parameters of the hollow cylinder piezoceramic transducer.

Resonance frequency (f_o)	Capacitance (C)	Generated voltage (V)	Displacement (Δd_m)
$\frac{N_c}{d_m}$	$\frac{2\pi K_3^T \epsilon_o l}{\ln(d_o/d_i)}$ [a]	$\frac{1}{2} g_{31} d_o p$	$\frac{d_{31} d_m}{t} V$

[a] ϵ_o is the dielectric permittivity for air: 8.85×10^{-12} (F/m).

The driving-point admittance of the piezoceramic transducer was measured in both air and water to determine its resonance frequency, blocked capacitance, and quality factor. As shown in Fig. 4, the resonance frequency was 16.8 kHz in water and 24.3 kHz in air. The in-air resonance frequency calculated from the material characteristics in Tables II and III was 24.9 kHz, which differs by only 2% from the measurements. The quality factor, Q, was determined to be 28 in water, and 35 in air. From analysis of the admittance circle, the blocked electrical capacitance was found to be 29.8 nF and the electrical-equivalent capacitance of the mechanical part was found to be 0.67 nF. Their ratio gives the effective coupling factor for this configuration as $\kappa_{\text{eff}}^2=0.0225$. The free capacitance (the capacitance that would be measured at the transducer terminals well below the mechanical resonance) is the blocked capacitance divided by $(1-\kappa_{\text{eff}}^2)$ or 30.5 nF. Direct calculation of the free capacitance using Tables II and III gives 30.2 nF.

The capacitance of the cabling or circuitry connected to the pressure sensor will reduce the output from the sensor. To minimize this adverse effect, a preamplifier with a gain of 10 and an input impedance of 10 MΩ was designed using a low-noise, BiFET operational amplifier, AD743 (Analog Devices, Norwood, MA; www.analog.com).

The response of the piezoceramic cylinder to acoustic pressure (including this preamplifier) was measured by the comparison technique in the Applied Research Laboratory calibration facility and determined to be 3.15 mV/Pa (-170 ± 0.5 dB re 1 V/μPa) from 10 Hz to well over 1000 Hz.

E. Accelerometer

The piezoceramic accelerometer uses a tungsten mass coupled to a sensitive, annular shear, piezoceramic element. It has an internal low-noise preamplifier that requires a constant current supply of between 0.3 and 0.6 mA. A constant-

TABLE II. Fundamental physical and electrical properties of the hollow cylinder piezoceramic transducer manufactured by Channel Industries, Inc. The material is Channel Type C5500. The cylinder has an outside diameter of 3.81 cm, a length of 2.54 cm, and a wall thickness of 1.5 mm.

Coupling coefficient k_{31}	Piezoelectric constant $g_{31}(10^{-3}\,\text{Vm/N})$	Piezoelectric constant $d_{31}(10^{-12}\,\text{m/V})$	Free dielectric constant K_3^T	Frequency constant N_c (Hz m)	Density $\rho(10^3\,\text{kg/m}^3)$	Mean diameter d_m (cm)
−0.37	−11.9	−185	1750	910	7.6	3.73

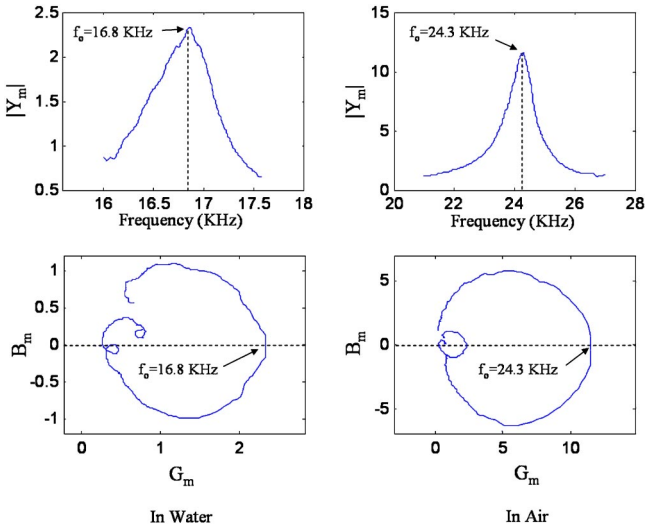


FIG. 4. Admittance ($Y_m = G_m + jB_m$) measurement of hollow cylinder piezoceramic transducer of length $l = 2.54$ cm, outer diameter $d_o = 3.81$ cm, and thickness $t = 1.5$ mm. For display, each component of admittance and its magnitude is multiplied by 1000. The blocked electrical capacitance C_0 has been subtracted from these data. The admittance circle in water is overlaid (small circle in lower right panel) on the admittance circle in air for better comparison.

current diode is incorporated between the voltage supply (5 Vdc) and the sensor to accomplish this. The sensor power and signal share the same line; the dc bias is removed with a blocking capacitor.

The responses of the accelerometers mounted in the cylinder were measured (in air) using a Wilcoxon F4 shaker (Wilcoxon Research, Gaithersburg, MD; www.wilcoxon.com) and a PCB Piezotronics W353366 reference accelerometer (PCB Piezotronics, Depew, NY; www.pcb.com). The cylinder was mounted on the shaker using steel clamps on both sides of the end caps. The pair of accelerometers was wired as shown in Fig. 2. They were calibrated twice—once with their sense axes in the same direction and once with their sense axes in the opposite direction. The response curves for the individual and combined accelerometer outputs are shown in Fig. 5. The data indicate that reasonable cancellation (25–35 dB) of the radial vibration of the cylinder is possible from the accelerometers wired in series. The degree of cancellation depends not only on the matching of accelerometer magnitude response, but also on the relative phase response between them.

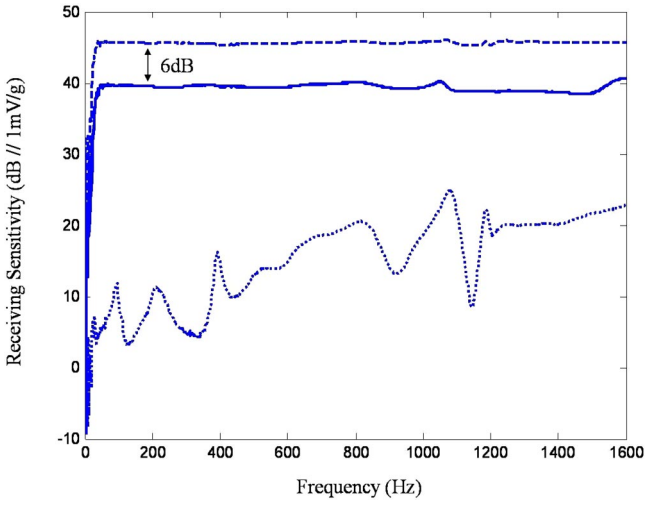


FIG. 5. Calibration of accelerometers mounted in the cylinder using a Wilcoxon F4 shaker and a PCB Piezotronics W353366 reference accelerometer. The dashed line represents the doubled response curve for combined accelerometer output. The solid lines are response curves for each individual accelerometer and they are almost identical. The dotted line at the bottom of the plot indicates reasonable cancellation (25–35 dB) of the radial vibration.

III. RESPONSE OF THE INTENSITY PROBE

The pressure sensor and the accelerometer of the intensity probe produce voltages, not pressure and acceleration. Therefore, a voltage cross spectrum must be converted to a pressure-acceleration cross spectrum using the sensor responses. In the frequency domain, the voltage output, V, is related to the acoustic quantity by the complex response, H, of the sensor:

$$V_{ps} = pH_{ps}, \tag{7}$$

$$V_{as} = aH_{as}, \tag{8}$$

where the subscript ps relates to the pressure sensor and the subscript as relates to the acceleration sensor. The relationship between the measured cross spectrum between voltage outputs and the desired cross spectrum between acoustical quantities is then

$$\frac{G_{V_{ps}V_{as}}}{G_{pa}} = \frac{2E(V_{ps}V_{as}^*)/T}{2E[pa^*]/T} = H_{ps}H_{as}^*. \tag{9}$$

(The response factors are not random so we do not need to carry the averages through to the right side of this equation.) This product of responses is, in effect, the calibration factor for the intensity probe since

$$G_I = \frac{j}{\omega} \cdot \frac{G_{V_{ps}V_{as}}}{H_{ps}H_{as}^*}. \tag{10}$$

One approach to response measurement is to use a calibrated reference pressure sensor (a standard hydrophone) and a well-known sound field. If we measure the ratio of the voltage output of the intensity sensor hydrophone to the voltage output of the reference hydrophone and the ratio of the voltage output of the accelerometer to the voltage output of the reference hydrophone, we can write

$$H_{ps}H_{as}^* = \frac{1}{pa^*} \cdot \frac{V_{ps}}{V_{pr}} \cdot \frac{V_{as}^*}{V_{pr}^*} \cdot V_{pr}V_{pr}^*, \tag{11}$$

where the subscript, pr, refers to the reference pressure hydrophone. Since

$$V_{pr}V_{pr}^* = pp^* \cdot H_{pr}H_{pr}^* = pp^* \cdot |H_{pr}|^2 \tag{12}$$

the intensity-sensor calibration product becomes

$$H_{ps}H_{as}^* = \left(\frac{p}{a}\right)^* |H_{pr}|^2 \left(\frac{V_{ps}}{V_{pr}}\right)\left(\frac{V_{as}}{V_{pr}}\right)^*. \tag{13}$$

This result defines what is required for this variety of intensity sensor response measurement and what is not required.

The complex ratios of sensor output voltage to reference output voltage are required for both the intensity-sensor hydrophone and the accelerometer. These measurements are straightforward. Also, the complex ratio of acoustic pressure to acoustic particle acceleration for the acoustic field in which the response measurement is performed must be known. The magnitude of the reference hydrophone's response must be known but the phase response of the reference is *not* required.

In order to determine the separate complex response of either the intensity-sensor hydrophone or accelerometer, we would need to know the complex response of the reference hydrophone. But, for the intensity sensor, we need only the product of H_{ps} and the complex conjugate of H_{as}. Since we know that the phase of the reference sensor cancels out in this product, we can assume any phase for the reference hydrophone. The most convenient assumption is that the phase of the reference is zero.

With this assumption, the response of the intensity-sensor hydrophone is determined as an ordinary comparison measurement:

$$H_{ps} = |H_{pr}| \cdot \frac{V_{ps}}{V_{pr}}. \tag{14}$$

The ratio of acoustic acceleration to acoustic pressure is

$$\frac{a}{p} = \frac{V_{as}}{H_{as}} \cdot \frac{H_{pr}}{V_{pr}}, \tag{15}$$

therefore,

$$H_{as} = |H_{pr}| \cdot \frac{p}{a} \cdot \frac{V_{as}}{V_{pr}}. \tag{16}$$

The response of the accelerometer should not be a function of the nature of the acoustic field so any field variations in the ratio of pressure to acceleration should be offset by the measured voltage ratio between the accelerometer output and the reference output. This gives a check of the validity of the p/a model for the acoustic field. If the model is wrong, then features associated with the acoustic field will appear in the calculated response.

Over most of the frequency range considered in this paper, the reference hydrophone used has a phase response of very nearly zero. (That is, the voltage output of the reference hydrophone is zero with respect to the phase of the acoustic pressure at the hydrophone.) But this is not essential for determination of the intensity sensor response. What must be understood is that, using this procedure, the phase of the individual sensor responses may be unknown but the phase errors introduced by the assumption of zero phase for the reference cancel out in the response product [Eq. (13)].

IV. RESPONSE MEASUREMENT IN THE SLOW-WAVE TUBE

The simplest field in which to perform an intensity-sensor response measurement is a one-dimensional traveling plane-wave field but this is a difficult field to generate accurately, especially at low frequency. The low-frequency measurement of the p-a probe described here was performed in a slow-wave tube; a fluid-filled, compliant-wall tube driven at one end. Compliant-walled waveguides have a lower longitudinal sound speed compared to a similar rigid-walled tube; therefore, the standing wave pattern at a given frequency in a slow-wave tube will exhibit more nodes and antinodes than in the rigid tube of the same length. The slow waveguide is convenient for evaluating the performance of intensity sensors because the sensor can be traversed along the length of the tube and the measured pressure and particle

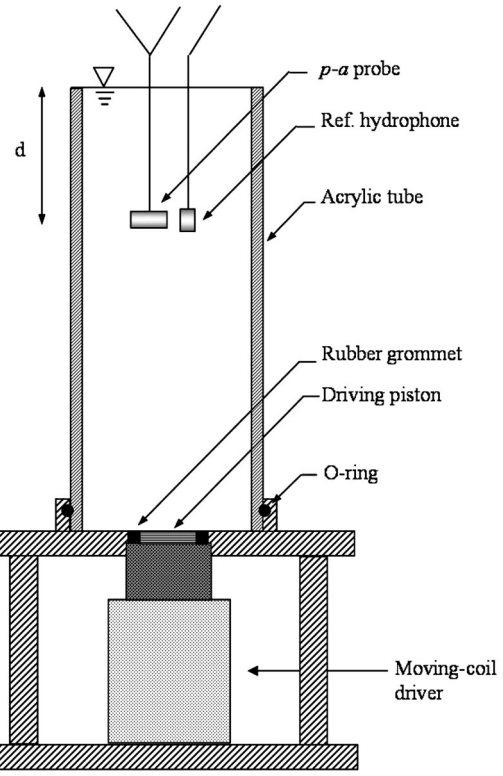


FIG. 6. Experimental setup for the calibration in SWC. The p-a intensity probe and reference hydrophone scan the sound field in the vertical direction.

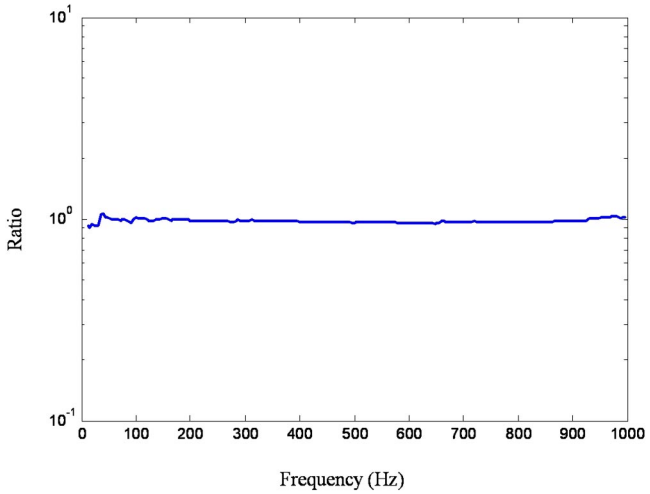


FIG. 7. Response comparison between intensity sensor pressure element and reference hydrophone. The magnitude of the ratio of the pressure element output to the reference hydrophone output is plotted from 16 to 1000 Hz. The nominal response of the reference hydrophone (including the conditioning electronics) is 3.16 mV/Pa.

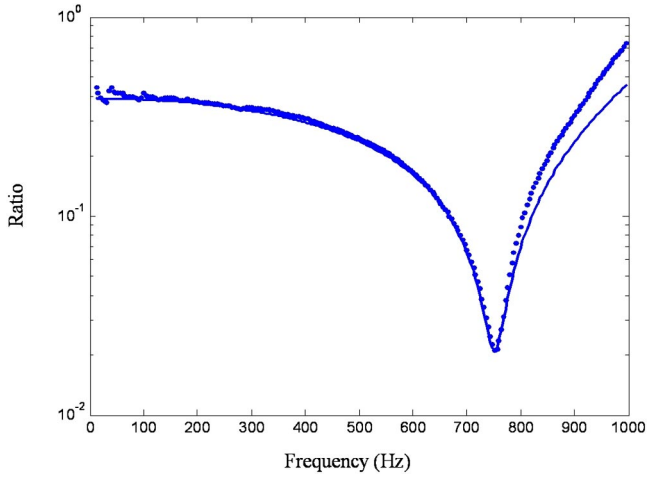


FIG. 8. Comparison between the measured ratio of acoustic pressure to acoustic acceleration (dots) and the lossy standing wave model (solid line). The model is fit to the measurement at only one point: the null at 750 Hz. The magnitude is shown here but both the real and imaginary parts compare equally well.

velocity should mimic the rich modal structure expected, if the sensor is designed properly. The design and operation of the slow-wave "calibrator" used in this research are discussed in detail in Bastyr *et al.*[28] and will not be repeated here. Acoustic wave fronts were measured and found to be plane (except near the walls) for frequencies up to 900 Hz.

The p-a intensity probe was ballasted for neutral buoyancy and suspended in the water. The physical arrangement is shown schematically in Fig. 6. The centerline of the sensor cylinder was perpendicular to the axis of the 20-cm-diameter slow-wave tube and the sensor cylinder was positioned so that the acceleration sense axis was parallel to the axis of the tube. The intensity probe and the reference hydrophone were placed at a depth of 11.4 cm below the water surface and spaced less than 3 cm apart. A USRD J-9 projector (USRD, Newport, RI; www.npt.nuwc.navy.mil/USRD/) drove the slow-wave tube with white noise up to 1600 Hz. The source level was raised until the measured signal was significantly above the sensor noise floors.

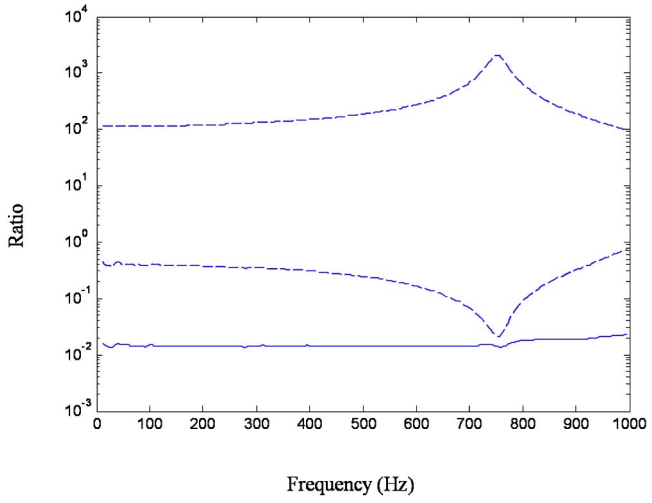


FIG. 9. Magnitudes of several of the factors in Eq. (16) plotted from 16 to 1000 Hz. The upper dashed curve is the predicted p/a. The lower dashed curve is the measurement: V_{as}/V_{pr}. Their product times the magnitude of the reference hydrophone response [see Eq. (16)] is the accelerometer response, H_{as}, and is plotted as the solid line. The accelerometer response function should not show the field variations evident in either dashed curve.

The response measurement results are reported here to 1 kHz. From a few Hz to 1 kHz, the reference hydrophone, a Reson TC4013 (Reson, Goleta, CA; www.reson.com/TC4013.html), has a nominally flat response of 31.6 μV/Pa (-210 dB re 1 V/μPa) and the reference hydrophone was followed by a preamplifier with a gain of 100. The absolute response is probably within 20% based on comparison to a laboratory standard hydrophone. More relevant for the present purposes are the additional uncertainties introduced by the specific measurement procedure for the intensity sensor and these uncertainties will be estimated.

From direct comparison of the amplified voltage outputs of the intensity sensor hydrophone and the reference hydrophone (see Fig. 7):

$$\frac{V_{ps}}{V_{pr}} = 0.98 \pm 0.03, \tag{17}$$

so the intensity sensor hydrophone response is about 3.1 $\pm$0.1 mV/Pa including the preamplifier. The uncertainty indicated is the additional uncertainty over the uncertainty in the reference.

The measurement of the ratio of the accelerometer output to the reference hydrophone output (Fig. 8) shows the character of a lossy standing wave. A reasonable model for the field in the slow-wave calibrator is

$$p = p_0 \sin(kx) e^{j\omega t} \tag{18}$$

and

$$a = \frac{\omega p_0}{\rho c} \cos(kx) e^{j\omega t}, \tag{19}$$

where k is a complex wave number

$$k = \frac{\omega}{c}(1 + j\alpha), \tag{20}$$

c is the effective sound speed in the water in the tube, ρ is the density, and α is a dimensionless loss parameter. Fitting this model to the measurement at the null near 750 Hz yields values of $c = 344$ m/s and $\alpha = 0.022$. This model fits the measurement well to about 800 Hz (see Fig. 8) and is consistent with the effective sound speed observed and calculated in other studies.[35,36] A model that includes dispersion fits the measurement well even above 800 Hz but the added complexity is not justified here.

Using Eq. (13) and the measured voltage ratio and Eqs. (18) and (19) for p/a, we find the response of the accelerometer to be constant at 14 ± 0.5 mV/(m/s^2). Figure 9 shows that the derived accelerometer response shows very little of the strong variation in the acoustic field just below 800 Hz. Only the magnitude is plotted in Fig. 9; the imaginary part of the accelerometer response is very small indicating that the relative phase between the sensors is negligible.

A detailed set of near- and far-field measurements was made with this intensity probe of the sound field created by single and multiple sources in a large water reverberant

tank.[28] The results compared very favorably with theoretical predictions performed for these sound fields. These findings will be published separately.

V. CONCLUSIONS

An inertial-sensor based acoustic vector sensor that functions underwater as an intensity sensor has been demonstrated. Some useful design parameters were determined for this p-a intensity sensor. The response of the p-a intensity sensor was determined by comparison with a reference hydrophone in a predominantly reactive acoustic field. Below 800 Hz where the interference of acoustic field due to the sensor itself is negligible (ka for the sensor body less than 0.3, where $a \approx 1.9$ cm, the radius of the piezoelectric cylinder), the response of the intensity sensor reasonably follows the lossy plane-wave model for slow-wave calibrator.

Miniature accelerometers provide adequate response to measure acoustic particle acceleration (from which velocity can be computed) of underwater sound waves over a reasonably wide band below the inherently high resonance frequency of the accelerometer. However, such accelerometers have a high noise floor compared to low ocean ambient conditions. The placement of these accelerometers within a cylindrical piezoelectric hydrophone element is reported here. It is critical that opposing accelerometers be used in order to cancel the breathing mode of the cylinder, while retaining the rigid body response of the near neutrally buoyant shell to acoustic waves. Any nonsymmetric mode of the cylinder, which may be induced by adding a mass locally, will change the response of the accelerometer pair to the rigid body motion. It is important to balance the mass distribution to prevent spurious resonances of the ceramic. The metal housing of the accelerometer as a ground will interfere electrically with the cylinder surface as an electrode. Electrical and mechanical isolation of the accelerometers from the piezoceramic element is also important. Routing of the electrical wires from the intensity probe must be carefully performed to assure that the individual sensors making up the probe have symmetrical response. The inertial sensor is sensitive to all kinds of motion, so careful design of the suspension system is required.

Compared to conventional hydrophone arrays and other versions (e.g., p-p types) of intensity sensors, p-a intensity sensors can be designed to be more compact so they can be applied not only for directional measurements in the far field, but also for near-field intensity mapping and source localization on complex underwater structures. Although not discussed earlier, another very important advantage of a colocated pressure and particle acceleration (or velocity) sensor is the probe's ability to extract the active, propagating intensity of a source of interest from a reactive field due to reverberation and/or isotropic diffuse noise. This ability is because of the cross-spectral processing and direct measurement of active and reactive components of the intensity. The isotropic noise and reverberation have strong reactive components, but negligible active components. Such discrimination is impossible with an omnidirectional pressure sensor acting alone. Support for this comment may be found in articles by Jacobsen and his co-workers[37,38] that show expressions for the coherence function between a pressure sensor and a velocity sensor. This coherence approaches zero in an isotropic noise field as the sensor separation distance approaches zero. Lo and Junger[39] showed further that time-averaged intensity measurements discriminate against isotropic noise like that created by rain on the ocean surface.

ACKNOWLEDGMENTS

This work has been supported by several Grants from the Office of Naval Research (ONR), Code 321SS. They include an MURI Grant No. N00014-96-1-1173, and regular Grants Nos. N00014-00-1-0300 and N00014-01-1-0108. The support of Program Manager Dr. James F. McEachern is gratefully appreciated.

[1] F. J. Fahy, ''Measurement of sound intensity using the cross-spectral density of two microphone sensors,'' J. Acoust. Soc. Am. **62**, 1057–1059 (1977).
[2] J. S. Bendat and A. G. Piersol, *Random Data Analysis and Measurement Procedures*, 3rd. ed. (Wiley Interscience, New York, 2000), p. 129.
[3] H. F. Olson, U.S. Patent No. 1,892,644 (December 27, 1932).
[4] J. H. Enns and F. A. Firestone, ''Sound power density fields,'' J. Acoust. Soc. Am. **14**, 24–32 (1942).
[5] R. H. Bolt and A. A. Petrauskas, ''An acoustic impedance meter for rapid field measurement,'' J. Acoust. Soc. Am. **15**, 79(A) (1943).
[6] S. Baker, ''Acoustic intensity meter,'' J. Acoust. Soc. Am. **27**, 269–275 (1955).
[7] T. J. Schultz, ''Acoustic Wattmeter,'' J. Acoust. Soc. Am. **28**, 693–699 (1956).
[8] J. F. Burger, G. J. J. van der Merwe, G. B. van Zyl, and L. Joffe, ''Measurement of sound intensity applied to the determination of sound power,'' J. Acoust. Soc. Am. **53**, 1167–1168 (1973).
[9] B. F. van Zyl and F. Anderson, ''Evaluation of intensity method of sound power determination,'' J. Acoust. Soc. Am. **57**, 682–686 (1975).
[10] J. Y. Chung, ''Cross-spectral method of measuring acoustic intensity without error caused by instrument phase mismatch,'' J. Acoust. Soc. Am. **64**, 1613–1616 (1978).
[11] F. J. Fahy, *Sound Intensity*, 2nd ed. (E&FN, Spon, 1995).
[12] S. S. Corbett, M.S. thesis in Acoustics, The Pennsylvania State University, 1983.
[13] A. F. Seybert and D. F. Ross, ''Experimental determination of acoustic properties using a two-microphone random excitation technique,'' J. Acoust. Soc. Am. **61**, 1362–1370 (1977).
[14] J. Y. Chung and D. A. Blaser, ''Transfer function method of measuring in-duct acoustic impedance, (i) theory, (ii) experiment,'' J. Acoust. Soc. Am. **68**, 907–921 (1978).
[15] A. R. Clark and P. S. Watkinson, in *Shipboard Acoustics*, edited by J. Bruiten (Kluwer Academic, Dordrecht, The Netherlands, 1986), pp. 177–188.
[16] D. DeVilbiss, M.S. thesis in Acoustics, The Pennsylvania State University, 1989.
[17] J. P. Segota, M.S. thesis in Acoustics, The Pennsylvania State University, 1990.
[18] D. J. Van Tol, M.S. thesis in Acoustics, The Pennsylvania State University, 1996.
[19] K. W. Ng, U.S. Patent No. 4,982,375 (January 1, 1991).
[20] C. B. Leslie, J. M. Kendall, and J. L. Jones, ''Hydrophone for measuring particle velocity,'' J. Acoust. Soc. Am. **28**, 711–715 (1956).
[21] B. Bauer and A. DiMattia, ''Moving-coil pressure-gradient hydrophone,'' J. Acoust. Soc. Am. **39**, 1264(A) (1966).
[22] G. L. D'Spain, W. S. Hodgkiss, and G. L. Edmonds, ''The simultaneous measurement of infrasonic acoustic particle velocity acoustic pressure in the ocean by freely drifting swallow floats,'' IEEE J. Ocean. Eng. **16**, 195–207 (1991).
[23] T. B. Gabrielson, D. L. Gardner, and S. L. Garrett, ''A simple neutrally buoyant sensor for direct measurement of particle velocity and intensity in water,'' J. Acoust. Soc. Am. **97**, 2227–2237 (1995).
[24] T. B. Gabrielson, J. F. McEachern, and G. C. Lauchle, U.S. Patent No. 5,392,258 (February 21, 1995).

[25] J. A. McConnell, G. C. Lauchle, and T. B. Gabrielson, U.S. Patent No. 6,172,940 B1 (January 9, 2001).
[26] G. L. D'Spain, ''Relationship of underwater acoustic intensity measurements to beamforming,'' Can. Acoust. **22**, 157–158 (1994).
[27] G. L. D'Spain and W. S. Hodgkiss, ''Further comments on beamforming with acoustic measurements at a single point in the ocean,'' J. Acoust. Soc. Am. **93**, 2375 (1993).
[28] K. J. Bastyr, G. C. Lauchle, and J. A. McConnell, ''Development of a velocity gradient underwater acoustic intensity sensor,'' J. Acoust. Soc. Am. **106**, 3178–3188 (1999).
[29] K. Kim, ''Investigation of an underwater acoustic vector sensor,'' Ph.D. Thesis in Acoustics, The Pennsylvania State University, 2002.
[30] J. A. McConnell, ''Analysis of a compliantly suspended acoustic velocity sensor,'' J. Acoust. Soc. Am. **113**, 1395–1405 (2003).
[31] J. F. McEachern and G. C. Lauchle, ''Flow-induced noise on a bluff body,'' J. Acoust. Soc. Am. **97**, 947–953 (1995).
[32] D. E. Capone and G. C. Lauchle, ''Modeling the unsteady lift and drag on a finite-length cylinder in cross flow,'' J. Fluids Struct. **14**, 799–817 (2000).
[33] D. E. Capone and G. C. Lauchle, ''Modeling the unsteady axial forces on a finite-length circular cylinder in cross flow,'' J. Fluids Struct. **16**, 667–683 (2002).
[34] J. Wang, G. C. Lauchle, and M. S. Howe, ''Flow-induced force fluctuations on a sphere at high Strouhal number,'' J. Fluids Struct. **17**, 365–380 (2003).
[35] V. T. Grinchenko and G. L. Komissarova, ''Characteristics of normal waves in elastic-fluid cylindrical waveguides,'' Int. J. Fluid Mech. **29**, 742–759 (2002).
[36] M. C. Junger and D. Feit, *Sound, Structures, and Their Interaction* (Acoustical Society of America, Woodbury, NY, 1999), p. 39.
[37] F. Jacobsen and T. G. Nielsen, ''Spatial correlation and coherence in a reverberant sound field,'' J. Sound Vib. **118**, 175–180 (1987).
[38] F. Jacobsen and T. Roisin, ''Spatial coherence in reverberant sound fields,'' Proc. Inter-Noise 99, 1559–1564, Fort Lauderdale, FL (December 6–8, 1999).
[39] E. Y. Lo and M. C. Junger, ''Signal-to-noise enhancement by underwater intensity measurements,'' J. Acoust. Soc. Am. **82**, 1450–1454 (1987).

Experimental demonstration of remote, passive acousto-optic sensing

Lynn Antonelli[a] and Fletcher Blackmon
Naval Undersea Warfare Center, 1176 Howell Street, Newport, Rhode Island 02841

(Received 13 January 2004; revised 9 September 2004; accepted 9 September 2004)

Passively detecting underwater sound from the air can allow aircraft and surface vessels to monitor the underwater acoustic environment. Experimental research into an optical hydrophone is being conducted for remote, aerial detection of underwater sound. A laser beam is directed onto the water surface to measure the velocity of the vibrations occurring as the underwater acoustic signal reaches the water surface. The acoustically generated surface vibrations modulate the phase of the laser beam. Sound detection occurs when the laser is reflected back towards the sensor. Therefore, laser alignment on the specularly reflecting water surface is critical. As the water surface moves, the laser beam is reflected away from the photodetector and no signal is obtained. One option to mitigate this problem is to continually steer the laser onto a spot on the water surface that provides a direct back-reflection. Results are presented from a laboratory test that investigates the feasibility of the acousto-optic sensor detection on hydrostatic and hydrodynamic surfaces using a laser Doppler vibrometer in combination with a laser-based, surface normal glint tracker for remotely detecting underwater sound. This paper outlines the acousto-optic sensor and tracker concepts and presents experimental results comparing sensor operation under various sea surface conditions. [DOI: 10.1121/1.1811475]

PACS numbers: 43.30.Yj, 43.38.Zp, 07.60.Ly [YHB] Pages: 3393–3403

I. INTRODUCTION

Detecting underwater sound from the air, without deploying any hardware, would provide a remote, aerial sensing capability, presenting opportunities for use in a number of commercial and military applications. Such a sensor technology could enable passive sonar[1] to detect and monitor marine mammal sounds and enable aerial reception of underwater acoustic communication signals.[2]

Ceramic-based hydrophones have long been used to detect underwater sound and form the basis of modern day sonar systems. However, these sensors must be physically immersed in the water for efficient acoustic signal reception. Remote sensing of underwater sound from the air has been achieved using reconnaissance aircraft, equipped with sensors relaying information about the oceanic environment back to the Naval vessel. The aircraft's sensors need to be able to detect underwater sound from in the air and typically deploy sonobuoys with radio frequency links relaying information back to the aircraft.

The concept of using lasers to probe an air–water interface for measuring the surface velocity to detect underwater sound evolved from water flow velocity measurements.[3] The use of lasers has since been used to measure the *transverse* water surface velocity produced by the underwater acoustic pressure field incident upon the water surface.[4–6] Early trends in the detection of underwater sound were accomplished by measuring the amplitude modulation of the reflected light as sinusoidal variations of the surface directed portions of the incident laser beam back to the detector.[6] This technique may prove difficult when the water surface is in motion, such as with a turbulent water surface or while probing the ocean surface. The surface oscillations due to the underwater sound field may be masked by the amplitude modulation of the water surface.

Interferometric techniques using coherent laser radiation have been used to measure the Doppler frequency shift of a laser beam on the water surface vibrating in response to an incident acoustic wave.[7–10] A laser-based, acousto-optic sensing method has been devised that enables the detection of underwater sound from a remote position in the air, without requiring underwater hydrophone equipment. The acousto-optic sensor detects underwater acoustic signals by probing the water surface with a laser Doppler vibrometer (LDV), thereby achieving acoustic to optical signal conversion.

Acoustic pressure waves propagating within the water environment will cause vibrations on an air–water interface, which can be optically detected by observing the amplitude or phase characteristics of the reflected optical beam. Additional sea surface motion occurs as a superposition of wave components including wind driven gravity waves, shorter wavelength capillary waves and nonlinear wave interaction, leading to breaking waves and the production and dissipation of wave energy distributed in amplitude, slope and spatial frequency. However, the small amplitude and low velocity of the acoustically generated surface waves is not completely masked by the larger sea surface wave motion. Commercial LDVs, although not optimized to measure on hydrodynamic surfaces or as sensitive as the highly developed in-water ceramic transducer devices, are able to detect a range of sound pressure levels across a wide frequency band. The foremost issue governing acousto-optic sensor performance on ocean surfaces is not the detection sensitivity but rather the signal

[a]Electronic mail: antonellilt@npt.nuwc.navy.mil

dropout due to laser reflections deflected away from the LDV sensor system by the surface wave motion. The laser transmitter and receiver are typically co-located in interferometric based vibrometer systems; thus these devices require a direct back-reflection of the laser beam from the water surface. Since turbulent conditions prevail in realistic situations, significant intermittence of the received signal is expected as the slope of the wave surface changes relative to the laser beam angle.[11,12]

The accuracy and reliability of acousto-optic sensing and the subsequent signal processing depends upon the sensor to actually receive the reflected optical signal from the water surface. The traditional implementations of basic sonar signal processing functions, such as convolution, Fourier transformation, and interpolation, assume that the continuous time waveform is sampled at a uniform rate. However, since the acousto-optic sensors are prone to signal dropouts, it may not be possible to conventionally process this data. Although signal processing may salvage the sensor data to some extent, the problem still exists due to the intermittent absence of signal information.[13] This issue may become especially critical for acousto-optic sensing to capture time critical and sequential acoustic communication signals or sensor array data for time delay beamforming applications.[1,10,14]

Possible solutions to increase the probability of receiving a reflection from the water surface are to enlarge the laser beam diameter and/or the receiver's field of view to capture a wider range of angled reflections. Regardless, signal dropout may not be eliminated entirely. The approach taken to reduce the signal dropout through system hardware discussed in this paper was to append a laser-based glint tracking system onto the acousto-optic sensor. The tracking system directs a laser beam in a search pattern on the water surface to locate a surface normal such that the reflected laser beam from this surface normal is captured by the system. The tracker then locks onto this surface glint feature and continually provides a feedback control signal to adjust the laser deflection mirrors to maintain position on the surface glint location. The laser beam from the acousto-optic sensor would then be aligned to superimpose itself on the tracking beam, such that it is also continually positioned to interrogate the surface reflecting glint location thereby maintaining acousto-optic reception under hydrodynamic surface conditions.

This paper presents the results from an experiment conducted under laboratory conditions at the Naval Undersea Warfare Center, Acoustic Tank Facility to demonstrate the feasibility of a laser-based system for the aerial detection of underwater sound. The minimum discernible signal level capable of being detected by probing the water surface and the capabilities of the acousto-optic sensor system incorporating the surface normal glint tracker are presented. The results of this experiment supports and demonstrates the feasibility of the proposed aerial sensing of underwater sound under realistic ocean surface conditions.

II. PRINCIPLES OF ACOUSTO-OPTIC SENSING

The acousto-optic sensor provides a method for detecting underwater sound by measuring the velocity of the water surface vibrations generated as the underwater acoustic waves reach the water surface. The laser light from a LDV is focused from the air onto the specularly reflective air–water boundary to remotely measure the water surface velocity. The reflected light, phase modulated by the surface motion, is detected and demodulated to obtain information on the underwater sound field. The laser light reflected from the surface is Doppler frequency shifted by an amount proportional to the instantaneous surface velocity, which contains information directly related to the pressure field in the water incident on the surface.

Consider an acoustic plane wave of amplitude P_0 propagating in the water at an angle θ with respect to the normal to the water–air interface, assumed here to be perfectly flat and still. The LDV detects the normal component of the velocity at the air–water interface. At the air–water interface, the acoustic pressure vanishes and the particle velocity doubles.[15] Since the characteristic impedance (ρc) of water is much greater than that of air, the air–water interface presents an impedance mismatch, pressure release surface that vibrates in response to an impinging acoustic field. As a result, the acoustic pressure in the water, incident on and reflected from the air–water boundary, causes the surface to vibrate at the frequency of the acoustic pressure producing the vibrations. The amplitude of the normal component of the velocity, v_0, created by the acoustic plane wave is therefore given by Eq. (1):

$$v_0 = \frac{2P_0 \cos(\theta)}{\rho c}. \tag{1}$$

The output voltage $V_{out}(t)$ from the interferometer is proportional to the surface velocity by applying the m/s per volt scale factor of the device, C, as indicated in Eq. (2):[16]

$$V_{out}(t) = \frac{v(t)}{C}. \tag{2}$$

The Fourier transform (FT) of the output voltage provides the spectrum level of the detected signal, where N is the number of data samples used to calculate the FT using Eq. (3), as programmed in the Matlab software environment. The velocity spectrum is then converted to a sound pressure spectrum level in dB relative to a μPa at 1 m (dB re μPa) using Eq. (4), where f_s is the sample rate of the data. The last term in Eq. (4) is a bandwidth correction term:

$$\Lambda(\mathrm{f}) = \frac{2C}{N}|FT(V_{out}(t))|, \tag{3}$$

$$\mathrm{SPL_{LDV}} = 20\,\mathrm{Log}\left(\frac{\Lambda(\mathrm{f})\frac{\rho \mathrm{c}}{2}}{10^{-6}}\right) - 10\,\mathrm{Log}(f_s/N). \tag{4}$$

The light reflected from the water surface must be captured by the optical detector on the LDV. Applying geometrical optics, for an LDV positioned normal to the surface, the surface slope angle, ϕ_{slope}, can be calculated using Eq. (5),

$$|\phi_{slope}| < \frac{1}{2}\tan^{-1}\frac{O_a}{2h_d}, \tag{5}$$

where O_a is the detector's optical aperture and h_d is the height of the detector from the measurement surface. For an

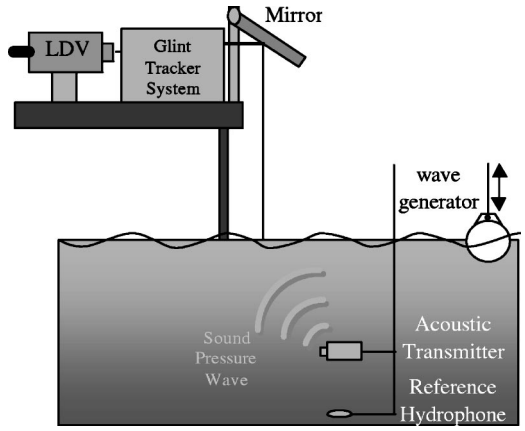


FIG. 1. Test setup used to evaluate the performance of the acousto-optic sensor to detect underwater sound by probing the water surface. The LDV was combined with a laser tracker system under various water surface conditions to improve detection on moving water surfaces.

aperture of 18 mm and a height of 1981 mm, a wave slope angle whose absolute value is greater than 0.13° would deflect the laser beam beyond the receiver aperture and therefore cause signal dropout.

The laser-based, interferometric, acousto-optic sensor presents a potentially wide bandwidth device for remote sonar reception without needing to be submerged. The bandwidth of the acousto-optic sensor extends over several decades (0.1 Hz to 150 MHz) and does not have a resonant operation frequency. Rather, the sensor's bandwidth is limited by the processing electronics such that reducing sensor bandwidth improves detection sensitivity within that band.

III. EXPERIMENTATION

A laboratory test for investigating the feasibility of acousto-optic sensing for remotely detecting underwater sound was conducted under various water surface conditions. Moving water surfaces tend to intermittently deflect the laser beam away from the sensor's photodetectors, resulting in noncontinuous signal acquisition. The objective of the test was to determine the improved acousto-optic sensor acquisition performance with the implementation of a surface reflectance tracking system relative to the detection performance achieved with no tracking on both hydrostatic and hydrodynamic surfaces with varying degrees of surface motion.

Two tests were conducted in an acoustic tank facility demonstrating acousto-optic sensing using the setup shown in Fig. 1, differentiated by the motion of the water surface. Surface velocity data, vibrations in the direction along the axis of the laser beam, were obtained on hydrostatic and hydrodynamic water surfaces using the acousto-optic sensor in combination with the surface reflection glint tracker system. The water surface fluctuations due to the reflected sound pressure field were measured using a single acousto-optic sensor, comprised of a laser Doppler vibrometer, model OFV-353 manufactured by Polytec PI.[16] The laser beam from the acousto-optic sensor was directed through the tracker system mirrors and out onto an angled mirror down onto the water surface. The laser beam was initially directed through the mirror setup straight down onto the water surface to maximize signal detection when the tracker algorithm was turned off. The total path length of the laser beam emitted from the LDV through the tracker mirror system onto the water surface was 78 inches ($h_d = 1981$ mm) due to physical constraints of the test setup area.

The glint tracker system identifies and maintains lock on a water surface reflection feature using a low-power (<0.5 mW), infrared (780 nm wavelength) beam and continually makes analog corrections to tracking mirrors in real time.[17,18] The tracking system directs a laser beam in a search pattern on the water surface to find a surface normal such that the reflected laser beam is captured by the system. The optical tracker monitors the reflection from the tracking beam's current position with a 10-mm beam diameter to optimize tracker performance with the wave slope conditions on the water surface. The tracker algorithm uses information on the detected tracker beam amplitude and shape to provide feedback on tracker mirror position to maintain lock on the surface glint.[19] The laser beam from the acousto-optic sensor is superimposed on the tracking beam such that it is also continually positioned at the surface reflecting glint location.

The underwater acoustic transmitter, AN/BQR-7, was submerged to a depth of 2 m, offset vertically such that the LDV laser beam did not extend onto the transmitter to avoid reading vibrations from the transducer surface rather than the air–water interface. The underwater acoustic signal was recorded using a traditional, ceramic-based reference hydrophone, model H-52 mounted 1 m directly below the acoustic transmitter. The reference hydrophone output voltage, V_{rms} was used to calculate the transmitted SPL at the water surface near the acousto-optic sensor using Eq. (6),

$$\mathrm{SPL}_{\mathrm{Hydrophone}} = 20\log(V_{\mathrm{rms}}) - \mathrm{RVS}_{10\mathrm{K}} + 20\log(d) \quad \mathrm{dB}\ re\ \mu\mathrm{Pa}, \tag{6}$$

where the receive voltage sensitivity of the H-52 hydrophone at 10 kHz ($\mathrm{RVS}_{10\,\mathrm{K}}$) was -178 dB re V/μPa, and $d = 1$ m is the distance between the transmitter and the hydrophone.

Initially, the transmitter was set to emit 2-ms acoustic pulses of a 10-kHz sinusoidal signal, providing for only short duration data captures of both the reference hydrophone and the acousto-optic sensor. This pulsing scheme was used primarily for investigating the hydrostatic surfaces where signal dropout was minimal. The short duration data acquisition scheme was found to be inadequate for analyzing the intermittent signal captures that would last for tens of milliseconds under hydrodynamic surface conditions, which are dominated by signal dropout. The underwater acoustic source transmitted a *continuous* 10-kHz signal into the water with an SPL of approximately 167 dB re μPa for all hydrodynamic tests.

Table I lists the test conditions for the recorded data files for the various surface condition tests with the tracker on and off.

TABLE I. Pulse/CW LDV data acquisition matrix.

Filename	Water surface	Tracker (on/off)	10 kHz transmit signal type	Data segment duration (s)
vSQTNTRQ1	hydrostatic	on	2 ms pulse	2.5 ms
vSQTNNTQ	hydrostatic	off	2 ms pulse	2.5 ms
DAT cut 3	dynamic: buoy—75 rpm	on	cw	26
DAT cut 7	dynamic: buoy—75 rpm	off	cw	6
DAT cut 9	dynamic: buoy—100 rpm	on	cw	45
DAT cut 8	dynamic: buoy—100 rpm	off	cw	45
DAT cut 10	dynamic: buoy—125 rpm	on	cw	45
DAT cut 11	dynamic: buoy—125 rpm	off	cw	45
DAT cut 13	dynamic: buoy—150 rpm	on	cw	45
DAT cut 6	dynamic: buoy—200 rpm	on	cw	30

A. Hydrostatic surface test

Previous tests of the acousto-optic sensor were conducted on hydrostatic surfaces prior to being integrated with the tracker hardware. Water tank test results demonstrated the feasibility of laser detection of sound on an air–water interface. Minimum discernible signal level performance using a commercial LDV resulted in detecting tonal signals having a sound pressure level (SPL) of 120 dB $re\ \mu$Pa at the water surface.[8] These early results conclusively demonstrated acousto-optic sensor detection of single tones in the water. The amplitude of the time series data was observed to be uniform throughout the duration of the signal pulse and the corresponding Fourier transform indicated a peak at the acoustic test frequency above the background noise level.[10] Additionally, acoustic communication signals were also detected error-free by the acousto-optic sensor probing the water surface while the underwater transmitter projected incoherently and coherently modulated signals.[20] Thus, there is a compilation of data to support hydrostatic operation of the acousto-optic sensor.

As tests progressed to measure acoustics on hydrodynamic surfaces, signal dropout due to surface reflections not being captured by the sensor optics became an issue for realizing this sensor concept in practical ocean surface conditions. The glint tracker system was devised to find a surface glint and steer the laser beam of the acousto-optic sensor onto that region to maintain a continuous acousto-optic sensor detection. It was then necessary to obtain new baseline sensor data as the acousto-optic sensor was combined with the glint tracker hardware.

The operation of the combined acousto-optic sensor (LDV) and surface reflection glint tracker system on a hydrostatic, nearly calm surface was tested to obtain baseline signal and noise and performance information. The test setup is shown in Fig. 1 with the exception that the buoy, later used to create surface waves, remained motionless during the hydrostatic tests.

Each data file contains the hydrophone and the LDV detection results from a single, 2-ms, 10-kHz transmit pulse, recorded at a sample rate of 819.2 kHz for 2.5 ms. The time series data for the reference hydrophone and the corresponding Fourier transform is provided in Figs. 2(a) and (b), respectively. The amplitude of the time series data in Fig. 2(a) is nearly uniform throughout the 2-ms pulse duration, with the 10-kHz oscillations clearly visible. The Fourier transform in Fig. 2(b) confirms the reception of the 10-kHz signal with minimal other frequency components present in the underwater environment or in the hydrophone detection system. The peak SPL of the 10-kHz in-water acoustic signal was calculated to be 166.5 dB $re\ \mu$Pa at the water surface from the reference hydrophone output voltage using Eq. (6) with $V_{\mathrm{rms}} = 267$ mV. The signal measured on the reference hydrophone was nearly constant throughout all of the hydrostatic and hydrodynamic tests.

Once the LDV was integrated with the tracker system, data was taken to ensure that the tracker did not adversely affect the acousto-optic sensor operation. Two data files were recorded under hydrostatic conditions with the tracker hardware in place and electrically powered, but with the tracking algorithm turned off, preventing tracker mirror steering. The LDV was set at a sensitivity level of $C = 1$ mm/s per volt. Figures 3(a) and (b) show the results from a single pulse detection on the acousto-optic sensor with the tracker algorithm off, while Figs. 3(c) and (d) show the corresponding

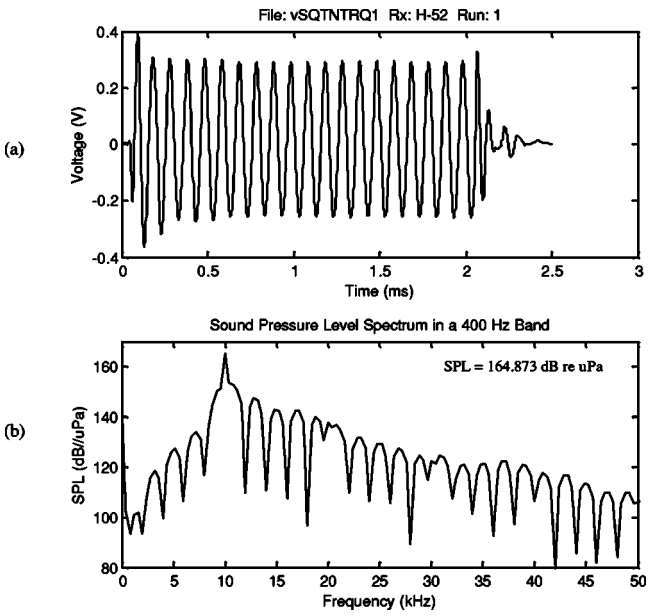


FIG. 2. The (a) time series data and (b) Fourier transform of the underwater acoustic signal recorded on the H-52 reference hydrophone. A 2-ms-long, 10-kHz acoustic pulse has been acquired on the underwater H-52 reference hydrophone.

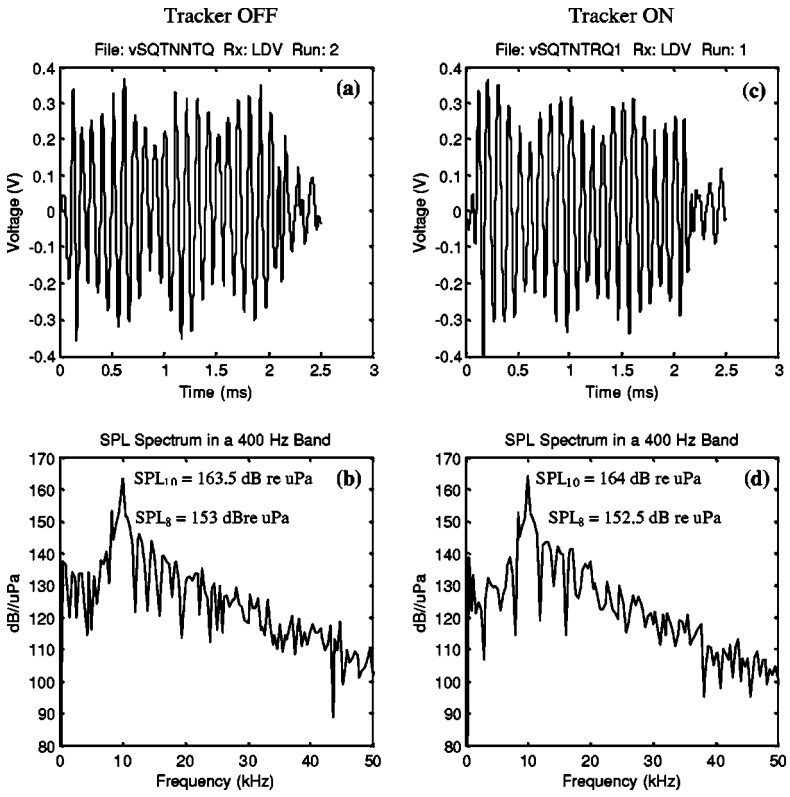


FIG. 3. A 2-ms-long, 10-kHz acoustic pulse acquired on the LDV probing the hydrostatic surface. (a) The time series data of the surface velocity recorded on the LDV acousto-optic sensor with the tracker algorithm off. (b) The Fourier transform spectrum of the signal recorded with the tracker algorithm off. (c) The time series data of the surface velocity recorded on the LDV acousto-optic sensor recorded with the tracker algorithm on. (d) The Fourier transform spectrum of the signal recorded with the tracker algorithm on.

results recorded with the tracker algorithm on. The LDV data recorded with and without the tracker algorithm were nearly identical for this hydrostatic surface condition.

Ideally, the acousto-optic sensor and tracker combination would detect a signal on the hydrostatic surface that imitates the waveform recorded on the underwater reference hydrophone. However, the time series data recorded on the acousto-optic sensor shows an amplitude modulation envelope of approximately 1.6 kHz, which is the difference frequency between the 10- and 8.4-kHz peaks observed in the Fourier spectrum shown in Figs. 3(b) and (d). The 10-kHz test signal produced vibrations on the water surface that were detected by the LDV. The 8.4-kHz signal results from the dither frequency signal of the resonant scanners inside the tracker system. The phase of the LDV laser beam was modulated by both the unwanted mirror motion and the desired water surface vibrations. Other than the presence of the 8.4-kHz tone, the tracker system did not affect the acousto-optic sensor operation or performance relative to the reference hydrophone reading, for the hydrostatic surfaces. As expected, no signal dropouts were observed by the acousto-optic sensor under hydrostatic conditions with the laser from the acousto-optic sensor properly aligned for perpendicular incidence on the water surface.

B. Hydrodynamic surface tests

The second test was designed to verify the operation of the combined LDV/tracker hardware on hydrodynamic surfaces to evaluate tracker performance versus surface condition as compared to acousto-optic sensor performance without the use of the tracker system. The surface waves for the hydrodynamic surface tests were generated using an 18-in.-diam. rubber buoy suspended in the water and connected to a rotation system such that the buoy was moved vertically up and down at a constant rate. The buoy motion created regularly spaced surface waves with a repeatable amplitude and frequency by adjusting the number of revolutions per minute (rpm) of the rotation motor. Acousto-optic sensor tests with motor rotation speeds at 75, 100, 125, 150, 175 and 200 rpm were conducted. A video camera was used to capture the water surface condition at each of the wave settings. A CCD camera embedded in the tracker system was used to monitor both the tracking laser and the red laser associated with the acousto-optic sensor.

Signal recordings as long as 1 min were taken to obtain time evolution recording of the acousto-optic sensor under the hydrodynamic surface conditions with the tracker algorithm on and off. The LDV velocity output and reference hydrophone signals were recorded on a DAT analog tape recorder and later digitized at 50 kHz for data analysis. A high-pass Butterworth filter with a 500-Hz cutoff frequency was applied to the data to attenuate dc and low-frequency signal components of the recorder. The reference hydrophone recorded the constant amplitude, 10-kHz test signal similar to that shown in Fig. 2 for all of the hydrodynamic tests with a peak SPL of approximately 167 dB $re\ \mu$Pa at the water surface, regardless of water motion.

1. Hydrodynamic—75 rpm

The rotating motor attached to the buoy was set for 75 rpm excitation. This generated a water surface that was observed to have a low amplitude (<5 mm, crest to trough), long wavelength (<1.5 m), standing wave pattern, with an oscillation frequency of approximately 1.1 Hz (17 wave oscillations in 15 s). Acousto-optic sensor data was first recorded with the tracking algorithm turned off to obtain base-

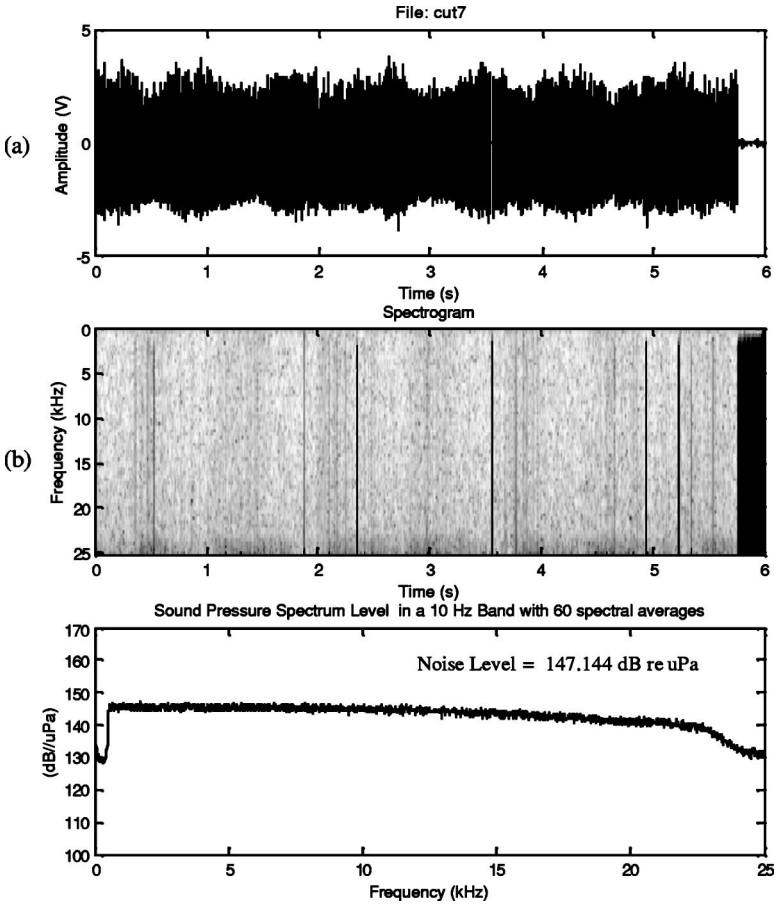


FIG. 4. (a) Acousto-optic sensor time series data recorded with the water excitation at 75 rpm and the tracker system off; (b) the spectrogram; and (c) the Fourier transform converted to units of sound pressure level, calculated in a 1-Hz band with 60 spectral averages.

line acousto-optic sensor performance on this surface condition. The level of signal dropout was compared to the acousto-optic sensor acquisition capabilities with the tracker algorithm activated.

The water surface motion due the 75 rpm excitation was sufficient to intermittently deflect the laser beam away from the acousto-optic sensor's photodetectors ($>0.13°$) and produce signal dropout as observed in Fig. 4, with the tracker algorithm off. The high amplitude signal observed in Fig. 4(a) indicates signal dropout due to the nature of the LDV which clips the output voltage when there is insufficient optical return to the device.

Since the tracker algorithm was shut off, the tracker mirrors were not adjusted to maintain lock on the surface glint. Therefore, signal acquisition where the light reflected from the water surface was redirected toward the sensor's photodetector was minimal. Correspondingly, the spectrogram in Fig. 4(b), which shows the spectral content of the detected signal as a function of time, is nearly filled with the whiter region containing a higher amplitude and small, yet significant, black strip regions that correspond to lower amplitude signal content. The spectrogram for the hydrodynamic surface case with the tracker algorithm off shows almost a continuum of broadband noise. The broadband noise indicates a high degree of signal dropout with little trace of the 10-kHz tone or the tracker's dither frequency present on either the spectrogram or the Fourier transform shown in Fig. 4(c), which displays the overall spectral content of the signal. This does not present a case of poor signal-to-noise ratio rather it presents a condition indicating the absence of optical signal detection. As the received optical signal decreases below the threshold of the vibrometer decoding electronics, the noise level increases. The background noise level (NL) with insufficient optical return was measured at 147 dB $re\ \mu$Pa in a 10-Hz spectral band, calculated using 60 spectral averages.

One 10-ms instance of the test signal was observed in the recording, as the natural surface conditions allowed for specular return into the sensor, occurring between 4.648 and 4.658 s in the trace shown in Fig. 4. An enlarged view of this time mark is shown in Fig. 5. The expanded time scale in Fig. 5(a) allows for the sinusoidal waveform of the test signal to be visible. During this data capture, the 10-kHz signal and the 8.4-kHz tracker dither frequency are observed on both the spectrogram in Fig. 5(b) and the Fourier transform in Fig. 5(c). The signal is observable with a SPL of 165 dB $re\ \mu$Pa, and the noise level was approximately 140 dB $re\ \mu$Pa in a 10-Hz band.

The tracker algorithm was then turned on, and a significant improvement is seen as a greatly reduced signal dropout shown in Fig. 6, for the 75 rpm case. The time series data of the acousto-optic sensor shown in Fig. 6(a) contains some large amplitude signal components associated with signal dropout or a high surface velocity, and some lower amplitude segments that contain the detected sinusoidal acoustic signal. Due to the large amplitude and time scales used in Fig. 6(a), the 10-kHz sinusoidal nature of the signal is not distinguishable throughout the 26-s recording. The spectrogram shown in Fig. 6(b) contains broadband signal components associ-

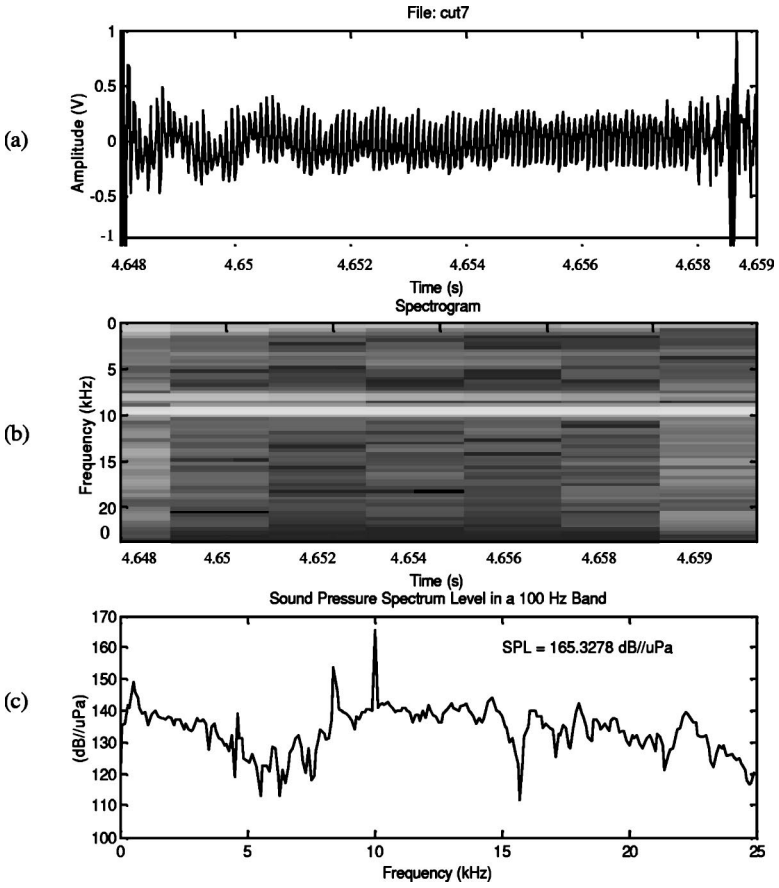


FIG. 5. (a) A subset of the recorded acousto-optic sensor time series data between 4.648 and 4.658 s; (b) the spectrogram; and (c) the Fourier transform calculated using the time series data in (a) recorded with the water excitation at 75 rpm and the tracker system off. This was the one instance of acousto-optic sensor signal detection for this test case occurring between times 4.648 and 4.658 s.

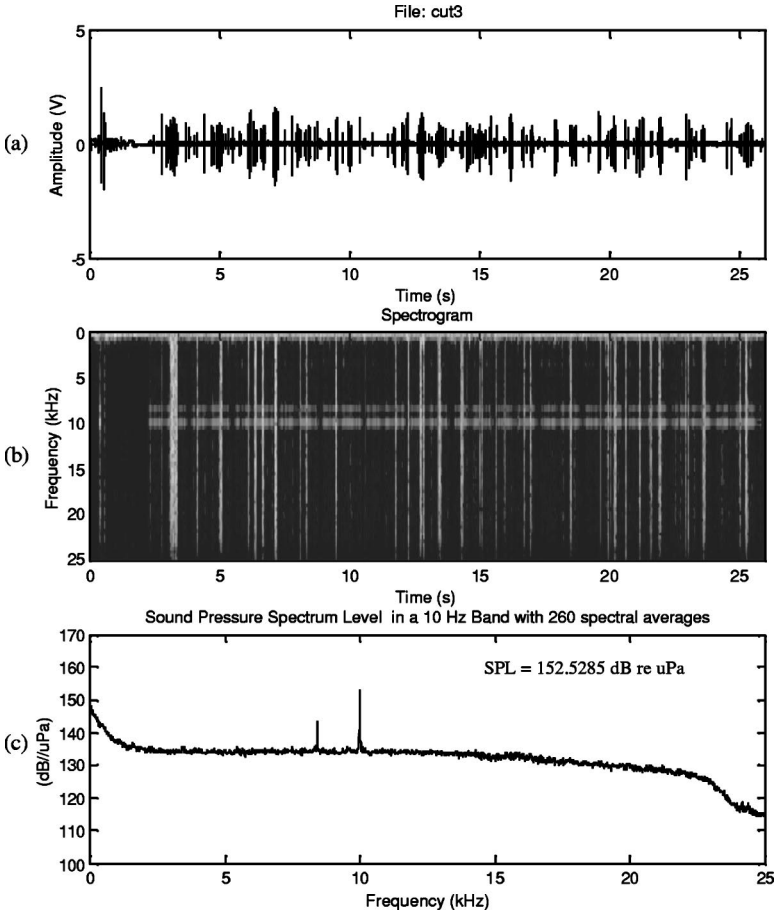


FIG. 6. (a) The recorded acousto-optic sensor time series data; (b) the spectrogram; and (c) the Fourier transform calculated using the time series data in (a) recorded with the water excitation at 75 rpm and the tracker algorithm on.

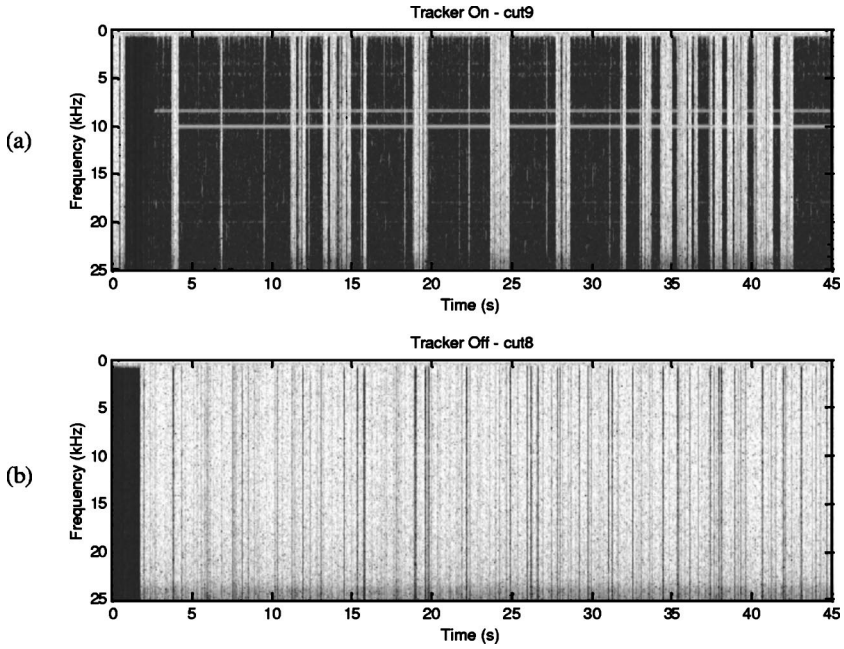


FIG. 7. Spectrogram of a 45-s long acousto-optic sensor data segment taken with a surface buoy excitation of 100 rpm with (a) the tracker algorithm on and (b) with the tracker algorithm off. Cut 9: peak SPL $=150.2$ dB $re\ \mu$Pa at 10 kHz; SPL $=135$ dB $re\ \mu$Pa at 8.4 kHz.

ated with signal dropout. The 10-kHz acoustic tone and the 8.4-kHz tracker dither frequency are observed throughout the spectrogram and are distinct in the Fourier transform shown in Fig. 6(c). The signal level measured was approximately 152.5 dB $re\ \mu$Pa and the noise level was reduced to 134 dB $re\ \mu$Pa in a 10-Hz band, since there was an increased incidence of optical return. The tracker video was reviewed with observed tracker lock established throughout the test duration, providing good optical signal detection from the water surface. Thus, it is believed that the broadband spikes observed in the spectrogram in Fig. 6(b) result from the optical Doppler shift of the laser beam due to steering of the beam on the moving surface.

2. Hydrodynamic—100 rpm

The acousto-optic sensor was tested with the wave surface excitation set at 100 rpm. The surface was observed to have a low amplitude (<5 mm), long wavelength (<1.5 m), standing wave pattern similar to the 75 rpm case but with a more rapid surface wave oscillation frequency of 1.5 Hz (22 wave oscillations in 15 s), due to the increased rotator speed controlling the buoy wave maker. Figures 7(a) and (b) show the spectrograms of the recorded acousto-optic sensor for the 100 rpm surface excitation with the tracker algorithm on and off, respectively.

Acousto-optic sensor data was first recorded with the tracker system electronics off to obtain baseline acousto-optic sensor performance on this surface condition. Surprisingly, the 100 rpm test results, shown in Fig. 7(b), indicate a higher incidence of intermittent signal detection than observed with the 75 rpm case. However, there is still a considerable amount of signal dropout preventing consistent detection for this test case with the tracker off during the 45 s of recorded data analyzed. The majority of the recording contains broadband signal indicative of signal dropout due to errant reflections away from the acousto-optic sensor. However, there are a number of short duration segments containing valid data observed in the spectrogram plot in Fig. 7(b). Data recorded at the time mark between 25.8 and 26.8 s contains several acousto-optic sensor detections of the 10-kHz test signal and the 8.4-kHz dither frequency between areas of broadband signal dropout. These positive signal detections typically last between 10 and 100-ms. There were enough valid data segments to create a peak at the 10-kHz test frequency in the Fourier transform with a SPL of approximately 150 dB $re\ \mu$Pa in a 10 Hz band. The noise floor has increased to 148 dB $re\ \mu$Pa consistent with the reduced incidence of optical return.

With the tracker off, there is a high incidence of signal dropout, as observed in Fig. 7(b), characterized with a broadband spectrum, indicated as the white sections of the graph. However, a dramatic increase in the valid data acquisition occurred with the tracker on. A set of acousto-optic sensor data was then recorded with the tracker algorithm turned on with the same 100 rpm agitated water surface condition. The distinct horizontal lines corresponding to the 10-kHz test signal and the 8.4-kHz tracker mirror dither frequency are observed throughout the spectrogram of the recorded acousto-optic sensor reading with the tracker algorithm as shown in Fig. 7(a). There were also shorter instances of high amplitude, broadband data segments indicating signal dropout throughout the 45-s data set. In fact, the test signal was detected for longer time spans lasting for several seconds as compared to the time that the signal was lost due to dropout. Activation of the glint tracker continuously updated the steering mirrors for directing the sensing laser beam onto the water surface glint. The glint tracker had a major impact on improving the duration of acousto-optic sensor signal detection of the water surface vibrations. The LDV and tracker combination provided a more consistent means for detecting the underwater sound for this surface condition.

3. Hydrodynamic—125 rpm

The excitation of the water surface standing wave pattern was increased to 125 rpm. The water surface was ob-

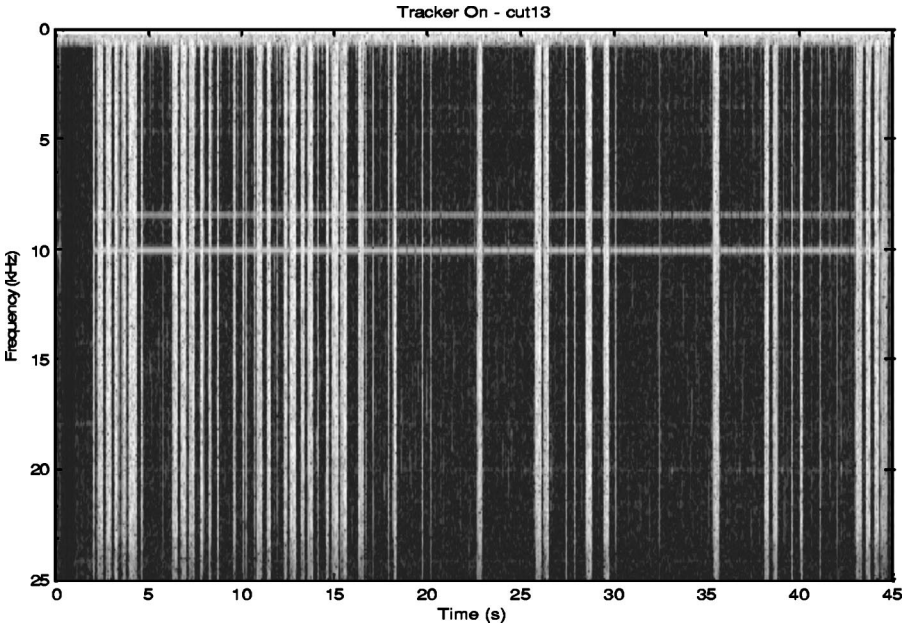


FIG. 8. Spectrogram of a 45-s long acousto-optic sensor data segment taken with a surface buoy excitation of 150 rpm with the tracker algorithm on.

served to have a low amplitude (<8 mm), long wavelength (<1 m), standing wave pattern, with an oscillation frequency of approximately 1.8 Hz (27 wave oscillations in 15 s). The detection capability of the acousto-optic sensor was observed to be greatly enhanced with the tracker algorithm on rather than off, but the spectrogram results of the 45-s data recording were similar to the 100 rpm wave case. The acousto-optic sensor was able to acquire valid data of the underwater acoustic field for a considerable amount of time, with several data segments lasting greater than 6 s, between intervals of broadband noise indicative of signal dropout or high surface Doppler velocity possibly due to the laser steering across the water surface. When it was observed that the tracker algorithm lost lock on the surface glint, the tracker system had to reestablish lock via a search pattern.

In contrast, the acousto-optic sensor experienced nearly continuous signal dropout when the tracker algorithm was shut off. However, there were still short duration occurrences in which the test signal was observed when the natural wave patterns specularly reflected the measurement laser beam back towards the optical detector. In particular, a 20-ms signal and another 15-ms long signal trace was observed. However, considering that the length of the data acquisition was 45 s, sparse appearances of 15- and 20-ms bursts are not long enough for communication signal detection with this sensor scheme on this and even rougher water surfaces.

4. Hydrodynamic—150 rpm

The excitation of the water surface was increased to 150 rpm. The surface waves were oscillating more frequently, at approximately 2.1 Hz (31 wave oscillations in 15 s), with the surface wave condition still characterized as a relatively low amplitude (<10 mm), long wavelength (<0.5 m), standing wave. The water surface roughness was such that while the tracker algorithm was turned off, the acousto-optic sensor did not appear to have any valid detection of the test signal and was completely dominated by signal dropout. Once the tracker algorithm was activated, there was an immediate improvement in acousto-optic sensor detection. However, the duration of valid detection started to degrade with the 150 rpm case as compared to the 125 rpm surface wave excitation case. The spectrogram of the water surface velocity detected by the acousto-optic sensor with the surface excitation at 150 rpm and the tracker algorithm on is shown in Fig. 8. The test signal was observed to last for several seconds. The tracker then lost lock on the surface glint, and broadband, signal dropout conditions persisted until relock was established. This pattern was repeated several times over the 45-s data segment where there was an approximately even distribution between valid signal detection and signal dropout. Regardless, the signal detection performance was greatly improved over test runs where tracking was not used.

5. Hydrodynamic—175 rpm

The excitation of the water surface was increased to 175 rpm. This water surface was still a standing wave pattern with higher frequency oscillations at approximately 2.4 Hz, but still with a relatively low yet increased amplitude (<15 mm) and long wavelength (<0.3 m) compared to previous test surfaces. The tracker system experienced great difficulty acquiring and maintaining lock on water surface reflective glint regions. It seemed that this surface condition approached the limit of tracker glint detection performance. Data taken at this water surface wave setting was invalid and appeared to neither have broadband signal dropout nor contain the test signal or dither frequencies.

6. Hydrodynamic—200 rpm

The most significant water surface wave condition was achieved by setting the rotating motor speed of the buoy to 200 rpm. The surface waves were oscillating more frequently, at approximately 2.7 Hz (40 wave oscillations in 15 s), with the surface wave condition still characterized as a relatively low amplitude (<15 mm), long wavelength (<0.2 m), standing wave. The vertical up–down buoy motion provided *traveling* waves on the water surface in con-

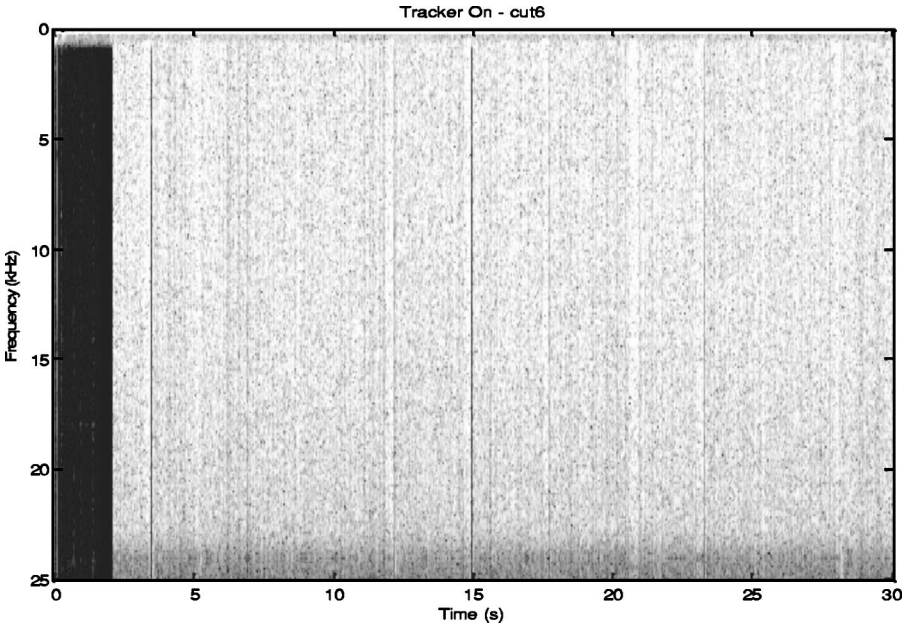


FIG. 9. Spectrogram of a 30-s long acousto-optic sensor data segment taken with a surface buoy excitation of 200 rpm with the tracker algorithm on.

trast to the standing waves that were produced at lower rpm. The tracker had difficulty at this water surface condition. The tracker algorithm works by searching for and locking onto a glint reflection from the water surface. For standing wave patterns, the glint region on the water surface moves around according to the wave oscillation amplitude and frequency. However, with traveling waves, the tracker will follow the glint until it travels beyond the search area prescribed by the tracker optics or the glint is extinguished by the physics of the wave action. The water surface roughness was such that while the tracker algorithm was turned off, the acousto-optic sensor did not appear to have any valid detections of the test signal and was completely dominated by signal dropout. Once the tracker algorithm was activated, there was little improvement in detection performance throughout the 30-s acousto-optic sensor data capture as shown in Fig. 9 compared to detection without using the glint tracker. Short duration (<10 ms) data segments were observed, but not attributed to glint tracking. This test case demonstrated that the tracking algorithm needs improvement to work on traveling wave patterns and, in general, on surfaces characterized by higher amplitude and higher wavenumber wave structures.

IV. CONCLUSION

A laboratory test for investigating the feasibility of acousto-optic sensing for remotely detecting underwater sound by detecting the vibrations on the water surface was conducted under hydrostatic and hydrodynamic conditions. As the water surface roughness increased, the acousto-optic sensor performance degraded as the narrow measurement laser beam was specularly reflected away from the sensor's optical detector, resulting in loss of signal detection. To mitigate the signal dropout condition, the acousto-optic sensor was integrated with a surface glint tracking system. The tracker system, by continually steering the measurement laser beam onto a specular glint region on the water surface, helps to increase the time during which the acousto-optic sensor is able to accurately detect the underwater acoustic signal. This experimentation was conducted to evaluate the performance capabilities of the tracker system by comparing the detection performance of the acousto-optic sensor operation under various water surface conditions.

The results indicate that the acousto-optic sensor is capable of detecting the underwater acoustic signal by probing the hydrostatic surface. Detection on hydrodynamic surfaces posed the greater challenge. Since the acousto-optic sensor was 78 in. from the water surface, even minor surface fluctuations produced signal dropout as the reflection angle of the narrow laser beam is deflected beyond the acceptance cone angle of the LDV-based acousto-optic sensor. The combined acousto-optic sensor and tracker system was tested under several, repeatable water surface conditions, including five tests conducted on standing wave patterns with increasing oscillation frequency and one test with a traveling wave pattern. The water surfaces were in general characterized as low-amplitude, low-wave-number surface waves, where optical glints are sustained on the water surface for possible tracker detection. Under such hydrodynamic surface wave conditions, activation of the glint tracker to continuously update the steering mirrors for directing the sensing laser beam onto the water surface glint had a major impact on the duration of positive acousto-optic sensor signal acquisition of the water surface vibrations and, thus, provides a more consistent means for detection of the underwater sound. The tracker algorithm was capable of finding and locking onto a surface reflection glint for low-amplitude, low-wave-number surfaces. However, the tracker was not robust enough to maintain glint lock under even low-amplitude, traveling wave conditions since the glint feature eventually disappeared, necessitating a new glint search process. Capillary waves were also found to be problematic in addition to the traveling waves since the tracker follows the surface wave glint until the waves travel beyond the laser steering area and tracker relock must be achieved on a different wave structure. Without the tracker, the LDV operated under predominantly dropout conditions. Tracker activation aided in main-

taining signal acquisition. However, the tracker would lose lock intermittently as the surface roughness increased.

The noise spectrum level of the acousto-optic sensor was found to be approximately 134 dB *re* μPa in a 10-Hz band on hydrodynamic surfaces with the tracker on across the measured frequencies between 0 and 25 kHz. However, with the tracker off, there was a reduction in the optical signal return and consequently an increase in the noise floor to 148 dB *re* μPa in a 10-Hz band, due to the velocity decoding electronics.

The acousto-optic sensor and tracker combination also produced an additional modulation tone due to the tracker system that was detected by the acousto-optic sensor. The 8.4-kHz dither frequency signal of the resonant scanning mirrors inside the tracker system interfered with the acousto-optic sensor detection. It should be noted that this effect can be mitigated by proper filtering and tracker design. This can be done by physically separating the LDV steering mirrors from the dither mirrors and increasing the dither frequency beyond the desired detection and filter frequency bands of the acousto-optic sensor. Increasing the dither frequency will also allow for an increase in the tracking bandwidth, which may improve tracking on high wave number surfaces.

Additionally, it was observed that the motion of the steering action following the glint on the hydrodynamic surface produces an additional Doppler shift to the recorded data. This Doppler shift, directly related to the steering of the measurement beam along the water surface, may affect the dynamic range of the velocity measurements from the acousto-optic sensor. Faster laser beam steering may help in reducing signal dropout. However, this can be achieved at the cost of inserting an additional high velocity Doppler term onto the low, micrometer per second velocities encountered from the acoustic excitation of the water surface as detected by the acousto-optic sensor. Mitigating this Doppler shift will also be the focus of future sensor system designs.

It is envisioned that the laser system would eventually be mounted on a surface ship, land-based mobile or tower platform, helicopter, or aircraft. The laser-based sonar system remains on the host platform and is therefore a reusable (nondisposable) resource for continual operation. This method for in-air optical sensing of underwater sound could be used for detecting in-water acoustic communication signals, tracking cetacean mammal migration, and detection and localization of in-water sound sources as a means of remote, aerial sonar. Further research in refining and improving the tracker system is necessary to realize an aerial implementation of the acousto-optic sensor system.

ACKNOWLEDGMENTS

This work was sponsored by the Naval Undersea Warfare Center, Strategic Initiative and Independent Applied Research (IAR) programs, supported by the Office of Naval Research.

[1] L. Antonelli and F. Blackmon, ''Experimental investigation of optical, remote, aerial sonar,'' Proc. Oceans 2002 MTS/IEEE Conference, October (2002).
[2] F. Blackmon, L. Antonelli, L., Estes, and G. Fain, ''Experimental investigation of underwater to in-air communications,'' UDT Europe 2002 Conference Proc., LaSpezia, Italy, June (2002).
[3] Y. Yeh and H. Cummings, ''Localized fluid flow measurements with a He-Ne laser spectrometer,'' Appl. Phys. Lett. **4**(**10**), 176–178 (1964).
[4] C. Zhou, K. Liu, J. He, J. Sun, R. Jiang, G. Chui, R. Li, T. Yun, J. He, G. Shang, and Z. Tian, ''Study on experimental method for underwater acoustic field detection from water surface using laser probe,'' Proceedings SPIE The International Society for Optical Engineering, Vol. 3558, Conference on Automated optical inspection for industry: theory, technology, and applications II, Beijing (1998), pp. 515–521.
[5] Z. Tian, T. Hun, J. He, G. Sang, G. Cui, and R. Li, ''Study for detection of underwater acoustic signals from undulant water surface by laser,'' Undersea Defense Technology Europe Conference Proceedings, Acropolis Convention Center, Nice France, June (1999), pp. 324–325.
[6] M. S. Lee, B. S. Bourgeois, S. T. Hsieh, A. B. Martinez, L. Hsu, and G. D. Hickman, ''A laser sensing scheme for detection of underwater acoustic signals,'' in Conference Proceedings: 1988 IEE SOUTHEASTCON, Knoxville, TN CAT NO. 88CH2571-8 (1998), pp. 253–257.
[7] R. H. Belansky and K. H. Wanser, ''Laser Doppler velocimetry using a bulk optic Michelson interferometer: A student laboratory experiment,'' Am. J. Phys. **61**, 1014–1019 (1993).
[8] L. Antonelli, K. Walsh, and A. Alberg, ''Laser interrogation of the air–water interface for in water sound detection: Initial feasibility tests,'' J. Acoust. Soc. Am. **106**, Pt. 2 (1999).
[9] A. Matthews and L. Arrieta, ''Acoustic optic hybrid (AOH) sensor,'' J. Acoust. Soc. Am. **108**, Pt. 1, 1089–1093 (2000).
[10] L. Antonelli and I. Kirsteins, ''Empirical acouto-optic sonar performance versus water surface conditions,'' Proc. Oceans 2001 MTS/IEEE Conference, November (2001), pp. 1546–1552.
[11] D. M. Phillips, ''Effects of the wavenumber spectrum of a sea surface on laser beam reflection,'' Aust. J. Phys. **32**, 469–489 (1979).
[12] J. Bufton, F. Hoge, and R. Swift, ''Airborne measurements of laser backscatter from the ocean surface,'' Appl. Opt. **22**, 2603–2618 (1983).
[13] L. Lading and R. Edwards, ''Laser velocimeters: lower limits to uncertainty,'' Appl. Opt. **32**, 3855–3866 (1993).
[14] L. Antonelli and I. Kirsteins, ''Bearing estimation error analysis using laser acoustic sensor measurements on a hydrodynamic surface,'' Proc. Oceans 2000 MTS/IEEE Conference, October (2000).
[15] L. E. Kinsler, A. R. Frey, A. B. Coppens, and J. V. Sanders, *Fundamentals of Acoustics*, 3rd Ed. (Wiley, New York, 1982), Chap. 6.
[16] Vibrometer Operator's Manual for Polytec Vibrometer Series 3000, VIB-MAN-3001(303)-9510-e01, Polytec GmbH., Waldbronn, Germany (1996).
[17] D. Ferguson, ''Servo tracking system utilizing phase-sensitive detection of reflectance variations,'' U.S. Patent number 5,767,941, 16 June 1998.
[18] D. Ferguson, ''Servo tracking system utilizing phase-sensitive detection of reflectance variations,'' U.S. Patent number 5,943,115, 24 August 1999.
[19] D. Ferguson, ''Automated ocean surface reflection tracking system,'' PSI Phase 2 Final Report – PSI-2761/TR-1933, NUWC contract No. N66604-03-M-1303, 25 June 2004.
[20] F. Blackmon, L. Antonelli, L. Estes, and G. Fain, ''Experimental investigation of underwater to in-air communications,'' UDT Europe 2002 Conference Proc., LaSpezia, Italy, June (2002).

Analysis and modeling of broadband airgun data influenced by nonlinear internal waves

Scott D. Frank[a)]
Department of Mathematical Sciences, Rensselaer Polytechnic Institute, 110 8th St., Troy, New York 12180

Mohsen Badiey
Ocean Acoustics Laboratory, College of Marine Studies, University of Delaware, Newark, Delaware 19716

James F. Lynch
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

William L. Siegmann
Department of Mathematical Sciences, Rensselaer Polytechnic Institute, 110 8th St., Troy, New York 12180

(Received 1 June 2004; revised 15 September 2004; accepted 15 September 2004)

To investigate acoustic effects of nonlinear internal waves, the two southwest tracks of the SWARM 95 experiment are considered. An airgun source produced broadband acoustic signals while a packet of large nonlinear internal waves passed between the source and two vertical linear arrays. The broadband data and its frequency range (10–180 Hz) distinguish this study from previous work. Models are developed for the internal wave environment, the geoacoustic parameters, and the airgun source signature. Parabolic equation simulations demonstrate that observed variations in intensity and wavelet time–frequency plots can be attributed to nonlinear internal waves. Empirical tests are provided of the internal wave-acoustic resonance condition that is the apparent theoretical mechanism responsible for the variations. Peaks of the effective internal wave spectrum are shown to coincide with differences in dominant acoustic wavenumbers comprising the airgun signal. The robustness of these relationships is investigated by simulations for a variety of geoacoustic and nonlinear internal wave model parameters.
[DOI: 10.1121/1.1819499]

PACS numbers: 43.30.Zk, 43.30.Bp, 43.30.Pc [AIT] Pages: 3404–3422

I. INTRODUCTION

Nonlinear internal waves often occur in coastal areas, including the Yellow Sea[1] and Mediterranean Sea,[2] and off both coasts of the United States.[3,4] They are now known to strongly influence acoustic propagation through them. Variations in acoustic field and travel time caused by nonlinear internal wave fields have been studied in the Barents Sea[5] and in the Gulf of Mexico.[6] Reference 1 proposed that unexpectedly large, frequency-dependent transmission losses observed in the Yellow Sea resulted from strong internal waves. It was suggested that the effect depended on an effective wavenumber of the internal waves and a classical wave–wave interaction phenomena. This leads to a resonance condition relating the internal wave wavelength and differences between the dominant acoustic mode wavenumbers.

Resonant interactions between linear internal waves and acoustic mode amplitudes were previously examined theoretically[7] in a random media setting. Throughout this paper, we focus mainly on acoustic influences from quasideterministic nonlinear internal waves, as opposed to a random internal wave field.[8] Fundamental questions arise concerning the ocean applicability of the resonance mechanism. For example, at moderate acoustic frequencies of 200–600 Hz, shallow-water channels have many propagating modes and many possibilities for wavenumber differences that resonate with the internal wave scales. What happens at lower frequencies? The classical resonance mechanism also requires an acoustic diffraction grating with well-defined environmental wavelengths. How are these wavelengths specified in realistic ocean environments, with trains consisting of variable-amplitude, irregularly spaced nonlinear waves? Significant acoustic mode coupling is now known to occur without the resonance mechanism; numerical studies show mode amplitude changes of 400 Hz signals due to the interaction with a single nonlinear internal wave[9] and an irregular internal wave packet.[10] Is the acoustic-nonlinear internal wave resonance condition viable in actual ocean environments?

Simulations support the resonance condition hypothesis as a possible coupling mechanism. For example, in the Yellow Sea study[1] nonlinear internal wave packets were represented by deterministic sinusoidal waves, so the effective wavenumber was easily specified, and simulations indicated that a resonant interaction could cause the large transmission losses observed. More recently, the role of effective internal wave wavenumbers has been shown for continuous wave (CW) transmissions at 240 Hz,[6] above 450 Hz,[11] and for broadband pulses centered at 224 Hz and above.[12,13] Nonetheless, we are not aware of combinations of a well-sampled ocean environment, research-quality acoustic data, and a comprehensive modeling study that demonstrates the occur-

[a)]Currently at the Department of Mathematics, Marist College, 3399 North Rd., Poughkeepsie, New York 12601.
Electronic mail: scott.frank@marist.edu

rence and effective operation of the resonance mechanism. Deterministic and random internal wave fields were found by simulations to have distinct effects on acoustic propagation[14] in an environment like SWARM. A more thorough review of relevant experimental and theoretical investigations has been prepared.[15] One important development is a formulation[16] in terms of the horizontal wavenumber spectrum of the nonlinear internal wave packet, as opposed to a single wavenumber. Resonance occurs when internal-wave wavenumber spectrum peaks correspond to differences between acoustic mode wavenumbers.

In the summer of 1995, the Shallow Water Acoustics in Random Media (SWARM) experiment was conducted off the coast of New Jersey[17] to obtain high quality environmental and acoustic data and study the influences of nonlinear internal waves on acoustic propagation. The main acoustic track in SWARM explored an across shelf geometry, which was perpendicular to the internal wave wavefronts. Broadband signals were also propagated in an along shelf geometry on two southwest tracks during the passage of at least one strong nonlinear internal wave packet. The along and across shelf geometry of these tracks allows the investigation of the azimuthal dependence of acoustic variations caused by nonlinear internal waves.[18]

In this paper we focus on additional observations and detailed modeling for the southwest tracks of the SWARM experiment. The airgun source used along this track produced signals with energy primarily in a low-frequency (10–180 Hz) band. This band contrasts with previous studies that focus on CW signals above 200 Hz and with other broadband simulations[12,13,18] at higher frequencies. Observed time variations in acoustic measurements, the presence of high quality environmental data, and the passage of a strong and coherent train of nonlinear internal waves provide an exceptional opportunity for testing the applicability of the resonance mechanism. However, low-frequency signals contain relatively few propagating modes and thus provide few opportunities for satisfying the resonance condition. Our study also uses time–frequency analysis to reveal frequency-dependent mode amplitude variations resulting from the nonlinear internal waves.

In this paper we also address modeling issues in the context of SWARM that should be relevant to other experiments. It is important for future simulations to develop an environmental model of that site and of the nonlinear internal waves that reproduce observed variations in data. Although questions arise about using weakly nonlinear Korteweg-deVries equation solutions to model nonlinear internal waves,[19] thermal records from SWARM show that these solutions are adequate for modeling spectral and temporal characteristics of the nonlinear wave packet observed. Moreover, the primary acoustic effects seen are reproduced without relying on contributions from a background field of linear internal waves. The level of uncertainty in seabed geoacoustic parameters, especially the upper sediment layers, is detailed for analyzing internal wave effects. An unexpected aspect of the modeling effort was the need to formulate an effective source representation, emphasizing the experimental information necessary to perform comparisons between data and simulations.

The paper is organized as follows. In Sec. II we provide details of the experimental tracks, including the airgun source and environmental data. In Sec. III we explain environmental observations, notably the occurrence of the nonlinear internal waves and their observed effects on acoustic signals. In Sec. IV we describe the full model of the southwest SWARM tracks obtained from available environmental, source, and acoustic data. In Sec. V we describe parabolic equation results that reproduce observed variations in pulse-averaged intensity and time–frequency behavior. The internal wave-acoustic resonance condition is discussed, and this track provides an effective empirical test of this condition. In Sec. VI, we investigate the sensitivity of pulse-averaged intensity variations to internal wave and geoacoustic model parameters. In Sec. VII we summarize our major findings.

II. EXPERIMENT DESCRIPTION

The SWARM experiment was performed off the New Jersey coast in the Mid-Atlantic Bight continental shelf region (Fig. 1, inset). The primary goal of this multi-institutional experiment was to study effects of nonlinear internal waves on acoustic signals, and a considerable amount of acoustic and environmental data was collected. A full technical overview of SWARM is in Ref. 17. The correlation of acoustic and internal wave data for the southwest tracks can be found in Ref. 18.

In the present work we focus on the southwest experimental tracks shown in Fig. 1. The two vertical linear arrays (VLAs), the primary acoustic receivers for SWARM, are indicated by black circles. The northerly one was a Woods Hole Oceanographic Institution (WHOI) telemetered array deployed in 70.5 m of water, consisting of 16 hydrophones spaced approximately 3.5 m apart, with top and bottom phones at depths of 14.9 and 67.5 m. The southerly one, a Naval Research Laboratory (NRL) telemetered array, was located approximately 9 km to the southeast of the WHOI VLA in 88 m of water and consisted of 32 elements spaced 2 m apart with top and bottom phones at 21 and 85 m.

Two thermistor strings (with five thermistors each) at the receiver sites provided ''end point'' observations of the nonlinear internal waves. The string positions are indicated by white squares in Fig. 1. The first, attached to the WHOI VLA, measured the temperature every 30 s at depths of 12.5, 22.5, 30.5, 50.5, and 60.5 m. The second, very close to the NRL VLA, took measurements every 60 s at depths of 19.8, 39.6, 51.5, 63.5, and 75.5 m. In addition, the R/V Oceanus was near the WHOI array and provided radar images of passing nonlinear internal wave packets.[17]

The R/V Cape Hatteras occupied the SWARM site for six days, 31 July to 5 August, 1995, and was positioned southwest of the two VLAs at the position denoted by the black square in Fig. 1. CTD data collected there also showed the passage of nonlinear internal waves and provided sound-speed profile data. The Cape Hatteras deployed two sources during this period; a J-15 for transmitting LFM sweeps and a 20 in.^{3} Bolt airgun; data from the latter source will be examined here. The airgun had a pulse signature that was found to

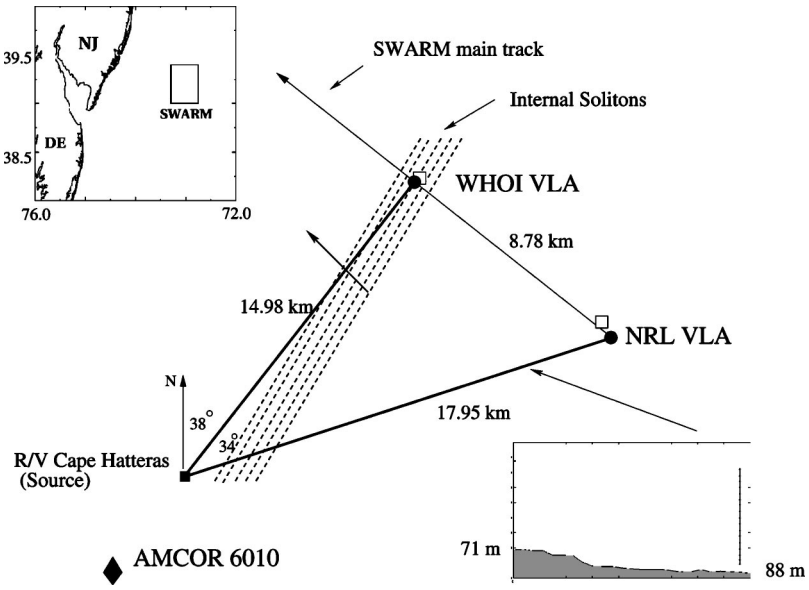


FIG. 1. Geometry of the southwest portion of SWARM experiment. Signals from an airgun at R/V Cape Hatteras (black square) received by WHOI and NRL VLAs (black circles). Thermistor strings (white squares) associated with each array. The AMCOR 6010 site (black diamond) located several km southwest of the source. Bathymetry of the NRL waveguide declines slightly toward VLA. Two acoustic tracks (dark lines) intersected by an internal soliton packet (dashed lines) observed on 4 August 1995.

be quite repeatable, based on measurements at a single monitor hydrophone[17,18] hung below the source. The airgun produced broadband pulses that had consistently repeatable spectra in the 10–180 Hz band (see Fig. 2 of Ref. 18). These spectra show significant peak frequency components centered near 32 Hz and several harmonics. The dependable source signature is essential for our subsequent analysis of waveguide-dependent broadband variations.

As shown in Fig. 1 the airgun signals propagated along two different tracks (heavy lines). The track to the WHOI array was about 15 km long, 70 m deep, and displayed very flat bathymetry. The NRL track was about 18 km long; the depth increased from approximately 70 m at the source to about 88 m at the VLA. GPS location data estimated the orientation of the WHOI and NRL tracks to be 34° and 74° from true north, respectively. The dashed lines in Fig. 1 represent a packet of strong nonlinear internal waves that were observed passing through the experimental region on August 4. The actual packet does not have the exactly linear wavefronts shown schematically here.

The southwest portion of the SWARM experiment is also near the AMCOR 6010 borehole site. Although this site was several km southwest of the Cape Hatteras, it still provides an initial approximation to the geoacoustic properties of the SWARM site. The core location is indicated by the black diamond in Fig. 1.

III. DATA ANALYSIS

A. Internal waves

Several episodes of nonlinear internal wave activity occurred while the R/V Cape Hatteras was on site.[17,18] Figure 2(a) shows records from the top three WHOI thermistors for August 4. Low-frequency tidal components were removed using a 5 min sliding highpass filter. The passages of several different packets of nonlinear internal waves are visible throughout the day. Figure 2(b) shows a specific group of internal waves that first crossed the WHOI thermistor string at approximately 1800 h GMT on August 4. By subtracting the DC component from these signals and using the thermistor sampling rate of 1/30 Hz, the spectrum of this temperature variability can be calculated. The normalized spectrum in Fig. 2(c) has the strongest peaks at 0.0013 and 0.0024 Hz, which correspond to periods of 12.8 and 6.9 min. This is consistent with the intervals between nonlinear internal waves in the time series of Fig. 2(b).

At 19:01 GMT, while this packet of internal waves was traversing both acoustic tracks, the airgun on the Cape Hat-

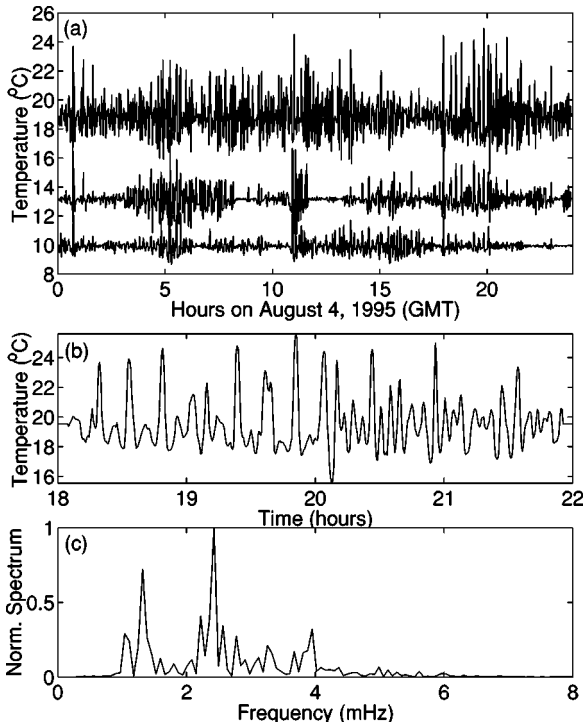


FIG. 2. (a) Records from top three thermistors at WHOI VLA on 4 August 1995. Internal soliton packets visible at all depths. (b) Smoothed record from the top (12 m) WHOI thermistor between 1800 and 2200 GMT. This packet overlapped acoustic tracks while the airgun fired. (c) The normalized frequency spectrum of the smoothed record in (b) with the DC component removed. Dominant frequency components correspond to periods of 12.82 and 6.94 min.

teras began firing once per minute. One valuable feature of this data is that the internal wave propagation direction intersects the two acoustic propagation directions at very different angles, as shown in Fig. 1. The bearing of this packet was estimated as 59° west of north from radar on the R/V Oceanus, which was near the WHOI VLA. Assuming the internal wave crests are linear, the angle between internal wave and acoustic propagation directions is about 49° along the NRL track and close to 90° along the other track. At an angle of 90° the nonlinear internal wave fronts would be parallel to the acoustic track.

B. Acoustics

1. Wavelet transform

The broadband character of the airgun signals allows analyzing changes in their time–frequency behavior after an interaction with nonlinear internal waves. To do this we have used a wavelet transform.[20] Wavelet transforms have been used, for example, in shallow water for sediment parameter inversion[21,22] but not for investigating effects of nonlinear internal waves. These transforms help show the frequency-dependent variations in arrival times, the strengths of signal components, and the interactions between acoustic modes. A summary of the time–frequency transforms we employed is in Ref. 20.

The time–frequency transform T_f of a signal $f(t)$ is given by the integral

$$T_f(u,\xi)=\int_{-\infty}^{\infty} f(t)\,\phi_{u,\xi}(t)\,dt, \tag{1}$$

where $\phi_{u,\xi}(t)$ is a basis function depending on two parameters, u and ξ. One familiar example is the Short Time Fourier Transform (STFT), for which the basis functions are sinusoids. For narrow band signals the STFT is effective, but if the signal has many different scales, the STFT can suffer from aliasing or a lack of resolution. For example, a finite time series record has only a few degrees of freedom at lower frequencies, but many at higher frequencies, which prompts the need for scaling. A wavelet basis introduces ξ as a variable scale. We use a Gabor (or Morlet) wavelet that is a Gaussian pulse translated by u, scaled by ξ, and modulated by chosen frequency η:

$$\phi_{u,\xi}(t;\eta)=\frac{1}{\sqrt{\xi}}e^{-[(t-u)/\xi]^2}e^{i\eta[(t-u)/\xi]}. \tag{2}$$

When a wavelet basis is used in Eq. (1), the graph of T_f is known as a scalogram. Wavelet analysis is not immune to aliasing and interference effects, which depend on the shape of the basis function. However, focusing on scale instead of frequency allows better resolution across a large frequency band, which is desirable since the airgun signal has a reliable frequency content between 10 and 180 Hz.

Figure 3 shows two scalograms of signals from phone 5 (29 m depth) of the NRL VLA. Phone 5 was chosen because its position in the water column is close to a peak in the third acoustic normal mode at $f=64$ Hz, and allows a clear observation of first, second, and third mode energy. Each signal is shown in three panels. The top panel is the time domain representation $f(t)$, the right panel is the Fourier spectrum of $f(t)$ over the full interval, and the large panel is the scalogram. The graphs have all been normalized by their maximum values, and the spectrum and scalogram are limited to the repeatable spectrum below 180 Hz. Figure 3(a) shows the signal corresponding to the 1904 GMT airgun shot. Its Fourier transform shows energy peaks near 32 Hz and its harmonics, with most energy concentrated at 32, 64, and 95 Hz. The scalogram shows the expected modal group velocity structure,[23] with two modes present at 32 Hz, three at 64 Hz, and at least four at 95 Hz. Mode arrivals at 32, 64, and 95 Hz are all well separated. Mode interaction causes some energy to travel at different group velocities, which causes energy to appear in the scalogram between otherwise well-defined modes. Some evidence of mode interaction can be seen between the second and third modes of the 120 Hz band. Figure 3(b), showing the 1910 GMT shot 6 min later, has a Fourier spectrum similar to that from the 1904 shot but with considerably more energy in higher frequencies around 120 and 150 Hz. However, the scalogram reveals even more striking differences between these signals. The modal group velocity curves are still visible at 1910, but a larger concentration of energy arrives early in higher frequencies. The modal interaction occurring between the first and second modes in the 95 Hz band is indicated by a strong blending between the corresponding peaks. The mode interaction is also visible above 95 Hz. Shot 1904 has stronger lower frequencies, while shot 1910 has significant high-frequency energy arriving at the start of the pulse. These differences can also be seen in the time series, but not so easily as on the scalograms.

The differences between Fourier transforms and scalograms of these signals are typical in the data from other hydrophones for the entire hour. The variations occur over 12 to 14 minute periods, and indicate the correlation between the nonlinear internal waves and variations in the acoustic signals.

2. Pulse-averaged intensity

Variations in pulse-averaged intensity were observed in deconvolved and filtered airgun data.[18] Since the source was extremely repeatable over 10–180 Hz, we expand the analysis of those variations to the entire reliable band. The decibel time-averaged intensity in W/m^2 of an acoustic pulse $p(t)$ is

$$I_T=10\log_{10}\left(\frac{1}{T}\int_0^T\frac{|p(t)|^2}{\rho c}dt\right)\ \text{dB re: 1 }\mu\text{Pa}, \tag{3}$$

where T is the interval of integration, ρ is water density, and c is an average sound speed. The interval T was chosen to contain a complete pulse and is typically 1.6 s for NRL VLA data and 1.2 s for WHOI data.

The quantity I_T is calculated for 60 pulses received during the hour beginning on 1901 GMT on August 4. Figure 4(a) shows the results from the WHOI VLA and Fig. 4(b) shows results from the NRL VLA. Multiple variations occur at several dB amplitudes in both datasets, and the quasiperiods of the variations are from 12 to 14 min. The presence of oscillations with similar periods in both datasets suggests

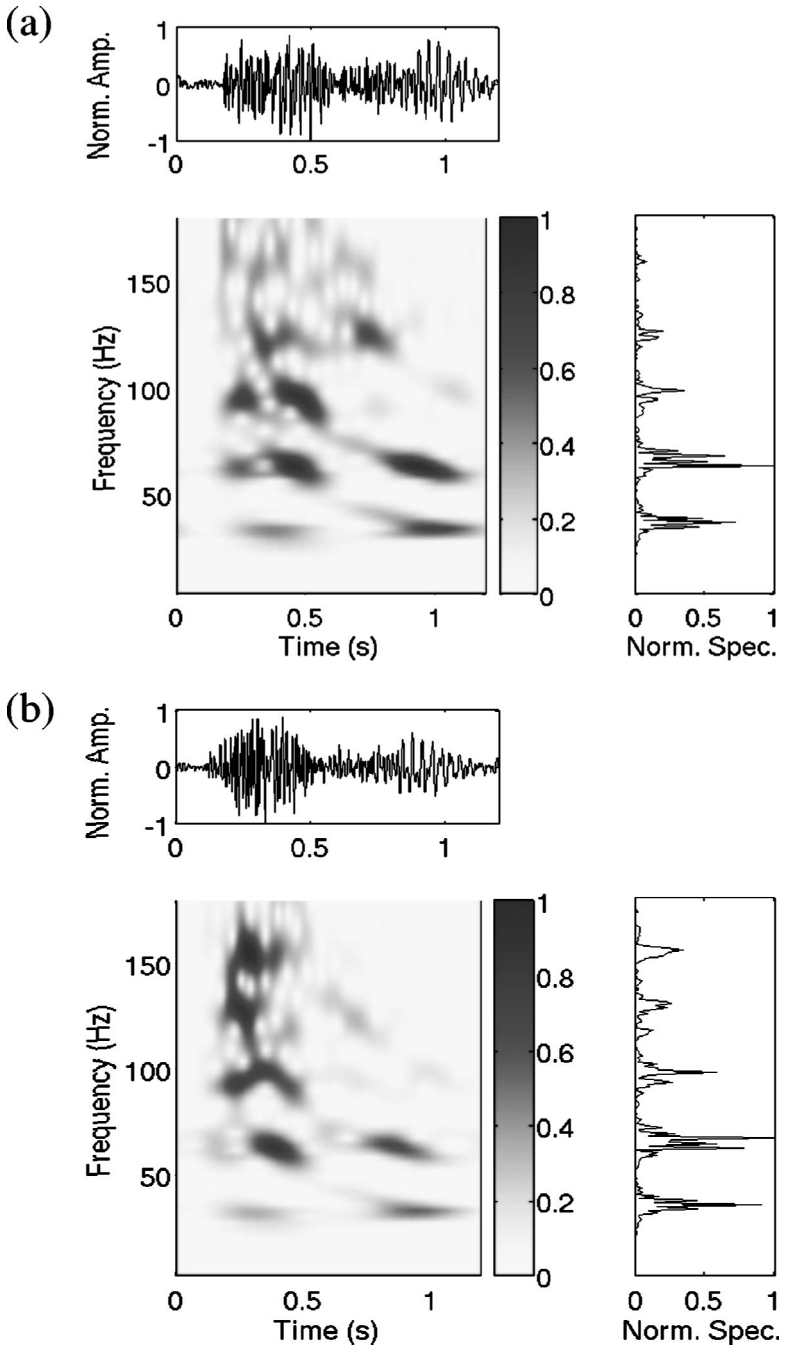


FIG. 3. A time–frequency analysis of NRL data at 29 m depth. (a) 1904 GMT shot, received time domain signal (top panel). The Fourier transform (side panel) shows dominant energy peaks at 32, 64, and 95 Hz. The scalogram (color panel) represents a Gabor wavelet analysis of the signal in the top panel. Group velocity curves are visible; two, three, and five modes revealed near 32, 64, and 95 Hz. (b) 1910 GMT shot (six minutes later). The spectrum shows more high-frequency energy than in (a). The scalogram shows substantial energy arriving early in frequency bands higher than 95 Hz. Evidence of the mode interaction visible, especially at 95 Hz.

these effects arise from the internal waves that traverse the tracks. The occurrence of these variations is largely independent of receiver depth, although the amplitudes and patterns do show some weak depth dependence. Consequently, a reliable measure of the overall variations can be obtained by examining the depth-averaged I_T, shown by thick lines in Fig. 4. The quasiperiod of the variations is preserved by the depth average, with amplitudes at the WHOI and NRL VLAs of 5.9 and 3.1 dB, respectively. The difference between these two values suggests that the internal waves have a larger effect on acoustic signals propagating nearly parallel to the internal wave fronts.

IV. MODELING

Several features of the southwest SWARM tracks are important in order to compare simulations with observations. The inherent range dependence of the water column due to the internal waves motivates the use of parabolic equation (PE) for calculating the waveguide transfer function. The code RAMGEO[24] was modified to output the received complex acoustic pressure at selected hydrophone depths, and these values are used to construct the transfer function between 10 and 180 Hz. Broadband pulses are then obtained by standard synthesis methods.[25] For the remainder of this study, we will focus on modeling the variations observed at the NRL VLA.

A. Internal waves

Several CTD casts are available from the R/V Cape Hatteras. Sound-speed profiles that were obtained from these are shown in Fig. 2 of Ref. 18. These records show that large internal waves passed under the ship and thus traveled past the source while it was firing. Enough samples were taken to

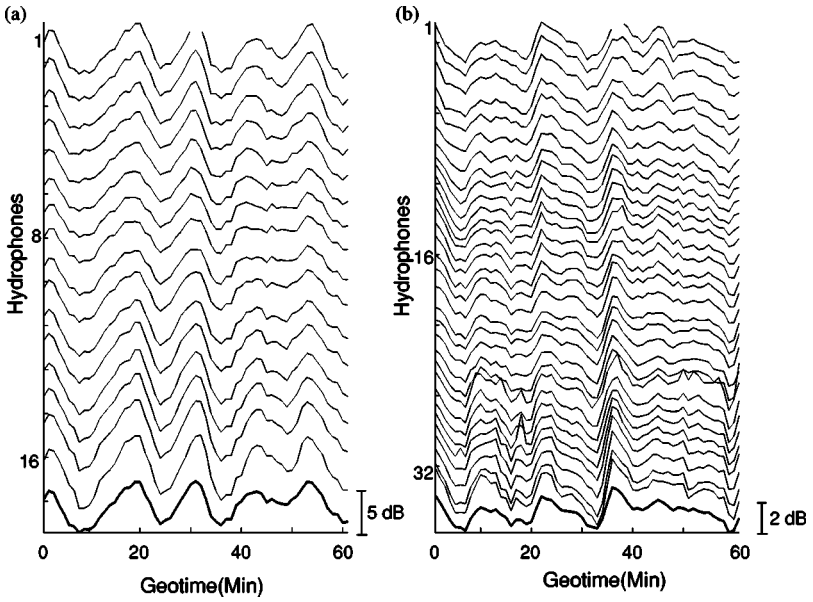


FIG. 4. Intensity I_T for each hydrophone at (a) WHOI and (b) NRL VLA versus geotime starting at 1901 GMT. The quasiperiod of variations at each phone is consistent with Fig. 2. Depth-averaged I_T (dark curves) shown at the bottom. Note different scales in the range of intensity variations for (a) and (b).

obtain a good representation of the average sound-speed profile at the source. This profile is used in the PE simulations. This average profile surely undergoes a modification along the propagation track, but in the absence of data this effect is not incorporated into the environmental model. Only range-dependent variations due to the nonlinear internal waves are included as changes to the average profile.

An early effort to model nonlinear internal waves in shallow water for acoustics work used sinusoidal wave packets.[1] Since an analysis of weak nonlinear oscillations in a two-layer fluid leads to the Korteweg-deVries (KdV) equation, many authors have used periodic KdV solutions for their models, including the cnoidal[6] and dnoidal[18,26] Jacobian elliptical integral solutions. These latter models are more physically correct but not so useful for this study, for two reasons. First, the temporal evolution of an internal wave packet is very slow compared to the passage of an acoustic signal, and second, the full nonlinear internal wave solution is unable to easily incorporate the randomly uneven spacings between wave peaks.[15] Consequently, a simple sum of nonperiodic KdV equation solutions is used to represent the horizontal variations due to nonlinear internal waves. Because the sampling rate of the WHOI thermistors is the only experimentally specified parameter, we first regard the packet as a time series:

$$\eta(t,z)=\Phi_1(z)\sum_{n=0}^{5} A_n \operatorname{sech}^2\left[\frac{2\pi(t-t_n)}{\tau_n}\right]. \tag{4}$$

Equation (4) represents a train of six KdV solitons where t represents time, and t_n, A_n, and τ_n represent the peak position (measured from $t=0$), amplitude, and width (in seconds) of each soliton. The first internal wave mode is represented by $\Phi_1(z)$ and is approximated by a continuous piecewise-linear function.[16] This model is similar to that used in Ref. 10. Note that τ_n is not a proper wavelength since the sech^2 solutions are not periodic.

We found in our modeling that it is necessary to obtain a reasonable spectral match between data and the modeled internal waves. This has important consequences for the internal wave-acoustic resonance condition in Sec. V, which depends on peaks of the wavenumber spectrum.[7,16] Due to the remarkably uniform peak spacing and height of the thermal variations in Fig. 2, our model employs evenly spaced internal waves. Figure 5 compares the internal wave packet frequency spectrum to that of the model from Eq. (4) using evenly spaced waves with $\tau_n=\tau=11.9$ samples, and $t_n=2.3n\tau$. Standard methods provide the frequency axis from the thermistor sampling rate. The principal features of the two spectra are quite similar. The largest component appears at $f_{iw}\approx 0.0024$ Hz, and two smaller components appear at 0.0013 and 0.0040 Hz. The feature match between these two spectra is evidence that our choices of internal wave model and parameters are appropriate for modeling, although these parameters by no means provide a unique representation of the packet. Differences in higher-frequency peaks are caused by slightly uneven spacing between the waves, fluctuations in the water column thermal properties such as the diffuse linear internal wave field, or other effects.

Because of the consistency between the two spectra in Fig. 5, we expect the time series of the nonlinear internal wave packet to be modeled reasonably well. The thermistor

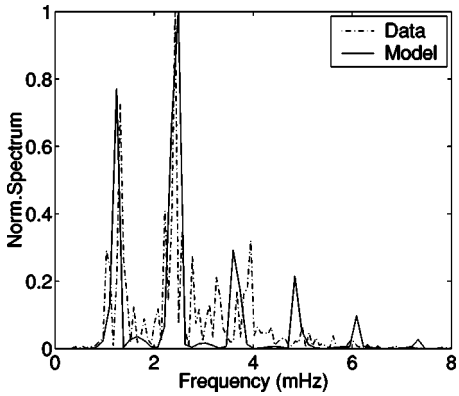


FIG. 5. The dashed curve is a normalized spectrum of an internal wave packet at the top (12 m) WHOI thermistor reproduced from Fig. 2. The solid curve is the normalized spectrum of the internal wave model in (4). Note the location and amplitude agreement of the dominant frequency components.

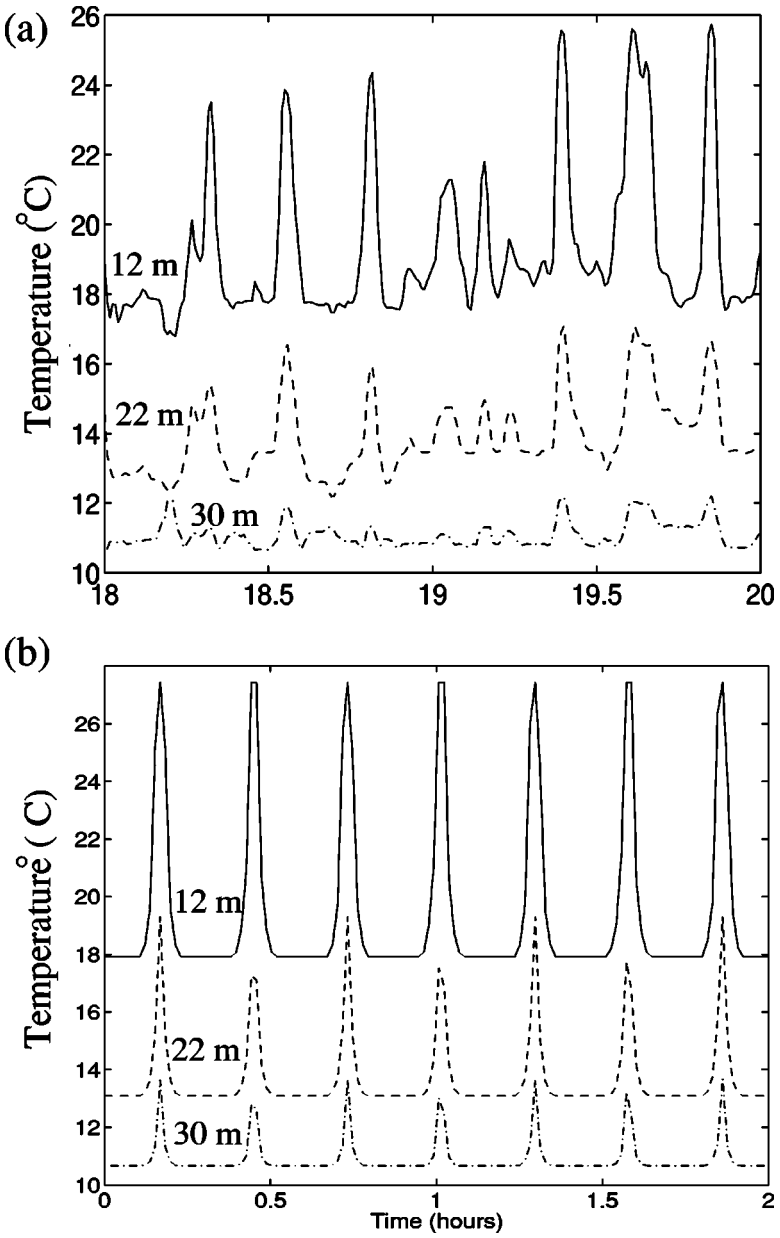


FIG. 6. (a) WHOI thermistor records between 1800 and 2000 GMT on 4 August 1995. The first internal wave mode suppresses the temperature change for increasing depth. Note the regular spacing between oscillations in these records. (b) Simulated thermistor records using the internal wave model in (4) with parameters that generated the spectrum in Fig. 5.

sampling rate permits a comparison of the model time series to data. Figure 6(a) shows three thermistor records of the type in Fig. 2(b) that focus on the nonlinear internal wave packet passing the WHOI thermistor string between 1800 and 2000 GMT. As the internal waves pass any location, they force warm water into the lower colder layer. The amount of temperature change decreases with depth, as seen by comparing the solid curve (12 m depth) in Fig. 6(a) to the dashed (22 m) and dashed-dotted (30 m) curves. From the model spectrum of the internal wave packet, the associated frequency domain representation is inverse Fourier transformed to obtain the time series, shown in Fig. 6(b). The first internal gravity wave mode used in the model expresses thermal variations as depth increases as in the data. Uniform peak spacing and height is reproduced in the model, as expected, although different parameters could be selected for a more detailed pattern match. Because of the relatively high symmetry in the data, we chose parameters for a symmetric model and thus a classical diffraction grating. Since diffraction gratings cause incident waves to break into bands of maxima and minima behind the grating,[27] this phenomenon provides an interpretation of the variations observed in our particular data. We emphasize that our conclusions do not show that either uniform spacing between waves or identical amplitudes is essential to model the observed spectrum. Indeed it has been shown that even a single wave can cause acoustic mode coupling,[9] although more than one wave is evidently necessary for a reasonable spectral match to this data.

The acoustic effects of linear internal wave field components of the internal wave field[28] are often distinct from the effects caused by nonlinear internal waves.[14,29] Calculations were performed to confirm that the environment under consideration is consistent with that conclusion, but these results are not shown here. Satellite images of coastal nonlinear internal waves[4,17] indicate that assuming essentially linear wavefronts in this region is reasonable. However, this as-

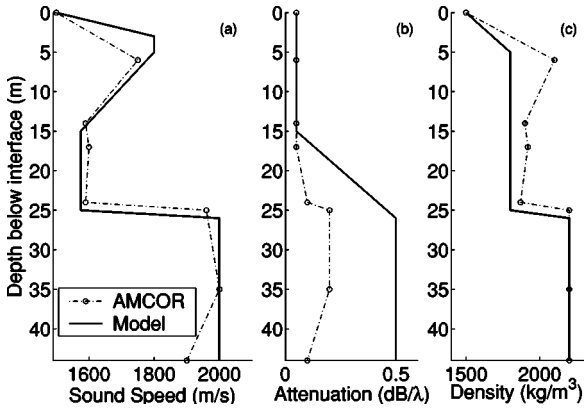


FIG. 7. AMCOR 6010 data (dashed curves with circles) and model geoacoustic profiles (solid curves) for (a) sound speed (m/s), (b) attenuation (dB/λ), and (c) density (kg/m^3). Principal characteristics including upper sediment layer, shallow reflector, low sound speed waveguide, and strong reflector correspond. The large attenuation values in the model account for elastic effects.

sumption introduces uncertainty into any environmental model of the SWARM region.

B. Geoacoustics

The NRL track bathymetry is slightly sloping from 71 to 86 m (inset, Fig. 1) according to depth values that were obtained from the National Geophysical Data Center.[30] We assume that sediment layer interfaces in the bottom follow the bathymetry, so that new layers are not revealed as the water depth increases. A PE implementation that handles bottom layers in this way, RAMGEO,[24] was modified to output complex pressure at given hydrophone depths and then used to calculate the waveguide transfer function.

Data from the AMCOR 6010 site (see Fig. 1) has significant features that we preserved in our bottom model. Core profiles of compressional sound speed, attenuation, and density are shown by dashed lines in Fig. 7, where small circles represent the actual data points.[17] One feature is a shallow reflector within 10 m of the water–sediment interface. A deep layer of dense fast material appears about 25 m from the water–sediment interface, with a channel of softer sediment between these two reflectors. The presence of the deep strong reflector has been well documented for this region.[31] We considered a variety of candidate bottom profiles that preserved these features and were numerically ''close'' to the data profiles. To evaluate a candidate bottom model, PE simulations were performed and the received pulses filtered using an order 10 noncausal Butterworth filter.[32] This filter was applied with 10 Hz bandwidth and center frequencies chosen at the dominant airgun signal frequencies (source details are shown in Sec. IV C). The goal was reasonable visual agreement between relative modal amplitudes and arrival times in data and simulations. This procedure for selecting geoacoustic profiles is similar to that used elsewhere,[33,34] and no strong claim can be made about the faithfulness of the result to the actual bottom profile (for example, the geoacoustic range dependence is not modeled) or about its uniqueness. Bottom profile inversions at the SWARM site are being performed by others.[35]

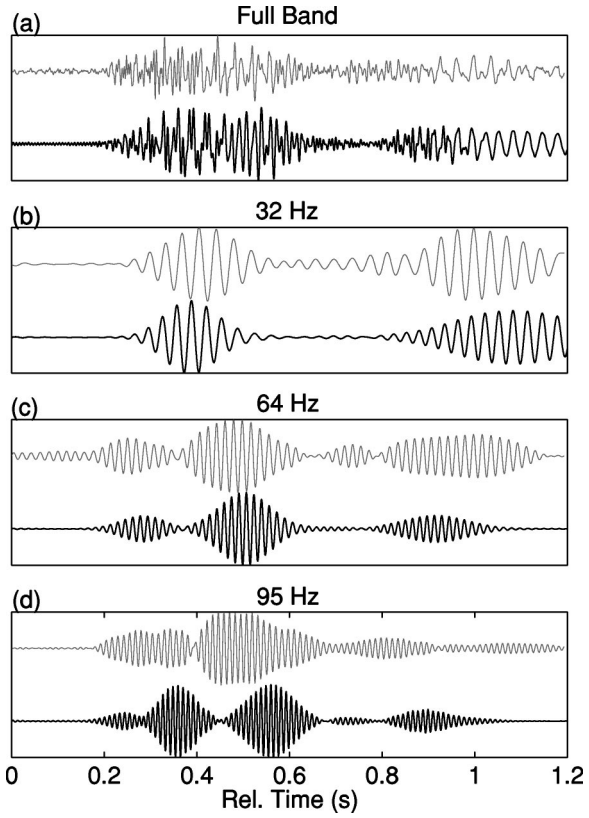


FIG. 8. Comparisons showing the similarity between modal characteristics (occurrence, dispersion, and strength) in pulse data (light upper curves) and simulations (heavy lower curves). (a) Full frequency band. (b) Light (and heavy) curve results from applying a 10 Hz bandwidth Butterworth filter centered at 32 Hz to a light (and heavy) broadband signal in (a). (c) The same, but with the filter centered at 64 Hz. (d) The same, but with the filter centered at 95 Hz.

The selected bottom profiles are shown by solid lines in Fig. 7. The upper reflective layer, slow channel, and strong reflector are preserved in Fig. 7(a). Since the strong reflector is probably hard material that supports shear waves, the increased attenuation in Fig. 7(b) is expected since elastic effects are not explicitly included.[36] Figure 8(a) compares a typical NRL data shot (light curve) with a 10–180 Hz full-band simulation (dark curve) at 29 m depth using this bottom. Figures 8(b), 8(c), and 8(d) compare Butterworth filtered data (light curves) and simulations (dark curves) in 32, 64, and 95 Hz bands of the curves in Fig. 8(a). Relative amplitude and arrival times of two modes at 32 Hz show good agreement. In the 64 Hz band, the first two modes are well reproduced, while the third is present but smaller than in data. In the 95 Hz band, the large initial pulse in the simulation represents a combination of the first and second modes, and the presence and arrival of higher modes show good agreement with data. Results at other depths show similar agreement, and 29 m is displayed to show as many modes as possible. The overall agreement is good considering the geoacoustic parameter uncertainty and neglected effects such as bottom elasticity and profile range dependence.

C. Source model

Broadband simulation is usually performed by calculating continuous wave solutions at many discrete frequencies

in bands of interest to obtain waveguide transfer functions.[25] The transfer function and the Fourier transform of the source signature are used to obtain the final broadband pulse. Thus, the accuracy of broadband simulations depends on a reliable frequency–domain representation of the source signal. The available source monitor is critical for assuring that the airgun produced repeatable signals, which means that an average of recorded source signatures deviates very little from an individual source signature over the frequency band. However, this monitor hydrophone suffered from apparent contamination in its output, which we explain next.

Figure 9 illustrates inconsistencies between the frequency spectrum of the monitor signal and received VLA data. Figure 9(a) shows a sample pulse recording from hydrophone 5 (29 m) of the NRL array at 1910 GMT. Figure 9(b) shows the amplitude-normalized spectrum of this data. Note the gradual decrease in the energy in the second and higher harmonics and that the energy in the first harmonic is slightly higher than in the fundamental. This spectral distribution is typical of all data from both the NRL and WHOI arrays. Consequently, we believe it represents the roughly ''correct'' distribution of source energy propagated down each waveguide. For comparison, the average of the signals received at the monitor hydrophone is shown in Fig. 9(c), and its normalized spectrum in Fig. 9(d). This spectrum shows significantly more energy in the fundamental band around 32 Hz than in any other. Peaks occur at 64, 95, and 150 Hz, but the largest of these is less than one-sixth the amplitude of the 32 Hz peak. The substantially different energy distribution in Fig. 9(d) from that in Fig. 9(b) is suspicious. If the true source spectrum looks like Fig. 9(d) then significantly more low-frequency energy than high-frequency energy would appear in Fig. 9(b). There is actually proportionally more high-frequency energy, Fig. 9(b)—the opposite of what is expected from standard medium attenuation or scattering effects.

Confirmation of the difficulty is provided by Fig. 9(e), which shows the broadband time series at the NRL VLA simulated using this source and previously described bathymetry, water sound speed, and geoacoustic profiles. The strong presence of the 32 Hz band is suggested by the character of the time series. The normalized spectrum shown in Fig. 9(f) verifies that the frequency composition of the monitor signal in Fig. 9(c) is preserved by the waveguide, as expected in the absence of significant frequency-dependent effects.

The cause of the mismatch between data in Figs. 9(a), 9(b) and the simulations in Figs. 9(e), 9(f) is apparently from signal corruption at the monitoring phone by echoes off the surface and seafloor interfaces. These echoes are not propagated down the waveguide, so a source representation that includes them will cause inaccurate broadband modeling. We next show that, with this particular configuration of the airgun in the waveguide, the echoes lead to abnormal amplification of key frequency components. Suppose two sources are in a homogeneous environment. The first source is 1 m above a receiver (which represents the monitor), and the second is at a distance above the receiver (25 m for the airgun at 12 m) that corresponds to the surface echo. At a given time,

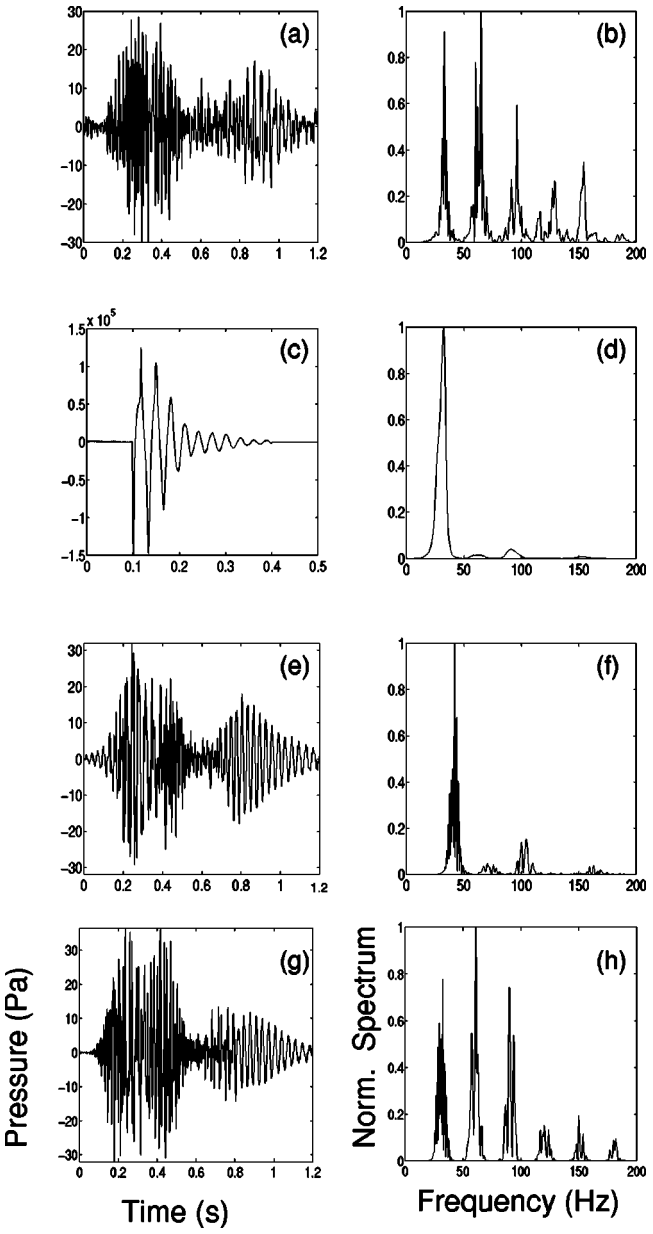


FIG. 9. An illustration of source modeling issues. (a) Sample NRL data pressure (Pa) data signal at 29 m; and (b) its normalized spectrum illustrates a gradual decrease in the energy level of higher-frequency harmonics. (c) The pressure obtained by averaging signals received at a source monitoring hydrophone. (d) Normalized spectrum of (c) contains a relatively large 32 Hz component. (e) The simulated signal using the source in (c), and (f) its normalized spectrum. Since the source frequency composition is essentially preserved by a waveguide, contamination has occurred in (c). (g) The simulated signal for the same waveguide environment as (e), but the source representation is Gaussian pulses in the frequency domain, and (h) its normalized spectrum. The spectral character matches (b) well.

both sources emit the same signal $f(t)$, for which $\tilde{F}(z)$ is the z-transform from sampling $f(t)$ at an interval T. If $g(t)$ is the system output, its z-transform is

$$\tilde{G}(z)=\tilde{F}(z)(1+z^{-k})\tilde{H}(z), \tag{5}$$

where $\tilde{H}(z)$ is the transfer function for the homogeneous environment and $z^{-k}\tilde{H}(z)$ is the (k sampling unit) delayed version of $\tilde{H}(z)$ corresponding to the second source. Range dependence everywhere has been ignored because of the

small distances involved. Equation (5) shows that the received signal is affected by additional poles at the kth roots of unity, which will amplify certain frequencies. We substitute $e^{i\omega T}$ for z and solve for ω, so that amplification is expected at

$$\omega_l = \frac{\pi}{kT}(1+2l) \Rightarrow f_l = \frac{1}{2kT}(1+2l), \quad l=0,1,2,..., \quad (6)$$

where kT is the total delay between the arrival of the first and second signals. With our configuration, $kT \approx 0.0167$ s, and hence $f_0 \approx 30.8$ Hz, so the fundamental in the source signature suffers substantial amplification from the surface echo return. Moreover, for $l=1$ (and $l=2$) the echo frequency is about 90 Hz (and 150 Hz), and thus anomalous amplification occurs in both frequency bands in Fig. 9(d), although they are weaker than the echo near 32 Hz. The conclusion is that signals recorded by the source hydrophone, while valuable for establishing the repeatability of the airgun cannot be used in modeling broadband propagation.

Such echoes or any unwanted frequency components can often be removed from a signal by standard filtering techniques.[32] However, when the corruption occurs in the same frequency bands as the signal of interest, it is not always possible to discriminate between the two receptions. One method is deconvolution techniques to estimate the impulse response of the waveguide.[18] These techniques require signal processing and assumptions that are not easily justified in this case. However, the repeatability of the source signature suggests that variations in the data are due to the environment and not the airgun source. Thus, we construct a model source signature that preserves the impulsive nature of the airgun, with frequency components, amplitudes, and bandwidths selected so reasonable agreement is obtained between VLA data and broadband simulations.

Since the frequency components in the received data signals are robust, we model the source by summing modulated Gaussian pulses in the frequency domain. The widths and locations of these pulses are chosen to match the dominant features in the spectra of the data from the NRL VLA. With a bandwidth of $2\sqrt{24\pi} = 17.34$ Hz, a model in the frequency domain is

$$S(\omega) = A\sum_{n=1}^{N} r^{n-1}[e^{-(\omega-2\pi q_n)^2/24\pi} + e^{-(\omega+2\pi q_n)^2/24\pi}], \quad (7)$$

where A is the amplitude, r is a number less than one that governs successive amplitudes of the harmonics, and q_n are frequency peak locations. To correspond with the experiment, peaks are chosen at 32 Hz and its harmonics. The harmonic frequencies need to be reduced by the observed factor of 15/16, because airgun signals are damped by surrounding water pressure (analogous to attenuation in oscillatory systems reducing a frequency to a quasifrequency), so the frequency peaks $q_n = 32(15/16)n$ Hz for $n=1,2,3,...$. While other parameters were obtained from the monitor signal, the selection $r=0.65$ is based on energy distributions at the VLA. The impulsive nature of the airgun signal is preserved, although with a quite different character than the monitor signal.[37]

TABLE I. Three strong peaks appear in the WHOI thermistor data frequency spectrum (Fig. 5). An internal wave speed of 0.42 m/s gives wavenumbers κ_{iw} and wavelengths λ_{iw} along the direction of internal wave propagation. The right two columns show values projected onto the NRL track.

WHOI thermistor data			NRL track parameters	
f_{iw} (Hz)	λ_{iw} (m)	κ_{iw} (rad/m)	λ_{NRL} (m)	κ_{NRL} (rad/m)
0.00132	319	0.0197	450	0.0140
0.00243	173	0.0364	244	0.0257
0.00395	106	0.0591	150	0.0418

Figure 9(g) shows a simulated time series using the model source. The spectrum of this time series is shown in Fig. 9(h) and clearly better matches the data. All subsequent simulations in this study are calculated using the source model of Eq. (7).

V. COMPUTATIONAL RESULTS

Using the environmental and source models from Sec. IV, broadband PE simulations are compared with one hour of airgun data. To obtain the internal wave model in a standard coordinate system of range r and depth z, we rewrite Eq. (4) for symmetric waves as

$$\eta(r,z) = \Phi_1(z)\sum_{n=0}^{5} A_n \operatorname{sech}^2\left[\frac{2\pi(r-r_n+v_n t)}{\Lambda_n}\right], \quad (8)$$

where $v_n = v_{iw}$ is a (uniform) internal wave speed, $A_n = A = 13$ m, $\Lambda_n = \Lambda$ is a uniform internal wave width, $r_n = sn\Lambda$, and s is called the spacing multiple. Varying v_n for each wave would introduce packet dispersion and was judged unnecessary for the current study. For models involving solutions of the KdV equation, v_{iw} can be estimated from environmental parameters as in Ref. 18. Here we choose $v_{iw} = 0.42$ m/s because when projected onto the NRL acoustic track (discussed below) we obtain $v_{\mathrm{NRL}} = 0.6$ m/s, which provides a consistent match with the period of acoustic I_T variations in Fig. 4 and is also close to the 0.579 m/s value in Ref. 18 for a similar packet of waves.

Column 1 of Table I contains the frequency locations of peaks in the spectrum of thermal data shown in Fig. 2(c). Values for wavelengths $\lambda_{iw} = v_{iw}/f_{iw}$ associated with these frequencies are shown in column 2, and column 3 shows wavenumbers $\kappa_{iw} = 2\pi/\lambda_{iw}$. To obtain internal wave parameters to model the NRL track, values in columns 2 and 3 are projected using an incidence angle estimate of 45°. Column 4 contains projected column 2 values $\lambda_{\mathrm{NRL}} = \lambda_{iw}/\cos(45°)$, and column 5 shows $\kappa_{\mathrm{NRL}} = 2\pi/\lambda_{\mathrm{NRL}}$. The location of the lowest frequency spectral peak in Fig. 5 corresponds to the periodic distance between evenly spaced nonlinear internal waves. Thus, the nonlinear internal wave width Λ in Eq. (8) can be found using $r_n = s\Lambda_{\mathrm{NRL}} = \lambda_{\mathrm{NRL}}$. Since $s=2.3$ from Sec. IV A, $\Lambda_{\mathrm{NRL}} = 450/2.3 = 195$ m.

The projected packet model is propagated toward the source for 60 ''minutes'' of geotime and a PE simulation is performed for each minute. Results at 29 m using the experimental bathymetry, water sound speed profiles at the source, and the bottom profiles shown in Fig. 7 are compared to data

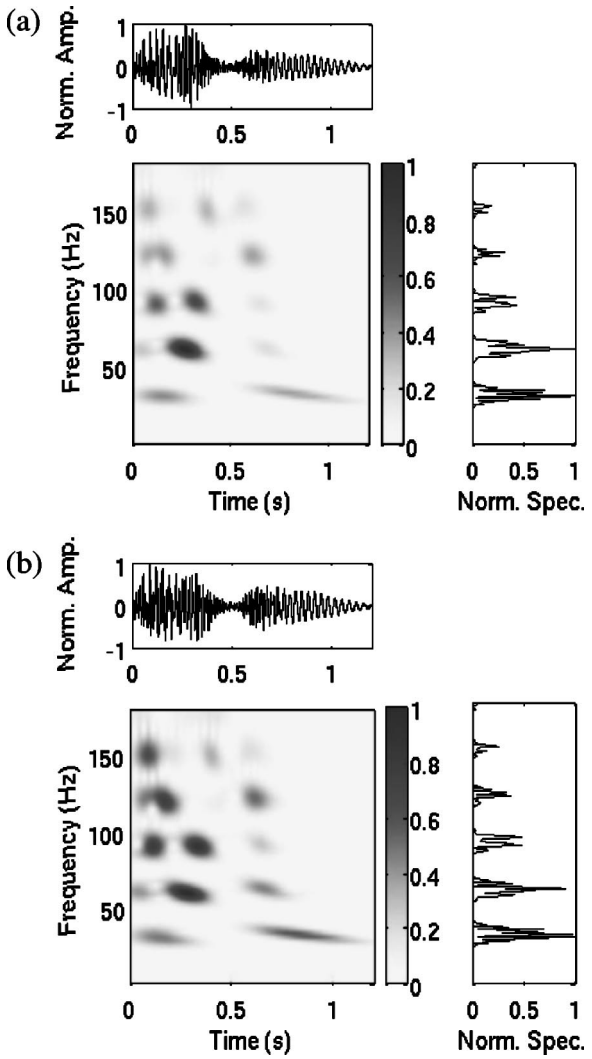


FIG. 10. A Gabor wavelet analysis of broadband signals from PE simulations. The same panels as in Fig. 3. (a) Simulation minute 12 concentrates acoustic energy near 32 Hz and its harmonics. (b) Simulation minute 18 (6 geotime minutes later) shows an increased higher-frequency content and variations in mode strengths.

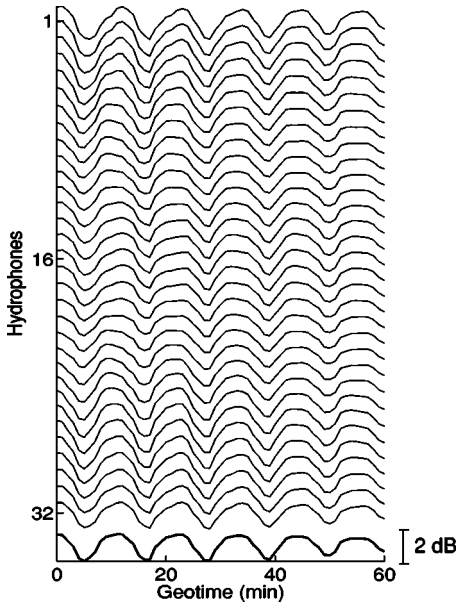


FIG. 11. Intensity I_T at each hydrophone for 60 min traversal of an internal soliton packet past the source. Variations occur clearly at all 32 NRL hydrophones, and the dynamic range of depth-averaged intensity variations closely matches observations (Fig. 4).

shot 1910 GMT in Fig. 8. Note that modes in bands around 32, 64, and 95 Hz are clearly present (and will thus be able to couple) and that their arrivals are consistent with data. Small discrepancies should not significantly affect scalograms or pulse-averaged intensity.

We next examine scalograms of the simulated signals. Similarity of group velocity curves and modal excitation between the data and simulations is also a good check of the broadband synthesis method. Figure 10 shows two signals that are 6 min apart at 29 m depth. Behavior comparable to that in the data scalograms in Fig. 3 is observable. Figure 10(a) shows simulation minute 12. The time domain pulse suggests, and its Fourier transform shows, significant energy peaks in expected frequency bands. Similar to Fig. 3(a), most signal energy arrives in the second mode of the 64 Hz band, while well-separated arrivals appear in the 95 and 120 Hz bands. Two modes appear at 32 Hz, three modes at 64 and 95 Hz, and up to four modes in bands above 95 Hz, consistent with Fig. 3(a). Figure 10(b) shows simulation minute 18 and notable differences from the minute 12 signal. Acoustic energy has been distributed to the third mode of the 64 Hz band and into higher-frequency bands. This is the same redistribution as is evident between the data scalograms in Fig. 3. Modes at 95 Hz and above are not as well separated as at minute 12, which again is similar to the data. The features are quasiperiodic in the computations with time scales close to those in the data.

Figure 11 shows pulse-averaged intensity I_T at each hydrophone of the NRL VLA for the simulation parameters above in a plot analogous to Fig. 4(b). A very close match between simulations and data is evident for the first 30 min. Approximately 3.5 periods occur in the calculations, similar to variations in the data. In the second 30 min, comparably strong variations occur in the data, but their period is longer and does not match the computations. This discrepancy is likely caused by evolution of the nonlinear internal wave packet and its deviations from the model. The overall depth dependence of the calculated I_T is very similar to that of the NRL data, with slightly larger oscillations at the top and bottom of the array. The depth-averaged amplitude is very close to that of the data.

Figure 12 displays depth-averaged I_T calculations for full-band and bandpass filtered data (solid curves) and computations (dashed–dotted curves). Figure 12(a) shows a very good match between data and computed depth-averaged I_T. Note a strong amplitude and phase correlation between data and computations for the first 30 min of geotime. Data variations have maximum amplitude slightly more than 2 dB for this time period, while the computations have 1.65 dB. The location of intensity maxima and minima is consistent with data, suggesting that the speed estimate is appropriate. Figures 12(b), 12(c), and 12(d) show depth-averaged I_T for order-10 Butterworth filtered signals with a 10 Hz bandwidth and center frequencies of 32, 64, and 95 Hz. At 32 Hz the data are somewhat noisy, although one large intensity mini-

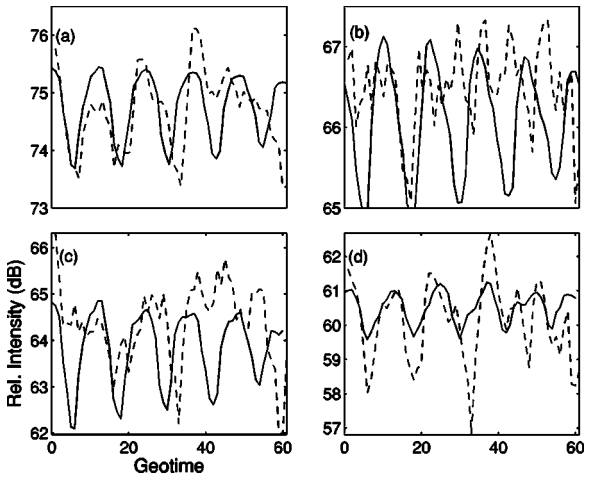


FIG. 12. Depth-averaged intensity variations for data (dashed curves) and PE simulations (solid curves): (a) full band, and Butterworth filtered bands centered at (b) 32 Hz, (c) 64 Hz, and (d) 95 Hz. The relative intensity represents I_T converted to dB re: 1 μPa. Mean sound levels, quasiperiods, and amplitudes of variations are reproduced well by simulations for the full frequency band and two lower bands.

mum is modeled well in location and amplitude by computations. For higher frequency bands, the intensity data are cleaner. The 64 Hz band shows two intensity minima with locations and amplitudes that are produced by computations, and the 95 Hz band comparison emphasizes the accuracy of the peak locations. The amplitude variation of computations in the 95 Hz band is smaller than in the data, which is a likely reason for the somewhat reduced amplitude of the computed full band variations. Also, the average intensity levels of the computations are very similar to data.

The physical mechanism that is evidently responsible for producing the observed acoustic variations is well described in the literature. Resonant coupling between acoustic modes is caused by range-dependent anomalies in the bottom[38] or water column[7,11] when the following condition is satisfied:

$$\kappa \approx k_n - k_m , \tag{9}$$

where κ is the effective spectral peak wavenumber of the internal wave anomalies (in our situation κ_{NRL}) and k_n and k_m are horizontal wavenumbers of acoustic modes n and m. Equation (9) also arises because mode coupling occurs when acoustic wavenumber differences nearly coincide with peaks in the spectrum of inhomogeneities.[16]

Table II shows horizontal wavenumbers k_n for the three lowest peak frequencies in the source spectrum. These wavenumbers were calculated using the program COUPLE[39] for the bottom parameters shown in Fig. 7 and sound-speed profile calculated from an average of CTD measurements taken by the R/V Cape Hatteras. Table III shows differences between selected wavenumbers from Table II. Specific wavenumber differences within 10% of effective internal wave wavenumber spectrum peaks (the fourth column of Table I) have been underlined. Numerous opportunities are available to satisfy resonance conditions at 32, 64, and 95 Hz. The number of possible resonance interactions is the probable reason this particular packet configuration provides comparable variations in data and computations. Also, the resonant coupling opportunities for all three dominant acoustic frequencies is a likely reason for the similarity of I_T variations seen in all panels of Fig. 12.

TABLE II. Acoustic wavenumbers calculated using COUPLE for the model environment at the source with no internal waves. Three, five, and seven propagating modes occur near 32, 64, and 95 Hz.

	32 Hz	64 Hz	95 Hz
1	0.131	0.268	0.399
2	0.117	0.258	0.391
3	0.102	0.243	0.381
4		0.232	0.366
5		0.216	0.355
6			0.343
7			0.316

Two aspects of this analysis bear mentioning. First, the speed of the nonlinear internal wave packet is lower than estimates obtained from other SWARM data,[40] but these estimates come from different days of the experiment. Second is the use of 45° as the incidence angle, which is slightly smaller than the 49° obtained from bearing estimates and experimental geometry. This variation can be acounted for by uncertainty in these estimates or by a slight curvature of the nonlinear internal wavefronts. Several combinations of internal wave parameters were tested with Λ ranging from 180 to 350 m and r_n from $2n\Lambda$ to $2.5n\Lambda$, but the formulation used above, with $\Lambda_{\mathrm{NRL}}=195$ m, $r_n=2.3n\Lambda_{\mathrm{NRL}}$, and $v_{\mathrm{NRL}}=0.6$ m/s, provided the best internal wave spectrum match. In addition, this configuration provided the largest acoustic variations and appropriate periods. This supports the hypothesis that the internal wave-acoustic resonance condition governs observed variations and a further analysis is presented in Sec. VI.

VI. PARAMETER SENSITIVITY

According to the resonance condition in Eq. (9), mode coupling will be affected by changes to internal wave spectral peak locations, governed by internal wave parameters, or acoustic wavenumbers, which depend on water column and geoacoustic parameters.[41] Ray theoretic intensity calculations[42] and scintillation index correlations[6] also indicate

TABLE III. Differences between selected acoustic wavenumbers from Table II. Comparisons with wavenumbers κ_{NRL} in Table I show multiple opportunities for acoustic mode coupling within the internal wave packet. Underlined values are within 10% of a κ_{NRL} peak.

	k_1-k_2	k_1-k_3	k_2-k_3	k_2-k_4	k_3-k_4	k_1-k_5	k_3-k_5	k_4-k_6
32 Hz	0.0140	0.0283	0.0143					
64 Hz	0.00928	0.0237	0.0144	0.0257	0.0113	0.0520	0.0271	
95 Hz	0.00752	0.0183	0.0107	0.0258	0.0151	0.0443	0.0260	0.0230

acoustic sensitivity to nonlinear internal wave parameters. Despite the quantity of environmental data and the consistent modeling results, uncertainty in conclusions about internal wave speed, v_n, width Λ_n, and spacing Δr_n in Eq. (8) are unavoidable. Limited knowledge of the water column and geoacoustic parameters also introduce ambiguity into computational results. Finally, the resonance condition itself cannot specify unique sets of nonlinear internal wave parameters.[7]

In this section, the dependence of acoustic variations on model parameters is described. In order to isolate resonance effects from range dependence, broadband (10–180 Hz) two-dimensional (2-D) PE simulations are performed in a 71 m deep range-independent waveguide using the source model in Sec. IV and geoacoustic parameters in Fig. 7. The signals are received at 16 depths, from 15 m and every 3.5 to 69 m. In all cases internal wave packets start 100 m from the source, have amplitudes $A_n = A = 15$ m, and propagate at speed $v_n = v = 0.65$ m/s toward the source for 30 min. Except where otherwise noted, the packets consist of evenly spaced waves with $\Delta r_n = \Delta r$. The computational evidence shows that the resonance condition can predict increases in acoustic variability with geotime, indicating that the actual variations result from mode coupling. The soliton amplitude will not affect peak locations in a packet's normalized spectrum, so soliton amplitude effects are not discussed.

A. Internal wave parameters

For a packet of nonlinear internal waves with equal amplitudes, widths, and spacings, the number of waves $N > 1$ does not affect the locations of the spectral peaks. Consequently, the occurrence of resonant mode coupling is not significantly affected. Geotime simulations performed that verify this assertion, for N between 3 and 9 waves, show that the amplitudes of I_T variations were always very similar.[37] Unless noted otherwise, PE simulations here use $N = 5$ sech^2 waves.

1. Width

In contrast to N, the width parameter Λ does affect the Fourier spectrum of Eq. (8). In Sec. V, acoustic variability consistent with data was produced with $\Lambda = 195$ m. Here nonlinear internal wave packets with larger widths are examined to determine if packet spectral peak locations can predict maxima in acoustic variability. Figure 13 shows normalized spectral peak behavior with Λ between 300 and 800 m for a $\Delta r = 2.3\Lambda$ nonlinear internal wave packet. The frequency axis is calculated using a sampling rate of 1/30 Hz. As the width Λ of each waveform increases, the dominant frequency of the internal wave packet decreases. For smaller Λ, the first (lowest-frequency) peak is the maximum of the internal wave packet spectrum. The frequency location of the first peak and its spectral height decreases as Λ increases. The location of the second peak also decreases with Λ, but its height increases until it becomes the spectral maximum. Further increases in Λ emphasize lower-frequency components of the packet and cause more spectral peaks to occur at frequencies below the maximum peak location.

Acoustic wavenumber differences for 32, 64, and 95 Hz

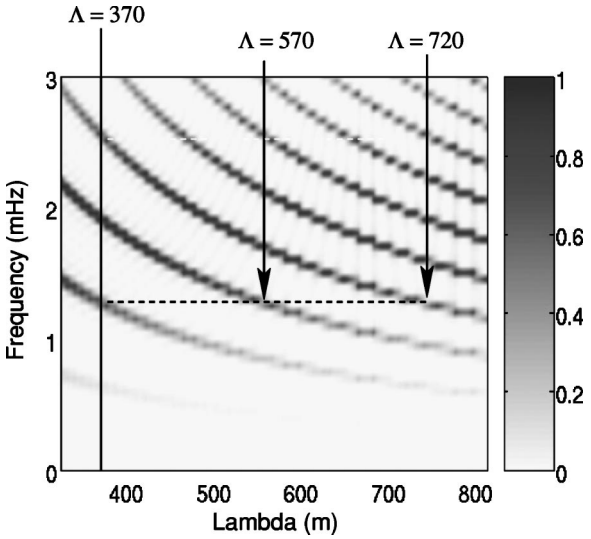


FIG. 13. Normalized internal wave spectrum for a packet with variable Λ and $r_n = 2.3\Lambda$. As Λ increases, dominant frequency components in the packet decrease. Resonance interactions occur for $\Lambda = 370$, with peaks at 0.0014 and 0.0024. Another set of peaks occur near these frequencies when $\Lambda \approx 570$ and $\Lambda \approx 720$.

are seen to correspond with internal wave packet frequencies near 0.0013 and 0.0024 Hz in Table I. By examining Fig. 13 we see that spectral peaks occur at these frequencies for several internal wave packets with larger Λ than that used to model the NRL track. To confirm resonant behavior consistent with Eq. (9), we examine acoustic variations for internal wave packets with Λ near 370 m, which has spectral peaks near both 0.0013 and 0.0024 Hz. Figure 14 shows results of broadband simulations for internal wave packets with five equally spaced values of Λ from 325 to 425 m, and Table IV shows dB amplitude values for I_T fluctuations of full band geotime simulations and bandpass filtered simulations with center frequencies at 32, 64, and 95 Hz. As predicted, maximum variations occur for $\Lambda = 375$ m, although $\Lambda = 350$ m also shows large-amplitude fluctuations, which can be seen in Figs. 14(b), 14(c), and in column 2 of Table IV. Amplitude degradation on either side of $\Lambda = 375$ m is evidence of a resonance peak for modal interaction. For the 32 and 64 Hz bands, $\Lambda = 375$ m causes maximum amplitude variations compared with $\Lambda = 350$ m in the 95 Hz band. This is consistent with Table III, which shows acoustic wavenumber differences occur for a slightly higher wavenumber (and thus a slightly higher frequency, near 0.0015 Hz) in the 95 Hz band.

By following the internal wave spectral frequencies associated with resonance across Fig. 13 (black dashed lines) it appears that resonance points should also occur when Λ is close to 570 and 720 m. Table V shows an amplitude minimum at $\Lambda = 675$ m and larger variations for $\Lambda = 575$ m and $\Lambda = 725$ m that confirm the occurrence of resonant peaks. Table V indicates that 32 and 64 Hz mode coupling is prevalent for $\Lambda = 575$ m, while 95 Hz coupling dominates variations caused by a packet with $\Lambda = 725$. This is consistent with Table III, which shows two occurrences of 95 Hz band wavenumber differences with 0.026 rad/m. Once again, this will occur at a slightly higher frequency than 0.0024 Hz; also the $k_1 - k_3$ value at 95 Hz is near 0.018 rad/m, which corre-

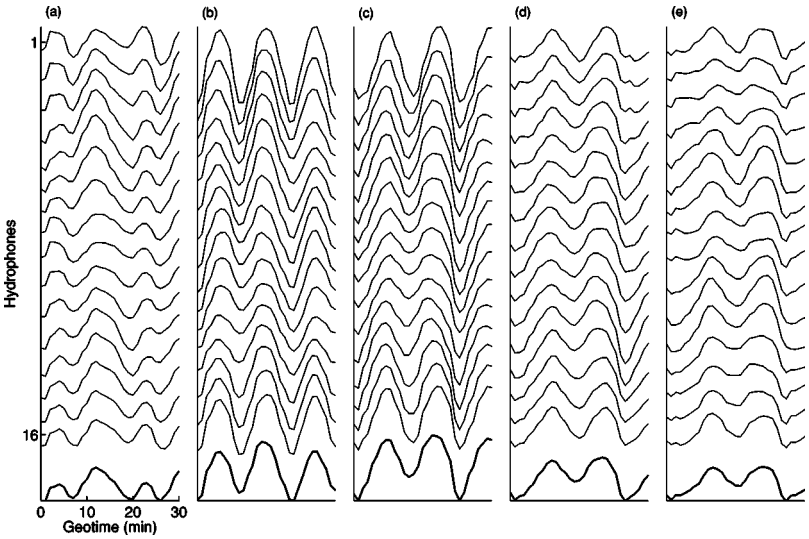


FIG. 14. Full band I_T variations for different Λ and $r_n = 2.3\Lambda$ as the nonlinear internal wave packet moves toward the source. (a) $\Lambda = 325$ m, (b) $\Lambda = 350$ m, (c) $\Lambda = 375$ m, (d) $\Lambda = 400$ m, and (e) $\Lambda = 425$ m. All variations show a similar phase and weak depth dependence. The largest amplitudes occur in (b) and (c) with amplitude decay visible on either side.

sponds to the peak for $\Lambda = 725$ m just above the lower white dashed line in Fig. 13.

Together these two demonstrations show that nonlinear internal wave packet parameters can predict acoustic frequencies where large acoustic variability will occur. They also indicate that for a region without large packet spectral peaks at resonance frequencies, the acoustic variability will be significantly reduced.

2. Spacing

Spacing variations between individual sech-squared waves also affect the nonlinear internal wave packet spectrum. For evenly spaced packets, $\Delta r = s\Lambda$ where s is called the spacing multiple. Due to the large acoustic variations observed (above), we focus here on internal wave packets with $\Lambda = 375$ m. Figure 15 shows a normalized spectral peak dependence on s. As s increases, the separation between internal wave peaks increases, decreasing the peak frequencies for the packet. Multiple spectral peaks arise for given s, and the dependence of the spectral maxima is similar to that in Fig. 13.

To illustrate resonant behavior, PE simulations are performed for broadband signals transmitted through nonlinear internal wave packets propagating toward the source with spacing multiples near 2.3. Table VI confirms the resonant behavior of I_T fluctuation amplitudes. For full band signals, the largest amplitude occurs at $s = 2.3$, with progressive amplitude decay for longer and shorter s. The $s = 2.3$ case has spectral peak frequencies locations near 0.0014 and 0.0025. Results centered at 95 Hz have maximum fluctuation amplitude for $s = 2.15$, due to slightly higher wavenumber differences in this frequency band (Table III).

An examination of Fig. 15 reveals packet spectra peaks near 0.0014 Hz when $s \approx 4.3$ and near 0.0025 when $s \approx 3.6$. Table VII shows the amplitude of I_T variations (in dB) caused by packets with spacing multiples from 3.5 to 4.4. For 10–180 Hz full band simulations, the amplitude peak at $s = 3.6$ is followed by decay and then an increase to another peak at $s = 4.4$. In addition, the 95 Hz band amplitude for $s = 4.2$ is relatively small given the resonance peak for $s = 4.3$. The small amplitude of depth-averaged oscillations at 64 and 95 Hz does not seem to support resonant condition behavior. However, the size of these depth-averaged values is affected by depth dependence of the pulse-averaged intensity curves. Figure 16 investigates bandpass filtered variations by a receiver for (a) 32 Hz, (b) 64 Hz, and (c) 95 Hz. Very little depth dependence is present in Fig. 16(a), while Fig. 16(b) shows that near 20 min, I_T curves in the middle of the array are completely out of phase with curves near the top of the array. The variations thus tend to cancel each other out when depth averaging. Figure 16(c) shows large-amplitude fluctuations at each receiver, which is expected for $s = 4.2$. Again strong depth dependence is present and oscillations change phase in alternating quarters of the water column. This depth dependence leads to a cancellation in depth-

TABLE IV. I_T variation amplitudes for different Λ with $r_n = 2.3\Lambda$. Resonance effects expected near $\Lambda = 375$ m. Amplitude maxima in 32 and 64 Hz bands for $\Lambda = 375$ m, in a 95 Hz band for $\Lambda = 350$ m.

Λ	Full	32 Hz	64 Hz	95 Hz
325	1.27	0.89	2.73	1.71
350	2.46	1.55	3.72	4.07
375	2.78	3.07	5.10	3.30
400	1.71	3.06	2.93	1.60
425	1.31	1.60	1.86	1.82

TABLE V. I_T variation amplitudes for different Λ with $r_n = 2.3\Lambda$. Peak locations in Fig. 13 predict resonance interaction near $\Lambda = 570$ m and $\Lambda = 720$. Column 2 shows 10–180 Hz full band simulations confirming resonant behavior. Interactions occur in all three frequency bands for $\Lambda = 570$ m, and are restricted to 64 and 95 Hz bands for $\Lambda = 720$ m.

Λ	Full	32 Hz	64 Hz	95 Hz
550	2.08	1.78	4.66	3.36
575	2.21	2.43	4.86	2.58
600	1.70	2.40	3.46	1.56
625	1.11	1.47	2.23	1.05
650	1.23	1.08	1.94	1.48
675	0.95	0.96	1.95	1.30
700	1.64	0.98	1.84	2.71
725	2.02	1.26	2.73	4.91

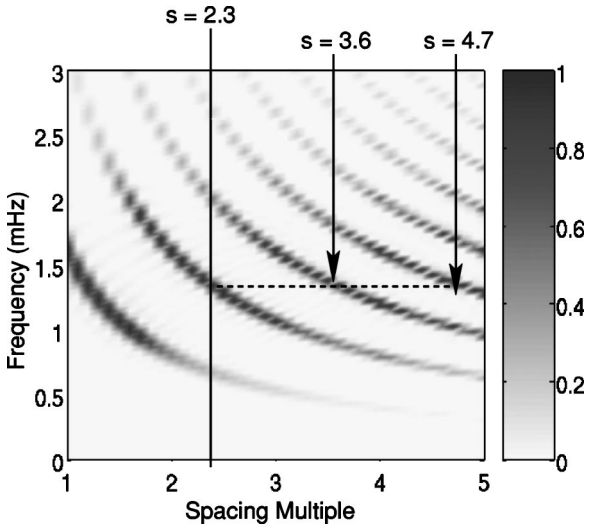


FIG. 15. Normalized internal wave spectrum dependence on spacing multiple s with $\Lambda=370$. As s increases, dominant frequency components in the packet decrease. Resonance interactions occur for $s=2.3$. Another set of peaks occur near $s=3.6$ and $s=4.2$.

averaged I_T and explains the small amplitude of the 95 Hz band in Table VII.

B. Geoacoustic parameters

A deep strong reflector can significantly impact wavenumbers and group speeds of low-frequency acoustic modes.[41] Figure 17 shows sound speed, attenuation, and density profiles for nine bottom models. In model R_1 through R_9 the reflector increases in steps of 5 m from 6 to 46 m below the water–sediment interface, respectively. Model R_5 is similar to that used in Sec. IV. Changes in wavenumbers from varying reflector depths cause differences in relative mode arrival times in each frequency band. These changes can force the wavenumber difference to depart from internal wave spectral peak locations, resulting in less modal interaction and smaller I_T variations.

Broadband simulations were performed using a nonlinear internal wave packet configuration with $\Lambda=370$ and $s=2.3$. Figure 18 shows I_T variations caused as the nonlinear internal wave packet moves toward the source for models R_3 through R_7. Large fluctuations occur at all hydrophones in Fig. 18(c), while significantly smaller-amplitude oscillations appear in the other panels. The variations in all cases are weakly depth dependent. Table VIII confirms that the I_T fluctuation amplitude peak occurs with model R_5 for the full band and all filtered signals. The amplitude decrease associated with raising and lowering the strong reflector from 26 m occurs because acoustic wavenumber differences for those configurations do not cause resonant interaction.

TABLE VI. I_T variation amplitudes for different s with $\Lambda=370$ m. Resonance interactions most significant near $s=2.3$ for full-band, 32, and 64 Hz band simulations. The resonant peak occurs for $s=2.15$ in a 95 Hz band.

s	Full	32 Hz	64 Hz	95 Hz
2.0	1.36	0.93	2.76	2.03
2.15	2.57	1.64	3.83	4.27
2.3	2.78	3.07	5.11	3.30
2.45	1.68	3.00	2.85	1.57
2.6	1.27	1.54	1.72	1.84

TABLE VII. I_T variation amplitudes for different s with $\Lambda=370$ m. Resonance interactions are expected near $s=3.6$ and $s=4.1$. Full-band and 32 Hz band simulations show decay from resonant peak at $s=3.6$ but do not indicate a second peak. An amplitude increase occurs for 64 and 95 Hz bands.

s	Full	32 Hz	64 Hz	95 Hz
3.5	1.86	1.92	5.28	1.64
3.6	1.87	2.50	4.37	1.70
3.7	1.59	2.26	3.13	1.66
3.8	1.33	1.74	2.26	1.36
3.9	1.14	1.32	1.57	1.49
4.0	1.05	1.03	1.52	1.16
4.1	1.08	0.90	1.90	1.21
4.2	1.13	0.95	1.87	1.23
4.3	1.45	0.93	2.01	2.36
4.4	1.74	0.85	1.58	4.36

The effect of the shallow reflector was also considered. Acoustically hard material is characterized by higher sound speed,[25] so decreasing the maximum sound speed of the shallow reflector could affect acoustic variability for several reasons. First, the amount of energy reflected back into the water column is reduced and energy is less effectively trapped between the shallow and deep reflectors. Second, reduced sound-speed gradients above and below the shallow reflector will also modify acoustic intensity. PE computations for several shallow sediment configurations were considered. The conclusion is that variations in the shallow reflector do not impact acoustic horizontal wavenumbers (and changes in I_T amplitude) as significantly as the location of the deep reflector (for the lowest frequencies, the shallow reflector may even be acoustically transparent).

C. Sensitivity estimation

If an internal soliton with a particular width $\Lambda_0(f_0)$ defined by Eq. (9) causes resonant modal interaction across a

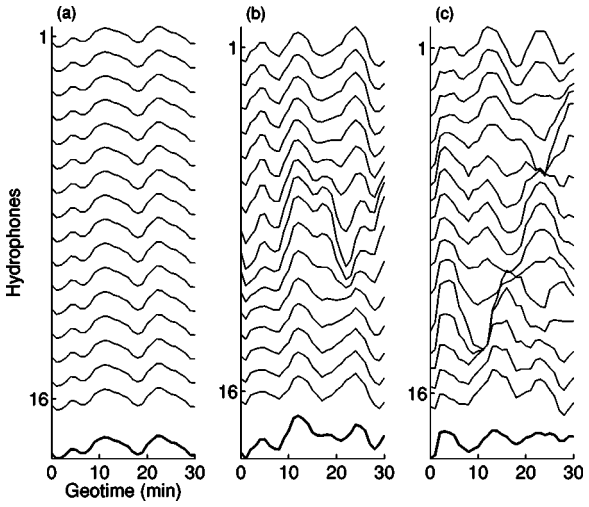


FIG. 16. Depth-dependent I_T variations for $\Lambda=370$ m and $s=4.2$: (a) the 32 Hz band shows little depth dependence; (b) the 64 Hz band shows depth dependence in the middle phones beginning near 20 min. (c) The 95 Hz band shows significant variations at each receiver. Depth-averaged variations in (c) show significant cancellation.

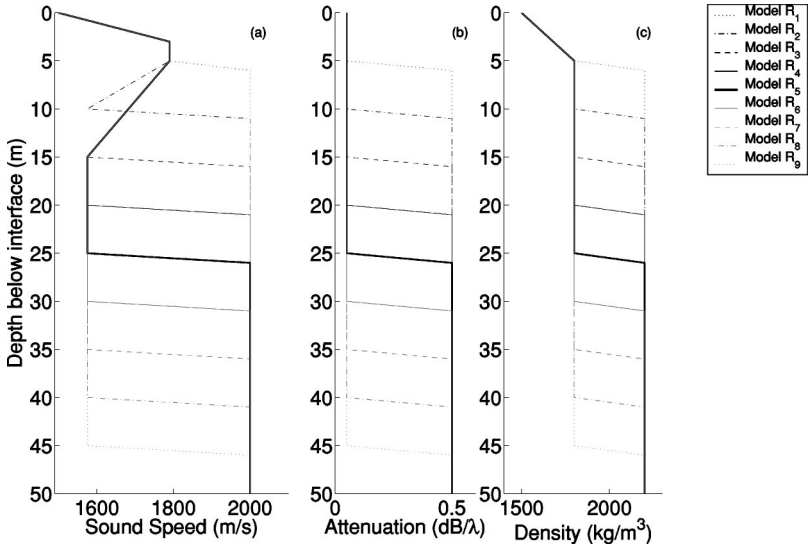


FIG. 17. (a) Sound speed (m/s), (b) attenuation (dB/λ), and (c) density (kg/m^3) for geoacoustic models. Model R_1 (light dotted curve), R_2 (light dashed-dotted curve), R_3 (light dashed curve), R_4 (light solid curve), R_5 (heavy black curve), R_6 (black solid curve), R_7 (black dashed curve), R_8 (black dashed-dotted curve), and R_9 (black dotted curve) have a 2000 m/s reflector at 6, 11, 16, 21, 26, 31, 36, 41, and 46 m from the water–sediment interface, respectively.

frequency band in oceanic waveguides, the associated internal wave spectral peak must have sufficient amplitude. The peak must also be broad enough to permit coupling for Λ reasonably close to Λ_0, even though the amount of coupling may be reduced. A related matter is to estimate the bandwidth around f_0 that can be expected to exhibit variability due to mode coupling.

To address these issues, assume Eq. (9) holds for a resonant soliton width Λ_0 associated with acoustic modes n and m at radian frequency $\omega_0 = 2\pi f_0$, where $k_i(\omega_0)$ is the ith acoustic horizontal wavenumber, so

$$\Lambda_0(\omega_0) = \frac{2\pi}{k_n(\omega_0) - k_m(\omega_0)}. \tag{10}$$

Is Λ_0 still resonant for an acoustic signal with frequency $\omega_0 + \Delta\omega$? Expanding the wavenumber in a Taylor series for a band near ω_0 and dropping higher-order terms gives

$$k_n(\omega_0 + \Delta\omega) \approx k_n(\omega_0) + k_n'(\omega_0)\Delta\omega. \tag{11}$$

Since

$$k_n(\omega_0) - k_m(\omega_0) = \omega_0[S_p^n(\omega_0) - S_p^m(\omega_0)] \equiv \omega_0 \Delta S_p^{nm}(\omega_0), \tag{12}$$

where $S_p^i(\omega)$ is the mode i phase slowness, and

$$k_n'(\omega_0) - k_m'(\omega_0) = \omega_0[S_g^n(\omega_0) - S_g^m(\omega_0)] \equiv \Delta S_g^{nm}(\omega_0), \tag{13}$$

where $S_g^i(\omega)$ is the mode i group slowness, Eq. (10) can be rewritten in terms of the Taylor series in Eq. (11) as

$$\Lambda_0(\omega_0 + \Delta\omega) \approx \frac{2\pi}{\omega_0 \Delta S_p^{nm}(\omega_0) + \Delta\omega \Delta S_g^{nm}(\omega_0)}. \tag{14}$$

Thus, in order for Λ_0 to remain a resonant wavelength, any change in $\Delta S_p^{nm}(\omega_0)$ must be offset by a change in $\Delta S_g^{nm}(\omega_0)$ that is opposite in sign. This requirement is reasonable in shallow waveguides where increased frequency usually leads to reduced phase velocity and increased group velocity (past the Airy phase).

A useful example of Eq. (14) is a rigid-bottom isovelocity waveguide of sound speed c and depth D, with wavenumbers for mode n at frequency f_0,

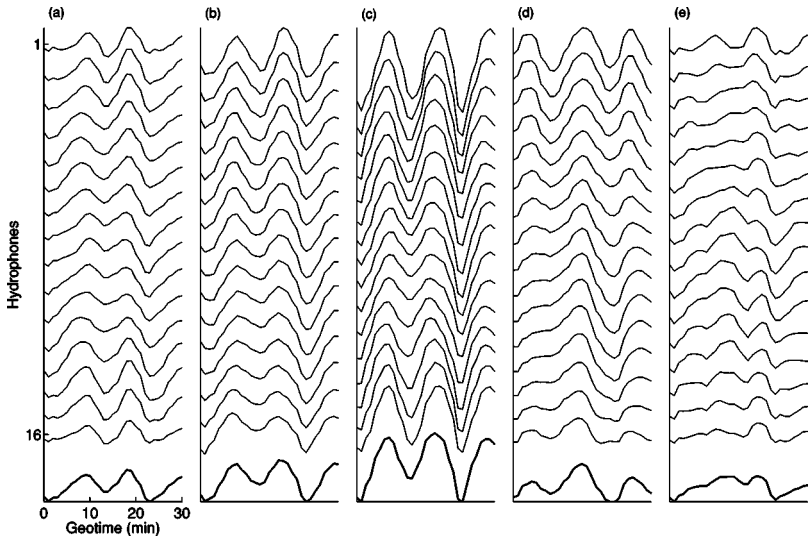


FIG. 18. I_T variations for different reflector depths using $\Lambda = 370$ and $s = 2.3$ for model (a) R_3, (b) R_4, (c) R_5, (d) R_6, and (e) R_7. All variations exhibit similar phase and weak depth dependence. Large-amplitude I_T fluctuations in (c) and decreasing amplitudes on either side is evidence of a resonant interaction.

TABLE VIII. I_T variation amplitudes for different geoacoustic models. The resonance peak occurs as the reflector passes through 26 m depth in all frequency bands.

Model	Full	32 Hz	64 Hz	95 Hz
R_1	0.84	1.20	1.26	1.35
R_2	1.08	1.68	1.67	1.34
R_3	1.23	1.65	1.93	1.15
R_4	1.66	2.31	2.28	1.68
R_5	2.95	2.82	5.00	3.91
R_6	1.47	1.48	2.40	2.40
R_7	1.00	1.32	1.71	1.46
R_8	1.02	1.37	1.57	1.43
R_9	1.12	1.02	1.56	1.39

$$k_n(f)=\sqrt{\left(\frac{2\pi f}{c}\right)^2-\left[\frac{\left(n-\frac{1}{2}\right)\pi}{D}\right]^2}. \tag{15}$$

Equation (11) in terms of f is

$$k_n(f_0+\Delta f)=k_n(f_0)+\left(\frac{f_0\,\Delta f}{k_n(f_0)}\right)\left(\frac{2\pi}{c}\right)^2, \tag{16}$$

and Eq. (14) becomes

$$\Lambda_0\approx\frac{2\pi}{k_n(f_0)-k_m(f_0)+\left(\frac{2\pi k_0 f_0\,\Delta f}{c}\right)\left(\frac{1}{k_n(f_0)}-\frac{1}{k_m(f_0)}\right)}. \tag{17}$$

This relation only occurs if $\Delta f=0$, $m=n$, or $k_m(f_0)=k_n(f_0)$. These are all trivial cases in the rigid bottom waveguide, which indicates that the resonant peaks are indeed sharp. However, other waveguides allow $k_n'(f_0)\approx k_m'(f_0)$, which can occur asymptotically at high frequencies, at the Airy phase, or in multilayered waveguides.[23]

Using this example we also examine the change in optimal soliton width for a given variation Δf from f_0:

$$\begin{aligned}\Delta\Lambda&=\Lambda(f_0+\Delta f)-\Lambda(f_0)\\&=\frac{2\pi}{k_n(f_0+\Delta f)-k_m(f_0+\Delta f)}-\frac{2\pi}{k_n(f_0)-k_m(f_0)}\\&=\frac{2\pi}{\underbrace{k_n(f_0)-k_m(f_0)}_{a}-\Delta f\underbrace{\left(\frac{4\pi^2 f_0}{c^2k_n(f_0)}-\frac{4\pi^2 f_0}{c^2k_m(f_0)}\right)}_{b}}\\&\quad-\frac{2\pi}{k_n(f_0)-k_m(f_0)}\\&\approx\frac{2\pi b\,\Delta f}{a^2}.\end{aligned} \tag{18}$$

Large values of Eq. (18) indicate that the resonant soliton wavelength must undergo significant changes for small changes in frequency. If this happens, then a model soliton width would be unable to cause a modal interaction across a frequency band, because Eq. (10) would not be satisfied. However, a relatively small value from Eq. (18) allows a soliton width selected for a particular center frequency to cause a modal interaction (though perhaps not at optimal resonance levels) over a frequency band Δf.

TABLE IX. Variability in $\Delta\Lambda_{ij}/\Delta f$ for modes i and j.

	$\Delta\Lambda_{12}$	$\Delta\Lambda_{13}$	$\Delta\Lambda_{23}$	$\Delta\Lambda_{24}$	$\Delta\Lambda_{34}$	$\Delta\Lambda_{15}$	$\Delta\Lambda_{35}$	$\Delta\Lambda_{46}$
32 Hz	14.5	9.8	6.2					
64 Hz	13.6	4.7	7.2	3.0	5.2	1.7	2.5	
95 Hz	13.5	4.6	6.9	2.8	4.8	1.4	2.1	1.7

Table IX shows values of $\Delta\Lambda/\Delta f$ calculated using Eq. (18) for mode combinations shown in Table III. Row one shows changes near 10 m/Hz for all mode combinations. For a 10 Hz bandwidth signal $\Delta f=5$ Hz so $\Delta\Lambda\approx 50$ m—about 20% of a soliton width. At 64 Hz, the largest value occurs for $\Delta\Lambda_{12}$. This value is about 50% of a typical wavelength, so it may be difficult to have interactions between these two modes. The other 64 Hz values are much smaller, especially for resonant mode combinations displayed in Table III, so interaction among these is feasible. Values of $\Delta\Lambda/\Delta f$ for 95 Hz follow the same pattern. These results are reasonable since Table IV and Table V demonstrate that for the 32 Hz band, significantly larger acoustic variability arises when Λ is within 50 m of the soliton width that causes the largest-amplitude variability. In the 64 and 95 Hz bands, the resonance peaks are sharper in Tables IV and V, which is confirmed by the reduced values for these frequency bands in Table IX. These estimates indicate the resonant soliton width is fairly sensitive to the acoustic frequency, especially in the lowest mode pairs, but the use of rigid bottom wavenumbers represents a worst-case scenario for the sensitivity. The values in Table IX are also consistent with results for nonlinear internal wave packet randomness since variability of about 20–30 m in Λ_{iw} do not significantly affect the internal wave spectrum.

Table X shows values of $\Delta\Lambda/\Delta D$ that result from a perturbation analysis similar to that for $\Delta\Lambda/\Delta f$. These values were calculated for 32 Hz using $D=96$, and for the higher frequencies using $D=71$. This difference is necessary to make isospeed wavenumbers correspond more closely to wavenumbers in Table I, and is expected since the 32 Hz band penetrates to the deep reflector while the higher-frequency modes are more effectively contained in the water column. Values in Table X are significantly larger than those obtained for $\Delta\Lambda/\Delta f$, which is consistent with the conclusion from Sec. VI B that resonant modal interaction is relatively sensitive to waveguide and deep reflector depths.

TABLE X. Variability in $\Delta\Lambda_{ij}/\Delta D$ for modes i and j.

	$\Delta\Lambda_{12}$	$\Delta\Lambda_{13}$	$\Delta\Lambda_{23}$	$\Delta\Lambda_{24}$	$\Delta\Lambda_{34}$	$\Delta\Lambda_{15}$	$\Delta\Lambda_{35}$	$\Delta\Lambda_{46}$
32 Hz	16.4	5.5	8.3					
64 Hz	24.2	8.1	12.1	4.9	8.2	2.5	4.7	
95 Hz	35.9	12.0	18.0	7.2	12.0	3.6	5.1	4.1

VII. SUMMARY AND DISCUSSION

An environmental model for the NRL acoustic track is formulated using thermistor data at the receiver, CTD data at the source, and perturbations of AMCOR 6010 borehole data. Due to the apparent contamination in the source monitor signal, a model source signature is developed for broadband acoustic modeling. Nonlinear internal wave packets that were observed between the acoustic source and receiver during transmission are modeled using a train of sech^2 solitons. This model allows the soliton width and spacing between solitons to be varied for each wave in the packet.

It has been found by comparing broadband SWARM acoustic data to broadband PE computations that the passage of nonlinear internal waves evidently cause the pulse-averaged intensity fluctuations in airgun signals recorded at the NRL/VLA. Wavelet transforms are used to analyze time–frequency variations of acoustic data caused by nonlinear internal waves. The transforms reveal that acoustic mode strength and arrival time variability of peak frequency components are correlated with the passage of the nonlinear internal waves.

The acoustic frequency bands near 32, 64, and 95 Hz studied here are significantly lower frequencies than have been previously examined in connection with nonlinear internal waves. While the internal wave-acoustic resonance condition is well known, the opportunity to relate observed nonlinear internal wave spectral peaks to acoustic wavenumbers is apparently novel. Significant acoustic effects arise despite the relatively low number of propagating modes at the frequencies of interest. The influence of nonlinear internal wave width and spacing parameters on acoustic variability is described, and the internal wave spectrum can predict packet configurations that cause resonant behavior for particular acoustic frequency bands.

Several aspects of the acoustic results from the NRL track remain open for future investigation. Time–frequency behavior observed in data and simulation scalograms was not fully analyzed and would be a rich area of investigation. New applications of time–frequency analysis, including inversions, may be possible with an understanding of the nonlinear internal wave effects. Range-dependent inversions along the NRL acoustic track could be performed for bottom or water-column properties. The influence of uncertainty modeled by random internal wave parameters deserves more extensive treatment.

ACKNOWLEDGMENTS

The authors would like to thank Yongke Mu for assistance reading the acoustic data and Steve Forsythe for insightful discussions involving signal processing related to the source monitor. This work was supported by an ONR Ocean Acoustics Graduate Traineeship Award and by ONR grants to Rensselaer, the University of Delaware, and Woods Hole Oceanographic Institution. This is WHOI contribution No. 11257.

[1] J. Zhou, X. Zhang, and P. H. Rogers, ''Resonant interaction of sound waves with internal solitons in the coastal zone,'' J. Acoust. Soc. Am. **90**, 2042–2054 (1991).

[2] C. O. Tiemann, P. F. Worcester, and B. D. Cornuelle, ''Acoustic scattering by internal solitary waves in the Strait of Gibraltar,'' J. Acoust. Soc. Am. **109**, 143–154 (2001).

[3] D. Rubenstein and M. H. Brill, ''Acoustic variability due to internal waves and surface waves in shallow water,'' in *Ocean Variability and Acoustic Propagation*, edited by J. Potter and A. Warn-Varnas (Kluwer Academic, Boston, 1991), pp. 215–228.

[4] A. K. Liu, ''Analysis of nonlinear waves in the New York Bight,'' J. Geophys. Res. **93**, 12317–12329 (1988).

[5] J. F. Lynch, G. Jin, R. Pawlowicz, D. Ray, A. J. Plueddermann, C.-S. Chiu, J. H. Miller, R. H. Bourke, A. R. Parsons, and R. Muench, ''Acoustic travel-time perturbations due to shallow-water internal waves and internal tides in the Barents Sea Polar Front: Theory and experiment,'' J. Acoust. Soc. Am. **99**, 803–821 (1996).

[6] D. Rubenstein, ''Observations of cnoidal internal waves and their effect on acoustic propagation in shallow water,'' IEEE J. Ocean. Eng. **24**, 346–357 (1999).

[7] L. B. Dozier and F. D. Tappert, ''Statistics of normal mode amplitudes in a random ocean. I. Theory,'' J. Acoust. Soc. Am. **63**, 353–365 (1978).

[8] X. Tang and F. D. Tappert, ''Effects of internal waves on sound pulse propagation in the Straights of Florida,'' IEEE J. Ocean. Eng. **33**, 245–255 (1997).

[9] J. C. Preisig and T. F. Duda, ''Coupled acoustic mode propagation through continental-shelf internal solitary waves,'' IEEE J. Ocean. Eng. **22**, 256–269 (1997).

[10] T. F. Duda and J. C. Preisig, ''A modeling study of acoustic propagation through moving shallow water solitary wave packets,'' IEEE J. Ocean. Eng. **24**, 16–32 (1999).

[11] S. A. Ching-Bing, A. Warn-Varnas, D. B. King, K. G. Lamb, M. Teixeira, and J. A. Hawkins, ''Analysis of coupled oceanographic and acoustic soliton simulations in the Yellow Sea: a search for soliton-induced resonances,'' Math. Comput. Simul. **62**, 11–20 (2003).

[12] R. H. Headrick, J. F. Lynch, J. N. Kemp, A. E. Newhall, K. von der Heydt, J. R. Apel, M. Badiey, C.-S. Chiu, S. Finette, M. H. Orr, B. Pasewark, A. Turgut, S. N. Wolf, and D. Tielbuerger, ''Modeling mode arrivals in the 1995 SWARM experiment acoustic transmissions,'' J. Acoust. Soc. Am. **107**, 221–236 (2000).

[13] D. Rouseff, A. Turgut, S. N. Wolf, S. Finette, M. H. Orr, B. H. Pasewark, J. R. Apel, M. Badiey, C.-S. Chiu, R. H. Headrick, J. F. Lynch, J. N. Kemp, A. E. Newhall, K. von der Heydt, and D. Tielbuerger, ''Coherence of acoustic modes propagating through shallow water internal waves,'' J. Acoust. Soc. Am. **111**, 1655–1666 (2002).

[14] D. Tielbuerger, S. Finette, and S. N. Wolf, ''Acoustic propagation through an internal wave field in a shallow water waveguide,'' J. Acoust. Soc. Am. **101**, 789–808 (1997).

[15] J. F. Lynch, M. H. Orr, and S. N. Wolf, ''Low frequency acoustic propagation through shallow water internal waves,'' in *Sound Propagation Through Internal Waves* (unpublished).

[16] B. G. Katsnel'son and S. A. Pereselkov, ''Resonance effects in sound scattering by internal wave packets in a shallow sea,'' Acoust. Phys. **44**, 684–689 (1998).

[17] J. R. Apel, M. Badiey, C.-S. Chiu, S. Finette, R. H. Headrick, J. Kemp, J. F. Lynch, A. E. Newhall, M. H. Orr, B. H. Pasewark, D. Tielbuerger, A. Turgut, K. von der Heydt, and S. N. Wolf, ''An overview of the 1995 SWARM shallow-water internal wave acoustic scattering experiment,'' IEEE J. Ocean. Eng. **22**, 465–499 (1996).

[18] M. Badiey, Y. Mu, J. F. Lynch, J. R. Apel, and S. N. Wolf, ''Temporal and azimuthal dependence of sound propagation in shallow water with internal waves,'' IEEE J. Ocean. Eng. **27**, 117–129 (2002).

[19] F. S. Henyey and A. Hoering, ''Energetics of nonlinear internal waves,'' IEEE J. Ocean. Eng. **102**, 3323–3330 (1999).

[20] S. Mallat, *A Wavelet Tour of Signal Processing*, 2nd ed. (Academic, San Diego, 1999).

[21] M. Badiey, I. Jaya, and A. H.-D. Cheng, ''Shallow water acoustic/geoacoustic experiments at the New Jersey Atlantic Generating Station site,'' J. Acoust. Soc. Am. **96**, 3593–3604 (1994).

[22] G. R. Potty, J. H. Miller, J. F. Lynch, and K. B. Smith, ''Tomographic inversion for sediment parameters in shallow water,'' J. Acoust. Soc. Am. **108**, 973–986 (2000).

[23] I. Tolstoy and C. S. Clay, *Ocean Acoustics: Theory and Experiment in Underwater Sound* (American Institute of Physics, Woodbury, NY, 1966).

[24] M. D. Collins, ''RAMGEO 1.5,'' URL ftp://albacore.nrl.navy.mil/RAM.

[25] F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computa-*

tional Ocean Acoustics (Springer-Verlag, New York, 2000).
[26] S. Finette, M. H. Orr, A. Turgut, J. R. Apel, M. Badiey, C.-S. Chiu, R. H. Headrick, J. N. Kemp, J. F. Lynch, A. E. Newhall, K. von der Heydt, B. Pasewark, S. N. Wolf, and D. Tielbuerger, ''Acoustic field variability induced by time evolving internal wave fields,'' J. Acoust. Soc. Am. **108**, 957–972 (2000).
[27] R. Guenther, *Modern Optics* (Wiley, New York, 1990).
[28] T. C. Yang and K. Yoo, ''Internal wave spectrum in shallow water: Measurement and comparison with the Garrett-Munk model,'' IEEE J. Ocean. Eng. **24**, 333–345 (1999).
[29] B. J. Sperry, J. F. Lynch, G. Gawarkiewicz, C.-S. Chiu, and A. Newhall, ''Characteristics of acoustic propagation to the eastern vertical line array receiver during the summer 1996 New England shelfbreak PRIMER experiment,'' IEEE J. Ocean. Eng. **28**, 729–749 (2003).
[30] National Geophysical Data Center, NOAA, *Hydrographic Survey Data*, Vol. 1, Version 3.3.
[31] J. D. Milliman, A. Jiezao, L. Anchun, and J. I. Ewing, ''Late quaternary sedimentation on the outer and middle New Jersey Continental shelf: Result of two local deglaciations?'' J. Geol. **98**, 966–976 (1990).
[32] S. D. Stearns, *Digital Signal Analysis* (Hayden, Rochelle Park, NJ, 1975).
[33] R. J. Cederberg, W. L. Siegmann, and W. M. Carey, ''Influence of geoacoustic modeling on predictability of low-frequency propagation in range-dependent shallow-water environments,'' J. Acoust. Soc. Am. **97**, 2754–2766 (1995).
[34] I. Rozenfeld, W. M. Carey, P. G. Cable, and W. L. Siegmann, ''Modeling and analysis of sound transmission in the Strait of Korea,'' IEEE J. Ocean. Eng. **26**, 809–819 (2001).
[35] A. Turgut and S. N. Wolf, ''Matched-field inversion of seabed geoacoustic properties complemented by chirp sonar surveys,'' J. Acoust. Soc. Am. **110**, 2661(A) (2001).
[36] Z. Y. Zhang and C. T. Tindle, ''Improved equivalent fluid approximations for a low shear speed ocean bottom,'' J. Acoust. Soc. Am. **98**, 3391–3396 (1995).
[37] S. D. Frank, ''Modeling nonlinear internal wave effects on broadband shallow water acoustics,'' Ph.D. thesis, Rensselaer Polytechnic Institute, 2003.
[38] S. T. McDaniel and D. F. McCammon, ''Mode coupling and the environmental sensitivity of shallow-water propagation loss,'' J. Acoust. Soc. Am. **82**, 217–223 (1987).
[39] R. B. Evans, ''A coupled mode solution for acoustic propagation in a waveguide with stepwise depth variations of a penetrable bottom,'' J. Acoust. Soc. Am. **74**, 188–195 (1983).
[40] R. H. Headrick, J. F. Lynch, J. N. Kemp, A. E. Newhall, K. von der Heydt, J. R. Apel, M. Badiey, C.-S. Chiu, S. Finette, M. H. Orr, B. Pasewark, A. Turgut, S. N. Wolf, and D. Tielbuerger, ''Acoustic normal mode fluctuation statistics in the 1995 SWARM internal wave scattering experiment,'' J. Acoust. Soc. Am. **107**, 201–220 (2000).
[41] R. J. Cederberg, W. L. Siegmann, and M. J. Jacobson, ''Predictability of acoustic intensity and horizontal wave numbers in shallow water at low frequencies using parabolic approximations,'' J. Acoust. Soc. Am. **94**, 1034–1043 (1993).
[42] L. Baxter II and M. H. Orr, ''Fluctuations in sound transmission through internal waves associated with the thermocline: A computer model for acoustic transmission through sound velocity fields calculated from thermistor chain, CTD, XBT, and acoustic backscattering,'' J. Acoust. Soc. Am. **71**, 61–66 (1982).

Bubble dynamics and size distributions during focused ultrasound insonation

Xinmai Yang[a)]
National Center for Physical Acoustics, University of Mississippi, 1 Coliseum Drive, University, Mississippi 38677

Ronald A. Roy and R. Glynn Holt
Department of Aerospace and Mechanical Engineering, Boston University, Boston, Massachusetts 02215

(Received 24 April 2004; revised 4 October 2004; accepted 5 October 2004)

The deposition of ultrasonic energy in tissue can cause tissue damage due to local heating. For pressures above a critical threshold, cavitation will occur, inducing a much larger thermal energy deposition in a local region. The present work develops a nonlinear bubble dynamics model to numerically investigate bubble oscillations and bubble-enhanced heating during focused ultrasound (HIFU) insonation. The model is applied to calculate two threshold-dependent phenomena occurring for nonlinearly oscillating bubbles: Shape instability and growth by rectified diffusion. These instabilities in turn are shown to place physical boundaries on the time-dependent bubble size distribution, and thus the thermal energy deposition. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1823251]

PACS numbers: 43.35.Wa, 43.80.Gx, 43.80.Sh [AJS] Pages: 3423–3431

I. INTRODUCTION

Research on the effects of therapeutic ultrasound dates back at least 70 years. Biological, physical, and chemical effects due to high intensity ultrasound, sometimes coupled with cavitation activity, have been observed by several researchers who essentially established the field of therapeutic ultrasound.[1–3] In addition to causing mechanical damage, cavitation has been observed to affect and, under the right conditions, enhance heat deposition from ultrasound.[4–9]

The preceding list of work on bubble-enhanced heating was reviewed in Holt and Roy,[10] who also observed bubble-enhanced heating in an agar-based tissue-mimicking phantom from focused MHz-frequency ultrasound. Their findings were characterized by a threshold dependence and enhanced, but often erratic heating above the threshold. They also modeled the two dominant mechanisms for bubble-enhanced heating: Viscous dissipation from bubble oscillations and absorption of the sound emitted by oscillating bubbles. Quantitative estimates for thermal power deposition from a single oscillating bubble were on the order of 10 mW, compared to ~1 W from absorption of the incident ultrasound waves. Though their findings established quantitatively that heat deposition from cavitation could account for their enhanced heating measurements, neither their model nor their experimental results (nor other groups results) could provide any guidance as to the bubble sizes responsible for the enhanced heating. We will return to this point.

Edson[11] extended the work of Holt and Roy by adding a passive cavitation detector (PCD)[12] to monitor the presence of bubbles. His experimental results showed that the PCD detected (by detecting broadband acoustic emissions above 2 MHz with a primary field of 1 MHz) inertial cavitation coincident with enhanced heating. In addition, he calculated the bubble-related power deposition in a reasonable parameter space, and, by estimating the number of bubbles in the gel to be on the order of 10–100, he could account for the observed cavitation-enhanced temperature rise.

In recent work, Bailey *et al.*[13] studied the effect of overpressure on focused ultrasound lesion shapes *in vitro*. They found at high intensity levels, lesions formed with no overpressure, were tadpole shaped, and shifted into the prefocal region, but lesions formed with high overpressure remained cigar shaped. They concluded this result supported the hypothesis that bubbles contributed to the tadpole-shaped lesion distribution.

An *in vivo* experimental investigation of bubble-enhanced heating is reported by Sokka *et al.*[14] These authors monitored lesion formation in rabbit thigh while monitoring temperature via magnetic resonance imaging (MRI). They observed a faster temperature rise within the first few seconds and higher overall temperature in the bubble-enhanced exposures than the sonications without bubble formation, consistent with the observations in Ref. 10. The lesions created with the gas-bubble-enhanced heating exposures were two to three times larger by volume, consistently more spherical in shape and closer to the transducer than the control exposures, in accordance with the observations reported in Ref. 13.

Hilgenfeldt *et al.*[15,16] modeled the energy deposited from bubbles of μm size and driven by diagnostic ultrasound pulses. They considered the effect of viscous dissipation and acoustic emissions from bubbles allowed to oscillate nonlinearly. They reported that the strong absorption of the acoustic radiation from nonlinearly oscillating bubbles results in a temperature rise as high as 100 K for driving pressure around 3 MPa at a distance of 1 μm from the bubble surface. Chavrier *et al.*[17] also investigated the emissions from nonlinearly oscillating bubbles and the resulting absorption of the

[a)]Author to whom correspondence should be addressed; electronic mail: xmyang@olemiss.edu

sound; a ''global attenuation coefficient'' is used in their energy equation to account for the presence of the bubble-enhanced heating owing to bubble acoustic emissions. They used their model to predict prefocal lesion growth when the bubble density exceeds 500 bubbles per cubic mm. Both of these numerical models employed an ideal gas assumption with a uniform temperature for the bubble interior dynamics. Neither of these models considered the effects of varying viscosity. Finally, neither of these models includes bubble shape instabilities or growth/dissolution via gaseous mass diffusion.

Though not the subject of the present paper, we note that our results here will be generally applicable to the occurrence of cavitation-related bioeffects other than enhanced heating. Thus, we note here that acoustically induced cavitation has been shown to play an important role in hemolysis,[18–21] hemostasis,[22–24] and drug transport,[25] especially in the context of thrombolysis.[26–31]

Though the importance of cavitation bubbles in the context of therapeutic ultrasound is thus well established, any attempt to quantitatively account for the contributions of bubbles to the heating (or any other effect) must rely on an inference of the bubble size distribution, as bubble dynamics at a fixed driving frequency is very sensitive to the bubble equilibrium radius. Since such a distribution is difficult to measure, we turn our focus toward calculating the bubble size distribution by incorporating shape and growth instabilities into a model for bubbles in therapeutic ultrasound. Taken together, these two threshold-dependent instabilities will define an allowed bubble size distribution.

II. BUBBLE DYNAMICS MODEL

A comprehensive review of bubble dynamics may be found in *The Acoustic Bubble*.[32] In what follows, a Newtonian fluid bubble dynamics model is assumed. We integrate the Newtonian model over a large viscosity range, from 0.001 Pa s to 0.5 Pa s. The choice of a large range of viscosity is to supplement this Newtonian fluid assumption; in reality we may expect tissue to behave as a viscoelastic material, and one of the features of such materials is shear thinning.

A. Radial motion equation

Throughout this work, the radial equation of motion we use is the so-called Keller–Miksis equation which was first derived by Keller and Miksis,[33,34] and has the form,

$$\left(1-\frac{\dot{R}}{c_L}\right)R\ddot{R}+\frac{3}{2}\left(1-\frac{1}{3}\frac{\dot{R}}{c_L}\right)\dot{R}^2 = \frac{1}{\rho}\left(1+\frac{\dot{R}}{c_L}+\frac{R}{c_L}\frac{d}{dt}\right)p_B - \frac{1}{\rho}\left(1+\frac{\dot{R}}{c_L}\right)\left(p_\infty - p_s\left(t+\frac{R}{c_L}\right)\right), \tag{1}$$

where dots denote time derivatives, R is the bubble radius, c_L is the sound speed in the surrounding medium, ρ is the density of surrounding medium, p_∞ is the ambient pressure, $p_s(t)$ is the applied acoustic field pressure, and p_B is the pressure on the surrounding medium side of the interface between the medium and the bubble. p_B is related to the internal pressure p of bubble by the following formula:

$$p = p_B + \frac{2\sigma}{R} + 4\mu\frac{\dot{R}}{R}, \tag{2}$$

where σ is the surface tension coefficient, and μ is the shear dynamic viscosity. The sound field we chose in this study is sinusoidal:

$$p_s(t) = p_a \cos \omega t, \tag{3}$$

where p_a is the acoustic amplitude and ω is the driving angular frequency. Most therapeutic applications of ultrasound rely on rather long tone bursts, approximating continuous-wave conditions. Our model is not limited to sinusoidal forcing, but experimental and numerical results show that with our experimental material parameters, sound fields in tissue and tissue-mimicking media are very weakly nonlinear, therefore the harmonics of ω are unimportant for the bubble dynamics for the current applications.

B. Internal pressure

To solve the above radial equation, the internal pressure of bubble is needed. A model for the internal temperature and pressure of the bubble is described in detail in Prosperetti *et al.*[35] and Kamath *et al.*[36] This model accounts for heat transport by convection and conduction to and from the surrounding medium. The model was derived by assuming an ideal gas and spatial uniformity of the gas pressure inside the bubble. The internal pressure p can be found by integrating:

$$\dot{p} = \frac{3}{R}\left[(\gamma-1)K\left.\frac{\partial T}{\partial r}\right|_R - \gamma p\dot{R}\right], \tag{4}$$

where γ is the ratio of the specific heats of the gas and K is the gas thermal conductivity. The gas temperature field $T(r,t)$ is obtained from

$$\frac{\gamma}{\gamma-1}\frac{p}{T}\left(\frac{\partial T}{\partial t}+v\frac{\partial T}{\partial r}\right)=\dot{p}+\frac{1}{r^2}\frac{\partial}{\partial r}\left(\kappa r^2\frac{\partial T}{\partial r}\right) \tag{5}$$

with

$$v=\frac{1}{\gamma p}\left[(\gamma-1)K\frac{\partial T}{\partial r}-\frac{1}{3}r\dot{p}\right], \tag{6}$$

where v is the radial velocity of the gas inside of the bubble. The temperature boundary condition for a relatively cold liquid, such as water at normal ambient temperature, is,

$$T[r=R(t),t]=T_\infty, \tag{7}$$

where T_∞ is the undisturbed liquid temperature.

The solution of this set of equations of motion (1)–(7) requires numerical integration. Kamath and Prosperetti developed a Galerkin spectral method which we employ in this work. The ODE solver ODE15s in MATLAB is used to solve the resulting set of ODEs.

C. Stability of a spherical bubble

An oscillating spherical bubble can also develop shape oscillations. An instability to wavy surface perturbations will cause shape oscillations of a spherical bubble. Plesset,[37] Eller and Crum,[38] Prosperetti,[39] Hilgenfeldt *et al.*,[40] and Hao and Prosperetti[41] have modeled the stability of a spherical surface using spherical harmonics to describe the perturbations. Consider a small axisymmetric distortion of the spherical interface $R(t)$,

$$r = R(t) + a_n(t) P_n(\theta), \tag{8}$$

where P_n is the Legendre polynomial of degree n and $a_n(t)$ is the amplitude. The shape distortion consists of a superposition of surface oscillations whose amplitudes are given by $a_n(t)$. The problem is to determine how the $a_n(t)$ behave.

Holt and Gaitan[42] and Gaitan and Holt[43] have reported experimental observations of the onset of shape instability for a bubble in water. A comparison of those experimental results with predictions (Hao and Prosperetti[41] and Prosperetti and Hao[44]) shows quantitative agreement regarding mode n observed and dependence on bubble size for all models of bubble oscillation. For the radial model Eq. (1), and with the following approximations (9)–(12), Hao and Prosperetti[41] document that good quantitative agreement can be expected in the single bubble sonoluminescence range, corresponding to inertially collapsing bubbles where the driving frequency is around 20 kHz and driving pressure is 0.8–1.4 atm.

The equation governing the time evolution of the $a_n(t)$ is

$$\ddot{a}_n + B_n(t)\dot{a}_n - A_n(t) a_n = 0 \tag{9}$$

with

$$A_n(t) = (n-1)\frac{\ddot{R}}{R} - \frac{\beta_n \sigma}{\rho R^3} + \frac{2\nu\dot{R}}{R^3}\left[-\beta_n + n(n+2)(n-1)\frac{1}{1+2\delta/R}\right] \tag{10}$$

$$B_n(t) = \frac{3\dot{R}}{R} + \frac{2\nu}{R^2}\left[-\beta_n + \frac{n(n+2)^2}{1+2\delta/R}\right], \tag{11}$$

where $\beta_n = (n-1)(n+1)(n+2)$, ν is the kinematic viscosity, a_n is the amplitude of the nth surface mode, and the boundary layer thickness $\delta = \min(\sqrt{\nu/\omega}, R/2n)$ (Prosperetti,[45] Brenner *et al.*,[46] and Hilgenfeldt *et al.*[40]) The ordinary differential Eq. (9) can then be solved numerically using numerical values of R, $\dot{R}$, and $\ddot{R}$ from Eq. (1). To finally determine the threshold for breakup plotted in the figures, the criterion $\max(|a_n(t)|/R(t)) \geqslant 1$ is used for breakup via shape instability.

D. Rectified diffusion

The shape instability threshold is not enough to yield the bubble size. For a shape-stable bubble, its equilibrium radius may grow, shrink, or in certain rather rare cases remain stable because of rectified diffusion (Eller and Flynn).[47] To determine the growth or dissolution of bubbles we integrate the following equation (Fyrillas and Szeri):[48]

$$\frac{dR_0}{dt} = R_0\omega_0 \frac{R_g T_\infty \rho}{M p_\infty}\frac{1}{Pe}\left(1 + \frac{4\sigma}{3R_0 p_\infty}\right)^{-1} \times \frac{(c_i - c_0\langle p_G^*(\tau)\rangle_{\hat{\tau}})}{\int_0^\infty \frac{dU}{\langle (3U + x^3(\tau))^{4/3}\rangle_{\hat{\tau}}}} \tag{12}$$

where R_g is the universal gas constant, R_0 is the equilibrium radius of bubble, ω_0 is the natural frequency of the bubble, T_∞ is the temperature of the surrounding liquid, ρ is the density of liquid, M is the molecule weight of gas inside the bubble, p_∞ is the ambient pressure, P_e is the *Péclet* number, σ is the surface tension, c_i is the dissolved gas concentration, c_0 is the saturation concentration in the liquid, p_G^* is the dimensionless pressure inside the bubble, τ is the dimensionless time defined as $t = \tau/\omega_0$, x is the dimensionless bubble radius defined as $R = xR_0$, and U is the dimensionless radial coordinate in the liquid, and angled brackets denote a fourth-order time average. The validity of this method depends on the *Péclet* number being large, which can be easily satisfied in most cases. The rectified diffusion threshold is given when the growth rate is zero

$$c_i - c_0\langle p_G^*(\tau)\rangle_{\hat{\tau}} = 0, \tag{13}$$

which is identical to Eller and Flynn's result.[47]

III. RESULTS

We chose the range of viscosity from 0.001 Pa s to 0.5 Pa s for simulations, which covered most of the possible viscosity coefficients encountered in human soft tissue. We investigate a bubble size range from 0.01 μm to 50 μm. The driving frequency used in all the simulations is 1 MHz. Other parameters used in the simulation are $c_L = 1600$ m/s, $\sigma = 0.0725$ N/m, $T_\infty = 20\,^\circ$C, $p_\infty = 1$ atm, and $\rho = 1100$ kg/m^3. These values correspond to the properties of tissue-mimicking phantoms used in our experiments.[10,11,49,50] For more details on the calculations and other results, the reader is referred to Ref. 50.

A. Stability boundaries

Figures 1–3 show overviews of the bubble dynamics for acoustic pressures $p_a = 1.0$ MPa and 2.5 MPa, at an excitation frequency of 1 MHz (for more details and calculations the reader is referred to Yang).[50] The threshold for breakup via shape instability (labeled SI, calculated for the most unstable mode at each bubble size and viscosity), the threshold for growth by rectified diffusion (labeled RD, calculated for a variety of dissolved gas concentrations), and the threshold for inertial cavitation (labeled IC, defined conventionally as an expansion of the bubble to a size greater than twice its equilibrium radius)[51,52] are all plotted on the same graph in order to facilitate identification of bubble behaviors.

Figure 1 plots all three thresholds for an acoustic pressure of 1.0 MPa at 1 MHz for a gas-saturated medium. Once nucleated, the only bubbles which will survive are those

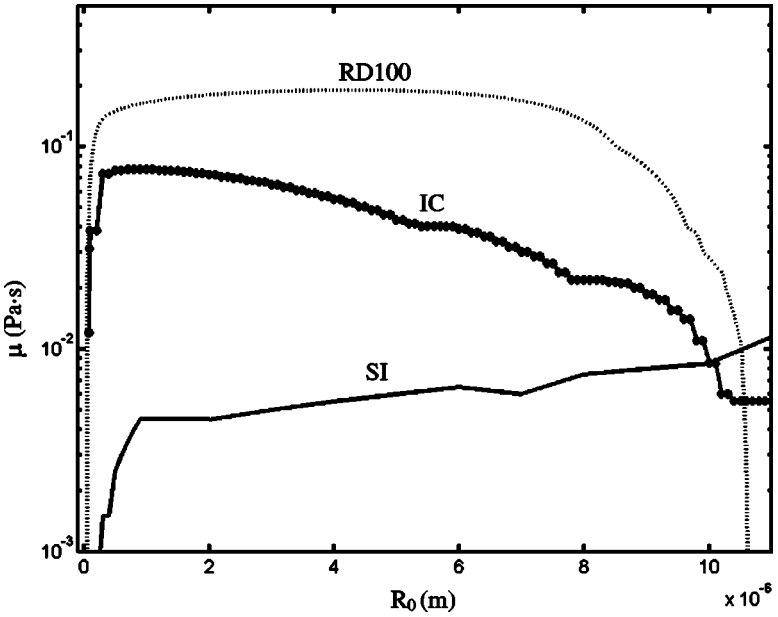


FIG. 1. $P_a = 1.0$ MPa, $f = 1$ MHz, and 100% gas saturation. Rectified diffusion (RD100, dashed line), shape instability (SI, thin solid line), and inertial cavitation (IC, thick solid line with stars) thresholds. Below RD bubbles grow. Below IC bubbles collapse inertially. Below SI bubbles break up from shape instabilities.

which are diffusively stable (on the RD threshold), or those which grow. In this space, a positive slope on the RD threshold line represents an unstable equilibrium, and a negative slope is stable. Bubbles below the RD threshold grow. Breakup via SIs bounds the region of surviving bubbles—all bubbles below the SI threshold line exceed the criterion for breakup via SI. Finally, since the experimental evidence of coincidence of broadband noise emissions detected by PCD always accompanies the observation of the onset of enhanced heating, we conclude that some fraction of the cavitation bubbles are collapsing inertially. The IC threshold line represents a minimal threshold for inertial collapse, and thus represents a boundary for the initial stages of cavitation activity. Thus, there is a wedge-shaped region in the parameter space for allowed bubble sizes, which in fact encompasses bubbles between 0.02 and 10 μm.

Figure 2 plots all three thresholds for an acoustic pressure of 2.5 MPa at 1 MHz for a gas-saturated medium. The situation is somewhat more complicated. The shape stable and shape unstable regions for this higher pressure are no longer simply connected; note the islands of instability for small bubbles between viscosity values 0.02 and 0.005 Pa s. The size of the allowed region has increased, and admits bubbles with radii as small as 10 nm and as large as 17 μm.

Figure 3 shows the effect of changing dissolved gas concentration. When the gas concentration decreases from 100% to 10%, the threshold does not shift very much, but from 10% to 0.1%, there is a large difference between the calculated thresholds, with the growth region dramatically shrinking for decreasing dissolved gas concentration. One expects to find nearly saturated media in the body. However, for violent enough bubble collapse, sonoluminescence results[42,40] have shown that only the noble gases participate in diffusion; N_2 and O_2 are dissociated during collapse and chemically bonded in water. This effect would expand the potential range of effective dissolved gas concentrations for

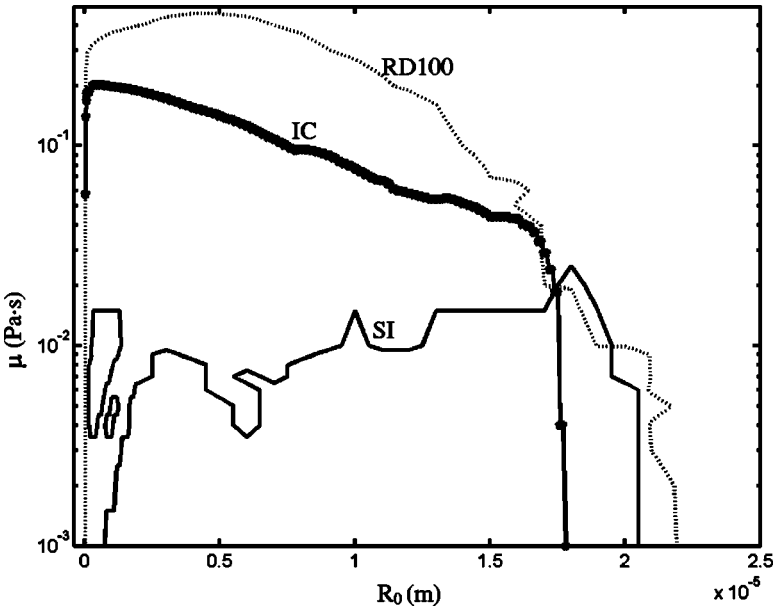


FIG. 2. $P_a = 2.5$ MPa, $f = 1$ MHz, and 100% gas saturation. Rectified diffusion (RD100, dashed line), shape instability (SI, thin solid line), and inertial cavitation (IC, thick solid line with stars) thresholds at $P_a = 1$ MPa, $f = 1$ MHz. Below RD bubbles grow. Below IC bubbles collapse inertially. Below and enclosed by SI bubbles break up from shape instabilities.

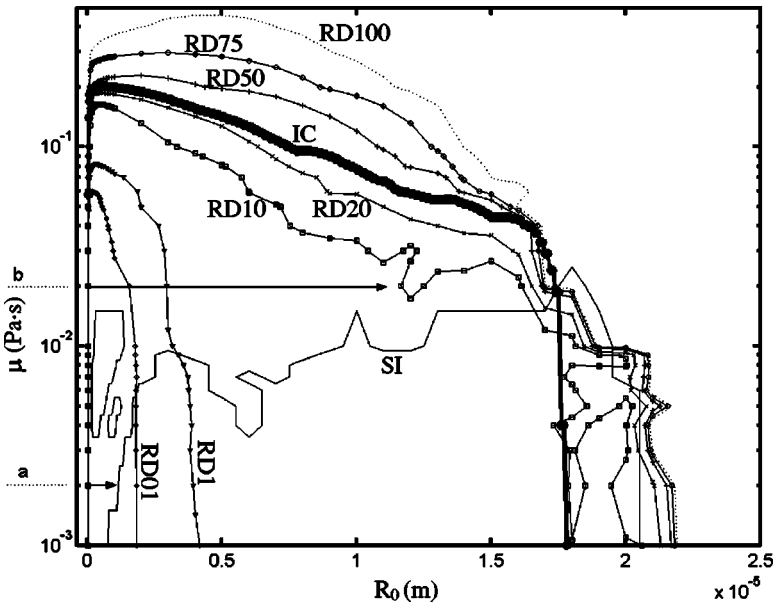


FIG. 3. $P_a = 2.5$ MPa, $f = 1$ MHz, and variable dissolved gas concentration. The inertial cavitation (IC) and shape instability (SI) thresholds are as in Fig. 2. All other lines are the RD thresholds for dissolved gas concentrations of 100% (RD100), 75% (RD75), 50% (RD50), 20% (RD20), 10% (RD10), 1% (RD1), and 0.1% (RD01) relative to saturation. The two arrows show two paths for qualitatively different bubble behavior. For these two paths, the viscosities are (a) 0.002 Pa s; and (b) 0.02 Pa s. The total growth times for each path when the dissolved gas concentration is 10% are (a) 1.38*e*-5 s; and (b) 1.57 s. The initial radius for both paths is $R_0(t=0) = 50$ nm.

cavitation zone bubble dynamics to the well-undersaturated values calculated in Fig. 3. Such an effect, if it plays a role, would thus further limit the allowed bubble sizes to smaller bubbles. Choosing the 0.1% case for a most dramatic case illustration, while the lower limit of allowed bubble sizes remains virtually unchanged, the upper limit decreases to (for any viscosity value) less than 2.5 μm, meaning that in this case all allowed bubbles are below resonant size at 1 MHz. Such a qualitative difference (from the saturated case) would profoundly affect the effective sound speed in the cavitation zone and thus would affect the primary HIFU propagation.

Figure 3 also shows the paths of two bubbles with initial radius of 50 nm in a 10% saturated medium which have two very different eventual fates. For the path labeled ''a'', we have chosen a viscosity of twice that of water. This bubble grows very quickly and in roughly 14 cycles becomes shape unstable. Such a bubble will break up, and likely one or more fragments will repeat this process of growth and breakup. For the path labeled ''b'', we have chosen a viscosity of 20 times that of water. This bubble grows more slowly, and eventually (after roughly 2 s) encounters the stable branch of the rectified diffusion threshold, where it will remain since it is shape stable.

B. Bubble size evolution

Based on the wide variety of bubble dynamics behaviors (from inertial collapse to sinusoidal oscillations, from periodic to chaotic motion) encountered in this parameter space, one would expect and in fact observes in computation, a large change in growth rates for different bubble sizes. Figure 4 plots growth rates for growing bubbles at 3 MPa and 100% saturation. Though the growth is rather modest for higher viscosity and larger bubbles sizes, the effect of inertial bubble dynamics (expansion of up to 80 times the initial

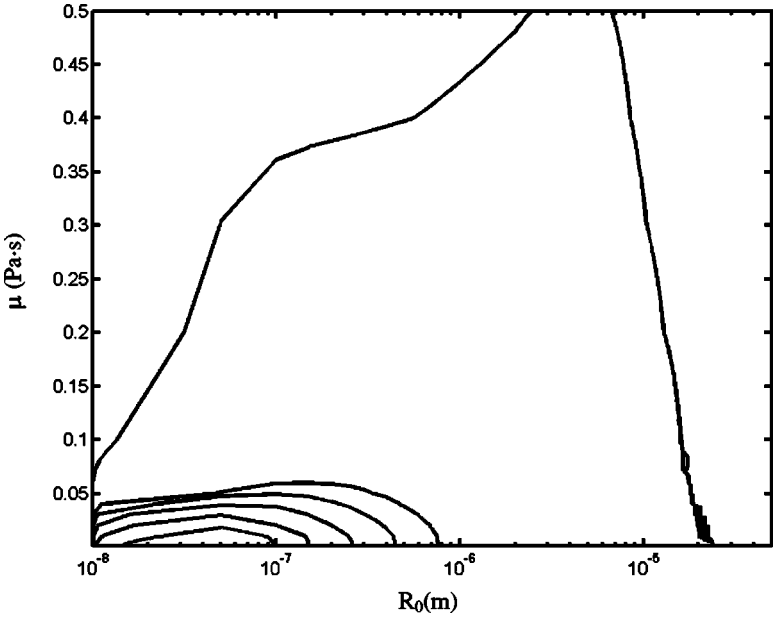


FIG. 4. Contour plot of bubble growth rates at 3 MPa, 1 MHz, and dissolved gas concentration 100%. Solid lines correspond to growth rates of, from top to bottom, 0, 1, 10, 100, 1000, and 10 000 m/s.

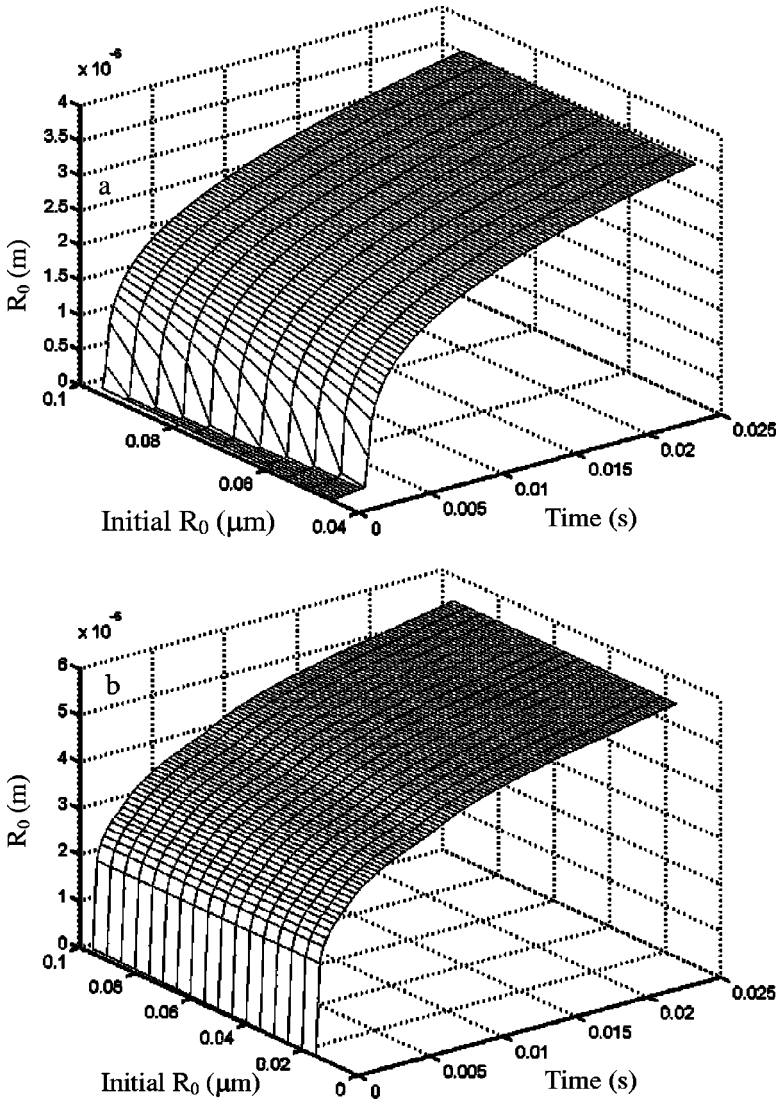


FIG. 5. Bubble equilibrium radius vs time for bubbles with initial radii ranging from 15 nm to 95 nm at 2 MPa for different viscosities: (a) 0.05 Pa s and (b) 0.005 Pa s. Dissolved gas concentration 20%.

radius followed by violent collapse in a single cycle) is clearly seen in the bunching of the growth rate contours for small (10 to 100 nm) bubbles and small (0.001 to 0.02 Pa s) viscosities.

The question arises as to whether or not this disparity of growth rates has implications for the temporal evolution of bubble size distributions. To investigate this question, we followed the growth of the bubble equilibrium radius for an array of initial bubble equilibrium sizes from 15 nm to 100 nm undergoing a 2 MPa insonation in a 20% saturated medium. A ''rectangular'' distribution [full width at half maximum (FWHM)$\sim$O(1)] corresponding to 1 bubble at each size in the range was used, and all initial R_0 are above the growth threshold. The results are shown in Fig. 5. In Fig. 5(a), when the viscosity is 0.05 Pa s, the first few time slices reveal that, owing to the different growth rates experienced by the small and large initial bubbles, a bimodal distribution has developed. However, soon thereafter (by 5 ms) the initially slow-growing bubbles have caught up, so that by the end of the 25 ms interval the profile is very nearly flat again. The mean radius is now, however, more than an order of magnitude greater at roughly 3.5 μm, so the FWHM at the end of the 25 ms interval is O(10^{-1}).

Figure 5(b) plots the same evolution when the viscosity is an order of magnitude lower (now five times that of water). The flat distribution evolves smoothly, and (though not seen on this plot) the differences between bubble equilibrium radii actually decrease from their initial value of $\Delta R_0 = 5$ nm during the 25 ms insonation, resulting in a FWHM at the end of 25 ms that is O(10^{-2}). Thus, for this range of viscosity values, an initial distribution of practically any character would narrow as the insonation progressed, so that any disperse distribution would be smoothed out by the growth dynamics into a nearly monodisperse distribution.

IV. DISCUSSION: WHICH BUBBLES WILL CONTRIBUTE?

The main purpose of this study was to investigate bubble dynamics in a large parameter space relevant to HIFU, and if possible, bound the bubble size distribution that contributes to the enhanced heating observed in experiments. Only bubbles that do not dissolve and that are shape stable will contribute to the heating. Bubbles which contribute to heat generation will thus be trapped in the wedge-shaped region between the two instability thresholds, perhaps further con-

strained by the IC threshold though our experimental knowledge is too incomplete to allow a stronger statement.

Two possible asymptotic behaviors naturally divide this allowed bubble region into two areas. In the lower area, characterized by viscosities on the order of water and illustrated by the path marked ''a'' in Fig. 3, bubbles will grow (rapidly) until they encounter the shape instability threshold and break up, whereupon a fragment may repeat the process. Such bubbles expand during one cycle to roughly 20 or more times their initial size and collapse inertially, breaking up typically via the Rayleigh–Taylor mechanism. To make any quantitative statement about the bubble sizes thus allowed one has to specify the acoustic pressure and frequency. For example, for 2.5 MPa and 1 MHz HIFU in a material whose viscosity coefficient is between 0.001 Pa s (water) and 0.008 Pa s (whole blood at low shear rates), we may expect bubbles to be distributed (unevenly, perhaps) between 10 nm and roughly 1 μm.

For higher values of viscosity in the upper area, illustrated by the path marked ''b'', bubbles will grow until they become trapped on the stable (since it has a negative slope) rectified diffusion threshold, and these bubbles will stably oscillate at that size. In this regime, the dissolved gas concentration becomes the dominant factor in determining the upper size limit. Once again one has to specify parameters to infer a bubble size distribution. For 2.5 MPa and 1 MHz HIFU in a material whose apparent viscosity coefficient is between 0.02 Pa s (synovial fluid at moderate shear rates) and 0.2 Pa s (SAE 20 oil), we may expect bubbles to be distributed between a common lower bound and an upper bound determined by the dissolved gas concentration. The lower the dissolved gas concentration, the lower the upper bound of bubble sizes. To cite two extreme cases, for 0.1% saturation, the bubble size distribution will range from about 10 nm to only 2 μm. For 100% saturation, bubbles could grow as large as 18 μm, many times the (linear) resonant size.

The disconnected shape instability regions in Fig. 3 are caused by the strong nonlinearity of the oscillating bubbles. In all simulations, we use the same initial conditions: $R(0)=R_0$ and $\dot{R}(0)=0$. For these strong nonlinear oscillating bubbles, different initial conditions may yield different radial motions; i.e., it is very likely that there are multiple coexisting steady-state solutions (or attractors).[53,54] Such solutions are only elicited (at fixed parameter values) if different initial conditions are used. Thus, one would expect small shifts in the calculated shape stability threshold for calculations imposing different initial conditions. Such shifts will typically be small compared to any experimental uncertainties in parameters and measurements.

The calculated growth rates for small bubbles just above the Blake threshold are dramatically high, implying that such bubbles survive only a few cycles. However, the calculation of the growth rate has used the equilibrium value of bubble oscillation, averaged over a cycle. These facts suggest that a steady-state time-average calculation cannot be applied to these small bubbles without large errors, because the fast growth ensures large departures from equilibrium. Though beyond the scope of this work, a more accurate treatment of these bubbles will be found by relaxing the implicit assumption that the smooth part of the mass diffusion problem is dominant (see Refs. 48 and 50 for details). For the larger viscosity regions of parameter space, the contribution from these small bubbles will be small since bubbles will grow and pass that region in a matter of a few cycles. For the small viscosity regions, whether a bubble survives one cycle or ten will also be unimportant as long as the problems of interest require hundreds or more cycles of oscillation. Thus, these calculations should be valid for therapeutic situations providing the ultrasound burst is at least 100 cycles or 100 μs at 1 MHz.

We used a Newtonian model to investigate the bubble dynamics. We know however that real tissue materials are viscoelastic. We investigated a large range of viscosity to try to cover many of the possibilities. However, how good (or bad) this approach is needs to be determined, and this requires a numerically tractable non-Newtonian bubble model.[55,56] Another practical difficulty related to this problem is that the viscoelastic material properties of tissue, in general, are difficult to obtain, and moreover are strain-type and strain-rate dependent. Bubbles present a dilatational strain, and nonlinearly varying strain rates (over a single cycle of motion in many cases). Ultimately, there is a need for an explicit experimental investigation of bubble dynamics in viscoelastic media, including tissue.

V. CONCLUSION

These bubble dynamics results over a broad parameter range allow an experimenter to narrow down the possibilities for bubble sizes as long as some information about the parameters (P_a, f, μ, and c_i/c_0) of a given experiment is available. The results may then be applied to calculate the bubbles contribution to the heat deposition during therapeutic ultrasound treatment. For example, in the series of tissue-mimicking experiments,[10,11,57] the investigators found that, by applying these results to the interpretation of their experiments, they could infer that only something on the order of 50 appropriately sized bubbles were required to account for the observed peak enhanced heating. Moreover, since the models as presented here are complete (they capture the dominant physical phenomena of nontranslating bubbles in a sound field) and numerically practical, they may be adopted as tools for investigation of bubble effects (including heating) for other situations requiring predictive characterization of ultrasound applications.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of NASA, DARPA (via a subcontract to the University of Washington Applied Physics Laboratories), and the U.S. Army (via a subcontract to the University of Mississippi National Center for Physical Acoustics). The suggestions and support of the PACLAB group at BU are also gratefully acknowledged. This research was supported through award No. DAMD17-02-2-0014, administered by the U.S. Army Medical Research Acquisition Activity, Fort Detrick, MD. The

content of this paper does not necessarily reflect the position of the policy of the Government, and no official endorsement should be inferred.

[1] E. N. Harvey, ''Biological aspects of ultrasonic waves, a general survey,'' Biol. Bull. **59**, 306–325 (1930).
[2] T. F. Hueter, H. T. Ballantine, and W. C. Cotter, ''Production of lesions in the central nervous system with focused ultrasound: A study of dosage factors,'' J. Acoust. Soc. Am. **28**, 192–201 (1956).
[3] F. J. Fry, R. F. Heimburger, L. V. Gibbons, and R. C. Eggleton, ''Ultrasound for visualization and modification of brain tissue,'' IEEE Trans. Sonics Ultrason. **SU-17**, 165–169 (1970).
[4] P. P. Lele, ''Effects of ultrasound on solid mammalian tissues and tumors in vivo,'' *Ultrasound: Medical Applications, Biological Effects and Hazard Potential* (Plenum, New York, 1987), pp. 275–306.
[5] K. Hynynen, ''The threshold for thermally significant cavitation in dog's thigh muscle in vivo,'' Ultrasound Med. Biol. **17**, 157–169 (1991).
[6] D. L. Miller and R. A. Gies, ''The interaction of ultrasonic heating and cavitation in vascular bioeffects on mouse intestine,'' Ultrasound Med. Biol. **24**, 123–128 (1998).
[7] S. Vaezy, X. Shi, R. W. Martin, E. Chi, P. I. Nelson, M. R. Bailey, and L. A. Crum, ''Real-time visualization of focused ultrasound therapy,'' Ultrasound Med. Biol. **27**, 33–42 (2001).
[8] N. I. Vykhodtseva, K. Hynynen, and C. Damianou, ''Histologic effects of high intensity pulsed ultrasound exposure with subharmonic emission in rabbit brain in vivo,'' Ultrasound Med. Biol. **21**, 969–979 (1995).
[9] G. R. ter Haar and S. Daniels, ''Evidence for ultrasonically induced cavitation *in vivo*,'' Phys. Med. Biol. **26**, 1145–1149 (1981).
[10] R. G. Holt and R. A. Roy, ''Measurements of bubble-enhanced heating from focused, MHz-frequency ultrasound in a tissue-mimicking material,'' Ultrasound Med. Biol. **27**, 1399–1412 (2001).
[11] P. L. Edson, ''The role of acoustic cavitation in enhanced ultrasound-induced heating in a tissue-mimicking phantom,'' Ph.D. dissertation, Boston University, 2001.
[12] R. A. Roy, S. I. Madanshetty, and R. E. Apfel, ''An acoustic backscattering technique for the detection of transient cavitation produced by microsecond pulses of ultrasound,'' J. Acoust. Soc. Am. **87**, 2451–2458 (1990).
[13] M. R. Bailey, L. N. Couret, O. A. Spaozhnikov, V. A. Khokhlova, G. R. ter Haar, S. Vaezy, X. Shi, R. Martin, and L. A. Crum, ''Use of overpressure to assess the role of bubbles in focused ultrasound lesion shape in vitro,'' Ultrasound Med. Biol. **27**, 695–708 (2001).
[14] S. D. Sokka, R. King, and K. Hynynen, ''MRI-guided gas bubble enhanced ultrasound heating in *in vivo* rabbit thigh,'' Phys. Med. Biol. **48**, 223–241 (2003).
[15] S. Hilgenfeldt, D. Lohse, and M. Zomack, ''Sound scattering and localized heat deposition of pulse-driven microbubbles,'' J. Acoust. Soc. Am. **107**, 3530–3539 (2000).
[16] S. Hilgenfeldt, D. Lohse, and M. Zomack, ''Response of bubbles to diagnostic ultrasound: A unifying theoretical approach,'' Eur. Phys. J. B **4**, 247–255 (1998).
[17] F. Chavrier, J. Y. Chapelon, A. Gelet, and D. Cathignol, ''Modeling of high-intensity focused ultrasound-induced lesions in the presence of cavitation bubbles,'' J. Acoust. Soc. Am. **108**, 432–440 (2000).
[18] A. A. Brayman, M. Azadniv, C. Cox, and M. W. Miller, ''Hemolysis of albunex-supplemented, 40% hematocrit human erythrocytes in vitro by 1 MHz pulsed ultrasound: Acoustic pressure and pulse length dependence,'' Ultrasound Med. Biol. **22**, 927–938 (1996).
[19] S. L. Poliachik, W. L. Chandler, P. D. Mourad, M. R. Bailey, S. Bloch, R. O. Cleveland, P. Kaczkowski, G. Keilman, T. Porter, and L. A. Crum, ''Effect of high-intensity focused ultrasound on whole blood with and without microbubble contrast agent,'' Ultrasound Med. Biol. **25**, 991–998 (1999).
[20] D. L. Miller and R. A. Gies, ''Enhancement of ultrasonically-induced hemolysis by perfluorocarbon-based compared to air-based echo-contrast agents,'' Ultrasound Med. Biol. **24**, 285–292 (1998).
[21] M. W. Miller, E. C. Everbach, C. Cox, R. R. Knapp, A. A. Brayman, and T. A. Sherman, ''A comparison of the hemolytic potential of optison and albunex in whole human blood in vitro: Acoustic pressure, ultrasound frequency, donor and passive cavitation detection considerations,'' Ultrasound Med. Biol. **27**, 709–721 (2001).
[22] S. Vaezy, R. Martin, U. Schmiedl, M. Caps, S. Taylor, K. Beach, S. Carter, P. Kaczkowski, G. Keilman, S. Helton, W. Chandler, P. Mourad, M. Rice, R. A. Roy, and L. A. Crum, ''Liver hemostasis using high-intensity focused ultrasound,'' Ultrasound Med. Biol. **23**, 1413–1420 (1997).
[23] S. L. Poliachik, W. L. Chandler, P. D. Mourad, R. J. Ollos, and L. A. Crum, ''Activation, aggregation and adhesion of platelets exposed to high-intensity focused ultrasound,'' Ultrasound Med. Biol. **27**, 1567–1576 (2001).
[24] M. L. Noble, S. Vaezy, A. Keshavarzi, M. Paun, A. F. Prokop, E. Y. Chi, C. Cornejo, S. R. Sharar, G. J. Jurkovich, R. W. Martin, and L. A. Crum, ''Spleen hemostasis using high-intensity ultrasound: Survival and healing,'' J. Trauma: Inj., Infect., Crit. Care **53**, 1115–1120 (2002).
[25] K. Y. Ng and Y. Liu, ''Therapeutic ultrasound: Its application in drug delivery,'' Med. Res. Rev. **22**, 204–223 (2002).
[26] C. Lauer, R. Burge, D. B. Tang, B. G. Bass, E. R. Gomez, and B. B. Alving, ''Effect of ultrasound on tissue-type plasminogen activator-induced thrombolysis,'' Circulation **86**, 1257–1264 (1992).
[27] C. W. Francis, P. T. Onundarson, E. L. Carstensen, A. Blinc, R. S. Meltzer, K. Schwarz, and V. J. Marder, ''Enhancement of fibrinolysis in vitro by ultrasound,'' J. Clin. Invest. **90**, 2063–2068 (1992).
[28] W. C. Culp, T. R. Porter, F. Xie, T. C. Goertzen, T. C. McCowan, B. N. Vonk, and B. T. Baxter, ''Microbubble potentiated ultrasound as a method of declotting thrombosed dialysis grafts: Experimental study in dogs,'' Cardiovasc. Intervent Radiol. **24**, 407–412 (2001).
[29] B. Devcic-Kuhar, S. Pfaffenberger, M. Grö Schl, C. Kollmann, E. Benes, and M. Gottsauner-Wolf, ''In vitro thrombolysis enhanced by standing and traveling ultrasound wave fields,'' Ultrasound Med. Biol. **28**, 1181–1187 (2002).
[30] Y. Adler, A. Dagan, G. Golovchiner, Z. Iakobishvili, I. Matz, E. Lev *et al.*, ''Augmentation of low-frequency ultrasound-induced clot disruption by hydroxyethyl starch is dependent on the duration and intensity of ultrasound exposure; An in vitro study,'' Ultrasound Med. Biol. **29**, 483–486 (2003).
[31] A. V. Alexandrov, A. M. Demchuk, M. D. Hill, R. A. Felberg, O. Chernyshev, Z. Garami *et al.*, ''CLOTBUST: Phase I data on ultrasound enhanced thrombolysis for stroke,'' Stroke **33**, 72 (2002).
[32] T. G. Leighton, *The Acoustic Bubble* (Academic, New York, 1994).
[33] J. B. Keller and M. J. Miksis, ''Bubble oscillations of large amplitude,'' J. Acoust. Soc. Am. **68**, 628–633 (1980).
[34] A. Prosperetti and A. Lezzi, ''Bubble dynamics in a compressible liquid. I: First-order theory,'' J. Fluid Mech. **168**, 457–478 (1986).
[35] A. Prosperetti, L. A. Crum, and K. W. Commander, ''Nonlinear bubble dynamics,'' J. Acoust. Soc. Am. **83**, 502–514 (1988).
[36] V. Kamath and A. Prosperetti, ''Numerical integration methods in gas-bubble dynamics,'' J. Acoust. Soc. Am. **85**, 1538–1548 (1989).
[37] M. S. Plesset, ''On the stability of fluid flows with spherical symmetry,'' J. Appl. Phys. **25**, 96–98 (1954).
[38] A. Eller and L. A. Crum, ''Instability of the motion of a pulsating bubble in a sound field,'' J. Acoust. Soc. Am. **47**, 762–767 (1970).
[39] A. Prosperetti, ''On the stability of spherically symmetric flows,'' Atti Accad. Naz. Lincei, Rend. Cl. Sci. Fis. Mat. Nat. **62**, 196–203 (1977).
[40] S. Hilgenfeldt, D. Lohse, and M. P. Brenner, ''Phase diagrams for sonoluminescing bubbles,'' Phys. Fluids **8**, 2808–2826 (1996).
[41] Y. Hao and A. Prosperetti, ''The effect of viscosity on the spherical stability of oscillating gas bubbles,'' Phys. Fluids **11**, 1309–1317 (1999).
[42] R. G. Holt and D. F. Gaitan, ''Observation of stability boundaries in the parameter space of single bubble sonoluminescence,'' Phys. Rev. Lett. **77**, 3791–3794 (1996).
[43] D. F. Gaitan and R. G. Holt, ''Experimental observations of bubble response and light intensity near the threshold for single bubble sonoluminescence in an air–water system,'' Phys. Rev. E **59**, 5495–5502 (1999).
[44] A. Prosperetti and Y. Hao, ''Modeling of spherical gas bubble oscillations and sonoluminescence,'' Philos. Trans. R. Soc. London, Ser. A **357**, 203–223 (1999).
[45] A. Prosperetti, ''On the stability of spherically symmetric flows,'' Atti Accad. Naz. Lincei, Rend. Cl. Sci. Fis. Mat. Nat. **62**, 196–203 (1977).
[46] M. P. Brenner, D. Lohse, and T. F. Dupont, ''Bubble shape oscillations and the onset of sonoluminescence,'' Phys. Rev. Lett. **75**, 954–957 (1995).
[47] A. Eller and H. G. Flynn, ''Rectified diffusion during nonlinear pulsations of cavitation bubbles,'' J. Acoust. Soc. Am. **37**, 493–503 (1965).
[48] M. M. Fyrillas and A. J. Szeri, ''Dissolution or growth of soluble spherical oscillating bubbles,'' J. Fluid Mech. **277**, 381–407 (1994).
[49] J. Huang, ''Heating in vascular and flow-through tissue phantoms induced by focused ultrasound,'' Ph.D. dissertation, Boston University, 2002.
[50] X. Yang, ''Investigation of bubble dynamics and heating during focused ultrasound insonation in tissue-mimicking materials,'' Ph.D. dissertation, Boston University, 2003.

[51] H. G. Flynn, ''Cavitation dynamics. II: Free pulsations and models for cavitation bubbles,'' J. Acoust. Soc. Am. **58**, 1160–1170 (1975).
[52] H. G. Flynn and C. C. Church, ''Transient pulsations of small gas bubbles in water,'' J. Acoust. Soc. Am. **84**, 985–998 (1988).
[53] U. Parlitz, V. Englisch, C. Scheffczyk, and W. Lauterborn, ''Bifurcation structure of bubble oscillations,'' J. Acoust. Soc. Am. **88**, 1061–1077 (1990).
[54] W. Lauterborn and J. Holzfuss, ''Acoustic chaos,'' Int. J. Bifurcation Chaos Appl. Sci. Eng. **1**, 13–26 (1991).
[55] J. S. Allen and R. A. Roy, ''Dynamics of gas bubbles in viscoelastic fluids. I: Linear viscoelasticity,'' J. Acoust. Soc. Am. **107**, 3167–3178 (2000).
[56] J. S. Allen and R. A. Roy, ''Dynamics of gas bubbles in viscoelastic fluids. II: Nonlinear viscoelasticity,'' J. Acoust. Soc. Am. **108**, 1640–1650 (2000).
[57] R. G. Holt, R. A. Roy, P. A. Edson, and X. Yang, ''Bubbles and HIFU: The good, the bad, and the ugly,'' *Proceedings of the Second International Symposium on Therapeutic Ultrasound*, edited by L. Crum (2002), p. 120.

Explicit secular equations for piezoacoustic surface waves: Shear-horizontal modes

Bernard Collet and Michel Destrade
Laboratoire de Modélisation en Mécanique, CNRS, Université Pierre et Marie Curie (UMR 7607), 4 place Jussieu, case 162, 75252 Paris Cedex 05, France

(Received 7 April 2004; revised 9 September 2004; accepted 29 September 2004)

Attention is given to surface waves of shear-horizontal modes in piezoelectric crystals permitting the decoupling between an elastic in-plane Rayleigh wave and a piezoacoustic antiplane Bleustein–Gulyaev wave. Specifically, the crystals possess $\bar{4}$ symmetry (inclusive of $\bar{4}2$m, $\bar{4}3$m, and 23 classes) and the boundary is any plane containing the normal to a symmetry plane (rotated *Y* cuts about the *Z* axis). The secular equation is obtained explicitly as a polynomial not only for the metallized boundary condition but, in contrast to previous studies on the subject, also for other types of boundary conditions. For the metallized surface problem, the secular equation is a quadratic in the squared wave speed; for the unmetallized surface problem, it is a sextic in the squared wave speed; for the thin conducting boundary problem, it is of degree 16 in the speed. The relevant root of the secular equation can be identified and the complete solution is then found (attenuation factors, field profiles, etc.). The influences of the cut angle and of the conductance of the adjoining medium are illustrated numerically for AlAs ($\bar{4}3$m), $BaLaGa_3O_7$ ($\bar{4}2$m), and $Bi_{12}GeO_{20}$ (23). Indications are given on how to apply the method to crystals with 222 symmetry. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1819503]

PACS numbers: 43.35.Pt, 43.38.Fx, 43.38.Rh [ANN] Pages: 3432–3442

I. INTRODUCTION

The nature and properties of a piezoacoustic surface wave depend heavily on the crystallographic and anisotropic properties of the piezoelectric substrate, on the direction of propagation, and on the orientation of the cut (boundary) plane. For certain choices, the in-plane components of the mechanical displacement decouple from the antiplane component, leading to two different types of surface waves namely, the Rayleigh wave, elliptically polarized in the plane containing the direction of propagation and the normal to the substrate surface, and the shear-horizontal (SH) wave, linearly polarized in the direction normal to the direction of propagation and parallel to the free surface. Moreover, either wave or both waves may be coupled to the electromagnetic fields. Of special interest are the configurations allowing for a piezoacoustic SH wave, decoupled from a purely elastic Rayleigh wave. Indeed, the former type of wave, also known as Bleustein–Gulyaev[1,2] wave, penetrates more deeply into the substrate than the latter type; consequently, the acoustic energy is less localized and the power can be increased significantly before damage occurs (Tseng[3]) although, as pointed out by a referee, there are exceptions to this behavior.[4] Moreover, an SH wave-based resonator is smaller than a resonator based on the propagation of a sagittally polarized surface wave (see Kadota and collaborators[5,6]); this feature results in a downsizing of design, an attractive point for the miniaturization of mobile phones for instance, where SAW devices are used as filters. Koerber and Vogel[7,8] identified all the cuts and rotations of axes leading to piezoacoustic SH modes; they exist for some suitable cuts and transformations in the following crystal classes: 2, 23, $\bar{4}3$m, 222, 2mm, 4, $\bar{4}$, 6, 4mm, 6mm, 32, $\bar{4}2$m, $\bar{6}$m2, 422, and 622. The main purpose of this paper is to derive explicitly the secular equation for piezoacoustic SH surface waves uncoupled from purely elastic Rayleigh waves polarized in a plane of symmetry, for a crystal in the $\bar{4}$ symmetry class (and thus for the $\bar{4}2$m, $\bar{4}3$m, and 23 classes).

Several workers addressed this topic in the wake of the seminal papers by Bleustein and by Gulyaev, but explicit results remained limited either to propagation in special directions for which one of the piezoelectric constant is zero[3,9–13] or to the case where the free surface of the substrate is metallized,[9] or in the weak piezoelectric coupling approximation.[9,14] Here, the crystal is cut along a plane containing the *Z* axis and making *any* angle with the *XY* crystallographic plane. Moreover, the surface may be metallized, or in contact with the vacuum, or in contact with a thin conducting layer with arbitrary finite conductance. Also, no approximation is made about the strength of the piezoelectric effect. Attention is, however, limited to crystals with tetragonal $\bar{4}$ symmetry (inclusive of the tetragonal $\bar{4}2$m, cubic $\bar{4}3$m, and cubic 23 symmetries). This limitation is not essential but simplifies the notation to a certain extent; it is nevertheless possible to extrapolate the method presented hereafter to crystals with lower symmetries such as orthorhombic 222, as pointed out at the end of the paper. Note that the study of piezoacoustic SH surface waves in potassium niobate (2-mm symmetry) was recently undertaken by Mozhaev and Weihnacht,[15] who showed that for that class, the secular equation is a cubic in the squared wave speed.

The constitutive equations and the piezoacoustic equations for the class of crystals listed above are recalled in Secs. II A and II B, respectively, and some fundamental equations, which encapsulate the whole boundary value problem

and its resolution, are quickly derived in Sec. II C.

These fundamental equations are applied in Sec. III to the consideration of a piezoacoustic SH surface wave. A method based on the resolution of the propagation condition for partial modes first, and the resolution of the boundary value conditions next, would lead to quite an involved analysis, including the analytical examination of a quartic polynomial for the coefficients of attenuation. The method based on the fundamental equations derived in Sec. II C circumvents the stage of the quartic and delivers directly the secular equation in polynomial form. In particular it is seen that for metallized (also known as short-circuit) boundary conditions (Sec. III A), this equation is just a quadratic in the squared wave speed whose relevant root is readily identified. For open-circuit boundary conditions (Sec. III B), the secular equation is also a quadratic in the squared wave speed, but it is not valid (the corresponding solution does not satisfy the boundary conditions.) For the free-(nonmetallized) boundary condition (Sec. III C), the secular equation is a sextic in the squared wave speed. For the conductive thin layer boundary condition (Sec. III D), the secular equation is a polynomial of degree 16 in the wave speed.

Once the secular equation is solved for the speed, the complete description of the wave follows naturally (Sec. IV A), including the attenuation coefficients and the profiles for the mechanical displacement, electric potential, traction, and electrical induction. A simple check for the validity of the solution is proposed in Sec. IV B.

The results are illustrated numerically and graphically using experimental data available for AlAs (cubic $\bar{4}3$m), $BaLaGa_3O_7$ (tetragonal $\bar{4}2$m), and $Bi_{12}GeO_{20}$ (cubic 23) in Sec. V. The range of existence of the free-SH wave with respect to the angle of cut, the speeds of propagation, the amplitude of the profiles, etc. are all quantities which can be obtained numerically with as high a degree of numerical accuracy as is needed.

The paper concludes with a discussion of the merits and possible applications of the method presented, and with a chronological account of the several advances in the field, without which the results of this paper could not have been established.

II. PRELIMINARIES

A. Constitutive equations

Consider a piezoelectric crystal with mass density ρ, possessing at most the tetragonal $\bar{4}$ symmetry (this symmetry includes the tetragonal $\bar{4}2$m, cubic $\bar{4}3$m, and cubic 23 cases.) Let $\hat{c}_{ijkl}$, $\hat{e}_{ijk}$, and $\hat{\epsilon}_{ij}$ be its respective elastic, piezoelectric, and dielectric constants with respect to the coordinate system XYZ along the crystallographic axes. Now, cut the crystal by a plane containing the Z axis and making an angle θ with the XY plane. The new coordinate system $x_1x_2x_3$ (say), obtained after rotation of XYZ about Z, is defined by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \tag{1}$$

and so, the plane of cut is defined by $x_2=0$, as shown in Fig. 1.

In the new coordinate system, under the electrostatic approximation for the electrical field, the stress tensor components σ_{ij} and the electric induction components D_i are related to the gradients of the mechanical displacement **u** and of the electrical potential ϕ by the constitutive relations

$$\sigma_{ij}=c_{ijkl}u_{l,k}+e_{ijk}\phi_{,k}, \quad D_i=e_{ikl}u_{l,k}-\epsilon_{ik}\phi_{,k}, \tag{2}$$

where the comma denotes partial differentiation (here, with respect to the x_k coordinates). Using the Voigt contracted notation for the c, e, and ϵ, these relations are written in matrix form as

$$\begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{23} \\ \sigma_{31} \\ \sigma_{12} \\ D_1 \\ D_2 \\ D_3 \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & c_{16} & 0 & 0 & e_{31} \\ c_{12} & c_{11} & c_{13} & 0 & 0 & -c_{16} & 0 & 0 & -e_{31} \\ c_{13} & c_{13} & c_{33} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 & e_{14} & -e_{15} & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 & e_{15} & e_{14} & 0 \\ c_{16} & -c_{16} & 0 & 0 & 0 & c_{66} & 0 & 0 & e_{36} \\ 0 & 0 & 0 & e_{14} & e_{15} & 0 & -\epsilon_{11} & 0 & 0 \\ 0 & 0 & 0 & -e_{15} & e_{14} & 0 & 0 & -\epsilon_{11} & 0 \\ e_{31} & -e_{31} & 0 & 0 & 0 & e_{36} & 0 & 0 & -\epsilon_{33} \end{bmatrix} \begin{bmatrix} u_{1,1} \\ u_{2,2} \\ u_{3,3} \\ u_{2,3}+u_{3,2} \\ u_{3,1}+u_{1,3} \\ u_{1,2}+u_{2,1} \\ \phi_{,1} \\ \phi_{,2} \\ \phi_{,3} \end{bmatrix}. \tag{3}$$

Explicitly, the c_{ij} and e_{ij} are deduced from the $\hat{c}_{ij}$ and $\hat{e}_{ij}$ in XYZ by well-known relationships. In particular

$$c_{44}=\hat{c}_{44}, \quad \epsilon_{11}=\hat{\epsilon}_{11}, \quad e_{14}=\hat{e}_{14}\cos 2\theta-\hat{e}_{15}\sin 2\theta,$$

$$e_{15}=\hat{e}_{15}\cos 2\theta+\hat{e}_{14}\sin 2\theta. \tag{4}$$

B. Piezoacoustic equations

Now, consider the propagation of an antiplane (SH) inhomogeneous wave in the half-space $x_2\geqslant 0$, traveling with speed v and wave number k in the x_1 direction, with attenuation in the x_2 direction. It is known[16] that for the crystals

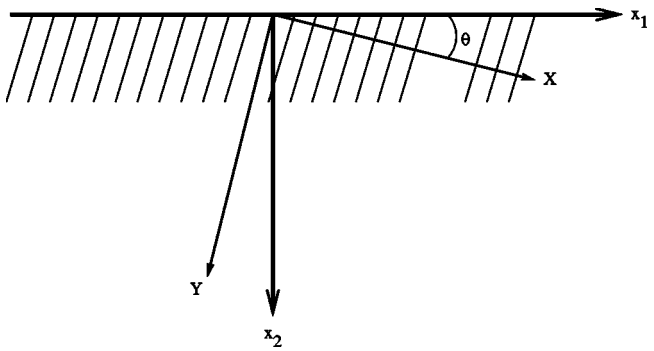


FIG. 1. Rotated Y cut about the Z axis.

under consideration this wave decouples entirely from its in-plane counterpart, a purely elastic two-component Rayleigh wave. Thus, the wave is modeled as $u_1=u_2=0$, and

$$\{u_3,\phi\}=\{U_3(kx_2),\varphi(kx_2)\}e^{ik(x_1-vt)}, \tag{5}$$

for some yet unknown functions U_3 and φ of kx_2. Accordingly, the constitutive equations Eq. (2) lead to similar forms for the stress and electrical components

$$\{\sigma_{ij},D_i\}=ik\{t_{ij}(kx_2),d_i(kx_2)\}e^{ik(x_1-vt)}, \tag{6}$$

where $t_{11}=t_{22}=t_{12}=d_3=0$, and

$$\begin{aligned} &t_{33}=c_{13}U_1-ic_{11}U_2', \quad t_{32}=-ic_{44}U_3'+e_{14}\varphi+ie_{15}\varphi',\\ &t_{31}=c_{44}U_3+e_{15}\varphi-ie_{14}\varphi', \quad d_1=e_{15}U_3-ie_{14}U_3'-\epsilon_{11}\varphi,\\ &d_2=e_{14}U_3+ie_{15}U_3'+i\epsilon_{11}\varphi'. \end{aligned} \tag{7}$$

Here and hereafter, the prime denotes differentiation with respect to the variable kx_2.

Using the above-introduced functions U_3, φ, t_{ij}, d_i, the classical equations of piezoacoustics

$$\sigma_{ij,j}=\rho u_{i,tt}, \quad D_{i,i}=0, \tag{8}$$

reduce to

$$-t_{13}+it_{32}'=-XU_3, \quad -d_1+id_2'=0, \tag{9}$$

where $X:=\rho v^2$.

When the region $x_2<0$ above the crystal is the vacuum (permeability: ϵ_0) and the surface $x_2=0$ remains free of tractions, then the boundary value problem corresponding to Eq. (9) is that of piezoacoustic (Bleustein–Gulyaev) SH surface waves.

C. Fundamental equations for the resolution

The method of resolution rests on the property that the piezoacoustic equations (9) can be written in the form,

$$\boldsymbol{\xi}'=i\mathbf{N}\boldsymbol{\xi}, \quad \text{where} \quad \boldsymbol{\xi}(kx_2)=[U_3,\varphi,t_{32},d_2]^{\mathrm{T}}, \tag{10}$$

and $\mathbf{N}$ is a real 4×4 matrix which, brought up to any positive or negative integer power n, has the following block structure:

$$\mathbf{N}^n=\begin{bmatrix} \mathbf{N_1^{(n)}} & \mathbf{N_2^{(n)}} \\ \mathbf{K^{(n)}} & (\mathbf{N_1^{(n)}})^{\mathrm{T}} \end{bmatrix}, \quad \text{with} \quad \mathbf{K^{(n)}}=(\mathbf{K^{(n)}})^{\mathrm{T}},$$

$$\mathbf{N_2^{(n)}}=(\mathbf{N_2^{(n)}})^{\mathrm{T}}, \tag{11}$$

where $^{\mathrm{T}}$ denotes the transpose, and where the submatrices of $\mathbf{N}^n$ are 2×2 matrices. As kindly pointed out by a referee, the first-order differential form Eq. (10) of the piezoacoustic equations dates back to Kraut,[17] and before that, to Stroh[18] and others (see Fahmy and Adler[19] for references) for the purely elastic case. The subsequent analysis below builds upon several crucial contributions, which are listed and put into context at the end of this article.

Now, because the wave amplitude must vanish away from $x_2=0$, $\boldsymbol{\xi}(kx_2)$ is such that

$$\boldsymbol{\xi}(\infty)=\mathbf{0}. \tag{12}$$

Clearly, premultiplication of $\mathbf{N}^n$ by $\hat{\mathbf{I}}$ defined as

$$\hat{\mathbf{I}}=\begin{bmatrix} 0 & \mathbf{1} \\ \mathbf{1} & 0 \end{bmatrix}, \quad \text{where} \quad \mathbf{1}=\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{13}$$

produces a symmetric matrix,

$$\hat{\mathbf{I}}\mathbf{N}^n=\begin{bmatrix} \mathbf{K^{(n)}} & (\mathbf{N_1^{(n)}})^{\mathrm{T}} \\ \mathbf{N_1^{(n)}} & \mathbf{N_2^{(n)}} \end{bmatrix}=(\hat{\mathbf{I}}\mathbf{N}^n)^{\mathrm{T}}, \tag{14}$$

so that taking the scalar product on both sides of Eq. (10) by $\hat{\mathbf{I}}\mathbf{N}^n\overline{\boldsymbol{\xi}}$, where the overbar denotes the complex conjugate, leads to $\overline{\boldsymbol{\xi}}\cdot\hat{\mathbf{I}}\mathbf{N}^n\boldsymbol{\xi}'=i\overline{\boldsymbol{\xi}}\cdot\hat{\mathbf{I}}\mathbf{N}^{n+1}\boldsymbol{\xi}$, the right-hand side of which is purely imaginary. Taking the real part and integrating yields $\overline{\boldsymbol{\xi}}\cdot\hat{\mathbf{I}}\mathbf{N}^n\boldsymbol{\xi}=\text{const.}=0$ [by Eq. (12)], and in particular

$$\overline{\boldsymbol{\xi}}(0)\cdot\hat{\mathbf{I}}\mathbf{N}^n\boldsymbol{\xi}(0)=0. \tag{15}$$

These *fundamental equations*[20–22] allow for a completely analytical derivation of the secular equation, for a great variety of boundary conditions. Note that because $\mathbf{N}$ is a 4×4 matrix, there are at most three independent fundamental equations according to the Cayley–Hamilton theorem. Any choice of three different integers n is legitimate, although the choice $n=-1,1,2$ seems to yield the most compact expressions for the components of $\mathbf{N}^n$.

III. SECULAR EQUATIONS

For the piezoacoustic (Bleustein–Gulyaev) shear-horizontal wave, the matrix $\mathbf{N}$ in Eq. (10) is written in compact form using the following quantity:

$$\kappa^2=\frac{e_{14}^2}{\epsilon_{11}c_{44}+e_{15}^2}, \tag{16}$$

as

$$\mathbf{N}=\begin{bmatrix} \frac{e_{15}}{e_{14}}\kappa^2 & -\frac{\epsilon_{11}}{e_{14}}\kappa^2 & \frac{\epsilon_{11}}{e_{14}^2}\kappa^2 & -\frac{e_{15}}{e_{14}^2}\kappa^2 \\ \frac{c_{44}}{e_{14}}\kappa^2 & \frac{e_{15}}{e_{14}}\kappa^2 & -\frac{e_{15}}{e_{14}^2}\kappa^2 & -\frac{c_{44}}{e_{14}^2}\kappa^2 \\ X-c_{44}(1+\kappa^2) & -e_{15}(1+\kappa^2) & \frac{e_{15}}{e_{14}}\kappa^2 & \frac{c_{44}}{e_{14}}\kappa^2 \\ -e_{15}(1+\kappa^2) & \epsilon_{11}(1+\kappa^2) & -\frac{\epsilon_{11}}{e_{14}}\kappa^2 & \frac{e_{15}}{e_{14}}\kappa^2 \end{bmatrix}. \tag{17}$$

Note that $\mathbf{N}$ is indeed of the form Eq. (11). To express explicitly integer powers of $\mathbf{N}$, it proves convenient to use scalar multiples of the matrix $\hat{\mathbf{I}}\mathbf{N}^n$, for which the fundamental equations Eq. (15) are also valid. For instance, the fundamental equations (15) also hold at $n=2,-1$ when $\hat{\mathbf{I}}\mathbf{N}^n$ is replaced with $\mathbf{M}^{(2)}$, $\mathbf{M}^{(-1)}$ respectively, defined by

$$\mathbf{M}^{(2)}=\frac{e_{14}}{e_{15}\kappa^2}\hat{\mathbf{I}}\mathbf{N}^2, \quad \mathbf{M}^{(-1)}=\left(\frac{\epsilon_{11}c_{44}}{e_{14}^2}\kappa^2 X-1\right)\hat{\mathbf{I}}\mathbf{N}^{-1}. \tag{18}$$

These symmetric matrices are given explicitly by their components,

$$\begin{aligned}
M_{11}^{(2)}&=2[X-2c_{44}(1+\kappa^2)],\\
M_{12}^{(2)}&=-\frac{\epsilon_{11}}{e_{15}}X+2\frac{\epsilon_{11}c_{44}-e_{15}^2}{e_{15}}(1+\kappa^2),\\
M_{22}^{(2)}&=4\epsilon_{11}(1+\kappa^2),\\
M_{13}^{(2)}&=\frac{\epsilon_{11}}{e_{14}e_{15}}X-\frac{\epsilon_{11}c_{44}-e_{15}^2}{e_{14}e_{15}}(1+2\kappa^2),\\
M_{23}^{(2)}&=-2\frac{\epsilon_{11}}{e_{14}}(1+2\kappa^2), \quad M_{33}^{(2)}=4\frac{\epsilon_{11}}{e_{14}^2}\kappa^2,\\
M_{14}^{(2)}&=-\frac{1}{e_{14}}X+2\frac{c_{44}}{e_{14}}(1+2\kappa^2),\\
M_{24}^{(2)}&=-\frac{\epsilon_{11}c_{44}-e_{15}^2}{e_{14}e_{15}}(1+2\kappa^2),\\
M_{34}^{(2)}&=2\frac{\epsilon_{11}c_{44}-e_{15}^2}{e_{14}^2e_{15}}\kappa^2, \quad M_{44}^{(2)}=-4\frac{c_{44}}{e_{14}^2}\kappa^2,
\end{aligned} \tag{19}$$

and

$$\begin{aligned}
M_{11}^{(-1)}&=\left(1+\frac{\epsilon_{11}c_{44}}{e_{14}^2}\right)\kappa^2X-c_{44}(1+\kappa^2),\\
M_{12}^{(-1)}&=-\frac{\epsilon_{11}e_{15}}{e_{14}^2}\kappa^2X+e_{15}(1+\kappa^2),\\
M_{22}^{(-1)}&=-\frac{\epsilon_{11}^2}{e_{14}^2}\kappa^2X+\epsilon_{11}(1+\kappa^2), \quad M_{13}^{(-1)}=\frac{e_{15}}{e_{14}}\kappa^2,\\
M_{23}^{(-1)}&=\frac{\epsilon_{11}}{e_{14}}\kappa^2, \quad M_{33}^{(-1)}=\frac{\epsilon_{11}}{e_{14}^2}\kappa^2, \quad M_{14}^{(-1)}=\frac{X-c_{44}}{e_{14}}\kappa^2,\\
M_{24}^{(-1)}&=\frac{e_{15}}{e_{14}}\kappa^2, \quad M_{34}^{(-1)}=\frac{e_{15}}{e_{14}^2}\kappa^2, \quad M_{44}^{(-1)}=\frac{X-c_{44}}{e_{14}^2}\kappa^2.
\end{aligned} \tag{20}$$

Now all the required equations and quantities are in place to treat various electrical boundary value problems.

A. Metallized (short-circuit) boundary condition

Here, the surface of the crystal is coated with a thin metallic film, with thickness negligible when compared to the wavelength, and brought to a zero electrical potential. Moreover, the coating still allows the surface to remain free of mechanical tractions. Then

$$\sigma_{23}=0, \ \phi=0, \ \text{at} \ x_2=0, \quad \text{so that}$$
$$\boldsymbol{\xi}(0)=[U_3(0),0,0,d_2(0)]^{\mathrm{T}}. \tag{21}$$

Writing $\boldsymbol{\xi}(0)=U_3(0)[1,0,0,\alpha]$, where $\alpha=d_2(0)/U_3(0)$ is complex, the fundamental equations (15) for $\hat{\mathbf{I}}\mathbf{N}$ $(n=1)$, $\mathbf{M}^{(2)}$ $(n=2)$, and $\mathbf{M}^{(-1)}$ $(n=-1)$ lead to the following homogeneous system of equations:

$$\begin{bmatrix} N_{31} & N_{21} & N_{24} \\ M_{11}^{(2)} & M_{14}^{(2)} & M_{44}^{(2)} \\ M_{11}^{(-1)} & M_{14}^{(-1)} & M_{44}^{(-1)} \end{bmatrix}\begin{bmatrix} 1 \\ \alpha+\overline{\alpha} \\ \alpha\overline{\alpha} \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{22}$$

For a nontrivial solution to exist, the determinant of the system's matrix must be zero. Factoring out common factors, this condition reads

$$\begin{vmatrix} X-c_{44}(1+\kappa^2) & c_{44}\kappa^2 & -c_{44} \\ 2[X-2c_{44}(1+\kappa^2)] & -X+2c_{44}(1+2\kappa^2) & -4c_{44} \\ \left(1+\frac{\epsilon_{11}c_{44}}{e_{14}^2}\right)\kappa^2X-c_{44}(1+\kappa^2) & (X-c_{44})\kappa^2 & X-c_{44} \end{vmatrix}=0. \tag{23}$$

When $X=0$, the first column in the determinant becomes proportional to the third, and the determinant is zero. Hence, X is a factor of the determinant; the remaining factor is a quadratic in X

$$X^2-c_{44}\left(\frac{3\epsilon_{11}c_{44}+4e_{14}^2+4e_{15}^2}{\epsilon_{11}c_{44}+e_{15}^2}\right)X + 2c_{44}^2\left(\frac{\epsilon_{11}c_{44}+2e_{14}^2+2e_{15}^2}{\epsilon_{11}c_{44}+e_{15}^2}\right)=0, \tag{24}$$

that is, the *explicit secular equation for piezoacoustic (Bleustein–Gulyaev) antiplane surface waves on a metallized tetragonal* $\bar{4}$ *(or tetragonal* $\bar{4}2m$*, or cubic* $\bar{4}3m$*, 23) crystal, cut along any plane containing the Z axis.*

This equation being a quadratic in X, it is solved explicitly and it yields *a priori* two roots. The selection is made by considering the known speed of a Bleustein–Gulyaev surface wave in a cubic $\bar{4}3$m or 23 crystal when $\theta=45°$. Then, the root of the quadratic corresponding to the plus sign is $X=\rho v^2=2c_{44}$, and the root corresponding to the minus sign is

$$\rho v^2=c_{44}\left(1+\frac{e_{15}^2}{c_{44}\epsilon_{11}+e_{15}^2}\right)=c_{44}\left(1+\frac{\hat{e}_{14}^2}{c_{44}\epsilon_{11}+\hat{e}_{14}^2}\right), \tag{25}$$

in accordance with Tseng[3] (see also Koerber and Vogel,[7] Alburque and Chao,[10] Velasco,[11] Bright and Hunt[13]). By continuity with the other cases ($\bar{4}$, $\bar{4}2$m, $\bar{4}3$m, 23, $\theta\neq 45°$), the *speed of the generic Bleustein–Gulyaev wave* found from the secular equation (24) is v_{BGm}, given by

$$\frac{\rho v_{\mathrm{BGm}}^2}{c_{44}}=\frac{3\epsilon_{11}c_{44}+4e_{14}^2+4e_{15}^2-[(\epsilon_{11}c_{44}+4e_{14}^2)^2+(4e_{14}e_{15})^2]^{1/2}}{2(\epsilon_{11}c_{44}+e_{15}^2)}. \tag{26}$$

For tetragonal $\bar{4}2$m, cubic $\bar{4}3$m, or cubic 23 crystals, $\hat{e}_{15}=0$ and by Eq. (4), this expression reduces to

$$\frac{\rho v_{\mathrm{BGm}}^2}{c_{44}}=\frac{3+4\hat{\chi}^2-[(1+4\hat{\chi}^2\cos^2 2\theta)^2+4\hat{\chi}^4\sin^2 4\theta]^{1/2}}{2(1+\hat{\chi}^2\sin^2 2\theta)}, \tag{27}$$

as proved by Braginskiĭ and Gilinskiĭ[9] using a different method. In Eq. (27), $\hat{\chi}^2=\hat{e}_{14}^2/(\epsilon_{11}c_{44})$ is the "piezoelectric coupling coefficient" for bulk waves.

B. Electrically open boundary condition

The substrate is said to be "mechanically free, electrically open" (Ingebrigsten,[23] Lothe and Barnett[24]) when

$$\sigma_{32}=0, \quad D_2=0, \text{ at } x_2=0, \quad \text{so that} \quad \boldsymbol{\xi}(0)=[U_3(0),\varphi(0),0,0]^{\mathrm{T}}. \tag{28}$$

Writing $\boldsymbol{\xi}(0)=U_3(0)[1,\alpha,0,0]$, where $\alpha=\varphi(0)/U_3(0)$ is complex, the fundamental equations (15) for $\hat{\mathbf{I}}\mathbf{N}$ $(n=1)$, $\mathbf{M}^{(2)}$ $(n=2)$, and $\mathbf{M}^{(-1)}$ $(n=-1)$ lead to the following homogeneous system of equations:

$$\begin{bmatrix} N_{31} & N_{32} & N_{42} \\ M_{11}^{(2)} & M_{12}^{(2)} & M_{22}^{(2)} \\ M_{11}^{(-1)} & M_{12}^{(-1)} & M_{22}^{(-1)} \end{bmatrix}\begin{bmatrix} 1 \\ \alpha+\bar{\alpha} \\ \alpha\bar{\alpha} \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{29}$$

For a nontrivial solution to exist, the determinant of the system's matrix must be zero. Its components are given in Eq. (17), Eq. (19), and Eq. (20). As in the short-circuit configuration, at $X=0$ the first and third columns in the determinant both become proportional to $[1,4,1]^{\mathrm{T}}$ and so X is a factor of the determinant; the remaining factor is a quadratic in X

$$\frac{\epsilon_{11}^2}{e_{14}^2}\kappa^2X^2-\epsilon_{11}(1+\kappa^2)\left[3-\left(1+4\frac{e_{15}^2}{e_{14}^2}\right)\kappa^2\right]X + 2(1+\kappa^2)^2(\epsilon_{11}c_{44}-e_{14}^2-e_{15}^2)=0. \tag{30}$$

At this stage, an important point must be raised. Although this secular equation might seem legitimate at first sight, it must be recalled that it was obtained through a process based on the fundamental equations (15) which, due to the involvement of integer powers of the matrix $\mathbf{N}$, might generate spurious secular equations. In fact, it has been proved[9,24] that the Bleustein–Gulyaev wave *does not exist for open-circuit boundary conditions*. Hence, the secular equation Eq. (30) is not valid. It is given here for completeness and to illustrate one limitation of the fundamental equations approach. However, as is seen in Sec. IV B, a simple check can be done to realize whether the speed given by an explicit secular equation is valid or not.

C. Free-boundary condition

In the general case of a nonmetallized, mechanically free boundary, the tangential component of the electric field and the normal component of the electric induction are continuous across the substrate/vacuum interface. These continuities lead to the relationship (e.g., Dieulesaint and Royer,[25] p.288),

$$d_2(0)=i\epsilon_0\varphi(0), \quad \text{so that} \quad \boldsymbol{\xi}(0)=\varphi(0)[\alpha,1,0,i\epsilon_o]^{\mathrm{T}}, \tag{31}$$

where $\alpha=\alpha_1+i\alpha_2=U_3(0)/\varphi(0)$ is complex. Now, the fundamental equations (15) for $\hat{\mathbf{I}}\mathbf{N}$ $(n=1)$, $\mathbf{M}^{(2)}$ $(n=2)$, and

$\mathbf{M}^{(-1)}$ $(n=-1)$ lead to a nonhomogeneous linear system of equations

$$\begin{bmatrix} N_{32} & \epsilon_0 N_{21} & N_{31} \\ M_{12}^{(2)} & \epsilon_0 M_{14}^{(2)} & M_{11}^{(2)} \\ M_{12}^{(-1)} & \epsilon_0 M_{14}^{(-1)} & M_{11}^{(-1)} \end{bmatrix} \begin{bmatrix} 2\alpha_1 \\ 2\alpha_2 \\ \alpha_1^2+\alpha_2^2 \end{bmatrix} = \begin{bmatrix} -N_{42}-\epsilon_0^2 N_{24} \\ -M_{22}^{(2)}-\epsilon_0^2 M_{44}^{(2)} \\ -M_{22}^{(-1)}-\epsilon_0^2 M_{44}^{(-1)} \end{bmatrix}. \quad (32)$$

By Cramer's rule, the unique solution to this system is

$$2\alpha_1=\Delta_1/\Delta, \quad 2\alpha_2=\Delta_2/\Delta, \quad \alpha_1^2+\alpha_2^2=\Delta_3/\Delta, \quad (33)$$

where Δ is the determinant of the 3×3 matrix in Eq. (32), with components given in Eq. (17), Eq. (19), and Eq. (20), and the Δ_k are the determinants obtained by replacing this matrix's kth column with the vector on the right-hand side of Eq. (32). It follows from Eq. (33) that

$$\Delta_1^2+\Delta_2^2-4\Delta\Delta_3=0, \quad (34)$$

which is the *explicit secular equation for piezoacoustic (Bleustein–Gulyaev) antiplane surface waves on a nonmetallized, mechanically free* $\bar{4}$ *(or* $\bar{4}2m$, $\bar{4}3m$, 23*) crystal, cut along any plane containing the Z axis*.

The expansions of the determinants $\Delta,\ldots,\Delta_3$ are lengthy and are not displayed here, but they are easily computed by using the components Eq. (17), Eq. (19), and Eq. (20). It turns out that Δ factorizes into the product of ϵ_0 and a cubic in X which is independent of ϵ_0, while Δ_1 (respectively, Δ_2, Δ_3) factorizes into the product of $\epsilon_0 X$ (respectively, X, ϵ_0) and a polynomial which is quadratic in X and linear in ϵ_0^2. The resulting secular equation (34) is a sextic in X and a cubic in ϵ_0^2, with the coefficient of the ϵ_0^6 term proportional to the ''metallized secular equation'' (24) and the coefficient of the ϵ_0^0 term proportional to the (nonvalid) ''open-circuit secular equation'' (30).

It is emphasized again that, as in Sec. III B, great care must be taken to ensure that the speed given by the secular equation Eq. (34) leads to a valid solution. This point is discussed in Sec. IV B.

D. Thin conducting layer boundary condition

As a final type of boundary condition, consider that the semi-infinite substrate is covered with a metallic film with thickness h and conductance γ, where h is assumed to be so small with respect to the acoustic wavelength that the effects of mechanical loading can be neglected. Then, the permeability of the region $x_2<0$ close to the interface is changed from ϵ_0 (see the previous subsection) to $\epsilon_0-i\epsilon$ (see Royer and Dieulesaint,[26] p. 301), with $\epsilon=\gamma h/v$. Replacing the former quantity with the latter in the previous subsection leads to similar results as above, except that the 3×3 matrix and the right-hand side of Eq. (32) are now replaced with

$$\begin{bmatrix} N_{32}+\epsilon N_{21} & \epsilon_0 N_{21} & N_{31} \\ M_{12}^{(2)}+\epsilon M_{14}^{(2)} & \epsilon_0 M_{14}^{(2)} & M_{11}^{(2)} \\ M_{12}^{(-1)}+\epsilon M_{14}^{(-1)} & \epsilon_0 M_{14}^{(-1)} & M_{11}^{(-1)} \end{bmatrix}, \text{ and}$$

$$\begin{bmatrix} -N_{42}-2\epsilon N_{22}-(\epsilon_0^2+\epsilon^2)N_{24} \\ -M_{22}^{(2)}-2\epsilon M_{24}^{(2)}-(\epsilon_0^2+\epsilon^2)M_{44}^{(2)} \\ -M_{22}^{(-1)}-2\epsilon M_{24}^{(-1)}-(\epsilon_0^2+\epsilon^2)M_{44}^{(-1)} \end{bmatrix}, \quad (35)$$

respectively. Then, the secular equation is Eq. (34), where $\Delta,\ldots,\Delta_3$ are appropriately changed. Because here the quantity ϵ depends upon v, the secular equation is a polynomial in the speed of degree higher than in the previous subsection, namely it is a polynomial of degree 16 in v.

IV. CONSTRUCTION OF THE SOLUTIONS

A. Description of the wave

Once a wave speed is determined from the secular equation, it is a rather straightforward matter to construct the corresponding complete solution. Indeed, the antiplane mechanical displacement u_3, the electrical potential ϕ, the shear stress σ_{32}, and the electric induction D_2 are given by

$$[u_3,\phi,\sigma_{32},D_2](x_1,x_2,t) = \Re\{[U_3,\varphi,ikt_{32},ikd_2](kx_2)e^{ik(x_1-vt)}\}. \quad (36)$$

Here, U_3, φ, t_{32}, d_2 are the components of $\boldsymbol{\xi}$, solution to $\boldsymbol{\xi}'=i\mathbf{N}\boldsymbol{\xi}$. Taking $\boldsymbol{\xi}$ in exponential form leads to the following decaying solution:

$$\boldsymbol{\xi}=\beta_1\boldsymbol{\xi}^{\mathbf{1}}e^{ikq_1x_2}+\beta_2\boldsymbol{\xi}^{\mathbf{2}}e^{ikq_2x_2}, \quad (37)$$

where β_1, β_2 are constants, q_1, q_2 are the two roots with positive imaginary part to the *inhomogeneous wave propagation condition*: $\det(\mathbf{N}-q\mathbf{1})=0$, and the $\boldsymbol{\xi}^{\mathbf{i}}$ satisfy: $\mathbf{N}\boldsymbol{\xi}^{\mathbf{i}}=q_i\boldsymbol{\xi}^{\mathbf{i}}$.

Explicitly, the q_i are roots of the quartic

$$(\epsilon_{11}c_{44}+e_{15}^2)q^4-4e_{14}e_{15}q^3-[\epsilon_{11}(X-2c_{44})-4e_{14}^2+2e_{15}^2]q^2+4e_{14}e_{15}q+e_{15}^2-\epsilon_{11}(X-c_{44})=0, \quad (38)$$

and the $\boldsymbol{\xi}^{\mathbf{i}}$ are proportional to any column vector of the matrix adjoint to $\mathbf{N}-q_i\mathbf{1}$, the third one say. Hence

$$\boldsymbol{\xi}^{\mathbf{i}}=\left[\frac{e_{14}}{c_{44}}a_i,b_i,e_{14}^2f_i,\frac{e_{14}^2}{c_{44}}g_i\right]^{\mathrm{T}}, \quad (39)$$

where the nondimensional real quantities a_i, b_i, f_i, g_i $(i=1,2)$ are given by

$$a_i=-\frac{\epsilon_{11}c_{44}}{e_{14}^2}(q_i^2+1), \quad b_i=\frac{e_{15}}{e_{14}}(q_i^2-1)-2q_i,$$

$$f_i=\frac{e_{15}}{e_{14}}(3q_i^2-1)-\frac{e_{15}^2}{e_{14}^2}q_i(q_i^2-1)-\frac{\epsilon_{11}c_{44}}{e_{14}^2}q_i(q_i^2+1)-2q_i, \quad (40)$$

$$g_i=\frac{\epsilon_{11}c_{44}}{e_{14}^2}\left(q_i^2+2\frac{e_{15}}{e_{14}}q_i-1\right).$$

Finally, the ratio β_2/β_1 comes from the condition that the third component of $\boldsymbol{\xi}(0)$, proportional to $t_{23}(0)$, must be zero, so that $\beta_2/\beta_1=-f_1/f_2$.

B. Validity of the solution

Once a wave solution has been constructed, its validity must be checked, that is, it must satisfy the boundary conditions.

Thus, for a solution to the *short-circuit* problem (Sec. III A), it must be checked that $\varphi(0)=\beta_1 b_1+\beta_2 b_2$ and $t_{32}(0)=\beta_1 e_{14} f_1+\beta_2 e_{14} f_2$ are indeed equal to zero. These conditions are equivalent to checking that

$$\begin{vmatrix} b_1 & b_2 \\ f_1 & f_2 \end{vmatrix} = 0. \tag{41}$$

For a solution to the *open-circuit* problem (Sec. III B), it must be checked that $t_{32}(0)=\beta_1 e_{14} f_1+\beta_2 e_{14} f_2$ and $d_2(0)=\beta_1(e_{14}^2/c_{44})g_1+\beta_2(e_{14}^2/c_{44})g_2$ are equal to zero. These conditions are equivalent to checking that

$$\begin{vmatrix} f_1 & f_2 \\ g_1 & g_2 \end{vmatrix} = 0.$$

As expected,[9,24] this condition is never satisfied.

For a solution to the *free-boundary* problem (Sec. III C), it must be checked that $t_{32}(0)=\beta_1 e_{14} f_1+\beta_2 e_{14} f_2=0$ and that $d_2(0)=i\epsilon_0\varphi(0)$, that is, $\beta_1(e_{14}^2/c_{44})g_1+\beta_2(e_{14}^2/c_{44})g_2=i\epsilon_0(\beta_1 b_1+\beta_2 b_2)$. These conditions are equivalent to checking that

$$\begin{vmatrix} f_1 & f_2 \\ \dfrac{e_{14}^2}{c_{44}\epsilon_0} g_1 - i b_1 & \dfrac{e_{14}^2}{c_{44}\epsilon_0} g_2 - i b_2 \end{vmatrix} = 0. \tag{42}$$

In general, this condition is met only for a limited range of the cut angle θ.

Finally, for a solution to the *thin conducting layer boundary* problem (Sec. III D), it must be checked that $t_{32}(0)=0$ and that $d_2(0)=(\epsilon+i\epsilon_0)\varphi(0)$. These conditions are equivalent to checking that

$$\begin{vmatrix} f_1 & f_2 \\ \dfrac{e_{14}^2}{c_{44}(\epsilon_0 - i\epsilon)} g_1 - i b_1 & \dfrac{e_{14}^2}{c_{44}(\epsilon_0 - i\epsilon)} g_2 - i b_2 \end{vmatrix} = 0. \tag{43}$$

V. EXAMPLES

In this section, the secular equations derived in Sec. III and the tests presented in Sec. IV are employed to find numerically the wave speed and its range of existence for three crystals, one with cubic $\bar{4}3$m symmetry, one with tetragonal $\bar{4}2$m symmetry, and one with cubic 23 symmetry. Data collected from the specialized literature are used for the values of the mass densities, of the stiffnesses, and of the piezoelectric and dielectric constants. In order to graph the depth profiles, a frequency of 100 MHz and a mechanical displacement of 10^{-13} m at $x_2=0$ are picked to fix the ideas. Note that at a 45° angle of cut, the profiles present pure (nonoscillating) exponential decay, because the propagation condition (38) is a biquadratic and the corresponding roots are purely imaginary; they are essentially similar to those displayed by Bright and Hunt.[13] Here, the profiles are computed at angles $\neq 45°$.

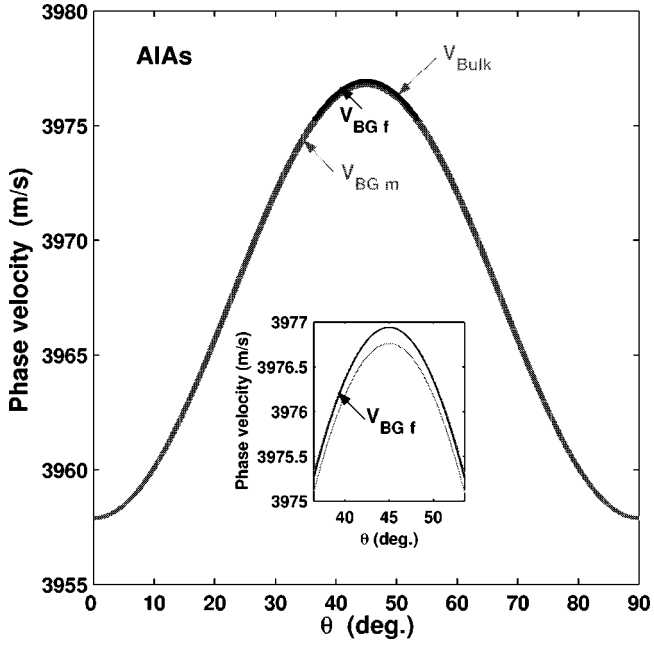


FIG. 2. Speeds of piezoacoustic waves in a AlAs ($\bar{4}3$m symmetry) crystal, as a function of the cut angle: bulk shear wave (upper curve), SH surface wave for free-(unmetallized) boundary conditions (intermediate curve), and SH surface wave for metallized boundary conditions (lower curve).

A. AlAs

For aluminum arsenide ($\bar{4}3$m symmetry) the physical quantities of interest are[27] $\rho=3760\ \mathrm{kg\cdot m^{-3}}$, $c_{44}=58.9\times10^9\ \mathrm{N\cdot m^{-2}}$, $\hat{e}_{14}=-0.225\ \mathrm{C\cdot m^{-2}}$, and $\epsilon_{11}=10.06\epsilon_0$.

Using the results of Secs. III and IV, it is found that the speeds of the piezoacoustic SH surface wave with metallized (Sec. III A) and with free (Sec. III C) boundary conditions are almost indistinguishable on a graph from the speed of the bulk shear wave. For instance at 45°, these speeds are ($\mathrm{m\cdot s^{-1}}$): 3976.784, 3976.965, and 3976.966, respectively. Note however that the SH surface wave for the metallized (''shorted'') boundary condition exists for all values of θ (within the range delimited above by the speed at 45° and below by the speed at 0° and at 90°, which is $3957.890\ \mathrm{m\cdot s^{-1}}$), whereas the SH surface wave for the unmetallized (''free'') boundary condition exists only within a limited range, delimited above by the speed at 45° and below by the speed at $45.0°\pm8.51°$, which is $3975.276\ \mathrm{m\cdot s^{-1}}$.

Figure 2 displays the variations of the three speeds as a function of θ. The speed of the bulk shear wave is always above the speed of the SH surface wave for the metallized boundary condition; they are both defined everywhere. The speed of the SH surface wave for the unmetallized boundary condition is intermediate between these two speeds, but exists only in the range $[36.49°, 53.51°]$. A zoom is provided for this range. In that zoom, the curve for the bulk shear wave almost coincides with the curve for the SH surface wave corresponding to the unmetallized boundary condition; together, they form the upper curve while the lower curve represents the variations of the SH surface wave speed corresponding to the metallized boundary condition.

Note that the simple test for the solution's validity presented in Sec. IV B works perfectly here and hence forward.

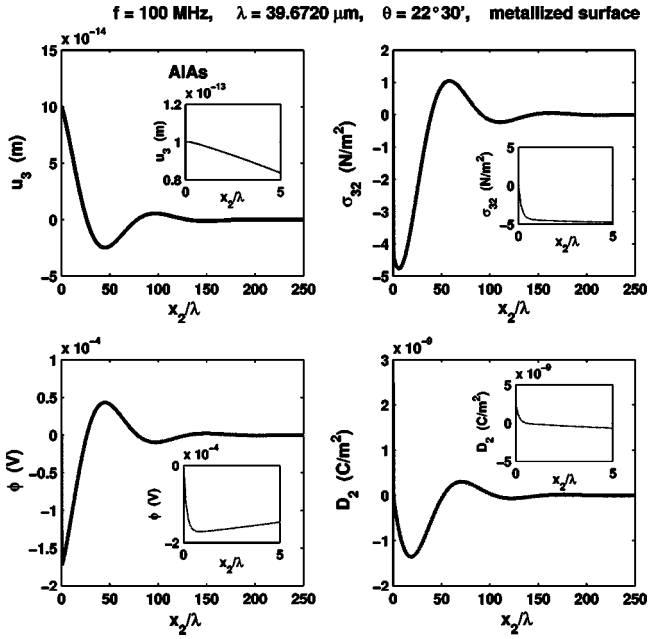


FIG. 3. Depth profiles of the SH surface wave for metallized boundary conditions in a AlAs crystal cut at 22°30′: mechanical displacement, shear stress, electrical potential, electric induction.

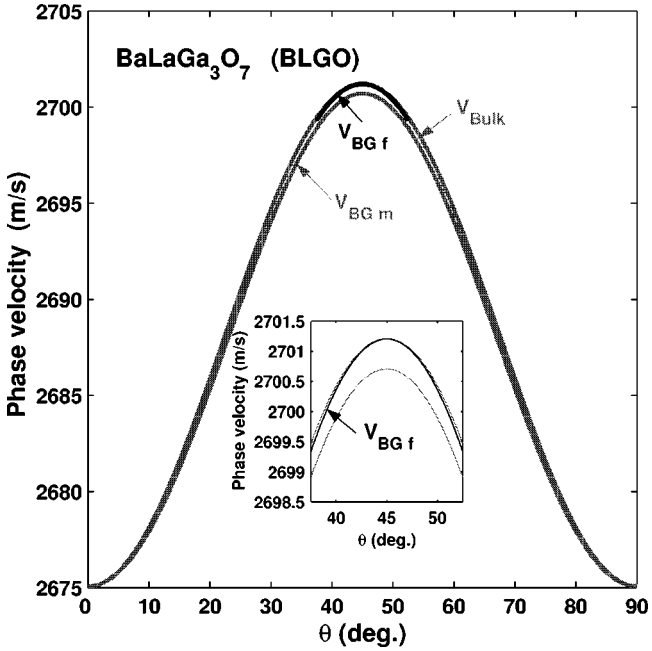


FIG. 4. Speeds of piezoacoustic waves in a $BaLaGa_3O_7$ ($\bar{4}2$m symmetry) crystal, as a function of the cut angle: bulk shear wave (upper curve), SH surface wave for free (unmetallized) boundary conditions (intermediate curve), and SH surface wave for metallized boundary conditions (lower curve).

Thus, using a 40-digit precision under MAPLE for AlAs, the modulus of the determinant in (42) is found to be less than 10^{-19} at $45.0° \pm 8.5125°$ and more than 0.2 at $45.0° \pm 8.5129°$.

Figure 3 shows the variations with depth of the fields of interest (mechanical displacement, shear stress, electrical potential, electric induction) for the SH surface wave corresponding to the metallized boundary condition at $\theta = 22°30'$. The variations of the fields are presented over 250 wavelengths, and zooms are provided for the [0, 5] wavelengths range, where ϕ, σ_{32}, and D_2 undergo rapid changes.

B. BLGO

Soluch *et al.*[28] measured experimentally the elastic, piezoelectric, and dielectric properties of $BaLaGa_3O_7$ ($\bar{4}2$m symmetry, mass density: $\rho = 5450\ \mathrm{kg \cdot m^{-3}}$) as: $c_{44} = 39 \times 10^9\ \mathrm{N \cdot m^{-2}}$, $\hat{e}_{14} = 0.29\ \mathrm{C \cdot m^{-2}}$, and $\epsilon_{11} = 12.4\epsilon_0$.

Here, the speeds of the piezoacoustic SH surface wave with metallized (Sec. III A) and with free-(Sec. III C) boundary conditions differ more notably than in the previous example from the speed of the bulk shear wave. For instance at 45°, these speeds are ($\mathrm{m \cdot s^{-1}}$): 2700.739, 2701.239, and 2701.242, respectively. The range of values for the wave speed v_{BGm} of the metallized shorted boundary condition is delimited above by the speed at 45° and below by the speed at 0° and at 90°, which is $2675.063\ \mathrm{m \cdot s^{-1}}$. For the unmetallized free-boundary condition, the corresponding (limited) range for the wave speed v_{BGf} is bounded above by the speed at 45° and below by the speed at $45.0° \pm 7.485°$, which is $2699.369\ \mathrm{m \cdot s^{-1}}$. The difference between the two speeds is the largest at $\theta = 45°$; there, the ratio $2(v_{\mathrm{BGf}} - v_{\mathrm{BGm}})/v_{\mathrm{BGf}}$ is equal to 3.70×10^{-4}.

Figure 4 displays the variations with θ of the speeds for the bulk shear wave, for the SH surface wave corresponding to the unmetallized boundary condition, and for the SH surface wave corresponding to the metallized boundary condition. Figure 5 shows the variations with depth of the fields of interest (mechanical displacement, shear stress, electrical potential, electric induction) for the SH surface wave corresponding to the metallized boundary condition at $\theta = 22°30'$. Similar comments to those made for Fig. 3 apply.

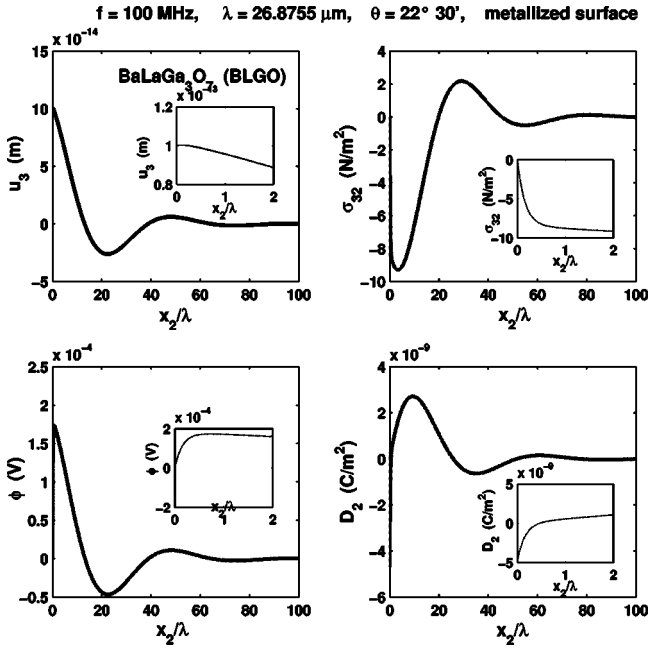


FIG. 5. Depth profiles of the piezoacoustic SH surface wave for metallized boundary conditions in a $BaLaGa_3O_7$ crystal cut at 22°30′: mechanical displacement, shear stress, electrical potential, electric induction.

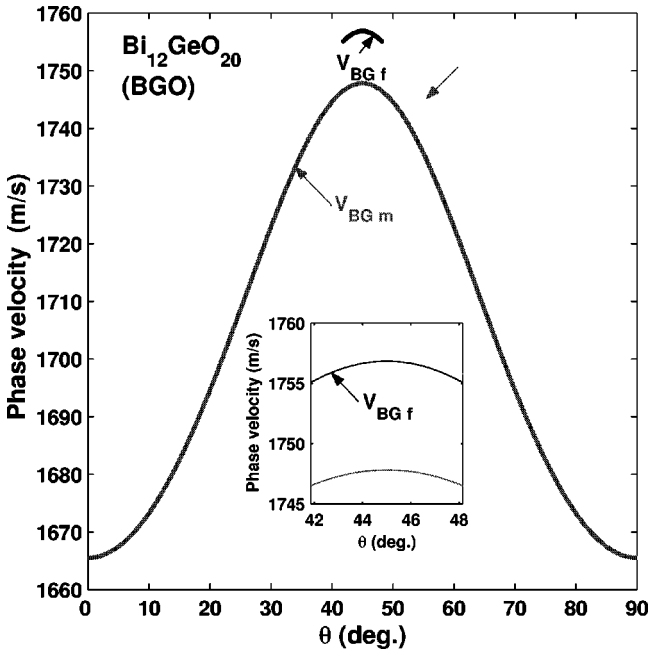


FIG. 6. Speeds of piezoacoustic waves in a $Bi_{12}GeO_{20}$ (23 symmetry) crystal, as a function of the cut angle: bulk shear wave (upper curve), SH surface wave for free (unmetallized) boundary conditions (intermediate curve), and SH surface wave for metallized boundary conditions (lower curve).

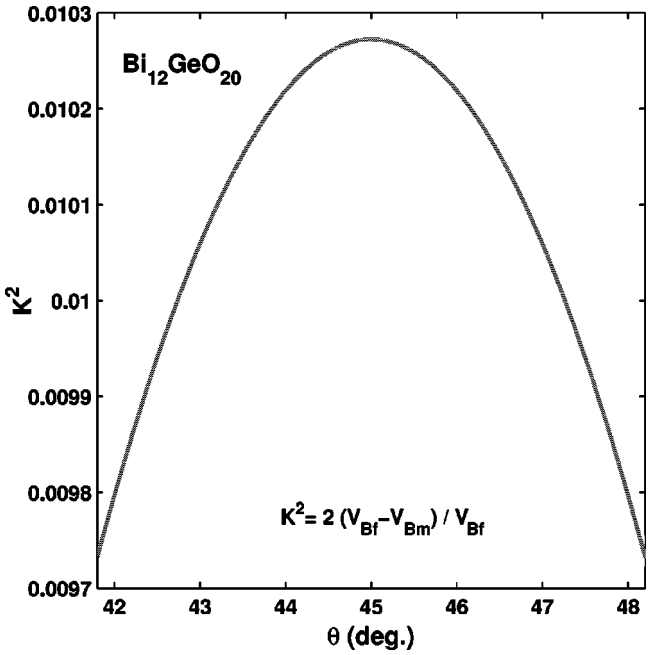


FIG. 7. Variations of $2(v_{BGf}-v_{BGm})/v_{BGf}$ with θ in a $Bi_{12}GeO_{20}$ (23 symmetry) crystal.

C. BGO

The relevant physical quantities of bismuth germanium oxide ($Bi_{12}GeO_{20}$, 23 symmetry) are[29] $\rho=9200\ \mathrm{kg\cdot m^{-3}}$, $c_{44}=25.52\times10^{9}\ \mathrm{N\cdot m^{-2}}$, $\hat{e}_{14}=0.983\ \mathrm{C\cdot m^{-2}}$, and $\epsilon_{11}=38.0\epsilon_0$.

Here, the differences between the speeds of the bulk shear wave, of the SH surface wave corresponding to the metallized boundary condition, and of the SH surface wave corresponding to the unmetallized boundary condition are more marked than in the previous example. At 45°, these speeds are ($\mathrm{m\cdot s^{-1}}$) 1747.812, 1756.836, and 1756.846, respectively. At 0° and at 90°, the speeds of the bulk shear wave and of the SH surface wave corresponding to the metallized boundary condition are both equal to $1665.507\ \mathrm{m\cdot s^{-1}}$. The SH surface wave for the unmetallized boundary condition exists only in the range $45.00\pm3.149°$, and at the extremities of this range, its speed is $1755.068\ \mathrm{m\cdot s^{-1}}$.

Figure 6 displays the variations of the three wave speeds as a function of θ. Figure 7 shows the variations of the quantity $2(v_{BGf}-v_{BGm})/v_{BGf}$ with the angle of cut; its largest (smallest) value is 1.027×10^{-2} at 45.00° (0.975×10^{-2} at 48.149°).

Figure 8 shows the variations with depth of the fields of interest (mechanical displacement, shear stress, electrical potential, electric induction) for the SH surface wave corresponding to the metallized boundary condition at $\theta=22°30'$. Similar comments to those made for Figs. 3 and 4 apply.

Figure 9 shows the variations with depth of the same fields for the SH surface wave corresponding to the unmetallized boundary condition at $\theta=42°30'$. By comparison with the previous figure, it can be seen that the SH surface wave for the unmetallized boundary condition penetrates far more deeply than the SH surface wave for the metallized boundary condition. The electrical potential and the electric induction are plotted inside the crystal for the range [0, 500] wavelengths, and also in the vacuum over the crystal for the range [−10,0] wavelengths; the continuity of these fields across the interface is made apparent with a zoom for the range [−5,5] wavelengths.

VI. CONCLUDING REMARKS

The method of resolution for the title problem of this paper is based on the fundamental equations (15). This

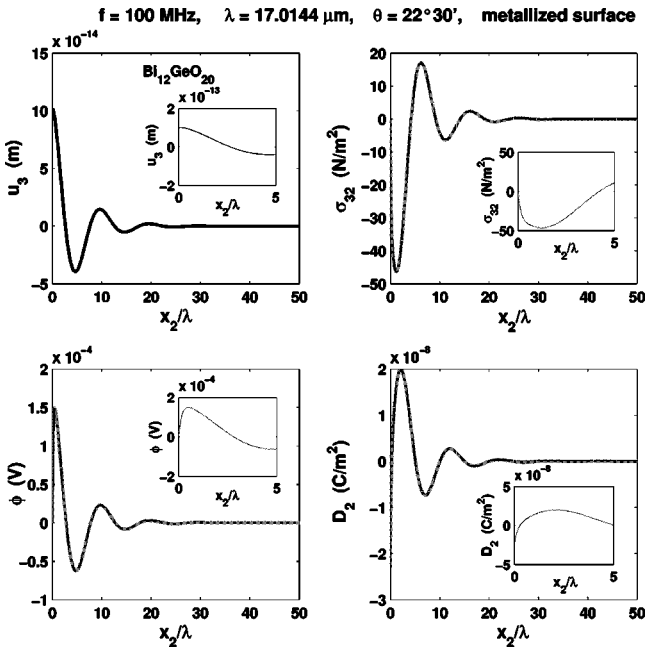


FIG. 8. Depth profiles of the piezoacoustic SH surface wave for metallized boundary conditions in a $Bi_{12}GeO_{20}$ crystal cut at 22°30′: mechanical displacement, shear stress, electrical potential, electric induction.

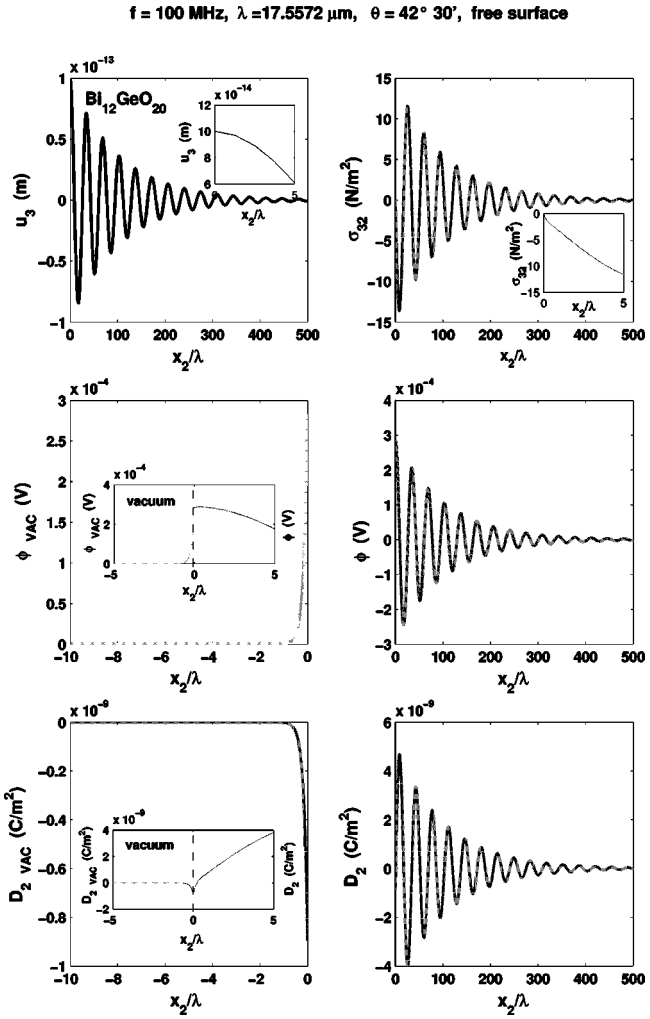


FIG. 9. Depth profiles of the piezoacoustic SH surface wave for free (unmetallized) boundary conditions in a $Bi_{12}GeO_{20}$ crystal cut at 42°30′: mechanical displacement, shear stress, electrical potential, electric induction.

method has proved itself to be very effective and versatile. The end result is the complete analytical elucidation of the problem, for a great variety of surface impedance problems in a piezoelectric half-space, i.e., problems where the electric induction is proportional to the electrical potential: $k\Phi = ivZD_2$ (say) at the boundary plane. The method can be followed through when the impedance Z is zero (short-circuit), infinite (open-circuit), pure imaginary (free-boundary), or complex (thin conducting layer). It has already been used for other types of surface impedance problems for elastic interface waves (Stoneley waves,[20,22] Scholte waves[21]) and could be adapted to configurations[30] with a resistance force proportional to the normal velocity, a mass concentrated in a thin surface layer, a system of elastic oscillators resting on an elastic half-space, a thin elastic layer longitudinally deformable, etc. The method can also accommodate a coupling between elastic and piezoelectric fields, in situations such as the one treated here or for instance, the case of interface acoustic waves at a domain boundary.[31]

Here, attention was restricted to Bleustein–Gulyaev waves in tetragonal $\bar{4}$ piezoelectric crystals. The extension to the classes of orthorhombic 222 or monoclinic 2 crystals is straightforward and only requires the computation of the elements of the matrix $\mathbf{N}$ in Eq. (10). Their general expression is given for instance by Abbudi and Barnett.[32] As an illustration, they are now presented for rhombic 222 crystals.

For such a crystal, the relevant nonzero piezoacoustic constants in the crystallographic coordinate system are $\hat{c}_{44}$, $\hat{c}_{55}$, $\hat{e}_{14}$, $\hat{e}_{25}$, $\hat{\epsilon}_{11}$, and $\hat{\epsilon}_{22}$. In the coordinate system obtained after the rotation Eq. (1), they are

$$
\begin{aligned}
&c_{44}=\hat{c}_{44}\cos^2\theta+\hat{c}_{55}\sin^2\theta, \quad c_{45}=(\hat{c}_{44}-\hat{c}_{55})\cos\theta\sin\theta,\\
&c_{55}=\hat{c}_{55}\cos^2\theta+\hat{c}_{44}\sin^2\theta, \quad \epsilon_{12}=(\hat{\epsilon}_{22}-\hat{\epsilon}_{11})\cos\theta\sin\theta,\\
&\epsilon_{11}=\hat{\epsilon}_{11}\cos^2\theta+\hat{\epsilon}_{22}\sin^2\theta, \quad \epsilon_{22}=\hat{\epsilon}_{22}\cos^2\theta+\hat{\epsilon}_{11}\sin^2\theta,\\
&e_{14}=\hat{e}_{14}\cos^2\theta-\hat{e}_{25}\sin^2\theta, \quad e_{15}=(\hat{e}_{14}+\hat{e}_{25})\cos\theta\sin\theta,\\
&e_{25}=\hat{e}_{25}\cos^2\theta-\hat{e}_{14}\sin^2\theta, \quad e_{24}=-\hat{e}_{15}.
\end{aligned}
\tag{44}
$$

The equations of motion can be cast in the form Eq. (10), where the matrix $\mathbf{N}$ is defined by its 2×2 blocks $\mathbf{N}_1^{(1)}$, $\mathbf{N}_2^{(1)}$, $\mathbf{K}^{(1)}$ in Eq. (11). The components of $-\mathbf{N}_1^{(1)}$ and $\mathbf{N}_2^{(1)}$ are given here by

$$
\begin{bmatrix}
\left(\dfrac{\epsilon_{22}c_{45}}{e_{14}e_{15}}-\dfrac{e_{25}}{e_{14}}\right)\kappa^2 & \left(\dfrac{\epsilon_{22}}{e_{15}}+\dfrac{\epsilon_{12}}{e_{14}}\right)\kappa^2\\
-\left(\dfrac{c_{44}e_{25}}{e_{14}e_{15}}+\dfrac{c_{45}}{e_{14}}\right)\kappa^2 & \left(\dfrac{\epsilon_{12}c_{44}}{e_{14}e_{15}}-1\right)\kappa^2
\end{bmatrix},
$$

$$
\begin{bmatrix}
\dfrac{\epsilon_{22}}{e_{14}e_{15}}\kappa^2 & -\dfrac{1}{e_{14}}\kappa^2\\
-\dfrac{1}{e_{14}}\kappa^2 & -\dfrac{c_{44}}{e_{14}e_{15}}\kappa^2
\end{bmatrix},
\tag{45}
$$

respectively, and those of $\mathbf{K}^{(1)}$ are

$$
\begin{aligned}
K_{11}^{(1)}&=X-c_{55}-\left(c_{44}\frac{e_{25}^2}{e_{14}e_{15}}+2c_{45}\frac{e_{25}}{e_{14}}-c_{45}\frac{\epsilon_{22}c_{45}}{e_{14}e_{15}}\right)\kappa^2,\\
K_{12}^{(1)}&=-e_{15}+\left[\frac{c_{44}\epsilon_{12}}{e_{14}}+\frac{c_{45}\epsilon_{22}}{e_{15}}+e_{25}\left(\frac{\epsilon_{12}c_{44}}{e_{14}e_{15}}-1\right)\right]\kappa^2,\\
K_{22}^{(1)}&=-\epsilon_{11}+\left[\epsilon_{12}\left(\frac{\epsilon_{12}c_{44}}{e_{14}e_{15}}-2\right)-\epsilon_{22}\frac{e_{14}}{e_{15}}\right]\kappa^2.
\end{aligned}
\tag{46}
$$

Here, the quantity κ^2 is defined by

$$
\kappa^2=\frac{e_{14}e_{15}}{\epsilon_{11}c_{44}+e_{15}^2}.
\tag{47}
$$

Finally, in the guise of a conclusion, the main relevant advances toward the full resolution of the problem presented in the paper are recapitulated. In the purely elastic case, the secular equation for Rayleigh waves polarized in a plane of symmetry was derived by Currie[33] who, using an algebraic approach based on the Stroh formalism, obtained the equations

$$
\bar{\mathbf{U}}(0)\cdot\mathbf{K}^{(\mathbf{n})}\mathbf{U}(0)=0,
\tag{48}
$$

where $\mathbf{U}(0)$ is the mechanical displacement on the free surface. Although these equations are also valid in generally anisotropic crystals, his derivation of the secular equation for triclinic (no symmetry) crystals apparently leads to a trivial identity. This problem was later corrected by Taylor and

Currie[34] and by Taziev[35] (see also Ting[36]). In contrast to these approaches based on the formulation of the equations of motion as a first-order differential system for the displacement-traction vector, Mozhaev[37] wrote the equations of motion as a second-order differential system for the displacement vector

$$\alpha \mathbf{U}'' - i\beta \mathbf{U}' + \gamma \mathbf{U} = \mathbf{0}, \tag{49}$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, are real symmetric matrices. Then, using first integrals, he quickly derived the secular equation for orthorhombic crystals. Destrade[38] rewrote the equations of motion, this time in the form

$$\hat{\alpha} \mathbf{t}'' - i\hat{\beta} \mathbf{t}' + \hat{\gamma} \mathbf{t} = \mathbf{0}, \tag{50}$$

for the tractions, where $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, are real symmetric matrices. Adapting Mozhaev's first integrals, he rederived (unaware of Currie's result) the secular equation for Rayleigh waves polarized in a symmetry plane. He also mentioned (and the proof was later given in the review article by Ting[39]) that the method of first integrals could not be used for arbitrary anisotropy when the equations of motion are written as Eq. (49) or Eq. (50). Recently[20–22] he made the connection between Currie's and Taziev's use of integer powers of the Stroh matrix **N** and Mozhaev's first integrals, as shown here also in Sec. II C. Note that Mozhaev and Weihnacht[15] were able to solve the problem of SH surface modes of a 2-mm crystal using first integrals of the piezoacoustic equations written as a second-order differential system.

ACKNOWLEDGMENTS

The learned and helpful comments of the referees are gratefully acknowledged, as well as the generous guidance of V. G. Mozhaev.

[1] J. L. Bleustein, "A new surface wave in piezoelectric materials," Appl. Phys. Lett. **13**, 412–413 (1968).
[2] Yu. V. Gulyaev, "Electroacoustic surface waves in piezoelectric materials," JETP Lett. **9**, 37–38 (1969).
[3] C.-C. Tseng, "Piezoelectric surface waves in cubic crystals," J. Appl. Phys. **41**, 2270–2276 (1970).
[4] K. Nakamura and M. Oshiki, "Theoretical analysis of horizontal shear mode piezoelectric surface acoustic waves in potassium niobate," Appl. Phys. Lett. **71**, 3203–3205 (1997).
[5] M. Kadota, T. Yoneda, K. Fujimoto, T. Nakao, and E. Takata, "Very small-sized resonator IF filter using shear horizontal wave on quartz substrate," IEEE Ultrason. Symp. Proc. **1**, 65–68 (2001).
[6] M. Kadota, J. Ago, H. Horiuchi, and M. Ikeura, "Longitudinally coupled resonator filter using edge reflection of Bleustein–Gulyaev–Shimizu and shear horizontal waves with various bandwidths realized by selecting substrates," Jpn. J. Appl. Phys., Part 1 **40**, 3722–3725 (2001).
[7] G. Koerber and R. F. Vogel, "Generalized Bleustein modes," IEEE Trans. Sonics Ultrason. **SU-19**, 3–8 (1972).
[8] G. G. Koerber and R. F. Vogel, "SH-mode piezoelectric surface waves on rotated cuts," IEEE Trans. Sonics Ultrason. **SU-20**, 10–12 (1973).
[9] L. S. Braginskiĭ and I. A. Gilinskiĭ, "Generalized shear surface waves in piezoelectric crystals," Sov. Phys. Solid State **21**, 2035–2037 (1979).
[10] E. L. Alburque and N. C. Chao, "On the propagation of elastic surface waves in piezoelectric crystals," Phys. Status Solidi B **104**, K11–K14 (1981).
[11] V. R. Velasco, "Surface and interface Bleustein–Gulyaev waves along symmetry directions of cubic crystals," Surf. Sci. **139**, 63–74 (1984).
[12] Yu. A. Kosevich and E. S. Syrkin, "Generalized surface shear waves in piezoelectric crystals," Sov. Phys. Solid State **28**, 134–139 (1986).
[13] V. M. Bright and W. D. Hunt, "Bleustein–Gulyaev waves in gallium arsenide and other piezoelectric cubic crystals," J. Appl. Phys. **66**, 1556–1564 (1989).
[14] G. G. Kessenikh and L. A. Shuvalov, "Bleustein–Gulyaev surface waves in monoclinic piezoelectric crystals," Sov. Phys. Crystallogr. **21**, 1–3 (1976).
[15] V. G. Mozhaev and M. Weihnacht, "Sectors of nonexistence of surface acoustic waves in potassium niobate," IEEE Ultrason. Symp. Proc. **1**, 391–395 (2002).
[16] G. W. Farnell and E. L. Adler, in *Physical Acoustics*, Vol. 9, edited by W. P. Mason and R. N. Thurston (Academic, New York, 1972), pp. 35–127.
[17] E. A. Kraut, "New mathematical formulation for piezoelectric wave propagation," Phys. Rev. **188**, 1450–1455 (1969).
[18] A. N. Stroh, "Dislocations and cracks in anisotropic elasticity," Philos. Mag. **3**, 625–646 (1958).
[19] A. H. Fahmy and E. L. Adler, "Propagation of acoustic surface waves in multilayers: A matrix description," Appl. Phys. Lett. **22**, 495–497 (1973).
[20] M. Destrade, "Elastic interface acoustic waves in twinned crystals," Int. J. Solids Struct. **40**, 7375–7383 (2003).
[21] M. Destrade, "Explicit secular equation for Scholte waves over a monoclinic crystal," J. Sound Vib. **273**, 409–414 (2004).
[22] M. Destrade, "On interface waves in misoriented prestressed incompressible elastic solids," IMA J. Appl. Math. (in press).
[23] K. A. Ingebrigsten, "Surface waves in piezoelectric," J. Appl. Phys. **40**, 2681–2686 (1969).
[24] J. Lothe and D. M. Barnett, "Integral formalism for surface waves in piezoelectric crystals. Existence considerations," J. Appl. Phys. **47**, 1799–1807 (1976).
[25] E. Dieulesaint and D. Royer, *Elastic Waves in Solids. Applications to Signal Processing* (Wiley, New York, 1980).
[26] D. Royer and E. Dieulesaint, *Ondes Elastiques dans les Solides. I. Propagation libre et guidée (Elastic Waves in Solids. Free and Guided Propagation.)* (Masson, Paris, 1996).
[27] Y. Kim and W. D. Hunt, "Acoustic fields and velocities for surface-acoustic-wave propagation in multilayered structures: An extension of the Laguerre polynomial approach," J. Appl. Phys. **68**, 4993–4997 (1990).
[28] W. Soluch, R. Ksiezopolski, W. Piekarczyk, M. Berkowski, M. A. Goodberlet, and J. F. Vetelino, "Elastic, piezoelectric, and dielectric properties of the $BaLaGa_3O_7$ crystal," J. Appl. Phys. **58**, 2285–2287 (1985).
[29] B. A. Auld, *Acoustic Fields and Waves in Solids* (Wiley, New York, 1973).
[30] S. Kaliski, "On the existence of transversal surface waves in an elastic half-space with surface impedance," Bull. Acad. Pol. Sci., Ser. Sci. Tech. **15**, 467–472 (1967).
[31] V. G. Mozhaev and M. Weihnacht, "Interface acoustic waves at a 180 degrees domain boundary in tetragonal barium titanate," J. Korean Phys. Soc. **32**, S747–S749 (1998).
[32] M. Abbudi and D. M. Barnett, "On the existence of interfacial (Stoneley) waves in bonded piezoelectric half-spaces," Proc. R. Soc. London, Ser. A **429**, 587–611 (1990).
[33] P. K. Currie, "The secular equation for Rayleigh waves on elastic crystals," Q. J. Mech. Appl. Math. **32**, 163–173 (1979).
[34] D. B. Taylor and P. K. Currie, "The secular equation for Rayleigh waves on elastic crystals. II. Corrections and additions," Q. J. Mech. Appl. Math. **34**, 231–234 (1981).
[35] R. M. Taziev, "Dispersion relation for acoustic waves in an anisotropic elastic half-space," Sov. Phys. Acoust. **35**, 535–538 (1989).
[36] T. C. T. Ting, "The polarization vector and secular equation for surface waves in an anisotropic elastic half-space," Int. J. Solids Struct. **41**, 2065–2083 (2004).
[37] V. G. Mozhaev, "Some new ideas in the theory of surface acoustic waves in anisotropic media," in *IUTAM Symposium on Anisotropy, Inhomogeneity and Nonlinearity in Solids*, edited by D. F. Parker and A. H. England (Kluwer, Dordrecht, 1995), pp. 455–462.
[38] M. Destrade, "The explicit secular equation for surface acoustic waves in monoclinic elastic crystals," J. Acoust. Soc. Am. **109**, 1398-1402 (2001).
[39] T. C. T. Ting, "Explicit secular equations for surface waves in an anisotropic elastic half-space: from Rayleigh to today," In *Proc. NATO Adv. Res. Workshop on Surface Waves in Anisotropic and Laminated Bodies and Defects Detection, Moscow, Russia, 7–9 February 2002*, edited by G. A. Maugin and R. V. Goldstein (Kluwer, Dordrecht, 2004).

An iterative effective medium approximation (IEMA) for wave dispersion and attenuation predictions in particulate composites, suspensions and emulsions

D. G. Aggelis, S. V. Tsinopoulos, and D. Polyzos[a)]
Department of Mechanical Engineering and Aeronautics, University of Patras, P.O. Box 1401, Panepistimioupolis, Rion 26504, Greece, and Institute of Chemical Engineering and High Temperature Chemical Processes (FORTH/ICE-HT), Patras 26500, Greece

(Received 10 May 2004; revised 6 September 2004; accepted 7 September 2004)

In the present work we deal with the scattering dispersion and attenuation of elastic waves in different types of nonhomogeneous media. The iterative effective medium approximation based on a single scattering consideration, for the estimation of wave dispersion and attenuation, proposed in Tsinopoulos *et al.*, [Adv. Compos. Lett. **9**, 193–200 (2000)] is examined herein not only for solid components but for liquid suspensions as well. The iterations are conducted by means of the classical relation of Waterman and Truell, while the self-consistent condition proposed by Kim *et al.* [J. Acoust. Soc. Am. **97**, 1380–1388 (1995)] is used for the convergence of the iterative procedure. The single scattering problem is solved using the Ying and Truell formulation, which with a minor modification can accommodate the solution of scattering on inclusions in liquid. Theoretical results for several different systems of particulates and suspensions are presented being in excellent agreement with experimental data taken from the literature.
[DOI: 10.1121/1.1810273]

PACS numbers: 43.38.Ja, 43.38.Kb [YHB] Pages: 3443–3452

I. INTRODUCTION

When a plane wave travels through a suspension of particles like particulate composites (solid particles in solids), liquid suspensions (solid particles in fluid), and emulsions (fluid inclusions in fluid), multiple scattering occurs and part of the incident energy is transferred to the scattered fields. Parameters such as the frequency of the incident wave, the relative position among the particles, the geometry of the particles and the material properties of both matrix and inclusions affect the amount of this energy. Thus, although matrix and particles can be nonattenuative, the amplitude of waves propagating through suspensions decays and the decay rate is frequency dependent. For a plane wave the decay of its amplitude is expressed via a frequency dependent exponential coefficient known as an attenuation coefficient. On the other hand, the size of the particles as well as the material mismatch between particles and surrounding medium imply that the dynamic behavior of the composite medium is strongly depended on the excitation frequency of the incident wave. Macroscopically this means that the phase velocity of a plane wave traveling through a suspension of particles is frequency dependent. This phenomenon is known in the literature as wave dispersion.

The quantitative determination of dispersion and attenuation of a plane wave, caused by a random distribution of inhomogeneities, is a problem which has been studied intensively either theoretically or experimentally by many investigators in the past. The first important theoretical work on the subject is that of Foldy[1] who, employing a configurational averaging procedure, derived a dispersion relation for scalar wave propagation through a medium containing isotropic scatterers. Later, Lax[2] extended the work of Foldy and proposed a new dispersion relation for multiple wave scattering by anisotropic scatterers. In both works the wave dispersion and attenuation was represented via a frequency dependent complex wave number expressed in terms of the particle concentration and the forward far field scattering amplitude taken from the solution of the single particle wave scattering problem. The results of Lax were further improved by Waterman and Truell,[3] Twersky,[4] Lloyd and Berry,[5] Varadan *et al.*[6] and Javanaux and Tomas[7] who derived dispersion relation expressed in terms of the particle concentration and the forward as well as the backward scattering amplitude of the single scattering problem inserting thus the contribution of the back-scattering to the multiple scattering process.

The above mentioned multiple scattering theories have been extensively exploited by many investigators in order to explain wave dispersion and attenuation observed in experiments dealing with wave propagation in nonhomogeneous fluids and solids. Here one can mention the representative works of Sayers and Smith,[8] Ledbetter and Datta,[9] Norris,[10] Anson and Chivers,[11] Shido *et al.*,[12] Lu and Liaw[13] and Challis *et al.*[14] for particulate composites, the works of Holmes *et al.*,[15] Mobley *et al.*,[16] Meulen *et al.*,[17] for elastic particles in liquid suspensions and the works of McClemens and Povey[18] and McClemens[19] for emulsions. In most of these articles, spherical inclusions are considered while the far field parameters of the single particle wave scattering problem, used in the dispersion and attenuation expressions, are mainly taken from the works of Epstein and Carhart[20] for emulsions, Allegra and Hawley[21] for elastic particles in a liquid continuum and Ying and Truell[22] for suspensions of

a)Electronic mail: polyzos@mech.upatras.gr

solids in solids. Comparisons showed that, for cases of particulate composites with a significant mismatch between the physical properties of particles and matrix, the aforementioned multiple scattering theories predict well only for very low concentrations of particles (less than 10%), while their prediction efficiency, in terms of particles concentration, can be improved in cases of nonhomogeneous solids with small differences in the physical properties of the material constituents. On the other hand, the simple multiple scattering theories of Foldy,[1] Waterman and Truell[3] and Lloyd and Berry[5] enhanced by the Epstein and Carhart[20] and Allegra and Hawley[21] models, where besides the interaction of the spherical particle with the incident wave, heat transport phenomena between particles and surrounding medium are taken into account, provide reasonable predictions for liquid suspensions and emulsions with concentrations up to 20%. This is an expected result since, due to mode conversion, multiple scattering effects are more pronounced in solid than in liquid suspensions. Eventually, one can say here that none of the so far mentioned theories is able to provide acceptable wave dispersion and attenuation predictions for all the types of suspensions and for a wide range of particle concentrations and wavenumbers.

Besides the aforementioned fundamental multiple scattering procedures, many analytical semi-analytical and numerical models for predicting wave dispersion and attenuation in nonhomogeneous media have been proposed in the literature. Among them, the methodologies applied for both solid and liquid suspensions can be grouped into two categories. In the first category belong the works that provide dispersion and attenuation expressions by means of the Kramers–Kroning relations. Representative works are those of Beltzer *et al.*[23] and Beltzer[24] for particulate composites and the works of Temkin,[25] Ye[26] and Leander[27] for suspensions, while an excellent mathematical description and derivation of Kramers–Kroning relations can be found in the work of Weaver and Pao.[28] However, as it is mentioned in the book of Zhang and Gross[29] and noted in the paper of Temkin,[30] Kramers–Kroning relations provide satisfactory results only for low concentrations of particles while one of the quantities phase velocity and attenuation coefficient should be known independently. The second category concerns the self-consistent theories. According to these theories, the frequency dependent wave velocity and attenuation coefficient are evaluated through self-consistent expressions most of which are based on scattering parameters taken from the solution of the single scattering problem where the microstructure of the composite medium is immersed into an infinitely extended effective medium. The self-consistent or effective medium theories appear in the literature with different versions and procedures depending on the type of suspensions they applied. Thus, for particulate composites one can mention the self-consistent models of Talbot and Willis,[31] Sabina and Willis[32] and Devaney,[33] the effective medium approximations of Kerr[34] and Kanaun *et al.*,[35] the dynamic self-consistent effective medium approximations of Berryman,[36] Kim *et al.*[37] and Tsinopoulos *et al.*[38] and the incremental self-consistent approach of Anson and Chivers[11] and Biwa *et al.*[39] For liquid suspensions, one can mention the effective medium approaches of Anson and Chivers,[40] Hemar *et al.*,[41] Cowan *et al.*,[42] McClements *et al.*[43] and Hipp *et al.*[44,45] and the coupled-phase models of Harker *et al.*,[46] Atkinson and Kytomaa[47] and Evans and Attenborough.[48] Comparisons with experimental results have shown that the self-consistent models are those which are able to predict satisfactory the behavior of a wave pulse propagating within a dense distribution of particle-scatterers.

Recently, Kim *et al.*[37] presented a modified version of the coherent potential approximation,[49–52] in order to predict the speed and the coherent attenuation of an elastic wave propagating in a medium containing randomly distributed, solid spherical inclusions. The frequency dependent effective stiffness and density of the composite are obtained by solving a system of three nonlinear volume-integral equations in which, however, the interior dynamic displacement field of a single inclusion immersed in an infinitely extended effective medium must be known *a priori*. Although in their theory correlations among the scatterers are neglected, their results were in a good agreement with experimental observations.

Kanaun *et al.*[35] claim that the application of this effective medium scheme for wave propagation problems is questionable, since matrix and inclusions play quite different roles in the process of wave diffraction. However, scattering occurs due to the interaction of the incident wave with the randomly distributed particles. Thus, considering inclusions and matrix as scatterers the surfaces of which have opposite unit normal vector, the self-consistent hypothesis of Kim *et al.*[37] seems to be quite reasonable

Later, Tsinopoulos *et al.*[38] proposed an iterative effective medium approximation (IEMA) combining effectively the self-consistent model of Kim *et al.*[37] and the simple multiple scattering theory of Foldy.[1] In their work, the evaluation of the wave speed and attenuation coefficient was accomplished through a practical and simple iterative procedure avoiding thus the solution of complex nonlinear systems of equations such those required in the approximation of Kim *et al.* Moreover, comparing the estimations provided by the two methods, IEMA appears to be more efficient and accurate in cases of highly concentrated elastic mixtures.

Here, the IEMA of Tsinopoulos *et al.*[38] is properly modified and improved in order to predict well wave dispersion and attenuation in particulate composites, particle suspensions and emulsions. Our aim in the current work is twofold: first to develop a single theoretical model that predicts well wave dispersion and attenuation for all types of suspensions and second to provide an iterative computational scheme that for the case of spherical particles is simple and easily implemented. The present new version of IEMA combines the self-consistent model of Kim *et al.*[37] with the quasicrystalline approximation of Waterman and Truell.[3] Considering the effective material properties of the composite medium being the same with the static elastic ones proposed by Christensen,[53] properly modified for liquid mixtures, and satisfying the single scattering self-consistent condition of Tsinopoulos *et al.*,[38] the effective and frequency dependent dynamic density of the nonhomogeneous medium is evaluated. The complex value of the effective density in conjunction with the static effective stiffness of the composite medium

determine both the velocity and the attenuation of an ultrasonic pulse propagating in the random particulate suspension. The single scattering problem is solved using the Ying and Truell[22] formulation, which with minor modification can accommodate solution of scattering problems dealing with inclusions suspended in liquid matrix. Several numerical results compared with experimental data taken from the literature demonstrate the IEMA efficiency on predicting wave dispersion and attenuation in all types of particle suspensions.

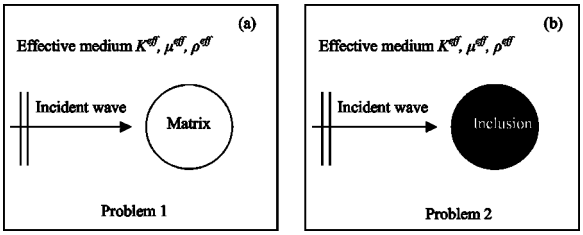


FIG. 1. A plane mean wave propagating in the effective medium and scattered by (a) a matrix inclusion (problem 1) and (b) a particle inclusion (problem 2).

II. THE IEMA FOR PARTICLE SUSPENSIONS

In this section, the IEMA proposed recently by Tsinopoulos *et al.*[38] and modified for the needs of the present work is presented.

The starting point of the IEMA is a self-consistent condition first considered in the coherent potential theory of Soven.[49] According to this theory, any wave propagating in a composite medium can be considered as a sum of a mean wave propagating in a medium having the dynamic effective properties of the composite and a number of fluctuating waves coming from the multiple scattering of the mean wave by the uniformly and randomly distributed material variations from these of the effective medium. On the average, the fluctuating field should be vanished at any direction within the effective medium, i.e.,

$$\langle \hat{\mathbf{k}} \cdot \tilde{\mathbf{T}} \cdot \hat{\mathbf{k}} \rangle = 0, \tag{1}$$

where $\langle\ \rangle$ denotes the average over the composition and the shape of the scatterers, $\tilde{\mathbf{T}}$ is a matrix corresponding to the total multiple scattering operator for the fluctuating waves and $\hat{\mathbf{k}}$ is the propagation direction of the mean wave. Equation (1) is well known as self-consistent condition and can be used to determine the dynamic effective properties of the composite material. However, due to the prohibitive computational cost of the evaluation of the operator $\tilde{\mathbf{T}}$ Soven[49] proposed, instead of Eq. (1), the use of the following simplified self-consistent condition:

$$\langle \hat{\mathbf{k}} \cdot \tilde{\mathbf{t}} \cdot \hat{\mathbf{k}} \rangle = 0, \tag{2}$$

with $\tilde{\mathbf{t}}$ being a single scattering operator coming from the diffraction of the mean wave by each composition, i.e. matrix and particles, embedded in an infinitely extended effective medium. Devaney[33] proved that Eq. (2) could also be written as a function of the far field scattering amplitudes in the forward direction. Thus, for identical homogeneous particles embedded in a homogeneous elastic or liquid matrix, Eq. (2) assumes the following form:

$$n_1 g^{(1)}(\hat{\mathbf{k}},\hat{\mathbf{k}}) + (1-n_1) g^{(2)}(\hat{\mathbf{k}},\hat{\mathbf{k}}) = 0, \tag{3}$$

where n_1 represents the volume fraction of the particles and $g^{(1)}(\hat{\mathbf{k}},\hat{\mathbf{k}})$, $g^{(2)}(\hat{\mathbf{k}},\hat{\mathbf{k}})$ are the forward scattering amplitudes taken by the solution of the two single wave scattering problems illustrated in Fig. 1. The solution of the single scattering problem is described in the next section.

According to the IEMA the self-consistent condition (3) is satisfied numerically through an iterative procedure, which can be summarized as follows.

Consider a harmonic elastic plane wave with circular frequency ω, either longitudinal (P) or transverse (SH or SV), traveling through the composite. Due to the presence of the particles, multiple scattering occurs and thereby the considered wave becomes both dispersive and attenuated and its complex wavenumber $k_d^{\text{eff}}(\omega)$ can be written as

$$k_d^{\text{eff}}(\omega) = \frac{\omega}{C_d^{\text{eff}}(\omega)} + i\alpha_d^{\text{eff}}(\omega), \tag{4}$$

with $C_d^{\text{eff}}(\omega)$ and $\alpha_d^{\text{eff}}(\omega)$ being the frequency dependent wave phase velocity and attenuation coefficient, respectively. The subscript d denotes either longitudinal $(d \equiv p)$ or transverse $(d \equiv s)$ wave.

Next, the composite material is replaced by an elastic homogeneous and isotropic medium with effective Lame' constants μ^{eff}, λ^{eff}, given by the static model of Christensen,[53]

$$\lambda^{\text{eff}} = \lambda_2 + \frac{n_1(\lambda_1-\lambda_2)\left(\lambda_2+\frac{4}{3}\mu_2\right)}{n_2(\lambda_1-\lambda_2)+\left(\lambda_2+\frac{4}{3}\mu_2\right)},$$
$$A\left(\frac{\mu^{\text{eff}}}{\mu_2}\right)^2 + 2B\left(\frac{\mu^{\text{eff}}}{\mu_2}\right) + C = 0. \tag{5}$$

Subscripts 1 and 2 indicate particle and matrix material properties, respectively, and A, B and C are functions of (μ_1, μ_2, n_1) given in the paper of Christensen.[53] Since the Lame' constant λ is usually referred to an elastic medium, in the present work where liquid suspensions and emulsions are considered the bulk modulus $K^{\text{eff}} = \lambda^{\text{eff}} + (2/3)gm^{\text{eff}}$ is used instead.

For the cases of a liquid matrix, the shear modulus, μ^{eff}, instead of being calculated through (5), is set to a very small value, since, even for high concentrations, the inclusions do not form an interconnected network that would effectively reinforce the shear rigidity of the mixture. In the present paper the shear modulus for all the considered liquid phases has been taken equal to 100 Pa.

In the first step of the IEMA, the effective density of the composite is assumed to be

$$(\rho^{\text{eff}})_{\text{step1}} = n_1\rho_1 + (1-n_1)\rho_2. \tag{6}$$

Then, the effective wave number $(k_d^{\text{eff}})_{\text{step1}}$ is evaluated as straightforward through the relations

$$(k_p^{\text{eff}})_{\text{step1}} = \omega \left[\frac{3K^{\text{eff}} + 4\mu^{\text{eff}}}{3(\rho^{\text{eff}})_{\text{step1}}} \right]^{-1/2}, \tag{7}$$

for a P-wave and

$$(k_s^{\text{eff}})_{\text{step1}} = \omega \left[\frac{\mu^{\text{eff}}}{(\rho^{\text{eff}})_{\text{step1}}} \right]^{-1/2}, \tag{8}$$

for a shear wave, respectively.

In the sequel, utilizing the material properties obtained from the first step, the two single wave scattering problems illustrated in Fig. 1 are solved. The solution of these problems is accomplished analytically from the matrix notation of the Ying and Truell formulation, as will be explained in Sec. III. Combining the evaluated forward scattering amplitudes $g_d^{(1,2)}(\hat{\mathbf{k}},\hat{\mathbf{k}})$, according to the self-consistent condition (3), i.e.,

$$g_d(\hat{\mathbf{k}},\hat{\mathbf{k}}) = n_1 g_d^{(1)}(\hat{\mathbf{k}},\hat{\mathbf{k}}) + (1-n_1) g_d^{(2)}(\hat{\mathbf{k}},\hat{\mathbf{k}}), \tag{9}$$

and making use of the dispersion relation proposed by Waterman and Truell,[3] one obtains the new effective wave number of the mean wave,

$$[(k_d^{\text{eff}})_{\text{step2}}]^2 = [(k_d^{\text{eff}})_{\text{step1}}]^2 + \frac{3n_1 g_d(\hat{\mathbf{k}},\hat{\mathbf{k}})}{a^3} + \frac{9n_1^2 (g_d^2(\hat{\mathbf{k}},\hat{\mathbf{k}}) - g_d^2(\hat{\mathbf{k}},-\hat{\mathbf{k}}))}{4k^2 a^6}, \tag{10}$$

where a is the radius of the smallest sphere including the particle.

The new complex wave number $(k_d^{\text{eff}})_{\text{step2}}$ of the mean wave propagating through the composite medium is the departure point of the second step. Keeping the same static material properties (5) for the effective medium and utilizing relations (7) and (8) for longitudinal and transverse incidence, respectively, one calculates the new effective density of the host medium $(\rho^{\text{eff}})_{\text{step2}}$, which due to $(k_d^{\text{eff}})_{\text{step2}}$ is now complex. Considering the new material properties $\lambda_{\text{eff}}, \mu_{\text{eff}}$ and $(\rho^{\text{eff}})_{\text{step2}}$, the two single wave scattering problems depicted in Fig.1 are solved again and the procedure is repeated until the self-consistent condition (3) is satisfied. This means that $(k_d^{\text{eff}})_{\text{step}(n-1)} = (k_d^{\text{eff}})_{\text{step}(n)}$. Finally, the evaluated k_d^{eff} in conjunction with Eq. 4 determines the frequency dependent, effective velocity $C_d^{\text{eff}}(\omega)$ and the attenuation coefficient $a_d^{\text{eff}}(\omega)$ of the propagating wave. The whole procedure is summarized in the flow chart of Fig. 2.

In the just described procedure, a point that needs further discussion is the use of the complex density throughout the iterations of the IEMA. From a physical point of view, one can say that the choice of using the density as the main parameter controlling the material properties of a particle suspension seems to be realistic, since both dispersion and attenuation are dynamic properties of the considered composite medium. On the other hand the idea of a complex density is not something new in the literature. As a representative example one can mention the works of Petculescu and Wilen,[54] Lee *et al.*[55] and Pan and Horne[56] dealing with sound

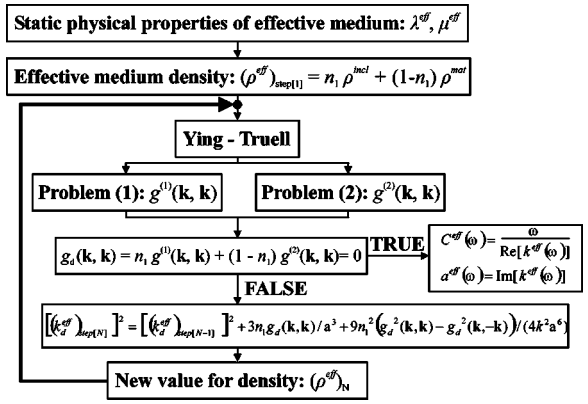


FIG. 2. A schematic representation of IEMA.

propagation and fluid flow in porous media and that employ complex densities in their frequency domain analysis. Usually, the imaginary and real part of a complex density come from the frequency domain transformation of first and second order time derivatives, respectively, involved in the differential operator of the problem. Thus, beyond the different type of explanations being available in the literature, the real and imaginary part of a complex density are directly related to the kinetic and the absorbing energy of the medium, respectively. This could explain why in the proposed here iterative methodology the complex density is responsible for the final evaluation of the frequency dependent velocity and attenuation coefficient of the particulate composite medium.

It should be also mentioned that an alternative IEMA procedure would be the consideration of a constant density [Eq. (6)] for all the steps and the use of either the complex values of the bulk modulus evaluated from Eq. (7) when longitudinal waves propagate through the composite medium, or the complex shear modulus obtained from Eq. (8) when shear waves are considered. However, although the two procedures seem to be equivalent, the use of the complex modulus instead of the complex density leads to dispersion and attenuation predictions that in many cases are in poor agreement with experimental observations. On the contrary and as it is evident in the sections after next, the use of the complex density in the IEMA procedure provides results being in a very good and sometimes in excellent agreement with the available experimental data.

III. FORMULATION AND SOLUTION OF THE SINGLE SCATTERING PROBLEM

In this section the formulation and solution of the single scattering problem is briefly described. The present approach is based on the Ying and Truell formulation[22] considering scattering of a plane wave on an elastic sphere embedded in an infinite elastic matrix. The subject of scattering on a spherical obstacle has been discussed extensively in literature so only general guidelines will be presented here.

When a compressional wave impinges on the particle it gives rise to both compressional and shear waves inside the particle, as well as the scattered compressional and shear waves outside the particle. Expressions for each of these waves are equated using the boundary conditions at the sur-

face of the particle, yielding four equations with four unknown scattering coefficients, A_n, B_n, C_n and D_n. In the present formulation temperature and heat transfer effects are not included. The equations, concerning namely the continuity of the normal and tangential velocity component as well as the continuity of the normal and tangential stress component are as follows:

$$A_n k_1 a h_{n+1}(k_1 a) + B_n n \kappa_1 a h_{n+1}(\kappa_1 a) - C k_2 a j_{m+1}(k_2 a) - D_n n \kappa_2 a j_{m+1}(k_2 a) = (-i)^{n-1}(2n+1)\frac{1}{k_1}[k_1 a j_{m+1}(k_1 a)], \quad (11)$$

$$A_n h_n(k_1 a) - B_n[(n+1)h_n(\kappa_1 a) - \kappa_1 a h_{n+1}(\kappa_1 a)] - C_n j_m(k_2 a) + D_n[(n+1)j_n(\kappa_2 a) - \kappa_2 a j_{n+1}(\kappa_2 a)] = (-i)^{n-1}(2n+1)\frac{1}{k_1} j_m(k_1 a), \quad (12)$$

$$A_n[(\kappa_1 a)^2 h_n(k_1 a) - 2(n+2)k_1 a h_{n+1}(k_1 a)] + B_n n[(\kappa_1 a)^2 h_n(\kappa_1 a) - 2(n+2)\kappa_1 a h_{n+1}(\kappa_1 a)] - C_n p[(\kappa_2 a)^2 j_n(k_2 a) - 2(n+2)k_2 a j_{n+1}(k_2 a)] - D_n p n[(\kappa_2 a)^2 j_n(\kappa_2 a) - 2(n+2)\kappa_2 a j_{n+1}(\kappa_2 a)] = (-i)^{n-1}(2n+1)\frac{1}{k_1}[(\kappa_1 a)^2 j_m(k_1 a) - 2(n+2) \times k_1 a j_{m+1}(k_1 a)], \quad (13)$$

$$A_n[(n-1)h_n(k_1 a) - k_1 a h_{n+1}(k_1 a)] - B_n\left[\left(n^2 - 1 - \frac{\kappa_1^2 a^2}{2}\right)h_n(\kappa_1 a) - \kappa_1 a h_{n+1}(\kappa_1 a)\right] - C_n p[(n-1)j_n(k_2 a) - k_2 a j_{n+1}(k_2 a)] + D_n p\left[\left(n^2 - 1 - \frac{\kappa_2^2 a^2}{2}\right)j_n(\kappa_2 a) - \kappa_2 a j_{n+1}(\kappa_2 a)\right] = (-i)^{n-1}(2n+1)\frac{1}{k_1}[(n-1)j_m(k_1 a) - k_1 a j_{m+1}(k_1 a)], \quad (14)$$

where k_1 and k_2 are the longitudinal wavenumbers in the matrix and inclusion, respectively, κ_1 and κ_2 are the shear wavenumbers in the matrix and inclusion, respectively, $p = \mu_2/\mu_1$, a is the particle radius, and j_n and h_n are the spherical Bessel and Hankel functions.

In case the modeling concerns a problem of scattering on particles suspended in liquid, the equations can be derived by a limiting process ($\mu_2 \to 0$).[59] As it is already mentioned in the present work the shear modulus for any liquid phase was taken to be 100 Pa.

In order to calculate velocity and attenuation for a given frequency the equations must be solved for the scattering coefficients for each value of n. In the case of the scattered longitudinal wave, the A_n coefficients are of interest. This system of equations, in matrix notation, is solved by a Matlab routine using standard inversion command. Therefore, the forward scattering amplitude, g, can be calculated through

$$g(0) = \frac{1}{ik_1}\sum_{n=0}^{\infty}(2n+1)A_n, \qquad g(\pi) = \frac{1}{ik_1}\sum_{n=0}^{\infty}(-1)^n(2n+1)A_n. \quad (15)$$

The appropriate order of n has been shown by O'Neil *et al.*[60] to be roughly equal to the dimensionless wavenumber, $k_1 a$, meaning that for higher frequency bands as the particle radius rises significantly compared to the wavelength, and in order to have a reliable calculation of the scattering amplitude and therefore velocity and attenuation, more scattering terms must be summed in Eq. (15).

Indeed, this is evident throughout the present work as can be seen in Fig. 3. There, two indicative examples concerning the order of n, necessary for the convergence of velocity via the Waterman and Truell dispersion relation are depicted. In Fig. 3(a) where a case of a lead/epoxy particulate composite is considered, it is seen that for frequency 1.9 MHz corresponding to $k_1 a = 3$, for particle radius 660 μm, the velocity obtains a constant value after about $n = 6$. In Fig. 3(b) the medium is a polystyrene in the saline suspension and the appropriate number of n for convergence is 16 while the $k_1 a$ equals 12.5 (radius 152 μm and frequency 20 MHz). It is seen that generally velocity and attenuation converges for order $k_1 a + 3$ while thereafter no detectable change is mentioned. Therefore n was set equal to the integer of $k_1 a + 7$ for the needs of the present study.

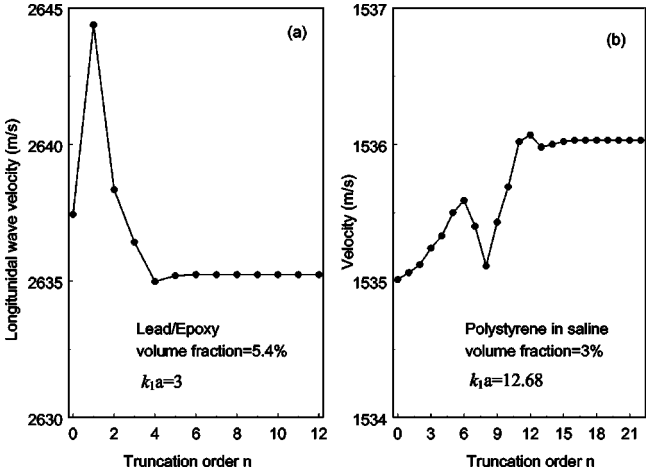


FIG. 3. (a) Velocity vs order n for 5.2% by volume composite of 660 μm radius lead spheres in epoxy. (b) Velocity vs order n for 3% by volume suspension of 152 μm radius polystyrene spheres in water.

IV. RESULTS AND DISCUSSION

In this section the prediction capability of the IEMA is examined. Although it has been proven to yield accurate results for particulate composites[38] for volume concentrations as high as 50%, herein the efficiency of IEMA is addressed also for liquid matrix systems. Calculations are carried out for several cases of systems for which experimental data are

TABLE I. Material properties.

Material	Cp (m/s)	Cs (m/s)	λ (Gpa)	μ (Gpa)	ρ (Kg/m^3)	Fig.
Iron	5941	3251	110.99	82.86	7840	4(a)
PMMA	2669	1305	4.37	2.00	1175	
TiC	10000	6200	113.28	188.36	4900	4(b)
Epoxy	2523	976	5.58	1.19	1250	
Lead	2210	860	38.48	8.36	11300	5
Epoxy (828z)	2640	1200	4.92	1.73	1202	
Al7091	6305	3066	59.5	26.7	2840	6
SiCp	12210	7707	100.0	196.0	3300	
Glass	5280	3240	17.15	26.16	2492	7
Epoxy (3012)	2541	1161	4.44	1.59	1180	
Polystyrene	2337	1098	3.21	1.27	1053	8, 9(b), 10
Water	1500	—	2.250	–	1000	8, 9(b), 10, 12, 13, 14(b)
Glass	6790	4167	27.36	41.76	2405	9(a)
ATB	1026	—	2.49	–	2365	
Glass	5600	3400	20.6	28.9	2500	11
Glycerol–water mixture	1840	—	4.063		1200	11
Xylene	1320	—	1.513	–	868.2	12
Bromoform	900	—	2.341	–	2890	13(a)
Benzene	1320	—	1.5263	–	876	13(b), 14(a)
Water/glycerine	1711	—	3.232	–	1104	14(a)
Carbon tetrachloride	968	—	1.536	–	1640	14(b)

available in the literature. The material properties of the constituent phases are summarized in Table I along with the corresponding figure number where experimental and theoretical curves are depicted. Although water and polystyrene takes part in more than one measurement, the properties do not exhibit remarkable differences from one case to other; therefore they are mentioned once.

A. Particulate composites

The first material studied is an Iron/PMMA composite. In Fig. 4(a) comparison between measured and calculated longitudinal velocity is depicted for a monochromatic wave ($k_1a=0.06$). The experimental data are obtained by the work of Piche and Hamel.[57] As observed the agreement is excellent while the Waterman–Truell model, as expected, predicts well only for low concentrations.

The other case, Fig. 4(b), concerns a titanium carbonate (TiC) in epoxy composite at the dimensionless frequency $k_1a=0.02$. It is obvious that the agreement for almost all volume fractions between experimental velocity[11] and IEMA predictions is very good. In this figure, the predictions made by the Waterman and Truell approach, without applying the iterative procedure, are also supplied. It is seen that the agreement is good but for high volume fractions the use of the self-consistent relation seems to more closely follow the trend at higher volume fractions.

The next case under consideration is a lead/epoxy (Epon 828z) composite with spherical particles of radius 660 μm and volume fractions 26% and 52%; see Figs. 5(a) and 5(b), respectively. The theoretical results are compared again to those taken directly from the Waterman–Truell dispersion relation. Although the discrepancy between IEMA results and experiment (taken from the work of Kinra and Rousseau[58]) seems to increase with volume content, it can be said that, qualitatively, the results are in good agreement. An

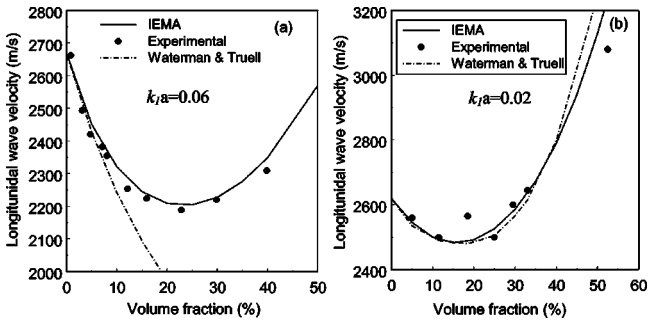


FIG. 4. A comparison between predicted and measured longitudinal velocity for (a) an iron/PMMA composite (experimental data from Ref. 52) and (b) for a titanium carbonate in epoxy composite (experimental data from Ref. 11).

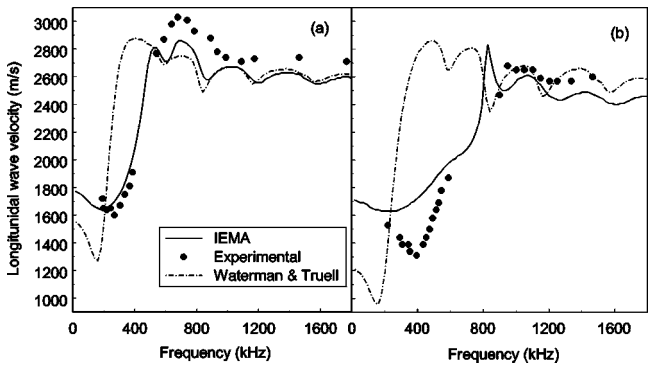


FIG. 5. A comparison between predicted and measured (Ref. 58) longitudinal velocity for a 660 μm radius lead spheres in an epoxy composite with volume fraction (a) 26% and (b) 52%.

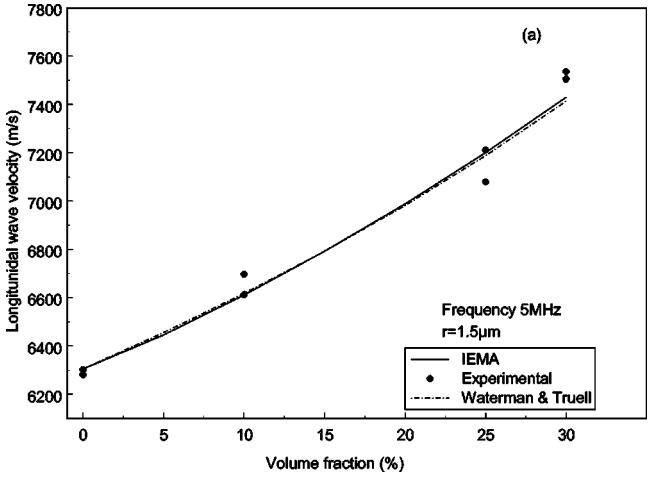


FIG. 6. A comparison between predicted and measured (Ref. 13) longitudinal velocity for a SiCp in aluminum composite, matrix type 7091.

important conclusion drawn by this figure is that the IEMA predicts the shift of the lowest and highest resonance frequencies to higher values as the volume fraction increases. The position of the resonant frequencies can be found in the diagrams of Fig. 5 since at these frequencies the velocity obtains maximum values.

Figure 6 depicts the longitudinal wave velocity of aluminum (Al) matrix composites containing silicon carbide (SiCp) particles. The increase of velocity with volume fraction is apparent for both cases and predicted values are quite close to the experimental ones concerning the random orientation of the particles. After microstructural characterization of the material,[13] the average SiCp size varied approximately from 2 to 4 μm. For the theoretical predictions the diameter was considered to be 3 μm.

The last particulate composite case concerns the attenuation of the glass/epoxy (Tra-cast 3012) system. The attenuation measurements carried out by Kinra *et al.*[61] for a 45% volume content of glass and the corresponding predictions of IEMA are presented in Fig. 7. It is apparent that the IEMA follows closely the experimental data for the frequency range

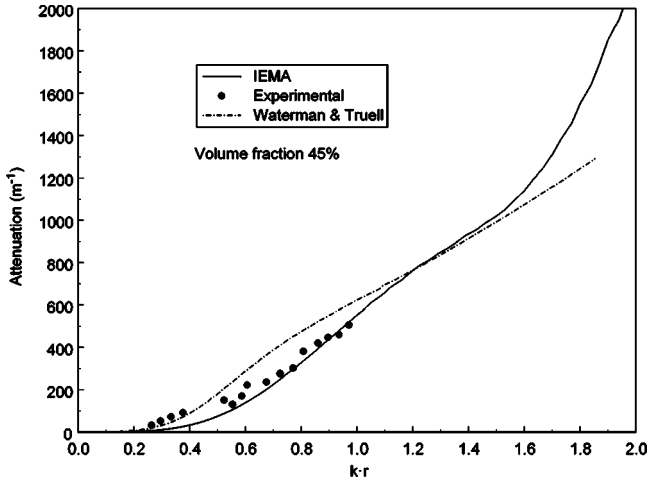


FIG. 7. A comparison between the predicted and measured (Ref. 61) longitudinal attenuation for a glass/epoxy (Tra-cast 3012) composite.

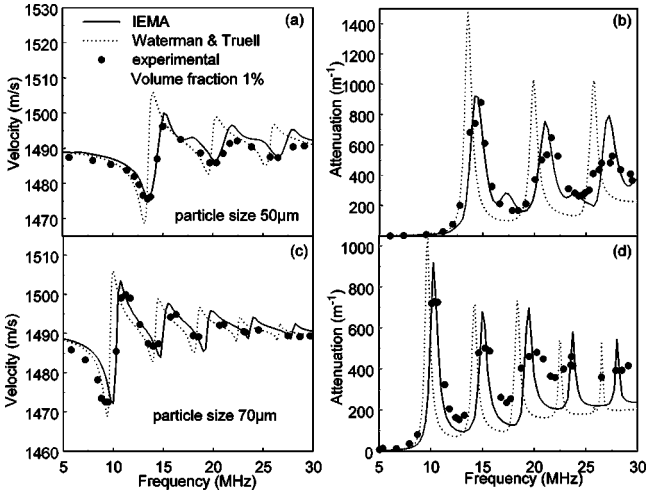


FIG. 8. A comparison between the predicted and measured (Ref. 16) sound velocity of a suspension of 1% polystyrene spheres in water with radius (a) 50 μm and (c) 70 μm and corresponding attenuation (b) and (d).

tested, while the Waterman and Truell approach seems suitable only for low frequencies.

B. Elastic–liquid suspensions

Liquid suspension modeling does not require a much different approach. As mentioned above, since liquids do not support shear waves, the shear modulus of the matrix obtained values very small (100 Pa).

The first case studied is a 1% by volume polystyrene in water suspension with particle size 50 μm interrogated experimentally[16] in the frequency band 3–30 MHz by means of ultrasonic spectroscopy. Overall, the predicted shape of the phase velocity and attenuation curves tracks the experimental results closely as seen in Figs. 8(a) and 8(b), respectively. The peaks and nadirs of the IEMA model coincide with the experimentally measured ones for both velocity and attenuation being closer than the original Waterman and Truell dispersion relation.

Another case of interest lies in Fig. 9(a). There, the velocity of glass in acetylene tetrabromide-benzene (ATB) is

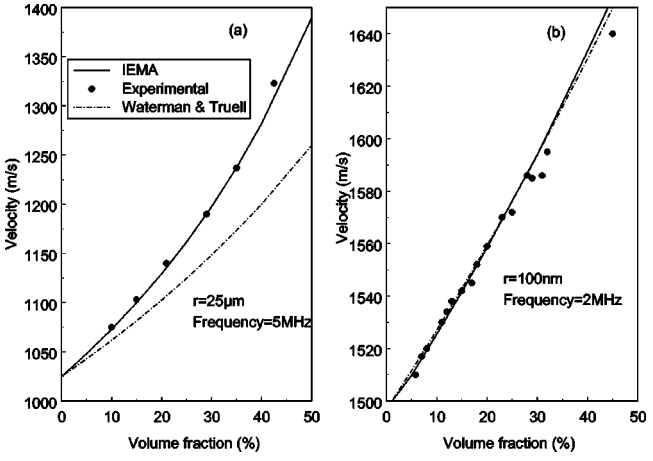


FIG. 9. A comparison between the predicted and measured sound velocity of a suspension of (a) glass in ATB (experimental data from Ref. 18) and (b) polystyrene in water (experimental data from Ref. 15).

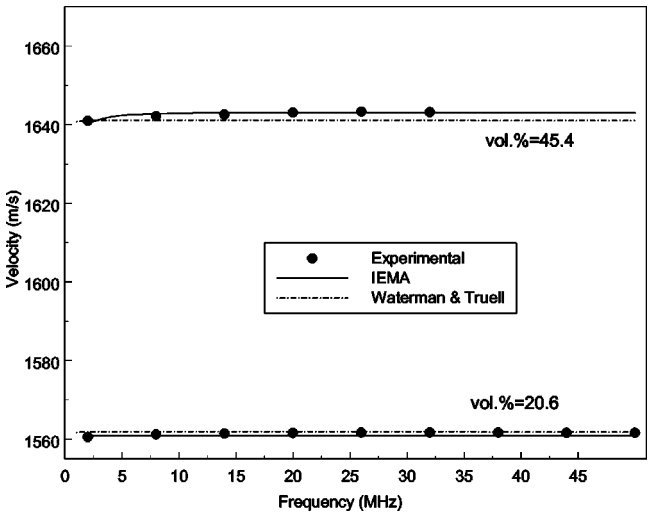


FIG. 10. A comparison between the predicted and measured (Ref. 15) sound velocity of a 308 nm polystyrene spheres in water suspension for different volume contents.

depicted vs the volume content. The frequency applied experimentally is 5 MHz while the particle radius is 12.5 μm, resulting in $k_1a=0.3829$. The experimental data is due to McClements and Povey[18] and they are in excellent agreement with IEMA predictions.

In Fig. 9(b) another case of polystyrene in water is examined concerning the effect of the volume fraction in velocity at 2 MHz. In such systems, due to the low-density contrast between the two phases thermal effects are expected to be dominant.[15] However, although the present formulation omits such effects, theoretical predictions are very close to experimental data.[15]

The same work of Holmes *et al.*[15] contains interesting comparisons between the dispersive behavior of different particle volume fraction suspensions sharing though the same particle size. In Fig. 10 the experimentally observed dispersion of 20.6% and 45.5% with a particle size of 308 nm between 2 and 50 MHz is depicted. It is apparent that both the IEMA and the Waterman–Truell model predict well for the present case.

The last case of elastic in a liquid suspension presented herein concerns a suspension of monodisperse (radius 0.438 mm) glass beads in a 75% glycerin–25% water mixture. The properties of the suspended and continuum media exhibit

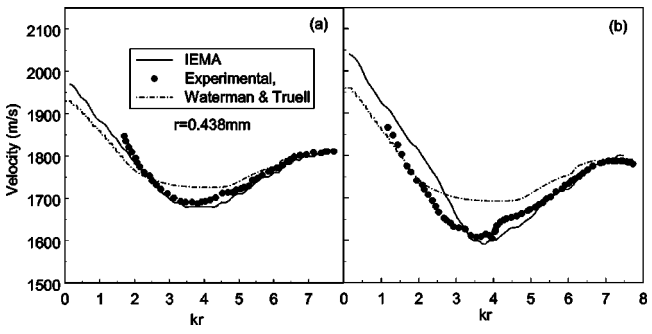


FIG. 11. A comparison between the predicted and measured (Ref. 42) sound velocity of a glass beads in glycerol–water mixture suspension for volume fraction (a) 34% and (b) 45%.

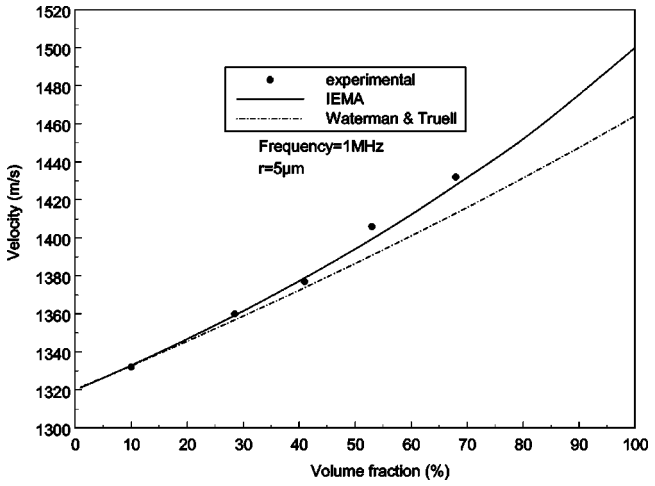


FIG. 12. A comparison between the predicted and measured (Ref. 18) sound velocity of water in a xylene emulsion.

large mismatch (the sound velocity of fluid is 1840 m/s while of glass beads 5600 m/s) resulting in strongly scattering behavior. IEMA succeeds in predicting very well the experimental behavior[42] up to about 5 MHz tested for the cases of 34% and 45% volume content of glass as seen in Figs. 11(a) and 11(b), respectively.

C. Liquid–liquid emulsions

Apart from the suspension of elastic particles in fluid, a separate category can be assumed for the liquid–liquid emulsions. Ultrasonic parameters of such systems as velocity and attenuation can also here be very closely predicted using $\mu=100$ Pa for both liquids and following the same iterative procedure. All experimental data concerning this section are taken again from McClements and Povey.[18]

In Fig. 12 a water in xylene emulsion is described with droplet size 5 μm, measured at the frequency of 5 MHz for a wide range of volume fractions. In the next Fig. 13 two cases of water based emulsions are presented. In Fig. 13(a) the dispersed phase is bromoform and in Fig. 13(b) it is benzene. The droplet size is 3 μm for the first case and 8 μm for the second while the frequencies used are 3 and 2 MHz, respectively. As can be seen, the increase of the dispersed liquid

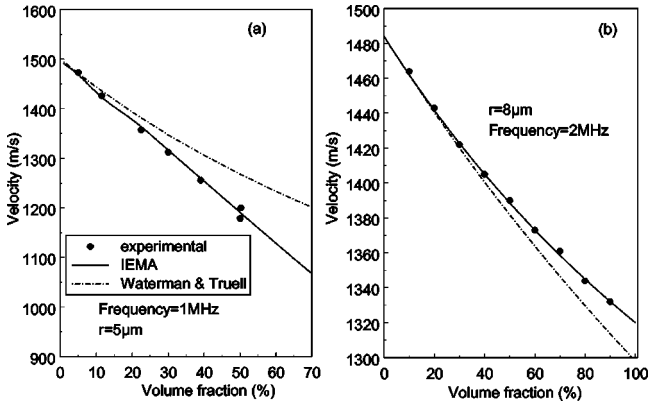


FIG. 13. A comparison between the predicted and measured (Ref. 18) sound velocity of (a) bromoform in water and (b) benzene in water emulsion.

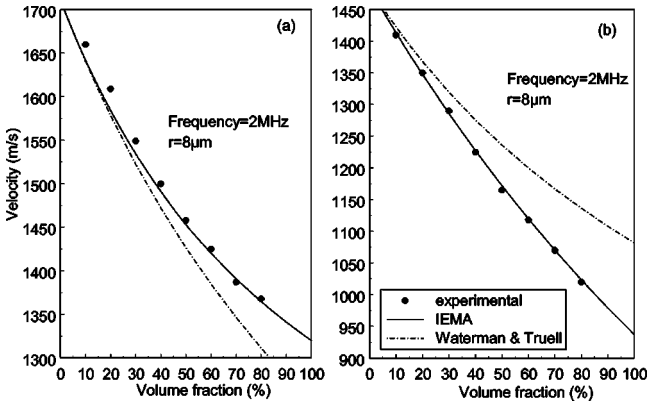


FIG. 14. A comparison between the predicted and measured (Ref. 18) sound velocity of benzene in water/40% glycerine and (b) carbon tetrachloride in water emulsion.

content causes a certain decrease in sound velocity, which is predicted exactly by the IEMA present herein.

The cases of the last figure concern a benzene in water/40% glycerin emulsion, see Fig. 14(a), and a carbon tetrachloride in water emulsion, Fig. 14(b). It is seen that the decrease in velocity with the increase of a droplet content is very closely predicted using the iterative procedure on the Ying and Truell formulation described herein. The properties of all different phases of the emulsions can be found in Table I.

V. CONCLUSIONS

A recently introduced iterative methodology for the quantitative estimation of wave dispersion and attenuation due to scattering is described here. Although successfully tested for particulate composites,[38] in this work its effectiveness is examined additionally on suspensions of solids in liquid and liquid in liquid emulsions systems. The obtained theoretical curves concerning velocity and attenuation predict very closely the experimental results, regardless of the nature of the phases. The elastic properties of the effective medium are given by the static mixture model of Christensen[53] while for the case of liquid matrix the shear rigidity is taken as 100 Pa. The dynamic behavior of the system is characterized by the complex effective material density, calculated through the iterative procedure discussed above. The IEMA being simpler than the one proposed by Kim *et al.*[37] is very powerful even for high volume fractions where it provides predictions close to experimental observations. The IEMA can be used as a practical tool for wave dispersion and attenuation prediction being very useful for applications in a nondestructive evaluation. Taking into consideration thermal effects is expected to improve the accuracy of dispersion and attenuation predictions and is proposed as a task for future work.

[1] L. L. Foldy, ''The multiple scattering of waves,'' Phys. Rev. **67**, 107–119 (1945).
[2] M. Lax, ''Multiple scattering of waves,'' Rev. Mod. Phys. **23**, 287–310 (1951).
[3] P. C. Waterman and R. Truell, ''Multiple scattering of waves,'' J. Math. Phys. **2**, 512–537 (1961).
[4] V. Twersky, ''On scattering of waves by random distributions. I. Free-space scatterer formalism and II. Two-space scatterer formalism,'' J. Math. Phys. **3**, 700–734 (1962).
[5] P. Lloyd and M. V. Berry, ''Wave propagation through an assembly of spheres IV. Relations between different multiple scattering theories,'' Proc. Phys. Soc. London **91**, 678–688 (1967).
[6] V. K. Varadan, Y. Ma, and V. V. Varadan, ''A multiple scattering theory for elastic wave propagation in discrete random media,'' J. Acoust. Soc. Am. **77**, 375–385 (1985).
[7] C. Javanaux and A. Thomas, ''Multiple scattering using the Foldy–Twersky integral equation,'' Ultrasonics **26**, 341–343 (1988).
[8] C. M. Sayers and R. L. Smith, ''Ultrasonics velocity and attenuation in an epoxy matrix containing lead inclusions,'' J. Phys. D **16**, 1189–1194 (1983).
[9] H. M. Ledbetter and S. K. Datta, ''Effective wave speeds in an SiC particle-reinforced Al composite,'' J. Acoust. Soc. Am. **79**, 239–248 (1986).
[10] A. N. Norris, ''Scattering of elastic waves by spherical inclusions with applications to low frequency wave propagation composites,'' Int. J. Eng. Sci. **24**, 1271–1282 (1986).
[11] L. W. Anson and R. C. Chivers, ''Utrasonic velocity in suspensions of solids in solids—a comparison of theory and experiment,'' J. Phys. D **26**, 1566–1575 (1993).
[12] Y. Shindo, H. Nozaki, and S. K. Datta, ''Effect of interface layers on elastic wave propagation in metal matrix composite reinforced by particles,'' J. Appl. Mech. **62**, 178–185 (1995).
[13] Y. Lu and P. K. Liaw, ''Effects of particle orientation in silicon carbide particulate reinforced aluminum matrix composite extrusions on ultrasonic velocity measurements,'' J. Comput. Math. **29**, 1096–1116 (1995).
[14] R. E. Challis, J. S. Tebbutt, and A. K. Holmes, ''Equivalence between three scattering formulations for ultrasonic propagation in particulate mixtures,'' J. Phys. D **31**, 3481–3497 (1998).
[15] A. K. Holmes, R. E. Challis, and D. J. Wedlock, ''A wide bandwidth study of ultrasound velocity and attenuation in suspensions: Comparison of theory with experimental measurements,'' J. Colloid Interface Sci. **156**, 261–268 (1993).
[16] J. Mobley, K. R. Waters, C. H. Hall, J. N. Marsh, M. S. Hughes, G. H. Brandenburger, and J. G. Miller, ''Measurements and predictions of phase velocity and attenuation coefficient in suspensions of elastic microspheres,'' J. Acoust. Soc. Am. **106**, 652–659 (1999).
[17] F. V. Meulen, G. Feuillard, O. Bou Matar, F. Levassort, and M. Lethiecq, ''Theoretical and experimental study of the influence of the particle size distribution on acoustic wave properties of strongly inhomogeneous media,'' J. Acoust. Soc. Am. **110**, 2301–2307 (2001).
[18] D. J. McClements and M. J. W. Povey, ''Ultrasonic velocity as a probe of emulsions and suspensions,'' Adv. Colloid Interface Sci. **27**, 285–316 (1987).
[19] D. J. McClements, ''Comparison of multiple scattering theories with experimental measurements in emulsions,'' J. Acoust. Soc. Am. **91**, 849–853 (1992).
[20] P. S. Epstein and R. R. Carhart, ''The absorption of sound in suspensions and emulsions. I. Water for in air,'' J. Acoust. Soc. Am. **25**, 553–565 (1953).
[21] J. R. Allegra and S. A. Hawley, ''Attenuation of sound in suspensions and emulsions: Theory and experiments,'' J. Acoust. Soc. Am. **51**, 1545–1564 (1972).
[22] C. F. Ying and R. Truell, ''Scattering of a plane longitudinal wave by a pherical obstacle in an isotropically elastic solid,'' J. Appl. Phys. **27**, 1086–1097 (1956).
[23] A. I. Beltzer, C. W. Bert, and A. G. Striz, ''On wave propagation in random particulate composites,'' Int. J. Solids Struct. **19**, 785–791 (1983).
[24] A. I. Beltzer, ''The effective dynamic response of random composites and polycrystals—A survey of the causal approach,'' Wave Motion **11**, 211–229 (1989).
[25] S. Temkin, ''Attenuation and dispersion of sound in bubbly liquids via the Kramers-Kroning relations,'' J. Fluid Mech. **211**, 61–72 (1990).
[26] Z. Ye, ''Acoustic dispersion and attenuation in many spherical scatterer systems and the Kramers-Kroning relations,'' J. Acoust. Soc. Am. **101**, 3299–3305 (1997).
[27] J. L. Leander, ''Comments on acoustic dispersion and attenuation in many spherical scatterer systems and the Kramers-Kroning relations,'' J. Acoust. Soc. Am. **104**, 1111–1115 (1998).
[28] R. L. Weaver and Y. H. Pao, ''Dispersion relations for linear wave propagation in homogeneous and inhomogeneous media,'' J. Math. Phys. **22**, 1909–1918 (1981).

[29] Ch. Zhang and D. Gross, *On Wave Propagation in Elastic Solids with Cracks* (Computational Mechanics Publications, Southampton, UK, 1998).
[30] S. Temkin, ''Attenuation and dispersion of sound in dilute suspensions of spherical particles,'' J. Acoust. Soc. Am. **108**, 126–146 (2000).
[31] D. R. S. Talbot and J. R. Willis, ''Variational estimates for dispersion and attenuation of waves in random composites,'' Int. J. Solids Struct. **22**, 1–11 (1987).
[32] F. J. Sabina and W. R. Willis, ''A simple self-consistent analysis of wave propagation in particulate composites,'' Wave Motion **10**, 127–142 (1988).
[33] A. J. Devaney, ''Multiple scattering theory for discrete, elastic, random media,'' J. Math. Phys. **21**, 2603–2611 (1980).
[34] F. H. Kerr, ''An effective medium approach to the study of plane wave propagation in an elastic matrix with spherical elastic inclusions,'' Int. J. Eng. Sci. **30**, 187–199 (1992).
[35] S. K. Kanaun, V. M. Levin, and F. T. Sabina, ''Propagation of elastic waves in composites with random set of spherical inclusions (effective medium approach),'' Wave Motion (in press).
[36] J. G. Berryman, ''Long-wavelength propagation in composite elastic media, I. Spherical inclusions, II. Ellipsoidal inclusions,'' J. Acoust. Soc. Am. **68**, 1809–1831 (1980).
[37] J.-Y. Kim, J.-G. Ih, and B.-H. Lee, ''Dispersion of elastic waves in random particulate composites,'' J. Acoust. Soc. Am. **97**, 1380–1388 (1995).
[38] S. V. Tsinopoulos, J. T. Verbis, and D. Polyzos, ''An iterative effective medium approximation for wave dispersion and attenuation predictions in particulate composites,'' Adv. Composite Lett. **9**, 193–200 (2000).
[39] S. Biwa, N. Ito, and N. Ohno, ''Elastic properties of rubber particles in toughened PMMA: Ultrasonic and micromechanical evaluation,'' Mech. Mater. **33**, 717–728 (2001).
[40] L. W. Anson and R. C. Chivers, ''Ultrasonic propagation in suspensions—a comparison of a multiple scattering and effective medium approach,'' J. Acoust. Soc. Am. **85**, 535–540 (1989).
[41] Y. Hemar, N. Herrmann, P. Lemaréchal, R. Hocquart, and F. Lequex, ''Effective medium model for ultrasonic attenuation due to thermo-elastic effect in concentrated emulsions,'' J. Phys. II **7**, 637–647 (1997).
[42] M. L. Cowan, K. Beaty, J. H. Page, L. Zhengyou, and P. Sheng, ''Group velocity of acoustic waves in strongly scattering media: Dependence on the volume fraction of scatterers,'' Phys. Rev. E **58**, 6626–6636 (1998).
[43] D. J. McClements, Y. Hemar, and N. Herrmann, ''Incorporation of thermal overlap effects into multiple scattering theory,'' J. Acoust. Soc. Am. **105**, 915–918 (1999).
[44] A. K. Hipp, G. Storti, and M. Morbidelli, ''Acoustic characterization of concentrated suspensions and emulsions. 1. Model analysis,'' Langmuir **18**, 391–404 (2002).
[45] A. K. Hipp, G. Storti, and M. Morbidelli, ''Acoustic characterization of concentrated suspensions and emulsions. 2. Experimental validation,'' Langmuir **18**, 405–421 (2002).
[46] A. H. Harker and J. A. G. Temple, ''Velocity and attenuation of ultrasound in suspensions of particles in fluids,'' J. Phys. D **21**, 1576–1588 (1988).
[47] C. M. Atkinson and H. K. Kytömaa, ''Acoustic wave speed and attenuation in suspensions,'' Int. J. Multiphase Flow **18**, 577–592 (1992).
[48] J. M. Evans and K. Attenborough, ''Coupled phase theory for sound propagation in emulsions,'' J. Acoust. Soc. Am. **102**, 278–282 (1997).
[49] P. Soven, ''Coherent-potential model of substitutional disordered alloys,'' Phys. Rev. **156**, 809–813 (1967).
[50] B. Velicky, S. Kirkpatrick, and H. Ehrenreich, ''Single-site approximation in the electronic theory of simple binary alloys,'' Phys. Rev. **175**, 747–766 (1968).
[51] J. E. Gubernatis and J. E. Krumhansl, ''Macroscopic engineering properties of polycrystalline materials: Elastic properties,'' J. Appl. Phys. **46**, 1875–1883 (1975).
[52] B. Kaelin and L. R. Johnson, ''Dynamic composite elastic medium theory. Part II. Three-dimensional media,'' J. Appl. Phys. **84**, 5458–5468 (1998).
[53] R. M. Christensen, ''A critical evaluation for a class of micromechanics models,'' J. Mech. Phys. Solids **38**, 379–404 (1990).
[54] A. Petculescu and L. A. Wilen, ''Lumped-element technique for measurement of complex density,'' J. Acoust. Soc. Am. **110**, 1950–1957 (2001).
[55] K. Lee, H. S. Roh, and S. W. Yoon, ''Acoustic wave propagation in bovine cancellous bone: Application of the modified Biot–Attenborough model,'' J. Acoust. Soc. Am. **114**, 2284–2293 (2003).
[56] Y. Pan and R. N. Horne, ''Generalized macroscopic models for fluid flow in deformable porous media I: Theories,'' Trans. Porous Media **45**, 1–27 (2001).
[57] L. Piche and A. Hamel, ''Ultrasonic evaluation of filled polymers 1: Techniques and models for the elastic moduli of resin filled with iron spherical inclusions,'' Polym. Compos. **7**, 355–361 (1986).
[58] V. K. Kinra and C. Rousseau, ''Acoustical and optical branches of wave propagation,'' J. Wave Mater. Interaction **2**, 141–152 (1987).
[59] Y. H. Pao and C. C. Mow, ''Scattering of plane compressional waves by a spherical obstacle,'' J. Appl. Phys. **34**, 493–499 (1963).
[60] T. J. O'Neill, J. S. Tebbutt, and R. E. Challis, ''Convergence criteria for scattering models of ultrasonic wave propagation in suspensions of particles,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 419–424 (2001).
[61] V. K. Kinra, M. S. Petraitis, and S. K. Datta, ''Ultrasonic wave propagation in a random particulate composite,'' Int. J. Solids Struct. **16**, 301–312 (1980).

Insertion loss of an acoustic enclosure

Ruisen Ming and Jie Pan
School of Mechanical Engineering, The University of Western Australia, Nedlands WA 6009, Australia

(Received 10 June 2004; revised 18 September 2004; accepted 24 September 2004)

Acoustical enclosures are the common arrangements in reducing airborne noise from shipboard machinery such as engines and generators. In this paper the theoretical models, established based on statistical energy analysis, are presented for predicting the insertion loss of acoustical enclosures in different frequency ranges. In addition to the consideration of resonant modal coupling between internal sound field and enclosure structural vibration, the nonresonant transmission though and the interaction between enclosure walls in the models are also included. It is shown that the insertion loss of enclosures is mainly controlled by the nonresonant modes in the intermediate frequency range. At high frequencies, the insertion loss of enclosures can be improved by increasing the sound absorption at the internal boundaries of enclosures. Experiments were carried out on two enclosures made of different materials. The measured results are compared with the predicted values and the good agreement between them is the initial demonstration of the validity and feasibility of the theoretical models. [DOI: 10.1121/1.1819377]

PACS numbers: 43.40.At, 43.50.Gf, 43.50.Fe [JGM] Pages: 3453–3459

I. INTRODUCTION

Noise reduction of acoustic enclosures has been extensively investigated by many authors. It is of importance because acoustic enclosures are used in many applications to reduce the noise emanating from acoustic sources such as running machines or noisy engines. The challenge in enclosure design is to achieve the required noise reduction with minimum cost and, at the same time, to fulfill functional requirements such as materials flow, ventilation, accessibility for repairs, shock-proof mounting, etc. The obvious action in meeting the challenge is to develop an accurate prediction model for the design.

Jackson[1,2] and Junger[3] studied the enclosure performance due to the excitation of an internal acoustic wave field by modeling the source and enclosure walls into two parallel flat plates separated by an air gap. One plate was assumed to be the source and the other was the enclosure wall. In Jackson's model both plates were assumed to be infinite, and the source plate generates plane waves, which are reflected and transmitted at the wall plate. In Junger's model, however, both plates are of finite sizes. The source plate was modeled as a rectangular uniform piston radiator generating a uniform pressure field in the gap, and the wall plate has simply support boundary conditions. Junger's model allows for absorbing material in the airgap, the higher order structural modes, and the coincidence effect in the enclosure wall. Tweed and Tree[4] have reviewed the Jackson and Junger's work and concluded from their measurements that neither method can give acceptable results for the prediction of the enclosure performance.

Oldham and Hillarby[5] extended Junger's model by assuming that the source plate undergoes forced vibration in typical mode shapes at low frequencies, and only symmetric mode shapes cause pressure effect on the enclosure plate. At high frequencies, statistical energy analysis (SEA) is employed by modeling the system into two subsystems. The interaction between enclosure walls and the nonresonance transmission through the enclosure walls were neglected. The comparison between the theoretical predictions and the experimental results[6] indicated that their model is only suitable for assessing the potential performance of an enclosure with simple source configurations.

Lyon[7] approached the enclosure problem in a different way based on the analysis model, consisting of a flexible panel and an enclosed acoustic volume, which was used later by Tweed and Tree.[4] He assumed that the flexible panel is excited by an external reverberant wave field rather than an interior noise source, and divided the analysis into three frequency regions: (1) at low frequencies, where both the flexible panel and interior acoustic volume are stiffness controlled; (2) at intermediate frequencies, where only the panel is resonant and the volume is stiff; and (3) at high frequencies, where both the panel and interior volume display resonant behavior. In the low-frequency region, Lyon suggested that the enclosure performance depends on the ratio of the air cavity and panel stiffness. In the intermediate and high-frequency regions, SEA was employed to calculate the noise reduction. At intermediate frequencies, Lyon modified his SEA model into two subsystems by excluding the interior acoustic volume from his statistical analysis. First, Lyon assumed one panel mode in the frequency band of interest to give a lower bound of noise reduction. Then, he used an average modal density to obtain the average value of noise reduction. Eichler[8] extended Lyon's analysis to a rectangular box with six flexible walls. He reduced his analysis model by assuming that the opposite walls vibrate symmetrically in phase with each other, but uncorrelated with respect to the other walls. His assumptions were not justified. Both Lyon and Eichler did not include the nonresonant transmission in their analysis but used it as a reference for comparison. Cole *et al.*[9] extended Eichler's high-frequency analysis further by treating each enclosure wall as a single SEA subsystem and by accounting for the nonresonant transmission between the reverberant room (subsystem 1) and the enclosed acoustic

volume (subsystem 8). Theoretical prediction has been compared with their experimental results. Good agreement was achieved for different conditions except at the intermediate frequencies and at the critical coincidence frequency. In addition, Cole *et al.* have studied, both theoretically and experimentally, the effects of apertures on the enclosure performance. However, Cole *et al.* did not make any attempt in the low and intermediate frequency ranges.

Ver[10] studied the noise reduction of acoustic enclosures in a similar way as Lyon did, but he placed the noise source inside the enclosure. Ver has detailed the noise reduction in the low-frequency range but did not present any analytical model for the intermediate frequency range, and stated: ''To date there is no analytical model available which would provide reliable engineering information for design of enclosures in this intermediate frequency range. Accordingly the design of enclosures in this frequency range is usually based on scale-model measurements or on experience.'' When the enclosed acoustic volume starts to exhibit resonances, Ver said, ''although no diffuse sound field exists in this frequency range, it has been found that the assumption of a diffuse sound field usually results in a conservative estimate of the insertion loss.'' In addition, Ver analyzed the enclosure performance due to the force excitation on the enclosure walls. Tweed and Tree[4] have concluded from their measurements that Ver's analytical model in the high-frequency range was inadequate for design purposes.

In SEA it is assumed that the response of a subsystem is determined by resonant modes. The uncertainty of SEA results becomes large when the number of modes is small in the frequency band of interest. It is shown[11,12] that this uncertainty is related to the mobility of the receiving subsystem. Craik *et al.* have given a semiempirical expression for correcting the coupling loss factors in the frequency bands where few modes are present.[11,12] Craik's correction is very useful for extending the SEA to the lower part of the intermediate frequency range.

The model studied in this paper is a sealed enclosure of any shape but with an enclosed noise source. In the analysis given, both the nonresonant sound transmission through walls and the interaction between walls are considered. The predicted values of insertion loss are compared with the experimental results. The agreement in all cases is very encouraging.

II. THEORETICAL ANALYSIS

We consider an airtight acoustical enclosure of any shape. Each wall of the enclosure is assumed to be flexible, homogeneous, isotropic, thin, and of uniform thickness. The ''thin panel'' assumption implies that the effects of shear deformation and rotational inertia are negligible. The classical linear theory for thin plates holds and the power transmitted by those forces through in-plane waves is negligible. Only bending waves are considered in the analysis of structural vibration on the enclosure walls.

A noise source is located inside the enclosure generating an interior pressure field, which excites the walls of the enclosure. The vibrating walls then radiate noise into the area outside of the enclosure. The enclosure is of such a size that the resonance frequency of the first structural mode in the walls is lower than that of the first acoustic mode of the enclosed volume.

The noise reduction in the low-frequency range has been detailed by Ver[10] and will not be repeated here. In most practical applications, the low-frequency sound transmission is not of much interest to the designer because the corresponding frequency band is usually narrow and at very low frequency range, where the noise radiation from the source is usually small.

A. Insertion loss

The insertion loss of an acoustical enclosure is defined as[13]

$$IL = L_{W0} - L_{Wc} \ \text{(dB)}, \tag{1}$$

where L_{W0} and L_{Wc} are the sound power levels radiated, respectively, by the acoustic source and by the vibrating enclosure walls.

In a given frequency band, two types of structural modes are present on the enclosure walls: resonant modes, which have their natural frequencies in the frequency band under consideration and hence have a high response, and nonresonant modes, which are excited such that their natural frequencies fall outside the frequency band under consideration. It is postulated that the nonresonant modes are responsible for the ''mass-law'' transmission of sound.[14–16] The sound power radiated by the nonresonant modes of enclosure walls, W_{non}, is given by

$$W_{\text{non}} = \tau W_0, \tag{2}$$

where W_0 is the sound power (watt) radiated by the acoustic source when it is within the enclosure and τ is the overall transmission coefficient[17] of the enclosure, which can be obtained by averaging over all the incident angles and enclosure walls.

On the other hand, the sound power radiated by the resonant modes of wall i, W_{resi}, can be estimated from the averaged vibration velocity, and is given by[13]

$$W_{\text{resi}} = \rho_0 c \sigma_i S_i \overline{v_i^2}, \tag{3}$$

where ρ_0 is the mass density of air; c is the sound speed in air; σ_i and S_i are, respectively, the radiation coefficient and surface area of wall i; v_i is the vibration velocity of wall i. The total sound power radiation from an enclosure is $W_c = \Sigma W_{\text{resi}} + W_{\text{non}}$.

B. Noise reduction at intermediate frequencies

In the intermediate frequency range, the air within the enclosure acts like a pure stiffness closely coupled to the walls, while the enclosure walls exhibit resonant behavior. The interior acoustic volume cannot be treated as a separate SEA subsystem in this frequency range. It only loads the walls reactively and increases some of their resonance frequencies slightly, but this increase may be neglected because it barely changes the modal density of the walls.[8] Each wall, however, may be treated as one SEA subsystem, but its modal density could be low at some frequencies and shows

significant fluctuations. The modal density of subsystem i could be more accurately predicted from the real part (Y_i) of its point mobility using the following equation:[13]

$$n_i = 4S_i\rho_{Si}Y_i, \tag{4}$$

where ρ_{Si} is the surface mass density.

For the case of low modal density, the coupling loss factor between plates also shows significant fluctuations. The fluctuations are very similar to those of the real part of the space-averaged point mobility of the receiving subsystem. The coupling loss factor (η_{ij}) needs to be modified using the following relationship:[11,12]

$$\eta_{ij} = \eta_{ij\infty}Y_j/Y_{j\infty}, \tag{5}$$

where $Y_{j\infty}$ is the conventional value of the point mobility obtained based on infinite plate theory (having the same thickness as the receiving subsystem); Y_j is the real part of the point mobility of the receiving subsystem, and $\eta_{ij\infty}$ is the predicted coupling loss factor based on two semi-infinite plates.

The enclosed noise source generates a uniform sound pressure to the interior wall surfaces. The power radiated by the enclosed noise source (W_0) is (1) dissipated inside the enclosed volume; (2) transmitted to and exciting the walls; and (3) transmitted through the walls due to the nonresonant modes. If there is little sound-absorptive material to dissipate sound in the enclosure, the dissipated power component becomes negligible. The power component transmitted to each enclosure wall will be proportional to the modal density, the damping loss factor, and the radiation coefficient of the wall. For an enclosure of m walls without the treatment of any acoustic absorptive material, the power input to wall i from the enclosed acoustic source is given by

$$W_i = \frac{n_i\eta_{ii}(1-\tau)W_0}{\Sigma_{j=1}^{m} n_j\eta_{jj}}, \tag{6}$$

where η_{jj} is the dissipation loss factor of wall j due to the structural damping and sound radiation. If wall i does not exhibit resonance behavior (n_i is very small) or has a small dissipation loss factor, W_i will be negligible or small in comparison with others.

The power balance equation on wall i ($i=1,2,...,m$) can be expressed as

$$\omega\eta_{i0}E_i - \omega\sum_{\substack{j=1\\ j\neq i}}^{m}\eta_{ji}E_j = W_i, \tag{7}$$

where $\omega = 2\pi f$ is radian frequency, $E_i = \rho_{Si}S_i\overline{v_i^2}$, and $\eta_{i0} = \eta_{ii} + \Sigma_{i=1,i\neq j}^{6}\eta_{ij}$ are, respectively, the total energy stored in wall i and the total loss factor of wall i. η_{ij} is the coupling loss factor from wall i to j. For the rectangular enclosure ($m=6$), all the joints are assumed as "corner joint" and η_{ij} is given as[18]

$$\eta_{ij} = \frac{2c_{Bi}L_{ij}}{\pi\omega S_i}\frac{2.754X}{1+3.24X}\frac{2\psi}{(1+\psi)^2}, \tag{8a}$$

where L_{ij} is the coupling length between walls i and j; $X = h_i/h_j$ is the thickness ratio; $\psi = \rho_i c_{Li}^{3/2}h_i^{5/2}/\rho_j c_{Lj}^{3/2}h_j^{5/2}$, c_{Bi}, and c_{Li} are, respectively, the bending and longitudinal wave speeds on wall i. If the enclosure walls are made of the same materials and have the same thickness, $X=\psi=1$, then

$$\eta_{ij} = \frac{0.2068c_{Bi}L_{ij}}{\omega S_i}. \tag{8b}$$

The solution of power balance, Eq. (7), gives the vibration level of each enclosure wall. Then, the power radiation due to the resonant modes of each enclosure wall can be estimated using Eq. (3).

C. Noise reduction at high frequencies

In the high-frequency bands, both the enclosure walls and the enclosed air volume are excited to have diffuse sound fields. The enclosed air volume can be treated as a separate SEA subsystem, and seven total subsystems exist for a rectangular enclosure. If the enclosed air volume is numbered as subsystem $m+1$, the power balance equations can be written as

$$\omega\eta_{i0}E_i - \omega\sum_{\substack{j=1\\ j\neq i}}^{m+1}\eta_{ji}E_j = 0, \quad i=1,2,...,m \tag{9a}$$

$$\omega\eta_{(m+1)0}E_{(m+1)} - \omega\sum_{j=1}^{m}\eta_{j(m+1)}E_j = W_0, \tag{9b}$$

where $E_{m+1} = V\overline{p^2}/\rho_0 c^2$ is the total acoustical energy stored in the enclosed air volume, V is the volume of the enclosure, and p is the sound pressure. The coupling loss factor from wall i to the enclosed air volume is given by[15]

$$\eta_{i(m+1)} = \begin{cases} \dfrac{2\rho_0 c\sigma_i}{\omega\rho_{Si}} & f < f_{ci}, \\ \dfrac{\rho_0 c\sigma_i}{\omega\rho_{Si}} & f \geqslant f_{ci}, \end{cases} \tag{10}$$

where f_{ci} is the critical frequency of the ith walls. The coupling loss factor from the enclosed acoustic volume to wall i can be obtained from the consistency relation:[13]

$$n_{m+1}\eta_{(m+1)i} = n_i\eta_{i(m+1)}, \tag{11}$$

where n_{m+1} is the modal density of the enclosed air volume and given by[19]

$$n_{m+1} = \frac{4\pi f^2 V}{c^3} + \frac{\pi f S}{2c^2} + \frac{P}{8c}, \tag{12}$$

V is the enclosed air volume, $S = \Sigma_{i=1}^{m}S_i$ is the total surface area of all enclosure walls, and P is the total edge length. The modal density of wall i is given by[13]

$$n_i = \frac{\sqrt{3}S_i}{c_{Li}h_i}. \tag{13}$$

Both the sound-pressure level inside the enclosure and the vibration levels on the walls can be obtained by solving Eqs. (9a) and (9b). Then, the power component radiated by the resonant modes of each enclosure wall can be calculated using Eq. (3).

III. EXPERIMENTAL STUDIES

A. Rectangular box enclosure

Experiments were carried out on two types of acoustical enclosures. One is a rectangular box made of aluminum. The box has a volume of $1.15\times1.0\times0.868$ m^3 and consists of six plates, which have the same thickness of 2.5 mm, and two rectangular frames (the cross dimensions are 35×6 mm^2) for clamping the lid plate. Five plates were welded together into a cavity with one opening (0.868×1.5 m^2) having a flange, which has the cross dimensions of 25×2.5 mm^2. The lid plate (0.918×1.55 m^2) was bolted to the flange between the two frames, and a thin rubber layer was placed between the lid and the flange to prevent sound leakage. The measurement was taken in a large open carpark in a sunny day. During the measurement, the box was supported by four steel balls of 50 mm in diameter at its corners. Each ball was placed on a short hollowed cylinder (the diameter is 40 mm), which was put on an adjustable stand. The outer surface of the balls is very smooth, and a point contact between the box and each ball is assumed. The power transmitted to the balls should be very small and negligible. The bottom plate of the box was about 1.55 m above the ground. The nearest building was at least 65 m away from the box surfaces.

A TOA TU-650 loudspeaker, fed with white-noise signals, was screwed into a hole located at one of the box corners. To avoid the acoustic radiation from the back, the loudspeaker was fitted into a sealed cylindrical case and only allowed to radiate sound power from the front opening. Before the measurement, the acoustic power radiation of the speaker was measured in an anechoic chamber. It was noticed that this speaker has a low output at frequencies below 160 Hz.

The dissipation loss factor of the single lid plate was measured using the decay method.[19] During the measurement, the plate was freely hung by two fishing wires and nine accelerometer positions were randomly selected. At each accelerometer position, five hammer impacts were made and a time-averaged decay time T_{60} was obtained. Then, a time-averaged dissipation loss factor was calculated from T_{60}. The dissipation loss factor, as shown in Fig. 1, is the average over nine individual time-averaged results. In this case, the dissipation loss factor is the sum of the structural damping factor and radiation loss factor. It is shown in Fig. 1 that the dissipation loss factor is dominated by structural damping at frequencies about one octave band below the critical frequency (the predicted critical frequency is 4.85 kHz) and by radiation loss at frequencies above the critical frequency. The structural damping factor can be represented by

$$\eta_s=\frac{0.7}{f^{0.9}}. \tag{14}$$

The above equation will be used to predict the structural damping factor of all the box walls. The measured dissipation loss factor will not be used in the prediction of insertion loss of the box enclosure.

The internal losses of the enclosed air volume are dominated by viscous and thermal losses at the walls, so that the corresponding loss factor is calculated by[8]

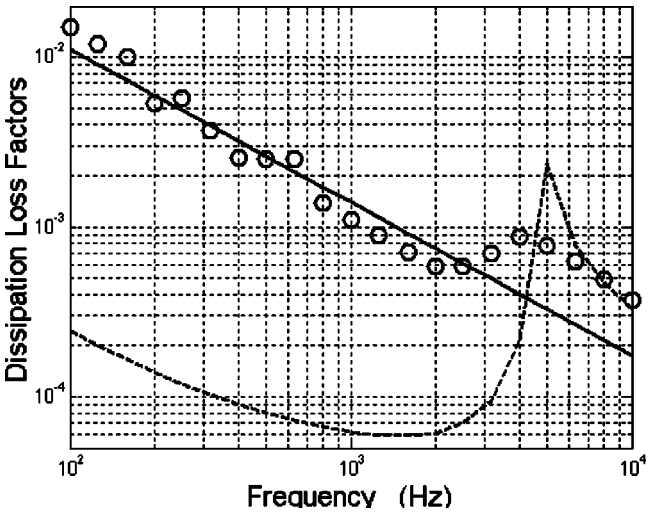


FIG. 1. Dissipation loss factors of the 2.5-mm aluminum plate. O: measured dissipation loss factor; ——: predicted structural damping factor using Eq. (14); - - -: predicted radiation loss factor.

$$\eta_{\text{acoustics}}=\frac{Sc\gamma}{4\omega V}, \tag{15}$$

where γ is the acoustic absorption coefficient, the lower limit of which is $\gamma=1.8\times10^{-4}\sqrt{f}$ for the case of air at room temperature.[19] No absorption material was present inside the box during measurement.

B. Enclosure with fibrous glass composite panels

Another tested enclosure was made from fibrous glass composite panels for the purpose of attenuating the conveyer belt noise in a mining site. It has an opening and has the length of 1.35 m, width of 1.25 m, and height from 0.65 to 0.75 m. All the walls have the same thickness of 3 mm. The top two plates are formed into a V-shape through a smooth arc, and the angle between them is 162°. For simplicity, these two (V-shape) plates are treated as a single flat plate (one SEA subsystem) in the analysis model. This enclosure is divided into five SEA subsystems and all the structural joints are assumed to be corner joint.

During the measurement, the opening was faced to the ground and the gap was sealed using a woolen layer. To ensure that there was no acoustic leakage in the sealed gap, an initial intensity measurement was carried out along the gap and then compared with the measurement over the top surface. Two measurements were carried out: one was the enclosure with a 50-mm rockwool layer and another without any acoustic absorption material. The statistical absorption coefficients of the 50-mm rockwool layer and a 3-mm enclosure wall sample, as shown in Fig. 2, were obtained from the measurement in an impedance tube and the chart for relating normal and statistical absorption coefficients.[20] During data processing, the mass of the 50-mm rockwool layer was taken into account in the calculation of the surface mass density ($\rho_{Si,\text{eff}}=\rho_{Si}+\rho_{S\text{rockwool}}$, where $\rho_{S\text{rockwool}}$ is the surface mass density of the rockwool layer).

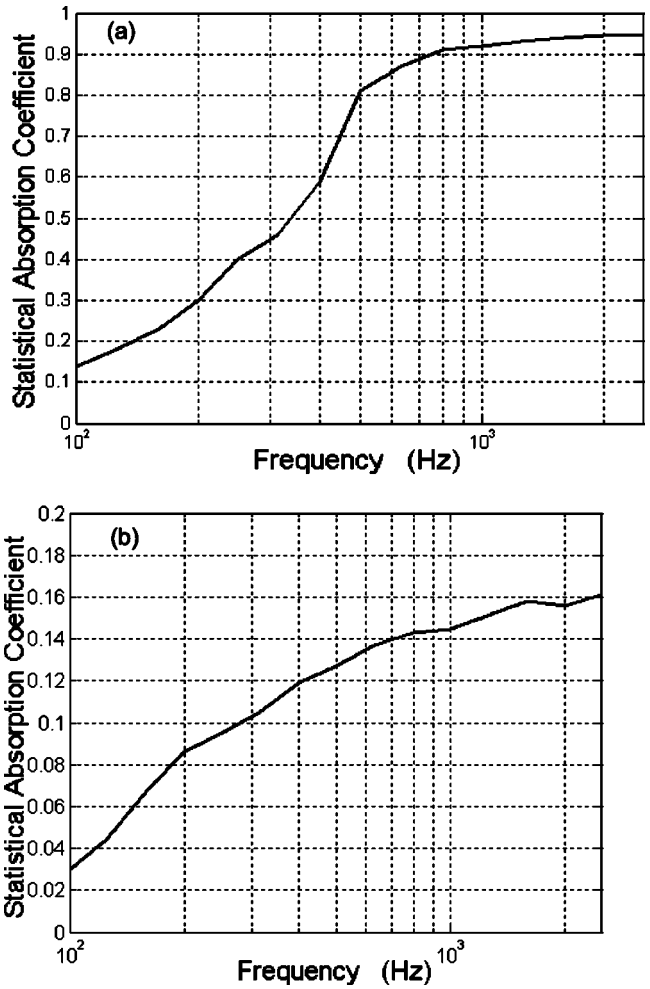


FIG. 2. Measured statistical sound absorption coefficients of a 50-mm rockwool layer(a) and the wall sample of the fibrous glass enclosure(b).

Two loudspeakers were used to generate high-level noise: one speaker(JL Audio 10WD3) emanated low-frequency noise (up to 800 Hz) and another (TOA TU-650) high-frequency noise. During the measurement, the loudspeakers were placed on the ground inside and close to the center of the enclosure. It was shown that the noise from the loudspeakers with and without the fibrous glass enclosure was at least 13 dB (mostly more than 20 dB) higher than the background noise.

The dissipation loss factor of the enclosure walls is quite high and is measurable using the decay method only at frequencies above 1 kHz. At all the frequency bands of interest, the dissipation loss factor is approximated by

$$\eta = \frac{0.45}{f^{0.3}}. \tag{16}$$

C. Sound intensity measurements

The measured insertion loss is obtained based on Eq. (1) from the averaged sound intensity measured over the enclosure walls and the acoustic power radiation of the speaker measured in an anechoic chamber. The BK2681 sound intensity probe was used to measure the sound intensity. Before and after the measurements, the microphones of the intensity probe were calibrated using the BK4231 acoustical calibrator. During the measurements, the BK2144 dual-channel real-time frequency analyzer was used to record both the sound-intensity data and the sound-pressure data.

Scanning technique was used and the moving speed of the intensity probe was very slow. The distance between the probe paths was about 0.1 m. In order to evaluate the accuracy of the intensity data, the pressure-intensity index Δ_{pI} was calculated during data processing[21]

$$\Delta_{pI} = L_{\bar{p}} - L_{|\bar{I}|} - 0.16, \tag{17}$$

where $L_{\bar{p}}$ and $L_{|\bar{I}|}$ are the simultaneously measured (time- and space-averaged) sound-pressure level and sound-intensity level with the reference values of 2.0×10^{-5} Pa and 1 pW/m^2, respectively. The results with an index value larger than a predetermined limit (10 dB for engineering level[22]) were rejected. It was shown that all the pressure-intensity indexes Δ_{pI} were less than 9.5 dB at frequencies above 250 Hz for the rectangular aluminum box, and above 100 Hz for the fibrous glass enclosure.

Each enclosure wall was divided into two scanning areas, and each scanning area was scanned three times with different spacers (the 50-mm spacer was used for the frequency range upto 1.6 kHz, the 12-mm spacer for the frequency range from 1.6 kHz to 3.2 Hz, and the 6-mm spacer for the frequency range above 3.2 kHz). During data processing the finite difference error was predicted and corrected using the following expression:[21]

$$I_{\text{corrected}} = \frac{k\Delta}{\sin(k\Delta)} I_{\text{measured}}, \tag{18}$$

where I_{measured} is the measured sound intensity; $I_{\text{corrected}}$ is the corrected sound intensity free from the finite difference error; k is the wave number of interest, and Δ is the spacing distance between the two microphones. The total sound power in each frequency band was calculated from the corrected sound intensity averaged over 12 individual data.

D. Results and discussions

For the rectangular aluminum box, the lowest acoustic frequency of the enclosed air volume is predicted at 98 Hz (diagonal mode). In the frequency bands below (including) 80 Hz, the intermediate frequency model should be used. Figure 3 shows the comparison of the measured and predicted insertion losses (dB) of the rectangular aluminum box. The measured results agree well with the prediction at frequencies above 315 Hz. At frequencies below 31.5 Hz the measured values are very large (out of the scale) because the measured sound intensity is residual (some are negative). Although there is no single acoustic mode within the intermediate frequency bands, the high-frequency model gives a comparable estimation as the intermediate frequency model. At the intermediate frequencies, the insertion loss is mainly controlled by the nonresonant modes because the resonant modes have very low radiation coefficient, although they carry high vibration energy. At most high frequencies, the insertion loss is underestimated. This is because the lower limit of the absorption coefficient ($\gamma = 1.8\times10^{-4}\sqrt{f}$) was

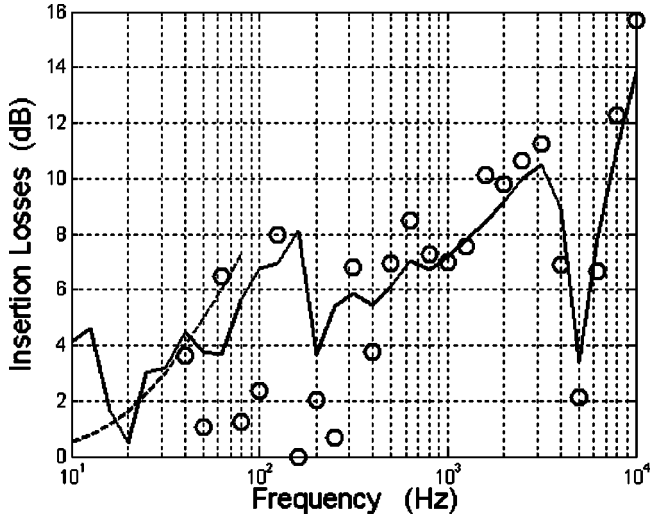


FIG. 3. Measured (O) and predicted (——: from the high-frequency model; – – –: from the intermediate frequency model) insertion losses (dB) of the rectangular aluminum enclosure.

taken for the internal boundaries during the analysis. The absorption coefficient inside the enclosure plays a very important role in the insertion loss improvement. For this rectangular aluminum box, the insertion loss can be increased up to 2.9 dB if the absorption coefficient is doubled ($\gamma = 3.6 \times 10^{-4}\sqrt{f}$), as shown in Fig. 4.

For the fibrous glass enclosure, the lowest acoustic frequency of the enclosed air volume is predicted at 88.1 Hz (diagonal mode). In the frequency bands below (including) 63 Hz, the intermediate frequency model should be used. Figure 5 shows the comparisons of the measured and predicted insertion losses (dB) of the fibrous glass enclosure with and without a 50-mm rockwool layer. It is shown that the absorption material layer greatly improves the performance at frequencies above (including) 250 Hz. For the fibrous glass enclosure without a rockwool layer, the prediction based on the high-frequency model has a better agreement than that on the intermediate frequency one,

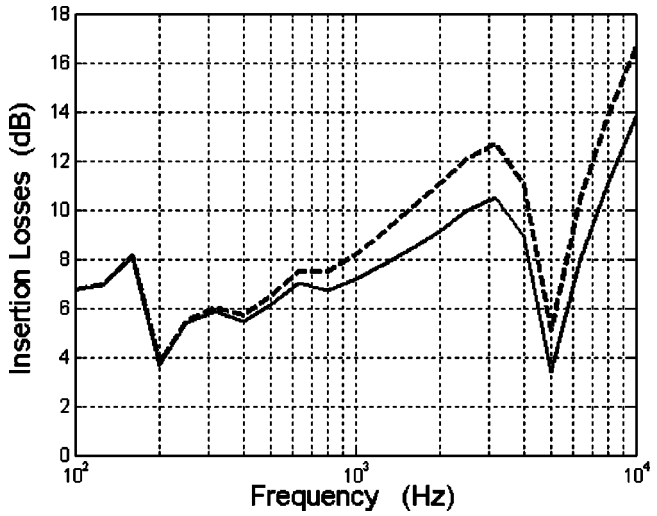


FIG. 4. Predicted insertion losses (dB) of the rectangular aluminum enclosure with different internal absorption coefficients (——: $\gamma = 1.8 \times 10^{-4}\sqrt{f}$; – – –: $\gamma = 3.6 \times 10^{-4}\sqrt{f}$).

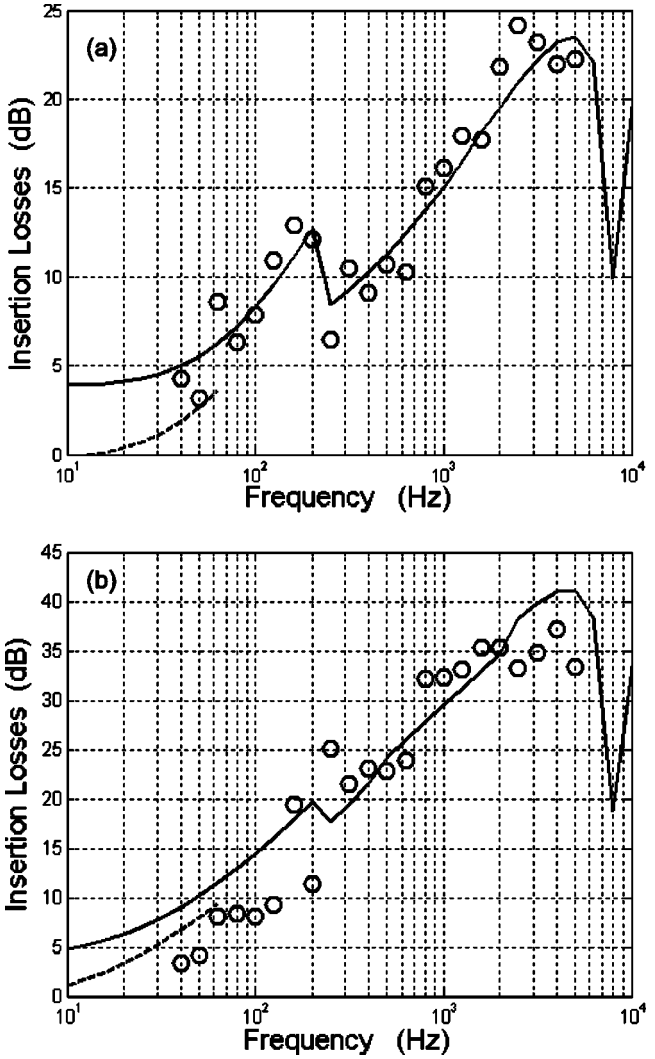


FIG. 5. Measured (O) and predicted (——: from the high-frequency model; – – –: from the intermediate frequency model) insertion losses (dB) of the fibrous glass enclosure without (a) and with (b) a 50-mm rockwool layer.

which underestimates the insertion loss. For the fibrous glass enclosure with a 50-mm rockwool layer, the measured results agree well with the prediction based on the high-frequency model at frequencies above 315 Hz. The intermediate frequency model overestimates the insertion loss. If the mass of the 50-mm rockwool layer was not taken into account in the prediction, however, the intermediate frequency model slightly underestimates the insertion loss and the high-frequency model prediction is much (from 5.8 to 10.6 dB) lower than the measured results at frequencies above 250 Hz but agrees well only at frequencies of 80, 100, 125, and 200 Hz. This indicates that the mass effect of the rockwool layer becomes significant at high frequencies. Again, the prediction indicates that the insertion loss is mainly controlled by the nonresonant modes in the intermediate frequency range.

IV. CONCLUSIONS

Two theoretical models, based on statistical energy analysis, have been presented in this paper to predict the performance of airtight acoustical enclosures in different fre-

quency ranges. Both the nonresonant sound transmission through and the interaction between enclosure walls are considered. It is shown that the insertion loss of enclosures is mainly controlled by the nonresonant modes in the intermediate frequency range. This is because at the intermediate frequencies the resonant modes have very low radiation capacity.

The validity and feasibility of the two theoretical models have been initially tested on two enclosures made of different materials. The absorption coefficient inside the enclosure plays a very important role in the insertion loss improvement. The addition of an acoustical absorption layer inside an enclosure not only increases the damping of enclosed air volume but also has an effect on the nonresonance transmission across enclosure walls. This effect should be considered in the prediction. The measured results suggests that for the case of enclosures without adding any acoustical absorption layer, the high-frequency model may be extended to the intermediate frequency bands where no single acoustical mode is presented in the enclosed air volume.

The insertion loss studied in this paper is the reduction in powers radiated from enclosed sources. By reciprocity, however, it is equal to the insertion loss (reduction in pressure) of enclosures, which are exposed to an acoustic wave field.

ACKNOWLEDGMENTS

Financial support for this work from the Australian Research Council and Strategic Marine Ltd is acknowledged.

[1] R. S. Jackson, ''The performance of acoustic hoods at low frequencies,'' Acustica **12**, 139–152 (1962).
[2] R. S. Jackson, ''Some aspects of the performance of acoustic hoods,'' J. Sound Vib. **3**, 82–94 (1966).
[3] M. C. Junger, ''Sound transmission through an elastic enclosure acoustically coupled to a noise source,'' ASME Paper No. 70-WA/DE-12 (1970).
[4] L. W. Tweed and D. R. Tree, ''Three methods for predicting the insertion loss of close fitting acoustical enclosures,'' Noise Control Eng. **10**, 74–79 (1978).
[5] D. J. Oldham and S. N. Hillarby, ''The acoustical performance of small close fitting enclosures. I. Theoretical models,'' J. Sound Vib. **150**, 261–281 (1991).
[6] D. J. Oldham and S. N. Hillarby, ''The acoustical performance of small close fitting enclosures. II. Experimental investigation,'' J. Sound Vib. **150**, 283–300 (1991).
[7] R. H. Lyon, ''Noise reduction of rectangular enclosures with one flexible wall,'' J. Acoust. Soc. Am. **35**(11), 1791–1797 (1963).
[8] E. Eichler, ''Thermal circuit approach to vibrations in coupled systems and the noise reduction of a rectangular box,'' J. Acoust. Soc. Am. **37**(6), 995–1007 (1965).
[9] V. Cole, M. J. Crocker, and P. K. Raju, ''Theoretical and experimental studies of noise reduction of an idealised cabin enclosure,'' Noise Control Eng. **20**(3), 122–132 (1983).
[10] I. L. Ver, ''Reduction of noise by acoustic enclosures,'' in Proceedings of the ASME Design Engineering Technical Conference, Cincinnati, Ohio, September 1973, Isolation of Mechanical Vibration, Impact and Noise, edited by J. C. Snowdon and E. E. Ungar, 192–220 (1973).
[11] R. J. M. Craik, J. A. Steel, and D. I. Evans, ''Statistical energy analysis of structure-borne sound transmission at low frequencies,'' J. Sound Vib. **144**(1), 95–107 (1991).
[12] J. A. Steel and R. J. M. Craik, ''Statistical energy analysis of structure-borne sound transmission by finite element methods,'' J. Sound Vib. **178**(4), 553–561 (1994).
[13] M. P. Norton, *Fundamentals of Noise and Vibration Analysis for Engineers* (Cambridge University Press, Cambridge, 1999).
[14] R. H. Lyon, *Random Noise and Vibration in Space Vehicles* (SVM-1, Shock and Vibration Information Center, U.S. Department of Defense, 1967).
[15] A. J. Price and M. J. Crocker, ''Sound transmission through double panels using statistical energy analysis,'' J. Acoust. Soc. Am. **47**(3), 683–693 (1970).
[16] M. J. Crocker and F. M. Kessler, *Noise and Noise Control* (Vol. II, CRC Press, Boca Raton, FL, 1982).
[17] L. L. Beranek, *Noise and Vibration Control* (McGraw-Hill, New York, 1971).
[18] D. A. Bies and S. Hamid, ''*In situ* determination of loss and coupling loss factors by the power injection method,'' J. Sound Vib. **70**(2), 187–204 (1980).
[19] H. Kuttruff, *Room Acoustics*, 2nd ed. (Applied Science Publishers, London, 1979).
[20] D. A. Bies and C. H. Hansen, *Engineering Noise Control* (E&FN Spon, 1996).
[21] R. S. Ming, *Sound Intensity Techniques* (in Chinese) (Zhejiang University Press, 1996).
[22] O. K. Petterson and M. J. Newman, ''An intensity method for determining sound power *in situ*,'' Noise Control Eng. J. **31**, 93–100 (1988).

Perception and annoyance due to wind turbine noise—a dose–response relationship

Eja Pedersen[a] and Kerstin Persson Waye
Department of Environmental Medicine, Göteborg University, P.O. Box 414, SE-405 30 Göteborg, Sweden

(Received 14 November 2003; revised 1 September 2004; accepted 18 September 2004)

Installed global wind power increased by 26% during 2003, with U.S and Europe accounting for 90% of the cumulative capacity. Little is known about wind turbines' impact on people living in their vicinity. The aims of this study were to evaluate the prevalence of annoyance due to wind turbine noise and to study dose–response relationships. Interrelationships between noise annoyance and sound characteristics, as well as the influence of subjective variables such as attitude and noise sensitivity, were also assessed. A cross-sectional study was performed in Sweden in 2000. Responses were obtained through questionnaires ($n=351$; response rate 68.4%), and doses were calculated as A-weighted sound pressure levels for each respondent. A statistically significant dose–response relationship was found, showing higher proportion of people reporting perception and annoyance than expected from the present dose–response relationships for transportation noise. The unexpected high proportion of annoyance could be due to visual interference, influencing noise annoyance, as well as the presence of intrusive sound characteristics. The respondents' attitude to the visual impact of wind turbines on the landscape scenery was found to influence noise annoyance.
 [DOI: 10.1121/1.1815091]

PACS numbers: 43.50.Qp, 43.50.Sr [LCS] Pages: 3460–3470

I. INTRODUCTION

Wind turbines generate renewable energy and thus contribute to sustainable development. However, disturbance from wind turbines may be an obstacle for large-scale production (Rand and Clarke, 1990; Ackerman and Söder, 2000). Few studies have so far been directed to the prevalence of disturbance, and existing knowledge of annoyance due to wind turbines is mainly based on studies of smaller turbines of less than 500 kW (Wolsink *et al.*, 1993; Pedersen and Nielsen, 1994).

Global wind power installed at the end of 2003 reached 39 GW according to American Wind Energy Association (2004), an increase of 26% in just one year. United States (7 GW) and Europe (29 GW) account for 90% of the cumulative capacity. In Sweden, more than 600 wind turbines are operating today with a total installed capacity of 0.4 GW, producing 600 GWh per year. They are placed in 84 of Sweden's 290 municipalities both along the coasts and in rural inland areas, concerning a number of people. The goal set up by the Swedish government for 2015 is 10 TWh, leading to an increase of 1600% from today. Most of these new turbines will probably be situated off shore, but as the cost for building on land is considerably lower, the development on land is expected to continue. Already, turbines are being erected near densely populated areas. Preliminary interviews conducted among 12 respondents living within 800 m of a wind turbine, and a register study of the nature of complaints to local health and environments authorities, indicated that the main disturbances from wind turbines were due to noise, shadows, reflections from rotor blades, and spoiled views (Pedersen, 2000).

All wind turbines in Sweden are upwind devices. The most common type is a 600 or 660 kW turbine with three rotor blades, rotor diameter 42–47 m, constant rotor speed 28 rpm (84 blade passages per minute, a blade passage frequency of 1.4 Hz), and hub height of 40–50 m. They often operate singly or in multiple units of 2 to 10. The noise emission at the hub is 98–102 dBA measured at wind velocity 8 m/s at 10 m height. Earlier turbines were often downwind devices and contained low-frequency noise (Hubbard *et al.*, 1983). In contrast to these, modern machines have the rotor blades upwind and the noise is typically broadband in nature (Fig. 1), (Persson Waye and Öhrström, 2002; Björkman, 2004). There are two main types of noise sources from an upwind turbine: mechanical noise and aerodynamic noise. Mechanical noise is mainly generated by the gearbox, but also by other parts such as the generator (Lowson, 1996). Mechanical noise has a dominant energy within the frequencies below 1000 Hz and may contain discrete tone components. Tones are known to be more annoying than noise without tones, but both mechanical noise and tones can be reduced efficiently (Wagner *et al.*, 1996). Aerodynamic noise from wind turbines has a broadband character. It originates mainly from the flow of air around the blades; therefore the sound pressure levels (SPLs) increase with tip speed. Aerodynamic noise is typically the dominant component of wind turbine noise today, as manufacturers have been able to reduce the mechanical noise to a level below the aerodynamic noise. The latter will become even more dominant as the size of wind turbines increase, because mechanical noise does not increase with the dimensions of turbine as rapidly as aerodynamic noise (Wagner *et al.*, 1996).

Previous international field studies of annoyance from wind turbines have generally found a weak relationship between annoyance and the equivalent A-weighted SPL (Rand

[a]Electronic mail: eja.pedersen@set.hh.se

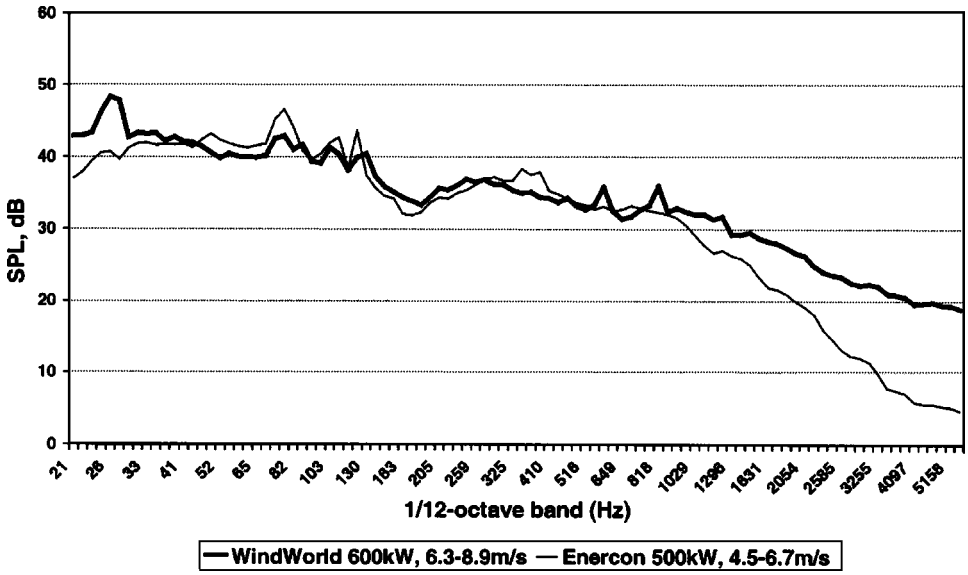


FIG. 1. Frequency spectra of two upwind three-bladed wind turbines recorded at down wind conditions; WindWorld 600 kW and Enercon 500 kW.

and Clarke, 1990; Wolsink *et al.*, 1993; Pedersen and Nielsen, 1994). It is possible that different sound properties, not fully described by the equivalent A-weighted level, are of importance for perception and annoyance for wind turbine noise. Support for such a hypothesis was given in a previous experimental study where reported perception and annoyance for five recorded wind turbine noises were different, although the equivalent A-weighted SPL were the same (Persson Waye and Öhrström, 2002). The results from that study and subsequent experiments suggested that the presence of sound characteristics subjectively described as lapping, swishing, and whistling was responsible for the differences in perception and annoyance between the sounds (Persson Waye and Agge, 2000). The descriptions swishing and whistling were found to be related to the frequency content in the range of 2000 to 4000 Hz (Persson Waye *et al.*, 1998) while the description lapping probably referred to aerodynamically induced fluctuations and was found to best be described by specific loudness over time (Persson Waye *et al.*, 2000). Sound characteristics such as described here could be of relevance for perception and annoyance, especially at low background levels.

It has been suggested that the perception of wind turbine noise could be masked by wind-generated noise. However, most of the wind turbines operating today have a stable rotor speed, and, as a consequence, the rotor blades will generate an aerodynamic noise even if the wind speed is slow and the ambient noise is low. Furthermore, noise from wind turbines comprises modulations with a frequency that corresponds to the blade passage frequency (Hubbard *et al.*, 1983) and is usually poorly masked by ambient noise in rural areas (Arlinger and Gustafsson, 1988).

It has also been shown in previous field studies that attitude to wind turbines is relevant to perceived annoyance (Wolsink *et al.*, 1993; Pedersen and Nielsen, 1994). Such a relationship, however, was not found in an experimental study where the participants were exposed to wind turbine noise (Persson Waye and Öhrström, 2002). The difference could be due to the fact that the subjects in the latter study had very little personal experience of wind turbines generally, or to their lack of visual impression during the noise exposure.

There is clearly a need for field studies to investigate the impact of wind turbines on people living in their vicinity and to further explore the presence of disturbances. In particular, dose–response relationships should be investigated to achieve a more precise knowledge of acceptable exposure levels. As noise annoyance may be interrelated to the presence of intrusive sound characteristics, ambient sound pressure level, and visual intrusion as well as individual variables, all these factors should be taken into account and their relative importance evaluated.

The aims of this study were to evaluate the prevalence of annoyance due to wind turbine noise and to study dose–response relationships. The intention was also to look at interrelationships between noise annoyance and sound characteristics, as well as the influence of subjective variables such as attitude and noise sensitivity.

II. METHOD

A. General outline

The investigation was a cross-sectional study comprising respondents exposed to different A-weighted sound pressure levels (SPL) from wind turbines. Five areas totaling 22 km^2 comprising in total 16 wind turbines and 627 households were chosen within a total area of 30 km^2 (Table I). Subjective responses were obtained through questionnaires delivered at each household and collected a week later in May and June 2000. The response rate was 68.4%. A-weighted SPLs due to wind turbines were calculated for each respondent's dwelling. Comparisons were made of the extent of annoyance between respondents living at different A-weighted SPLs.

B. Study area and study sample

The criteria for the selection of the study areas were that they should comprise a large enough number of dwellings at varying distances from operating wind turbines within a

TABLE I. Description of study areas.

Area	Square km	Wind turbines	Households	Study population	Responses	Response rate (%)
A	3.7	2	89	75	54	72.0
B	4.7	3	44	33	23	69.7
C	8.3	8	70	59	49	83.1
D	3.3	2	393	325	210	64.6
E	2.0	1	31	21	15	71.4
Total	22.0	16	627	513	351	68.4

comparable geographical, cultural, and topographical structure. Suitable areas were found in a municipality in the south of Sweden. More than 40 wind turbines are located in this region, either in small groups with two to five turbines or as single objects. The landscape is flat and mainly agricultural but small industries, roads, and railroads are also present. Most people live in privately owned detached houses in the countryside or in small villages. The wind turbines are visible from many directions. To define the study area, preliminary calculations of sound distribution were made so that the area would include dwellings exposed to similar A-weighted SPL irrespective of the number of wind turbines. Of the 16 wind turbines in the selected five areas, 14 had a power of 600–650 kW, the other two turbines having 500 kW and 150 kW. The towers were between 47 and 50 m in height. Of the turbines, 13 were WindWorld machines, 2 were Enercon, and 1 was a Vestas turbine. Figure 1 shows a $\frac{1}{12}$-octave band spectra of a WindWorld turbine sound recorded 320 m from a turbine in area A at 6.3–8.9 m/s and a spectra of an Enercon turbine sound recorded 370 m from the turbine in area E at 4.5–6.7 m/s. Both recordings were done under downwind conditions.

The study sample comprised one selected subject between the ages of 18 and 75 in each household in the area within a calculated wind turbine A-weighted SPL of more than 30 dB ($n=513$). The subject with birth date closest to May 20 was asked to answer a questionnaire.

C. Questionnaire

The purpose of the study was masked in the questionnaire; the questions on living conditions in the countryside also included questions directly related to wind turbines. The response of most questions was rated on 5-point or 4-point verbal rating scales. The key questions relevant for this paper were translated into English and are presented in the Appendix. The questionnaire was divided into four sections. The first section comprised questions regarding housing and satisfaction with the living environment, including questions on the degree of annoyance experienced outdoors and indoors from several sources of annoyance, wind turbines included. The respondent was also asked to rate his/her sensitivity to environmental factors, one being noise.

The second section of the questionnaire comprised questions on wind turbines, related to the respondent by the recent development of wind turbines in the community. The response to different visual and auditory aspects of wind turbines as noise and shadows were asked for, followed by questions on frequency of disturbances and experiences during certain activities and weather conditions. Respondents were also asked to describe their level of perception and annoyance related to the wind turbine sounds they could hear, using verbal descriptors of sound and perceptual characteristics. These descriptors were obtained from previous experimental studies were subjects initially verbally described their perception of annoying sound properties for five recorded wind turbine sounds (Persson Waye and Öhrström, 2002). This, together with some given adjectives, resulted in a total of 14 adjectives that were rated on unipolar scales with regard to annoyance. In this field study, the original descriptors were complemented with regionally used phrases. Several questions on attitude to wind turbines were also included.

The third section of the questionnaire concerned health aspects such as chronic illnesses (diabetes, tinnitus, cardiovascular diseases, hearing impairment) and general well-being (headache, undue tiredness, pain and stiffness in the back, neck or shoulders, feeling tensed/stressed, irritable). Respondents were asked questions about their normal sleep habits: quality of sleep, whether sleep was disturbed by any noise source, and whether they normally slept with the window open. The last section comprised questions on employment and working hours.

D. Calculations and measurements of noise exposure

For each respondent, A-weighted SPLs (dB) were calculated as the sum of contributions from the wind power plants in the specific area. The calculations were made with calculation points every fifth meter. The calculations followed the sound propagation model for wind power plants adopted by the Swedish Environmental Protection Agency (2001) and used as a basis for granting of building permission. The model assumes downward wind of 8 m/s at 10-m height. The calculation model is slightly different depending on the distance between the source and the receiver. For the cases in this study the following equation was used:

$$L_A = L_{WA,corr} - 8 - 20\lg(r) - 0.005r, \quad (1)$$

where r is the distance from the source to the receiver in meters. The atmospheric absorption coefficient is estimated to be 0.005 dB/m. $L_{WA,corr}$ is a modified sound power level of the wind power:

$$L_{WA,corr} = L_{WA} + k \cdot \Delta v_h. \quad (2)$$

TABLE II. Study sample, study population, and response rate related to sound category (dBA).

Sound category	<30.0	30.0–32.5	32.5–35.0	35.0–37.5	37.5–40.0	>40.0	Total
Study sample	25	103	200	100	53	32	513
Study population	15	71	137	63	40	25	351
Response rate	60.0%	68.9%	68.5%	63.0%	75.5%	78.1%	68.4%

L_{WA} is the A-weighted sound power level of the wind power plant, which in this study was given by the manufacturer; k describes how the sound power level varies with the wind speed at 10 m height and

$$\Delta v_h = v_h \left(\frac{\ln(H/z_0)}{\ln(h/z_0)} \frac{\ln(h/0.05)}{\ln(H/0.05)} - 1 \right), \tag{3}$$

where v_h is the wind speed at 10-m height, H the height of the hub, h is 10 m, and z_0 the surface roughness length. In these calculations, $z_0 = 0.05$ m (fields with few buildings) was used and therefore no value of k was needed. The SPL calculated this way is an estimate for the equivalent level for a hypothetical time period with continuous performance at downwind conditions 8 m/s at 10-m height.

To verify the calculations, to record frequency spectra, and to study background sound, a mobile caravan equipped with a sound level meter (Larson & Davis type 820), digital audio tape recorder (Sony TCD-D8 DAT), and meteorological instruments (Davis Weather Monitor type II) was used. The mobile station was placed on different sites of the study area. Both the meteorological instruments and the noise recording instruments were computer controlled and directed remotely via a cellular phone. The microphone was attached on a vertical hardboard facing the noise source. The equipment and procedures are thoroughly described by Björkman (2004). The sound pressure levels measured on the reflecting plane were corrected by −6 dB to present the free field value. The ambient sound pressure level varied from 33 dB $L_{\mathrm{Aeq,5\,min}}$ to 44 dB $L_{\mathrm{Aeq,5\,min}}$. The variations were mainly due to the amount of traffic within a 24-h time period. The lower background levels typically occurred during evening and nights.

The respondents were classified into six sound categories according to the calculated wind turbine A-weighted SPL at their dwelling. Table II shows the number of respondents living within each sound category and also the study sample and response rate for each sound category.

Data for the distance between the dwelling of the respondent and the nearest wind turbine were obtained from property maps, scale 1:10 000. The distance differed within each sound category, depending on the number of wind turbines in the area—the larger number of wind turbines, the shorter distance at the same A-weighted SPL. Table III shows the relationship between distance and A-weighted SPL. Two values are given for each category: the range and the median interval.

E. Statistical treatment of data

Due to the fact that most of the data were categorical (ordered or nonordered) and not continuos data, and therefore no assumptions on probability distribution could be made, nonparametric statistical methods were used, all described by Altman (1991). Data from verbal rating scales were calculated as proportions with 95% confidence intervals. When relevant, the two highest ratings of annoyance (rather annoyed and very annoyed) were classified as annoyed and the three lower ones as not annoyed (do not notice, notice but not annoyed, and slightly annoyed). In the analysis of attitude, negative and very negative were classified as negative; in the analysis of sensitivity, rather sensitive and very sensitive were classified as sensitive. More advanced statistical analyses were carried out using SPSS version 11.0. Relationships between variables were evaluated using Spearman's nonparametric rank correlation (r_s). Pearson's chi-square (chi2) was used to test that all sound categories contained the same proportion of observations. To evaluate differences between two unmatched samples of observations on an ordinal scale (e.g., comparing men and women's answers on a 5-graded verbal rating scale), the Mann–Whitney test was used (z_{MW}); a nonparametric test equivalent to the t test, but based on ranks (Altman, 1991). All significance tests were two-sided and p-values below 0.05 were considered statistically significant. When exploring several relationships at the same time, 1 out of 20 calculations would be classified as statistically significant by chance. This risk of mass significance was avoided using Bonferroni's method when appropriate, reducing the p-value considered statistically significant by dividing it with the number of correlations calculated at the same time (Altman, 1991).

Binary logistic multiple regression was used to study the impact of different variables on annoyance of wind turbine noise (annoyed–not annoyed). Sound category was used as the dose variable. Logistic regression is a method used to make a nonlinear function into a linear equation, using odds rather than straightforward probability. The equation is

TABLE III. Distance between dwelling and nearest wind turbine related to sound category (dBA).

Sound category	<30.0	30.0–32.5	32.5–35.0	35.0–37.5	37.5–40.0	>40.0
Range (m)	650–1049	550–1199	450–1099	300–799	300–749	150–549
Median interval (m)	850–899	750–799	550–599	450–499	350–399	300–349

TABLE IV. Characteristics of the respondents given as proportions of respondents in each sound category (dBA) and in total.

Sound category	<30.0	30.0–32.5	32.5–35.0	35.0–37.5	37.5–40.0	>40.0	Total
n	15	71	137	63	40	25	351
Gender: Male (%)	27	35	39	50	50	48	42
Residence: Detached houses/farms (%)	100	83	61	100	97	96	81
Occupation: Employed (%)	67	59	58	53	69	67	60
Sensitive[a] to noise (%)	62	44	49	53	58	50	50
Negative[b] to wind turbines (%)	8	10	11	18	20	8	13
Negative[b] to visual impact (%)	43	33	38	41	40	58	40
Long-term illness (%)	20	29	28	16	30	24	26
Age: Mean	46	47	47	50	48	48	48
(SD)	(13.3)	(13.7)	(14.3)	(14.6)	(13.1)	(14.3)	(14.0)

[a]Sensitive consists of the two ratings: rather sensitive and very sensitive.
[b]Negative consists of the two ratings: rather negative and very negative.

$$\ln\left(\frac{p}{1-p}\right)=b_0+b_1x_1+b_2x_2+\cdots, \quad (4)$$

where, in this case, p is the probability of being annoyed by noise from wind turbines, x_1-x_n are the variables put into the model, and b_1-b_n are the logarithmic value of the odds ratio for one unit change in the respective variable (Altman, 1991). A relevant measurement of explained variance using nonparametric statistics is Nagelkerke pseudo-R^2 (Nagelkerke, 1991).

To estimate how consistently the respondents answered to questions measuring similar response, Cronbach's alpha (Miller, 1995) was calculated as a testing of the internal consistency reliability of the questionnaire. Five of the questions regarding wind turbine noise were compared: annoyance outdoors, annoyance indoors, annoyance of rotor blades, annoyance of machinery, annoyance as a describing adjective. Demographic data on age and gender of the population in the four parishes in the study area were collected from local authorities. The study population was compared to these demographical data, parish-by-parish, and divided into 10-year categories for age and gender, as well as in total.

III. RESULTS

A. Study population

The overall response rate was 68.4%, ranging from 60.0% to 78.1% in the six sound categories (Table II). No statistically significant differences in variables related to age, gender, or employment were found among sound categories (Table IV). A statistically significant difference was found between sound categories as to whether respondents lived in apartments or detached houses (chi2=62.99, df=5, $p<0.001$). Overall, most of the respondents (80%) lived in privately owned detached houses or on farms. The remaining lived in tenant-owned or rented apartments. The latter were more frequent in sound category 32.5–35.0 dBA (Table IV). However, there was no statistically significant difference between the respondents living in privately owned detached houses or on farms, on one hand, and those living in tenant-owned or rented apartments, on the other hand, regarding subjective factors, when correcting for requirements to avoid mass significance. Most of the respondents did not own a wind turbine or share of a wind turbine (95%, $n=335$). No statistically significant differences in variables related to noise sensitivity, attitude, or health were found between the different sound categories.

The mean age in the study population was 48 years (SD $=14.0$) (Table IV) which did not differ statistically significantly from the demographic data (45 years, SD=15.2). The proportion of women in the study population was slightly higher than in the demographic data; in the study population, 58% women and 42% men (Table IV), compared to 49% women and 51% men in the demographic data. However, no statistically significant differences were found between men and women regarding perception and annoyance due to wind turbine noise, noise sensitivity, or attitude to wind turbines. Differences between genders were found regarding well-being. Women suffered more often from headache ($z_{MW}=-3.243$, $n=328$, $p<0.001$), undue tiredness ($z_{MW}=-3.549$, $n=327$, $p<0.05$), pain and stiffness in back, neck or shoulders ($z_{MW}=-3.312$, $n=331$, $p<0.001$), and tension/stress ($z_{MW}=-3.446$, $n=328$, $p<0.001$).

B. Main results

The proportion of respondents who noticed noise from wind turbines outdoors increased sharply from 39% ($n=27$, 95%CI: 27%–50%) at sound category 30.0–32.5 dBA to 85% ($n=53$, 95%CI: 77%–94%) at sound category 35.0–37.5 dBA (Table V). The proportion of those annoyed by wind turbine noise outdoors also increased with higher sound category, at sound categories exceeding 35.0 dBA. The correlation between sound category and outdoor annoyance due to wind turbine noise (scale 1–5) was statistically significant ($r_s=0.421$, $n=341$, $p<0.001$). No respondent self-reported as annoyed at sound categories below 32.5 dBA, but at sound category 37.5–40.0 dBA, 20% of the 40 respondents living within this exposure were very annoyed and above 40 dBA, 36% of the 25 respondents (Table V).

To explore the influence of the subjective factors on noise annoyance, binary multiple logistic regression was used (Table VI). Eight models were created, all containing sound category as the prime variable assumed to predict noise annoyance. The three subjective factors of attitude to visual impact, attitude to wind turbines in general, and sensitivity to noise were forced into the model one-by-one, two-by-two, and finally all together. In the first model only noise

TABLE V. Perception and annoyance outdoors from wind turbine noise related to sound exposure.

	<30.0 $n=12$ %(95%CI)	30.0–32.5 $n=70$ %(95%CI)	32.5–35.0 $n=132$ %(95%CI)	35.0–37.5 $n=62$ %(95%CI)	37.5–40.0 $n=40$ %(95%CI)	>40.0 $n=25$ %(95%CI)
Do not notice	75 (51–100)	61(50–73)	38(30–46)	15(3–23)	15(4–26)	4(19–57)
Notice, but not annoyed	25(1–50)	24(14–34)	28(20–36)	47(34–59)	35(20–50)	40(19–57)
Slightly annoyed	0	14(6–22)	17(10–23)	26(15–37)	23(10–35)	12(19–57)
Rather annoyed	0	0	10(5–15)	6(0–13)	8(−1–16)	8(19–57)
Very annoyed	0	0	8(3–12)	6(0–13)	20(8–32)	36(17–55)

exposure was used as the independent variable. The Exp(b) was 1.87, i.e., the odds for being annoyed by noise from wind turbines would increase 1.87 times from one sound category to the next. When adding the subjective factor of attitude to visual impact as an independent variable, the influence of the noise exposure decreased, but was still statistically significant. The pseudo-R^2 increased from 0.13 to 0.46, indicating that the new model explained 46% of the variance in annoyance. Adding the two remaining subjective factors did not improve the model as the coefficients did not reach statistical significance.

Noise from rotor blades was reported as the most annoying aspect of wind turbines. Of the respondents, 16% ($n=54$, 95%CI: 12%–20%) were annoyed by noise from rotor blades. Changed view (14%, $n=48$, 95%CI: 10%–18%), noise from machinery (9%, $n=33$, 95%CI: 6%–12%), shadows from rotor blades (9%, $n=29$, 95%CI: 6%–11%), and reflections from rotor blades (7%, $n=22$, 95%CI: 4%–9%) were also reported.

C. Attitude and sensitivity

Almost all respondents (93%, $n=327$, 95%CI: 91%–96%) could see one or more wind turbines from their dwelling or garden. When asked for judgments on wind turbines, the adjectives that were agreed on by most respondents were ''environmentally friendly'' (79%), ''necessary'' (37%), ''ugly'' (36%), and ''effective'' (30%). Only the word ''annoying'' (25%) was judged higher among those in higher sound categories than among those in lower sound categories ($z_{MW}=-3.613$, $n=351$, $p<0.001$).

The high judgment of the word ''ugly'' corresponds to the outcome of the attitude questions. Of the respondents, only 13% ($n=44$, 95%CI: 9%–16%) reported that they were negative or very negative to wind turbines in general, but 40% ($n=137$, 95%CI: 34%–44%) that they were negative or very negative to the visual impact of wind turbines on the landscape scenery (Table IV).

All correlations between sound category, noise annoyance, and subjective factors are shown in Table VII. Noise annoyance was correlated to both sound category and the three subjective factors, strongest to attitude to the wind turbines' visual impact on the landscape. The subjective factors were also correlated to each other, except for general attitude and sensitivity to noise. Of all the respondents, 50% ($n=169$, 95%CI: 45%–55%) regarded themselves as rather sensitive or very sensitive to noise (Table IV).

When comparing those annoyed by wind turbine noise and those not, no differences were found regarding the judgments of the local authorities, with the exception of perceived opportunity to influence local government ($z_{MW}=-2.753$, $n=300$, $p<0.005$). Those annoyed reported negative changes to a higher degree ($z_{MW}=-5.993$, $n=307$, p

TABLE VI. Results of multiple logistic regression analyses with 95% confidence intervals.

	Variables	b	p-value	Exp(b) (95%CI)	Pseudo-R^{2a}
1	Noise exposure	0.63	<0.001	1.87(1.47–2.38)	0.13
2	Noise exposure	0.55	<0.001	1.74(1.29–2.34)	0.46
	Attitude to visual impact	1.62	<0.001	5.05(3.22–7.92)	
3	Noise exposure	0.62	<0.001	1.86(1.45–2.40)	0.20
	Attitude to wind turbines	0.56	<0.001	1.74(1.30–2.33)	
4	Noise exposure	0.63	<0.001	1.88(1.46–2.42)	0.18
	Sensitivity to noise	0.56	<0.005	1.75(1.19–2.57)	
5	Noise exposure	0.55	<0.001	1.73(1.28–2.33)	0.46
	Attitude to visual impact	1.66	<0.001	5.28(3.26–8.56)	
	Attitude to wind turbines	−0.10	0.319	0.91(0.64–1.28)	
6	Noise exposure	0.57	<0.001	1.77(1.30–2.40)	0.47
	Attitude to visual impact	1.59	<0.001	4.88(3.08–7.72)	
	Sensitivity to noise	0.22	0.344	1.25(0.79–1.96)	
7	Noise exposure	0.63	<0.001	1.88(1.45–2.45)	0.24
	Attitude to wind turbines	0.58	<0.001	1.78(1.32–2.41)	
	Sensitivity to noise	0.59	<0.005	1.80(1.22–2.67)	
8	Noise exposure	0.56	<0.001	1.76(1.29–2.39)	0.47
	Attitude to visual impact	1.63	<0.001	5.11(3.10–8.41)	
	Attitude to wind turbines	−0.10	0.597	0.91(0.64–1.29)	
	Sensitivity to noise	0.21	0.373	1.23(0.78–1.94)	

[a]Nagelkerke (1991).

TABLE VII. Correlation between noise annoyance, sound category (dBA) and the subjective variables. Statistically significant correlations in boldface. To avoid the risk of mass significance $p<0.008$ were required for statistical significance.

	Sound category	Attitude to visual impact	Attitude to wind turbines	Sensitivity to noise
Noise annoyance	**0.421**	**0.512**	**0.334**	**0.197**
Sound category	···	**0.145**	0.074	0.069
Attitude to visual impact		···	**0.568**	**0.194**
Attitude to wind turbines			···	0.023
Sensitivity to noise				···

<0.001); 83% compared to 37% among those not annoyed. Of the 138 respondents who reported negative changes overall, 41% ($n=57$, 95%CI: 33%–50%) specified wind turbines in the response to an open question.

D. The occurrence of noise annoyance

Among those who noticed wind turbine noise ($n=223$), 25% ($n=47$, 95%CI: 18%–31%) reported that they were disturbed every day or almost every day and 17% ($n=33$, 95%CI: 12%–23%) once or twice a week. Annoyance was most frequently reported when relaxing outdoors and at barbecue nights.

Perception of wind turbine noise was influenced by weather conditions. Of the respondents who noticed wind turbine noise, 54% stated that they could hear the noise more clearly than usual when the wind was blowing from the turbines towards their dwelling. Only 9% reported that the noise was heard more clearly when the wind was from the opposite direction. The noise was also more clearly noticed when a rather strong wind was blowing (39%), but 18% reported that the noise was more clearly noticed in low wind. For warm summer nights, 26% noticed the noise more clearly than usual.

E. Sound characteristics

There was a statistically significant correlation between sound category and annoyance due to noise from rotor blades ($r_s=0.431$, $n=339$, $p<0.001$) and from the machinery ($r_s=0.294$, $n=333$, $p<0.001$). In all sound categories, a higher proportion of respondents noticed noise from rotor blades than from the machinery (Fig. 2). The proportion who noticed noise from rotor blades was similar to the proportion of respondents who noticed wind turbine noise in general. Noise from rotor blades was noticed in lower sound categories than noise from the machinery, i.e., it could be heard at a greater distance. However, comparing the numbers of annoyed with the numbers of those who could hear noise from the two sources, respectively, both noises were almost equally annoying. Of the 215 respondents who noticed noise from rotor blades, 25% ($n=54$, 95%CI: 19%–31%) were annoyed. Of the 101 respondents who noticed noise from the machinery, 30% ($n=30$, 95%CI: 21%–39%) were annoyed.

Among those who noticed noise from wind turbines, swishing, whistling, pulsating/throbbing, and resounding were the most common sources of annoyance according to verbal descriptors of sound characteristics (Table VIII). These descriptors were all highly correlated to noise annoyance. All other verbal descriptors of sound characteristics were also statistically significantly correlated to noise annoyance, but to a lower degree. When analyzing annoyance due to noise from rotor blades, the strongest correlated verbal descriptor of sound characteristics was swishing ($r_s=0.807$, $n=185$, $p<0.001$), which can be compared to noise annoyance due to noise from the machinery—which had the highest correlation with scratching/squeaking ($r_s=0.571$, $n=133$, $p<0.001$).

F. Indoor noise annoyance and sleep disturbance

A total of 7% of respondents ($n=25$, 95%CI: 5%–10%) were annoyed by noise from wind turbines indoors. Forty-five percent ($n=24$, 95%CI: 32%–59%) of those who were annoyed by noise from wind turbines outdoors were also

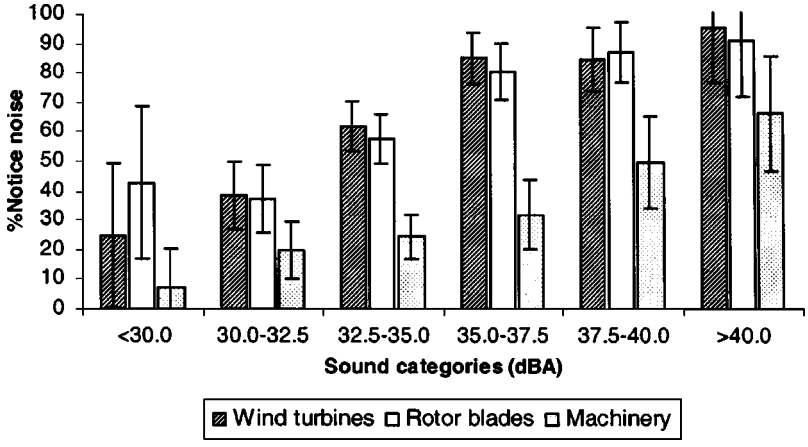


FIG. 2. Proportions with 95% confidence intervals of perception outdoors due to noise (notice but not annoyed, slightly annoyed, rather annoyed, very annoyed) from wind turbines, from rotor blades, and from machinery, related to sound categories.

TABLE VIII. Verbal descriptors of sound characteristics of wind turbine noise, based on those who noticed wind turbine sound ($n=223$). Statistically significant correlations in boldface. To avoid the risk of mass significance $p<0.0062$ were required for statistical significance.

	Annoyed by the specified sound character	Correlation to noise annoyance
Swishing	33%(27%–40%)	**0.718**
Whistling	26%(18%–33%)	**0.642**
Pulsating/throbbing	20%(14%–27%)	**0.450**
Resounding	16%(10%–23%)	**0.485**
Low frequency	13% (7%–18%)	**0.292**
Scratching/squeaking	12% (6%–17%)	**0.398**
Tonal	7% (3%–12%)	**0.335**
Lapping	5% (1%–8%)	**0.262**

annoyed indoors. There was a statistically significant correlation between indoor annoyance and sound category ($r_s=0.348$, $n=340$, $p<0.001$).

Regarding sleep disturbance, 23% ($n=80$, 95%CI: 18%–27%) of respondents stated that they were disturbed in their sleep by noise. Several sources of sleep disturbance, such as road traffic, rail traffic, neighbors, and wind turbines, were reported in an open question. At lower sound categories, no respondents were disturbed in their sleep by wind turbine noise, but 16% ($n=20$, 95%CI: 11%–20%) of the 128 respondents living at sound exposure above 35.0 dBA stated that they were disturbed in their sleep by wind turbine noise. Of those, all except two slept with an open window in the summer. No statistically significant correlations were found between sleep quality in general and outdoor noise annoyance, indoor noise annoyance, attitude to visual impact, attitude to wind turbines in general, or sensitivity to noise.

IV. DISCUSSION

A. Method

The results were based on the questionnaire survey and calculated A-weighted SPL. The purpose of the study was masked in order to avoid other factors such as attitude and ownership influencing the answers. The survey method is well established and has been used in several previous studies exploring annoyance due to community noise (e.g., Öhrström, 2004).

The results indicate a high validity for the questionnaire. The questions detected annoyance by odor from industrial plants in the area where the biogas plant is located [of those annoyed by odor from industrial plants, 83% ($n=19$) lived close to the biogas plant]; it also detected annoyance by noise from trains in the areas where the train passes [all of the respondents who reported that they were annoyed by noise from railway traffic ($n=12$) lived in areas where the railway passed]. There was a high correspondence between the responses to the general question of noise from wind turbines at the beginning of the questionnaire and the more specific questions later (alpha: 0.8850, $n=326$), also indicating high reliability of results.

The response rate at the different sound categories ranged from 60.0% to 78.1%, with the overall mean 68.4% and the dropout fairly equally distributed over sound categories. The distribution of age in the study population was similar to that of the demographic data for the area, but the proportions of women were somewhat higher than expected, especially in the lower sound categories. It has previously been shown that annoyance is not related to gender (Miedema and Vos, 1999) and as this study found no differences between men and women regarding noise annoyance and attitude to wind turbines, the higher proportion of women in the study population presumably had no impact on the results. A rather high proportion, 50%, of respondents self-reported as rather or very sensitive to noise. Other field studies in Sweden on annoyance due to road traffic noise in urban areas have found a lower proportion of noise-sensitive persons; for example, Matsumura and Rylander (1991) reported 25% of the respondents as noise sensitive in a road traffic survey ($n=805$). The difference might reflect preference of living environment, indicating that noise sensitive individuals prefer a more rural surrounding or that people living in areas with low background noise levels might develop a higher sensitivity to noise.

The calculated A-weighted SPL reflected downwind conditions assuming a wind speed of 8 m/s. Over a larger period of time, the direction and speed of the wind will vary and hence affect the actual SPL at the respondent's dwelling. It is likely that these variations, seen as an average over a longer period of time, in most cases will result in lower levels than the calculated SPL. Several unreliabilities related to the calculations might have led to an over- or underestimation of the dose levels. However, this error would not invalidate the comparison between respondents living at different SPL. Another source of error is that no account was taken of the physical environment around the respondent's house (e.g., location of patio or veranda, presence of bushes and trees in the garden). The actual SPL that the respondent experienced in daily life might therefore differ from the calculated, leading in most cases to an overestimation of the calculated dose.

B. Results

The results suggest that the proportions of respondents annoyed by wind turbine noise are higher than for other community noise sources at the same A-weighted SPL and that the proportion annoyed increases more rapidly. A comparison between established estimations of dose–response relationships for annoyance of transportation noise (Schultz, 1978; Fidell *et al.*, 1991; Miedema and Voss, 1998; Miedema and Oudshoorn, 2001; Fidell, 2003) and an estimation of a dose–response relationship for wind turbine noise, based on the findings in this study, are shown in Fig. 3. All curves are third order polynomials. The established curves describing annoyance from transportation noise are based on a large amount of data, and the wind turbine curve on only one study, so interpretations should be done with care. An important difference between studies of transportation noises and wind turbine noise is however where the main annoyance reaction is formed. For most studies of transportation noises

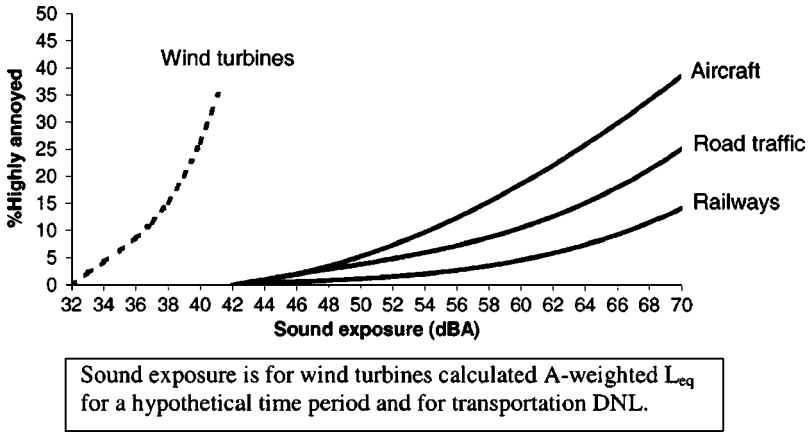


FIG. 3. A comparison between the dose–response relationship for transportation noise estimated by third order polynomials suggested by Miedema and Oudshoorn (2001) and wind turbine noise (dotted line). The latter ($\%HA = 4.38*10^{-2}(LEQ-32)^3 - 2.413*10^{-1}(LEQ-32)^2 + 2.4073(LEQ-32)$) were derived using regression based on five points interpolated from sound categories used in this study and the assumption that "very annoyed" in this study equals "highly annoyed" (Miedema and Voss, 1998).

it can be assumed that annoyance is formed mainly as a reaction to the sound pressure levels perceived indoors, and hence the actual noise dose should be reduced by the attenuation of the façade. For wind turbine noise the main annoyance reaction is formed when spending time outdoors. The actual difference in noise dose could therefore, at least partly, explain the comparatively higher prevalence of noise annoyance due to wind turbines. However, this factor does not explain the steep gradient.

Another factor that could be of importance for explaining the seemingly different dose–response relationships is that the wind turbine study was performed in a rural environment, where a low background level allows perception of noise sources even if the A-weighted SPL are low. Wind turbine noise was perceived by about 85% of the respondents even when the calculated A-weighted SPL were as low as 35.0–37.5 dB. This could be due to the presence of amplitude modulation in the noise, making it easy to detect and difficult to mask by ambient noise. This is also confirmed by the fact that the aerodynamic sounds were perceived at a longer distance than machinery noise.

Data obtained in this study also suggest that visual and/or aesthetic interference influenced noise annoyance. Support for this hypothesis can be found in studies evaluating auditory-visual interactions (Viollon *et al.*, 2002). In one field-laboratory study, subjects evaluating annoyance due to traffic noise were less annoyed if a slide of a visually attractive street was presented together with the noise, as compared to the same noise level presented together with a visually unattractive street. The difference in noise annoyance amounted to as much as 5 dBA (Kastka and Hangartner, 1986). The hypothesis was also supported by the logistic multiple regression analyses in the present study, where the visual variable attitude to visual impact had a significant impact on the model. However, although the inclusion of the variable increased the pseudo-R^2, the influence of noise exposure was still a significant factor for noise annoyance. A general prediction of the visual influence on noise annoyance, however, can not yet be made with any certainty as both attenuating (Kastka and Hangartner, 1986) and amplifying effects (e.g., Watts *et al.*, 1999) have been detected.

The high prevalence of noise annoyance could also be due to the intrusive characteristics of the aerodynamic sound. The verbal descriptors of sound characteristics related to the aerodynamic sounds of swishing, whistling, pulsating/throbbing, and resounding were—in agreement with this hypothesis—also reported to be most annoying. The results for the sounds of swishing and whistling agree well with results from previous experimental studies (Persson Waye *et al.*, 2000; Persson Waye and Agge, 2000; Persson Waye and Öhrström, 2002), while pulsating/throbbing in those studies was not significantly related to annoyance.

Most respondents who were annoyed by wind turbine noise stated that they were annoyed often, i.e., every day or almost every day. The high occurrence of noise annoyance indicates that the noise intrudes on people's daily life. The survey was performed during May and June when people could be expected to spend time outdoors, and the results therefore reflect the period that is expected to be most sensitive for annoyance due to wind turbine noise.

A low number of respondents were annoyed indoors by wind turbine noise. Some of the respondents also stated that they were disturbed in their sleep by wind turbine noise, and the proportions seemed to increase with higher SPL. The number of respondents disturbed in their sleep, however, was too small for meaningful statistical analysis, but the probability of sleep disturbances due to wind turbine noise can not be neglected at this stage.

Noise annoyance was also related to other subjective factors such as attitude and sensitivity. These results correspond well with the results from other studies regarding community noise (e.g., noise from aircraft, railways, road traffic, and rifle ranges). In a summary of 39 surveys performed in ten different countries, the correlation was 0.42 between dose and response, 0.15 between exposure and attitude, 0.41 between annoyance and attitude, −0.01 between exposure and sensitivity, and 0.30 between annoyance and sensitivity (Job, 1988). Corresponding numbers from this study are presented in Table VII and show a noteworthy similarity.

Two aspects of attitude were explored in the present study. Attitude to the visual impact of wind turbines on the landscape scenery was more strongly correlated to annoyance than the general attitude to wind turbines. The four most supported adjectives queried in the survey were environmentally friendly, necessary, ugly, and effective, thus giving the picture of a phenomenon that is accepted, but not regarded as a positive contribution to the landscape.

Previous studies of community noise have found that people who tend to be consistently negative could be predicted to be more annoyed by a new source of noise (Weinsten, 1980). More recent studies on community noise have included additional aspects and suggest conceptual models describing individual differences in the terms of stress, appraisal, and coping (Lercher, 1996). In the case of annoyance due to wind turbine noise, the findings suggest that individual differences others than attitude and sensitivity could influence the variation of noise annoyance. Respondents annoyed by wind turbine reported negative changes in their neighborhood to a higher degree than those not annoyed and stated that they had little perceived opportunity to influence local government. The importance of these parameters for noise annoyance due to wind turbines should be further studied.

C. Conclusions

A significant dose–response relationship between calculated A-weighted SPL from wind turbines and noise annoyance was found. The prevalence of noise annoyance was higher than what was expected from the calculated dose. It is possible that the presence of intrusive sound characteristics and/or attitudinal visual impacts have an influence on noise annoyance. Further studies are needed, including a larger number of respondents especially at the upper end of the dose curve, before firm conclusions could be drawn. To explore attitude with regard to visual impact, some of these studies should be performed in areas of different topography where the turbines are less visible. There is also a need to further explore the influence of individual and contextual parameters.

ACKNOWLEDGMENTS

We gratefully acknowledge the assistance of Agneta Agge; we also thank the Swedish Energy Agency P13644-1 and the Adlerbertska Research Foundation for funding the study. This article has benefited greatly from comments provided by Associate Editor Lou Sutherland and two anonymous reviewers.

APPENDIX: QUESTIONNAIRE

Key questions from the questionnaire used in the study. Questions with the main purpose to mask the intention of the questionnaire and standard questions on socio-economic status and health are not shown here. Translated from Swedish.

Section I

—How satisfied are you with your living environment? (very satisfied, satisfied, not so satisfied, not satisfied, not at all satisfied)

—Have there been any changes to the *better* in your living environment/municipality during the last years? (no, yes) State which changes.

—Have there been any changes to the *worse* in your living environment/municipality during the last years? (no, yes) State which changes.

—State for each nuisance below if you notice or are annoyed when you spend time *outdoors* at your dwelling: odor from industries, odor from manure, flies, noise from hay fans, noise from wind turbines, railway noise, road traffic noise, lawn mowers. (do not notice, notice but not annoyed, slightly annoyed, rather annoyed, very annoyed)

—State for each nuisance below if you notice or are annoyed when you spend time *indoors* in your dwelling: odor from industries, odor from manure, flies, noise from hay fans, noise from wind turbines, railway noise, road traffic noise, lawn mowers. (do not notice, notice but not annoyed, slightly annoyed, rather annoyed, very annoyed)

—How would you describe your sensitivity to the following environmental factors: air pollution, odors, noise, littering? (not sensitive at all, slightly sensitive, rather sensitive, very sensitive)

Section II

—Can you see any wind turbine from your dwelling or your garden? (yes, no)

—What is your opinion on the wind turbines' impact on the landscape scenery? (very positive, positive, neither positive nor negative, negative, very negative)

—Are you affected by wind turbines in your living environment with regard to: shadows from rotor blades, reflections from rotor blades, sound from rotor blades, sound from machinery, changed view? (do not notice, notice but not annoyed, slightly annoyed, rather annoyed, very annoyed)

—If you are annoyed by noise, shadows and/or reflections from wind turbines, how often does this happen? (never/almost never, some/a few times per year, some/a few times per month, some/a few times per week, daily/almost daily)

—If you hear sound from wind turbines, how would you describe the sound: tonal, pulsating/throbbing, swishing, whistling, lapping, scratching/squeaking, low frequency, resounding? (do not notice, notice but not annoyed, slightly annoyed, rather annoyed, very annoyed)

—Have you noticed if sounds from wind turbines sound different at special occasions: when the wind blows from the turbine towards my dwelling, when the wind blows from my dwelling towards the turbine, when the wind is low, when the wind is rather strong, warm summer nights? (less clearly heard, more clearly heard, no differences, do not know)

—Are you annoyed by sound from wind turbines during any of the following activities: relaxing outdoors, barbecue nights, taking a walk, gardening, other outdoor activity? (do not notice, notice but not annoyed, slightly annoyed, rather annoyed, very annoyed)

—Do you own any wind turbines? (no, yes I own one or more turbines, yes I own shares of wind turbines)

—What is your general opinion on wind turbines? (very positive, positive, neither positive nor negative, negative, very negative)

—Please mark the adjectives that you think are adequate for wind turbines: efficient, inefficient, environmentally friendly, harmful to the environment, unnecessary, necessary, ugly, beautiful, inviting, threatening, natural, unnatural, annoying, blends in.[1]

[1]Developed by Karin Hammarlund, Department of Human and Economic Geography, Göteborg University, Sweden, and used with her permission.

Ackermann, T., and Söder, L. (**2000**). ''Wind energy technology and current status—a review,'' Renew. Sustain. Energy Rev. **4**, 315–374.

Altman, D. G. (**1991**). *Practical Statistics for Medical Research* (Chapman & Hall, London).

American Wind Energy Association (**2004**). *Global wind energy market report 2004*, http://www.awea.org/pubs/documents/globalmarket2004.pdf (last viewed online 15 June, 2004).

Arlinger, S., and Gustafsson, H.-Å. (**1988**). ''Hur ett ljudband med konstant ljudnivå maskerar ett brusband med periodiskt varierande ljudnivå'' (''How a broadband noise with constant sound pressure level masks a broadband with periodically varying sound pressure levels''), Department of Technical Audiology, Linköping University, Sweden.

Björkman, M. (**2004**). ''Long time measurements of noise from wind turbines,'' J. Sound Vib. **277**, 567–572

Fidell, S. (**2003**). ''The Schultz curve 25 years later: A research perspective,'' J. Acoust. Soc. Am. **144**, 3007–3015.

Fidell, S., Barber, D. S., and Schultz, T. J. (**1991**). ''Updating a dosage-effect relationship for the prevalence of annoyance due to general transportation noise,'' J. Acoust. Soc. Am. **89**, 221–233.

Hubbard, H. H., Grosveld, F. W., and Shepherd, K. P. (**1983**). ''Noise characteristics of large wind turbine generators,'' Noise Control Eng. J. **21**, 21–29.

Job, R. F. S. (**1988**). ''Community response to noise: a review of factors influencing the relationship between noise exposure and reaction,'' J. Acoust. Soc. Am. **83**, 991–1001.

Kastka, J., and Hangartner, M. (**1986**). ''Machen hässlichen Strassen den Verkehrlärm lästiger'' (''Do ugly streets make traffic noise more annoying?''), Arcus **1**, 23–29.

Lercher, P. (**1996**). ''Environmental noise and health: an integrated research perspective,'' Environ. Int. **22**, 117–128.

Lowson, M. V. (**1996**). ''Aerodynamic noise of wind turbines,'' in *Proceedings of Internoise*, 30 July–2 August, Liverpool, England, pp. 479–484.

Matsumura, Y., and Rylander, R. (**1991**). ''Noise sensitivity and road traffic annoyance in a population sample,'' J. Sound Vib. **151**, 415–419.

Miedema, H. M. E., and Oudshoorn, C. G. M. (**2001**). ''Annoyance from transportation noise: relationships with exposure metrics DNL and DENL and their confidence intervals,'' Environ. Health Perspect. **109**, 409–416.

Miedema, H. M. E., and Vos, H. (**1998**). ''Exposure–response relationships for transportation noise,'' J. Acoust. Soc. Am. **104**, 3432–3445.

Miedema, H. M. E., and Vos, H. (**1999**). ''Demographic and attitudinal factors that modify annoyance from transportation noise,'' J. Acoust. Soc. Am. **105**, 3336–3344.

Miller, M. B. (**1995**). ''Coefficient alpha: a basic introduction from the perspectives of classical test theory and structural equation modeling,'' Struct. Eq. Model. **2**, 255–273.

Nagelkerke, N. J. D. (**1991**). ''A note on the general definition of the coefficient of determination,'' Biometrika **78**, 691–692.

Öhrström, E. (**2004**). ''Longitudinal surveys on effects of changes in road traffic noise-annoyance, activity disturbances, and psycho-social well-being,'' J. Acoust. Soc. Am. **145**, 719–729.

Pedersen, E. (**2000**). ''Störningsupplevelser från vindkraft, förstudie'' (''Experience of annoyance from wind turbines, a pilot study''), Halmstad University, Sweden.

Pedersen, T. H., and Nielsen, K. S. (**1994**). ''Genvirkning af støj fra vindmøller'' (''Annoyance by noise from wind turbines''), Report 150, DELTA Acoustic & Vibration, Lydtekniske Institut, Copenhagen, Denmark.

Persson Waye, K., and Agge, A. (**2000**). ''Experimental quantification of annoyance to unpleasant and pleasant wind turbine sounds,'' in *Proceedings of Internoise*, 27–30 August, Nice, France, pp. 3994–3997.

Persson Waye, K., and Öhrström, E. (**2002**). ''Psycho-acoustic characters of relevance for annoyance of wind turbine noise,'' J. Sound Vib. **250**, 65–73.

Persson Waye, K., Öhrström, E., and Björkman, M. (**1998**). ''Sounds from wind turbines–can they be made more pleasant?'' in *Proceedings of 7th International Congress on Noise as a Public Health Problem*, 22–26 November, Sydney, Australia, pp. 531–534.

Persson Waye K., Agge, A., and Björkman M. (**2000**). ''Pleasant and unpleasant characteristics in wind turbine sounds,'' in *Proceedings of Internoise*, 27–30 August, Nice, France, pp. 2327–2330.

Rand, M., and Clarke, A. (**1990**). ''The environmental and community impacts of wind energy in the UK,'' Wind Eng. **14**, 319–330.

Schultz, T. J. (**1978**). ''Synthesis of social surveys on noise annoyance,'' J. Acoust. Soc. Am. **64**, 377–405.

Swedish Environmental Protection Agency (**2001**). ''Ljud från vindkraftverk'' (''Noise from wind turbines''), Report 6241, Stockholm, Sweden.

Viollon, S., Lavandier, C., and Drake, C. (**2002**). ''Influence of visual setting on sound ratings in an urban environment,'' Appl. Acoust. **63**, 493–511.

Wagner, S., Bareiss, R., and Guidati, G. (**1996**). ''Wind turbine noise,'' EUR 16823 (Springer).

Watts, G., Chinn, L., and Godfrey, N. (**1999**). ''The effects of vegetation on the perception of traffic noise,'' Appl. Acoust. **56**, 39–56.

Weinsten, N. D. (**1980**). ''Individual differences in critical tendencies and noise annoyance,'' J. Sound Vib. **68**, 241–248.

Wolsink, M., Sprengers, M., Keuper, A., Pedersen, T. H., and Westra, C. A. (**1993**). ''Annoyance from wind turbine noise on sixteen sites in three countries,'' European Community Wind Energy Conference, 8–12 March, Lübeck, Travemünde, Germany, pp. 273–276.

The role of noise sensitivity in the noise–response relation: A comparison of three international airport studies

Irene van Kamp[a)]
National Institute of Public Health and the Environment, P.O. Box 1, 3720 BA, Bilthoven, The Netherlands

R. F. Soames Job and Julie Hatfield
Sydney University, School of Psychology, Australia

Mary Haines
Centre for Psychiatry, Institute of Community Health Sciences, Barts and the London, Queen Mary's School of Medicine and Dentistry, Queen Mary, University of London, UK
and Health Risk Management Practice, PricewaterhouseCoopers, Sydney

Rebecca K. Stellato
National Institute of Public Health and the Environment, The Netherlands

Stephen A. Stansfeld
Centre for Psychiatry, Institute of Community Health Sciences, Barts and the London, Queen Mary's School of Medicine and Dentistry, Queen Mary, University of London, United Kingdom

(Received 13 December 2003; revised 3 September 2004; accepted 7 September 2004)

In order to examine the role of noise sensitivity in response to environmental noise, this paper presents detailed comparisons of socio-acoustic studies conducted around international airports in Amsterdam, Sydney, and London. Earlier findings that noise sensitivity moderates the effect of noise on annoyance were examined to see if they could be replicated in each of the datasets, independent of the technique of measuring noise sensitivity. The relation between exposure to aircraft noise and noise annoyance was studied separately for groups of individuals with low, medium, and high noise sensitivity, with statistical adjustment for relevant confounders. Results support the previous findings that noise sensitivity is an independent predictor of annoyance and adds to the prediction of noise annoyance afforded by noise exposure level by up to 26% of explained variance. There is no evidence of a moderating effect, whereby the covariation between noise exposure level and annoyance is weak for people who score at the extreme high or low end of the sensitivity scale, and strong for people who score in the middle of the sensitivity scale. Generally, noise sensitivity appears to increase annoyance independently of the level of noise exposure after adjustment for relevant confounders. These findings were consistent across the three datasets. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1810291]

PACS numbers: 43.50.Qp, 43.66.Lj [DKW] Pages: 3471–3479

I. INTRODUCTION

Noise sensitivity is generally accepted as an important nonacoustical modifier of the noise–annoyance relation (reviews: Refs. 1–6). Very few studies have focused on the precise role of noise sensitivity.[7,8] Extensive reviews of Fields,[9] Job,[1] and Guski[3] found firm evidence that reaction to noise (annoyance, dissatisfaction) was associated with an expressed sensitivity to noise in general. Estimates of the amount of variation in annoyance and dissatisfaction with noise that is explained by noise sensitivity scores are reasonably consistent (10%–26% according to Job,[1] and 10%–16% according to Guski[3]) in comparison to 10% by noise level itself as primary stimulus. Noise sensitivity is not synonymous with annoyance. The association between noise sensitivity and noise exposure is consistently low, and the sensitivity–reaction relation is seen at all levels of noise exposure.

Noise sensitivity is also associated with other health variables such as sleep disturbance[10–12] and psychological distress and psychiatric disorders.[8,13–18] In the Caerphilly study[6] noise sensitivity predicted psychiatric disorder at follow-up, but after adjusting for trait anxiety the effect of noise sensitivity was no longer significant. This suggests that the influence of noise sensitivity on psychiatric disorder is confounded by anxiety: people who are anxious are more aware of threatening aspects of the environment (including noise) and more prone to psychiatric disorders.

Laboratory studies produced contradictory results in task performance situations under noisy conditions.[19] Self-reported noise sensitivity has been validated to some extent against physiological responses to noise in laboratory studies.[8] People with higher levels of noise sensitivity display higher tonic skin conductance, slower habituation, and greater changes in heart rate in response to noise.[20] Noise-sensitive people also show higher baseline arousal levels,[21] higher baseline blood pressure, and higher baseline psychophysiological arousal, which leads to more pronounced reactions to all stressors, including noise,[22] independent of residential noise exposure levels. Several studies of associations

a)Electronic mail: irene.van.kamp@rivm.nl

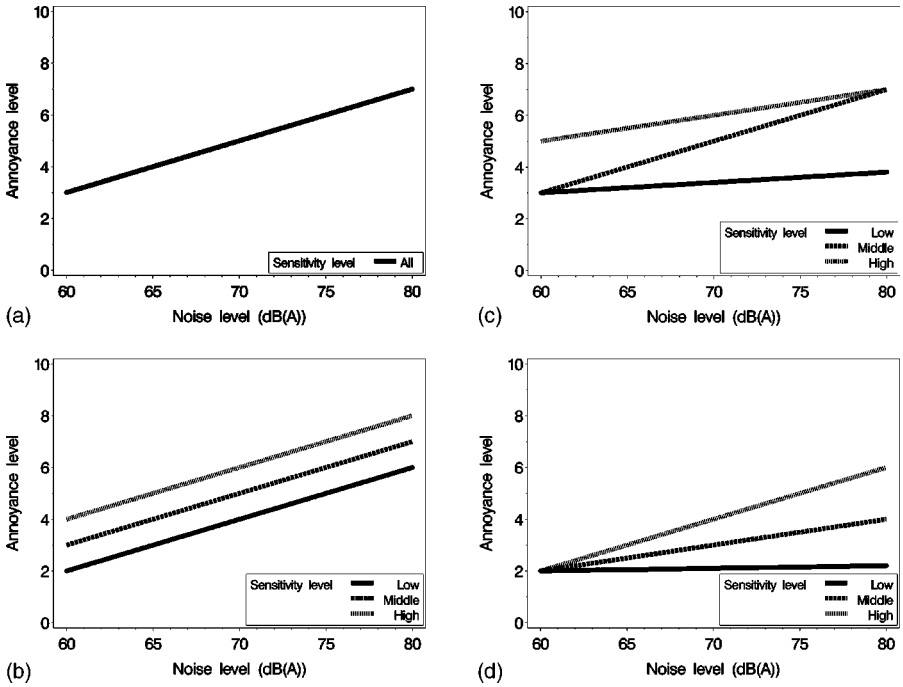


FIG. 1. (a)–(d). Four possible models of the relation noise, noise sensitivity, and reaction.

of noise sensitivity with anxiety, neuroticism, and extraversion suggest an association between noise sensitivity and personality trait measures,[23,24] but with the exception of the association with neuroticism, findings are inconsistent.[13,14,19,22,25–28] Although Field's synthesis[9] suggests no strong relation between sensitivity and demographic characteristics, noise sensitivity has been shown in other studies to increase with age and length of residency in noisy areas,[8,11,29–32] although again these results are not always consistent.[30] Age and length of residency are also related to extent of reaction to noise,[5,33,34] complicating analysis of the sensitivity–reaction relation.

Fields[9] found no evidence to support a selection effect of noise-sensitive people out of highly exposed areas, although this selection effect may well occur under certain circumstances. It is also conceivable that people who consider themselves as sensitive to noise do not move into noisy areas. The lack of an association between noise sensitivity and residential exposure is unsurprising given that noise sensitivity may be only one reason among many for moving from or staying in a particular area.

The role of noise sensitivity remains insufficiently understood partly due perhaps to differences in definition and measurement, partly to variations in research techniques,[3] and partly to the lack of detailed analysis of the interrelations among the relevant variables.

This paper describes an analysis of the relation between noise exposure, noise sensitivity, and noise reaction in three separate studies. Three alternative models of the influence of noise sensitivity on the noise–annoyance relation are presented and tested on three data sets, making use of a uniform analytical approach.

II. DEFINITIONS

The concept of noise sensitivity has been defined in several different ways. Definitions in the literature include: (1) a stable trait covering attitudes to a whole range of environmental sounds;[25] (2) internal states (be they physiological, psychological, attitudinal, or related to lifestyle or activities conducted) of an individual that increases his degree of reactivity to noise in general;[2] (3) an attitude towards noise in general;[8,35,36] and (4) susceptibility to noise or sound and susceptibility to annoyance, not necessarily by noise.[31,32] Definitions are often operational or *post hoc* by outcome (for a review, see Job[2]). Thus, sensitivity has been viewed as a trait, resulting from personal factors which increase reactivity to noise (e.g., attitudes), as a state and as a resultant of both traits and states. We focus here on a susceptibility to noise annoyance rather than a susceptibility to annoyance in general (general sensitivity). We distinguish between noise sensitivity and attitude towards the noise source.[2] In addition, we distinguish self-reported noise sensitivity from susceptibility to physiological and health effects of noise.

III. MODELS AND HYPOTHESES

Early surveys[7,37] indicated that noise sensitivity was not related to noise level[38] but rather acted as an effect modifier[39] in the noise-annoyance relation. An interaction effect was observed in which the extremes of the scale resulted in a low regression coefficient visualized in a flat curve, whereas the curve is steeper in the middle group, due to a larger regression coefficient [see Fig. 1(c)]. Recently, Miedema and Vos[40] concluded on the basis of a meta-analysis on 28 combined datasets that noise sensitivity modifies the relation between noise and annoyance. These studies all employed a single-item rating scale about the degree to which people consider themselves sensitive to noise, which is considered to be the least reliable measure of sensitivity.[25] Stansfeld[13,14] found that highly noise-sensitive subjects reported significantly higher noise annoyance in high-noise areas but not in low-noise areas, also indicative of an interaction effect. However,

TABLE I. Comparison of study methodology across the three centers.

	Sydney	Amsterdam	London
Data sets	Sydney Airport study 1994	Schiphol Study 1996	Children School Performance Study 1996
Sample size and selection	1005 Sampled on basis of exposure (dich.) and distance in a 2×2 design HH, LL, HL, LH In these analyses, only the HL group ($N=253$) was used	11601 Sampled on basis of exposure (continuous) and distance	193 Sampled on basis of childrens' school exposure levels. Four high [16-h outdoor L_{eq} 66 dB(A)] and four low schools [16-h outdoor $L_{eq}<57$ dB (A)] selected
Survey procedure	interview (every 7th resident)>18 yrs	Postal survey>18 yrs	Children took questionnaire home for mothers
Response rate	86%	39%	84%
Noise exposure measures	ANEI	KE	L_{eq} 16
Noise annoyance measures	Combination of three 0–10 numeric scales with verbal labels ranging from ''not at all'' to ''very much''	0–10 numeric scale with verbal labels ranging from ''not at all'' to ''very much''	1–4 point scale with verbal labels ranging from ''not at all'' to ''very much''
Noise sensitivity measures	6 noise situations *re*: loud noise 0–10 numeric scale and verbal labels ranging from ''not at all'' to ''very much''	Single item, 0–10 numeric scale	WNS, 10-item version
Socio-demographic measures	Age, gender, (non) native, education level, employment status, length of time living at the current address, and home ownership.	Age, gender, country of birth, education level, employment status, length of time living at the current address, and home ownership.	Age, gender, ethnicity, education level, employment status, household deprivation, length of time living at the current address, and home ownership.

the Sydney Airport Health Study results[23,24] did not support this finding: the effect of noise sensitivity was not higher in the areas with high noise exposure. While it is possible that these results genuinely reflect differences between cultures, the inconsistencies may be due to differences in measurement or analyses.

Job[2] describes three possible causal connections between reaction and noise sensitivity: (1) noise sensitivity influences reaction; (2) reaction influences noise sensitivity; and (3) a third variable directly influences reaction and noise sensitivity independently. To the models described by Job we could add an intervening variation in which noise influences annoyance indirectly via noise sensitivity (sensitization), or noise influences sensitivity indirectly via annoyance. This can be described as a transactional process where an increase in sensitivity leads again to an increase in annoyance. Within the framework of this paper the aspect of a third, underlying factor such as anxiety or general vulnerability will not be addressed. For considerations of this issue see Stansfeld.[6]

Figure 1(a)–(d) is a graphic representation of the three different models of the noise–annoyance relation that were tested in the current study. In Fig. 1(a), the model assumes no effect of noise sensitivity on reaction. In Fig. 1(b), noise sensitivity has a direct effect on reaction independent of noise, and in Figs. 1(c) and (d), sensitivity has a moderating effect on the relation between noise and reaction.

A. Aim

The major aim of the current analysis is to investigate the role of noise sensitivity in the noise–response relation, via detailed analysis of socio-acoustic data collected from three international airports: Schiphol Amsterdam Airport and London Heathrow as major airports in Europe and Sydney Airport as one of the major airports of the Asian and Pacific region. The primary question is whether noise sensitivity has a direct or a moderating effect on levels of annoyance. Within the framework of this paper the following issues were examined: first, what demographic variables are related to noise sensitivity? Do noise-sensitive and -insensitive people differ in terms of age, gender, length of residency, education, ethnicity, or home ownership? Second, does noise sensitivity influence reactions independent of noise exposure levels [as shown in Fig. 1(b)] or does it interact with noise exposure levels in the sense that noise sensitivity leads to different levels of reaction at differing levels of exposure [Figs. 1(c) and (d)]?

IV. METHOD

A comparison of the methodology of the three studies shows major differences in approach. Although the concepts used are highly comparable, the measurement techniques used varied somewhat (see Table I).

A. Datasets

For this paper three datasets were used: (1) the Sydney Airport (Kingsford Smith) study from 1994;[41] (2) the Amsterdam Airport (Schiphol) survey from 1996;[42] and (3) the Schools Environment and Health Study around London Airport (Heathrow) from 1996.[43] The three locations differ primarily in climate and possibly in culture (e.g., in terms of acceptance of noise). Although there are important differ-

ences in measurement, procedure, and other methodological aspects, the core aspects examined in the three datasets, such as noise exposure, annoyance, and noise sensitivity, were comparable (see Table I). The large sample sizes allow detailed analysis of the complicated relations between noise exposure, annoyance, and noise sensitivity.

B. Samples

In the Sydney study, 523 female and 482 male residents 18 years and older were selected on the basis of noise exposure and location relative to Sydney (Kingsford Smith) Airport to produce a 2×2 design; current noise exposure was ''high'' or ''low'' and noise exposure was projected to either decrease, increase, or remain unchanged due to flight-path changes. Due to repeated calls back by a highly motivated field team the refusal rate was less than 14%. For full details of method and design, see Hatfield *et al.*[41] Due to the manner in which noise was measured, there was a gap in the distribution of noise level (no levels were measured between 0 and 15 ANEI) for the two low areas and a very small range of noise levels in the ''high remains high'' area. Because we wanted to use noise as a continuous variable in all analyses, only the ''high decreases to low'' exposure group ($N=253$; 114 men and 139 women) was used in Sydney. This group had the largest range of noise exposure, and noise level was approximately normally distributed. In the Amsterdam study, 5524 female and 6077 male residents 18 years and older and living within a radius of 25 km around Schiphol Airport participated in the survey. The sample was randomly selected from 30 216 addresses, which were stratified by aircraft noise levels and distance to the airport. The response rate was 39%. For full details of methods and design, see Franssen *et al.*[42]

In the London study 193 mothers of children participating in the Schools Environment and Health study were included in the analyses. Children were selected on the basis of aircraft noise exposure at school, and all parents of participating children were given a self-administered questionnaire to complete. Parents were classified into home noise exposure by aircraft noise contours. The response rate was 84% of parents of children who took part in the study. For full details of methods and design, see Haines *et al.*[43]

C. Procedure

The data for the Sydney study were obtained from face-to-face interviews that were conducted among people aged 18 years and older before the changes to the configuration of Sydney Airport. Houses were selected at random within each census district, and within a selected house, one randomly selected adult was asked to participate. Full details of the selection procedure are described by Hatfield *et al.*[39]

The data for the Amsterdam study were collected in a cross-sectional survey making use of a postal questionnaire among people aged 18 years and older, sampled from 30 216 addresses in the area in a radius of 25 kilometers around Schiphol Airport. After a few weeks a reminder was sent to those who had not yet responded. To examine if the results could have been biased by selective nonresponse in the Schiphol sample, the postal questionnaire was supplemented with a short telephone survey among a small portion of the nonresponders. Full details are given by Franssen *et al.*[42]

The data for the London study were collected in a cross-sectional study among children, their parents, and their teachers. Children taking part in the study received a questionnaire for their mothers to complete at home and return to the school. The children (and parents) in the study were selected on the basis of the child's aircraft noise exposure at school high aircraft noise areas [high= 16-h outdoor equal or greater than L_{eq} 66 dB(A); low 16-h outdoor $L_{eq}<57$ dB(A)], and were matched on the following criteria: (a) age of the children; (b) sound level at the school from nonaircraft sources; (c) noise protection (double glazing) in the schools; (d) socio-economic group distribution and unemployment rate, and (e) proportion of ethnic groups in the area.[43]

D. Materials

In the Sydney study, a structured interview, based on previous socio-acoustic survey questionnaires[42] and revised on the basis of the results of a pilot study, assessed psychological reactions to noise, sensitivity to noise, and demographic variables, as well as aspects of physical and mental health, attitudes to the noise source, noise-induced activity disturbance, and personality.

In the Amsterdam study, a self-administered questionnaire based on validated questions and questionnaires from Dutch noise research and national health surveys was employed. The questionnaire contains questions about the residential environment, noise annoyance from several sources including aircraft noise, perceived environmental safety, Schiphol Airport, socio-demographics, living situation, general health, lifestyle, and quality of sleep and sleep disturbance.

In London, a self-administered questionnaire for the parents measured a range of outcomes including parental health, environmental perception, and socio-economic factors.

E. Noise exposure measures

Unfortunately, it was not possible to convert the exposure measures used in the three studies into one standard measure such as L_{den}.

In Sydney, data on aircraft noise exposure levels at each participating residential address were derived from the Integrated Noise Model (INM) program, at the time of the study verified by numerous measures at different locations around the airport. The INM was then employed to produce aircraft noise exposure data for the sample areas, and sample periods parallels the Australian Noise Exposure Index (NEF) with a modified evening penalty (based on Australian reaction data[44]) of 6dB between 7 p.m. and 7 a.m. and is a measure of what has occurred rather than being a forecast. These noise data were geo-coded to each participating residential address using GEOGRAPHIC INFORMATION SYSTEM (GIS) software.

In Amsterdam, aircraft noise exposure level was calculated by the National Aerospace Laboratory (NLR), using the Dutch standard calculation model for determining the annual exposure to aircraft noise around Schiphol. Aircraft noise

levels for 1996 were calculated for each x, y coordinate of the geometric center of the area covered by the 6-digit postal code for the respondent's residential address. The levels are expressed in Kosten units (Ke), a yearly average defined by the maximum noise levels during flights, the total number of flights, and the time at which these flights take place (night flights are weighted more heavily) as well as in La_{eq} measures. The model only includes that part of each aircraft movement during which the calculated noise level at ground level is higher than 65 dB(A). Of the different noise exposure measures available in the Amsterdam sample, KE was chosen in view of its comparability to the ANEI exposure measure used in Sydney.

In London, aircraft noise exposure levels at each participating parent's home were taken from the published Civil Aviation Authority dB(A) L_{eq}, contour maps indicating the average continuous equivalent sound level of aircraft noise within a particular area for 16-h daily periods during 15 June to 15 September (92 days) in 1991. These are the UK standard aircraft noise contours.

F. Noise sensitivity measures

At the three study locations, noise sensitivity was measured in three different ways. While these methods vary, there is strong evidence in the literature to indicate that they measure the same underlying construct. For a review of the value of these different instruments in terms of reliability, validity, and interrelations, see Zimmer and Ellermeier.[25]

In the Sydney study, noise sensitivity is derived from reactions (annoyance and dissatisfaction) to various noise situations other than aircraft noise were measured, in line with previous Australian studies.[45,46] When noise sensitivity is measured in this way, factor analysis indicates that noise sensitivity is not a unitary concept. Rather, two distinct factors appear: one related to loud, generally outdoor noises (road traffic, lawn mower) and the other related to generally indoor, quieter but nonetheless distracting noise situations (rustling papers at movies, people talking while watching television). For a detailed description of the strong reliability and construct validity of the scales, see Job *et al.*[2] In the current analyses only the sensitivity to loud noise scale was used.

In the Amsterdam survey a single-item 11-point scale was used to measure noise sensitivity, asking people the extent to which they consider themselves sensitive to noise.[15,37,47] The reliability of this method is questionable, in the sense that it violates the assumption of independence of noise level.[25]

The Schools Environmental and Health study in London administered the Weinstein noise sensitivity scale [WNS-10 Refs. 31, 32]. The WNS is comprised of statements describing responses to noise related situations (e.g., ''I get annoyed when my neighbors are noisy,'' and ''Even music I normally like will bother me if I'm trying to concentrate''). Respondents rate their agreement with every statement on a scale ranging from strong disagreement to strong agreement. The 21-item scale has shown to be sufficiently reliable (with alpha around 0.85), and to have construct validity in different languages.[25,28,31] Stansfeld[15] reports a Cronbach's alpha reliability coefficient of 0.78 on the 10-item version.

G. Statistical analysis

Age, length of residency, employment status, and education were categorised in order to allow comparisons across the three studies. The annoyance and sensitivity scales were converted to 0–100 scales for ease of comparability. For the regression models, the sensitivity scales in all three studies were recoded into tertiles, chosen to represent low, medium, and high levels of sensitivity to noise. Sensitivity was categorized in order to test the noise by sensitivity interaction. Three categories were chosen in order to test the hypotheses, represented by Figs. 1(a) to (d), of no effect, an additive effect, or a moderating effect (especially, but not limited to, the ''Z-form'' theory proposed by McKennell[7]). Tertiles were chosen as a manner of standardizing the different scales used in the different studies: an interaction term between two continuous variables is not only difficult to interpret, but also would make testing the Z-form theory impossible.

In preparation for regression analyses, the mean noise sensitivity levels and the associations of noise exposure with noise sensitivity were examined in the three samples. Moreover, the distribution of demographic factors in the different sensitivity groups was examined, to determine inclusion as confounder in the analysis.

In order to investigate whether the influence of noise sensitivity on annoyance depends on noise level, and whether in an additive and/or a multiplicative fashion, linear regressions were run on the three datasets separately, adjusting for potential confounding variables such as age, gender, length of residency, educational level, ethnicity, SES, and home ownership. Making use of a uniform methodological approach to the different data sets enables us to show whether different measurements of noise sensitivity and/or different cultural settings lead to different conclusions. On each of the three datasets, four models were fitted. In the first step (model 1) the simple exposure–response relations of noise and annoyance were studied. In the next step (model 1a) relevant confounders were introduced into the model, in order to adjust for differences in populations among the three studies. To the regression model were then added the three noise sensitivity categories (model 2). In the last step (model 3) the interaction terms of noise exposure level multiplied by noise sensitivity categories was added to the model. The best-fitting model was chosen based on the increase in R^2 and by examining residual diagnostics. Because the three sites studied very different populations, there was concern that the results may have been biased by the difference in background characteristics. While the Sydney and Amsterdam studies were population-based, the London study was based primarily on mothers of school-aged children, thus women between the ages of 18 and 49. To determine whether this unusual distribution of potentially confounding variables affected the results, the regression models described above were repeated for the three separate datasets, limiting the datasets to only women in the age range 18 to 49.

TABLE II. Characteristics of the three study populations.

	Sydney ($N=233$)	Amsterdam ($N=9767$)	London ($N=193$)
Noise level, mean ANEI (SD)	26.5 (3.1)	N/a	N/a
Noise level, mean KE (SD)	N/a	28.5 (10.5)	N/a
Noise level, mean DBA (SD)	N/a	N/a	60.4 (5.6)
Annoyance level, mean (SD)	66.5 (28.4)	63.3 (34.3)	63.5 (34.3)
Sensitivity level, mean (SD)	48.5 (22.6)	51.1 (24.1)	60.7 (17.4)
Sensitivity level, %[a]			
low	33.5	30.4	35.2
middle	33.9	34.7	32.6
high	32.6	34.9	32.1
Gender male, %	46.4	54.8	15.0
Age, %			
17–29	30.9	14.5	2.6
30–39	25.3	22.1	59.1
40–49	17.2	21.8	35.2
50–59	9.9	18.1	3.1
60 and over	16.7	23.5	0.0
Ethnic minority[b]	50.6	5.9	30.2
Employment status, %			
employed	70.8	60.8	73.6
retired, student, homemaker	26.2	33.5	20.7
unemployed	3.0	5.7	5.7
Education, %			
none/low	6.0	17.6	49.7
middle	52.4	32.6	30.1
high	41.6	49.8	20.2
Length of residency, %			
0–5 years	49.4	27.0	38.3
5–10 years	18.9	18.0	33.2
10 or more years	31.7	55.0	28.5
Rented house, %	34.8	40.0	18.7

[a]Categories defined for each site separately, dividing the sensitivity scale into three roughly equal groups.
[b]In Amsterdam and Sydney, members of an ethnic minority are those respondents not born in the Netherlands or Australia, respectively. In London, this was measured as self-reported ethnicity which is not dependent upon country of birth.

V. RESULTS

A. Demographic structure and noise sensitivity groups in each sample

Table II presents the demographic structure and noise sensitivity groups in each sample. Mean exposure levels are not fully comparable, since different noise metrics are used. The mean levels of annoyance and noise sensitivity are very similar in the three samples. The high mean sensitivity score in London is possibly due to the large percentage of women (85%) in the sample; some previous studies have shown that women generally score higher on noise sensitivity, although this finding is not consistent.[25] There are differences in demographic structure in terms of age, ethnicity, education, employment, and length of residency.

Analysis of the characteristics of subsamples (data not shown) over the three noise sensitivity groups indicates that a significant difference in noise exposure level is found between the three sensitivity groups only in the Schiphol sample. Noise annoyance increases significantly across noise sensitivity groups across all three studies. The sensitivity groups differ significantly with regard to gender and age in the Sydney and Amsterdam sample; the high noise sensitivity group contains significantly more women and people between 39 and 59 years. In all three samples the percentage of non-native people is significantly higher in the high noise sensitivity group. In the Amsterdam sample noise sensitivity is associated with high levels of education, whereas in Sydney and London the middle education group has higher sensitivity. Length of residency also shows an inconsistent pattern. Both in the Sydney and London sample, people who had lived 10 years or longer in the area are under-represented in the high NS group, although the difference is not significant. In the Amsterdam sample, the pattern of association is unclear. In the Amsterdam sample, the percentage of unemployed/retired people is significantly higher in the high noise sensitivity groups, while in the other samples such an association is not found. Finally, home ownership is also differentially related to noise sensitivity. These results are strongly a result of the socio-demographic features of the areas around the three different airports and the samples selected.

B. Exposure–response relations

The distribution of mean annoyance level was examined by noise exposure classes for each location, and is presented in Figs. 2(a)–(c).

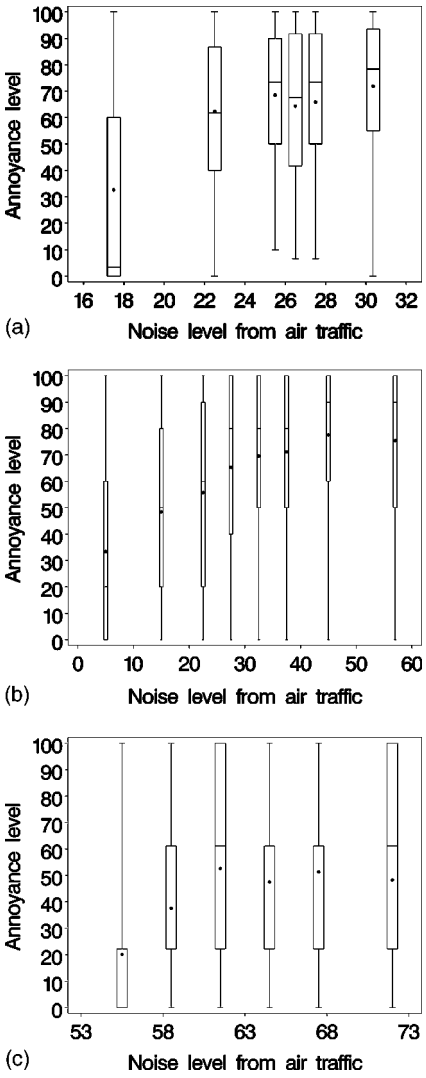


FIG. 2. (a)–(c). Distribution of annoyance level by air traffic noise level in Sydney (top) Amsterdam (middle), and London (lower). The upper, middle, and lower lines in each box represent the 75th, 50th, and 25th percentile, respectively, of annoyance level for each group; ● indicates the mean level; and the ends of the ''whiskers'' indicate the lowest and highest annoyance levels.

C. Noise sensitivity and noise annoyance: four models

Linear regression analysis was performed in the four stages described in Sec. IV G. Table III shows the results of this (stratified) linear regression analysis. Noise exposure explains 4%, 10%, and 12% of the variance in annoyance in Sydney, Amsterdam and London respectively, increasing to 18%, 12%, and 17% after inclusion of confounders (model 1a). The total amount of explained variance after adding noise sensitivity to the model (model 2) is 21%, 38%, and 23% in the three samples. A striking difference between the surveys is found in the extent to which variance is explained by noise sensitivity after noise level and confounders have been taken into account; the increase in R^2 from model 1a to model 2 is only 3% in Sydney and 6% in London, but is 26% in the Amsterdam sample.

Addition of the noise by sensitivity interaction (model 3) does not significantly increase the model fit (R^2), and examination of residual diagnostics (data not shown) confirms that the fit of model 3 is no better than that of model 2. The noise by sensitivity interaction thus adds nothing new to the model with only main effects.

Analyses performed on the datasets restricted to females between 18 and 49 displayed similar patterns of results (data not shown). Again, the model with only an additive effect of noise sensitivity (model 2) gave the best fit, and the noise by sensitivity interaction did not add significantly to the prediction afforded by the model.

VI. CONCLUSIONS AND DISCUSSION

This study examined the role of noise sensitivity in the relation between aircraft noise and annoyance. The results of analysis support the hypothesis that noise sensitivity increases annoyance independently from, and above, level of noise exposure, after adjustment for relevant confounders. It can be concluded that noise and sensitivity are important and independent predictors of annoyance in all three studies across a range of cultures and climates. No evidence was found for an interaction effect of noise sensitivity with noise. These findings were consistent across three studies, broadly suggesting that neither the cultural differences nor the different measures involved in these studies influenced the observed role of sensitivity. Of course, other cultures and measures may influence the observed role of sensitivity. This finding contrasts with the findings of Miedema and Vos,[40] who conclude, on the basis of a meta-analysis, that there is evidence for an interaction effect above an additive effect of noise sensitivity on annoyance. An explanation for divergent evidence has to be primarily sought in differences between statistical approach. The former study used noise sensitivity as a continuous variable, whereas in the current study noise sensitivity was categorized in order to test the noise–noise sensitivity interaction for low, medium, and high levels of noise sensitivity.

A. Strengths and limitations

The present study finds consistent patterns of results over three different studies in three countries. The differing characteristics of the samples and the varied measurements used made direct comparisons more difficult and may explain the inconsistencies in the associations between demographic characteristics and noise sensitivity. Nevertheless, the consistency of the association between noise, noise sensitivity, and annoyance despite the different samples suggests that these associations are robust. The weak association between noise level and noise sensitivity in Amsterdam may reflect the use of a single-item measurement of sensitivity. The noise source examined in these studies was limited to aircraft noise, but there is little reason to suppose the findings would be different with other sources of noise.

Noise sensitivity has previously been shown to increase with age and length of residency,[8,11,29,30,31,32] although not consistently.[25] These findings are only partly confirmed: a significant relation between length of residency and noise sensitivity was found only in the Amsterdam sample, while in the other samples people who lived less than 5 years in the area were over-represented in the high noise sensitivity group.

TABLE III. Results of linear regression models of annoyance in three cities.

	Sydney ($N=233$)		Amsterdam ($N=9767$)	London ($N=193$)		
Variable(s) in model	β (s.e.)	Model R^2	β (s.e.)	model R^2	β (s.e.)	model R^2
Model 1						
Noise level[a]	1.70 (0.58)	0.04	1.04 (0.03)	0.10	2.15 (0.42)	0.12
Model 1a						
Noise level[a]	1.60 (0.58)	0.18	1.03 (0.03)	0.12	2.31 (0.44)	0.17
+potential confounders						
Model 2						
Noise level[a]	1.48 (0.57)	0.21	0.90 (0.03)	0.38	2.09 (0.43)	0.23
Noise sensitivity						
low	(ref)		(ref)		(ref)	
medium	11.05 (4.27)		26.17 (0.68)		10.75 (5.66)	
high	12.79 (4.45)		44.31 (0.69)		21.27 (5.78)	
+potential confounders						
Model 3						
Noise level[a]	2.72 (0.93)	0.22	0.84 (0.05)	0.38	2.32 (0.81)	0.24
Noise sensitivity						
low	(ref)		(ref)		(ref)	
medium	10.63 (4.27)		26.29 (0.68)		10.36 (5.71)	
high	13.01 (4.46)		44.51 (0.69)		21.37 (5.86)	
Noise* sensitivity interaction						
low	(ref)		(ref)		(ref)	
medium	−1.89 (1.31)		0.22 (0.06)		0.12 (1.09)	
high	−2.21 (1.49)		−0.06 (0.06)		−0.64 (1.03)	
+potential confounders						

[a]Different definitions of noise level were used in the different cities: in Amsterdam, KE; in London, dBA; in Sydney, ANEI.
*Multiplied by

B. Implications

Noise sensitivity does not appear to increase annoyance levels only at high levels of noise exposure as a moderator of annoyance. Rather, it seems that noise sensitivity influences noise annoyance irrespective of noise level: at all levels of noise intensity, noise-sensitive people are likely to be more annoyed than the general population. Reducing only the highest levels will still mean that there are highly sensitive people highly annoyed at lower levels of noise.

C. Recommendations

Based on these outcomes, several recommendations can be put forward. To begin with, noise sensitivity is an important nonacoustical factor with a strong influence on annoyance, and therefore standardized measure of noise sensitivity should be included in studies of reaction to noise. At this time, no recommendations can be made for a standard noise sensitivity measure, because further research is needed to test the parametric properties of the different noise sensitivity scales in different noise situations. For this purpose the authors have designed a multicenter laboratory study to compare the four noise-sensitivity indices, in terms of their validity against physiological responses (heart rate and blood pressure), their test–retest reliability, and their internal consistency (where relevant) under two conditions: during task performance and during a control condition. The study will tentatively be conducted in London, Sydney, and Amsterdam. Much has already been achieved in terms of the international standardization of noise and reaction measures. A purpose-designed multicenter comparison employing standardized measures of noise, noise reaction, and noise sensitivity, would allow extension of the understanding of the influence of noise sensitivity on noise offered by the present meta-analysis. The running 5th Framework EU projects ''HEARTS'' and ''HYENA,'' as well as the Schiphol monitor, will allow for such comparisons.

ACKNOWLEDGMENTS

The Sydney Airport Health Study was funded by a grant from the Sydney Airports Authority. We acknowledge our collaborators on the study, Andrew J. Hede, Peter Peploe, Norman L. Carter, Richard Taylor, and Stephen Morrell. The Schiphol Airport study was funded by grants from the Dutch Ministry of Housing, Spatial Planning and the Environment (VROM), the Ministry of Transport, and the Ministry of Public Health. We acknowledge the research work of all people involved in designing, organizing and analyzing the survey around Schiphol Airport (1996), especially the contribution of Ellis Franssen and Ronald de Jong. The Schools Environment and Health study in London was funded by grants from a consortium of local authorities and health agencies around Heathrow Airport and was submitted as a public report for the Terminal 5 Public Inquiry. All our coauthors on the study are acknowledged: Birgitta Berglund, Soames Job, and Jenny Head. We thank Val Beale, Jane Boyd, Robert Gibson, and Patrick Shortt for help in data collection and noise measurement; and Maria Luz Herrero for coding and Staffan Hygge, Colin Cobbing, and Terry Gould for their support. The work for this meta-analysis was funded by the Dutch Ministry VROM; their financial support is gratefully acknowledged.

[1] R. F. S. Job, ''Community response to noise: A review of factors influencing the relation between noise exposure and reaction,'' J. Acoust. Soc. Am. **83**, 991–1001 (1988).
[2] R. F. S. Job, ''Noise sensitivity as a factor influencing human reactions to noise,'' Noise and Health **3**, 57–68 (1999).
[3] R. Guski, ''Personal and social variables as co-determinants of noise-annoyance,'' Noise and Health **3**, 45–56 (1999).
[4] J. M. Fields, ''The effect of numbers of noise events on people's reactions to noise. An analysis of existing survey data,'' J. Acoust. Soc. Am. **75**, 447–467 (1994).
[5] J. M. Fields and J. G. Walker, ''Comparing the relation between noise level and annoyance in different surveys,'' J. Sound Vib. **81**, 51–80 (1982).
[6] S. A. Stansfeld, M. M. Haines, and B. Brown, ''Noise and health in the urban environment,'' Noise and Health **2**, 43–82 (2000).
[7] A. C. McKennell, ''Aircraft noise annoyance around London (Heathrow) Airport,'' Central Office of Information: London, S.S 337 (1963).
[8] S. A. Stansfeld, ''Noise, noise sensitivity and psychiatric disorder: Epidemiological and psychophysiological studies,'' Psychol. Med. Monogr. Suppl. **22**, 1–44 (1992).
[9] J. M. Fields, ''Effects of personal and situational variables on noise annoyance with special reference to implications for en Route Noise,'' Washington DC: Federal Aviation Administration and NASA Report No: FAA-AEE-92-03 (1992).
[10] E. Ohrstrom, ''Effects of low levels of road traffic noise during the night: A laboratory study on number of events, maximum noise levels, and noise sensitivity,'' J. Sound Vib. **179**, 603–615 (1995).
[11] M. E. Nivison and I. M. Endresen, ''An analysis of relations among environmentalnoise, annoyance and sensitivity to noise, and the consequences for health and sleep,'' J. Behav. Med. **16**, 257–276 (1993).
[12] J. di Nisi, A. Muzet, J. Ehrhart, and J. P. Libert, ''Comparison of cardiovascular responses to noise during waking and sleeping in humans,'' Sleep **13**, 108–120 (1990).
[13] S. A. Stansfeld, C. R. Clark, and L. M. Jenkins, ''Sensitivity to noise a community sample: I. Measurement of psychiatric disorder and personality,'' Psychol. Med. **15**, 243–254 (1985).
[14] S. A. Stansfeld, C. R. Clark, G. Turpin, L. M. Jenkins, and A. Tarnopolsky, ''Sensitivity to noise in a community sample. II. Measurement of psychophysiological indices,'' Psychol. Med. **15**, 255–263 (1985).
[15] S. A. Stansfeld, D. Sharp, J. Gallacher, and W. Babisch, ''Road traffic noise, noise sensitivity and psychological disorder,'' Psychol. Med. **23**, 977–985 (1993).
[16] S. A. Stansfeld, J. Gallacher, W. Babisch, and M. Shipley, ''Road traffic noise and psychiatric disorder: Prospective findings from the Caerphilly Study,'' Br. Med. J. **313**, 266–267 (1996).
[17] A. Smith and S. A. Stansfeld, ''Aircraft noise exposure, noise sensitivity and everyday errors,'' Environ. Behav. **18**, 214–226 (1986).
[18] J. Gallacher and W. Babisch, ''Road traffic noise, noise sensitivity, and psychological disorder,'' Psychol. Med. **23**, 977–985 (1993).
[19] P. Bathia, I. S. Muhar, K. K Mahajan, and K. Chawia, ''Noise sensitivity as related to anxiety, intelligence and introversion–extraversion,'' J. Pers. Clin. Stud. **12**, 55–60 (1996).
[20] E. Ohrstrom, M. Bjorkman, and R. Rylander, ''Effects of noise during sleep with reference to noise sensitivity and habituation,'' Environ. Int. **16**, 477–482 (1990).
[21] K. S. Ann, ''Effects of aircraft noise and noise sensitivity on the physical and mental well-being of older adults,'' Health Sci. Nursing. **47(3)**, 991-B (1986).
[22] E. P. Sarafino, *Health Psychology, Biopsychosocial Interactions*, 2nd ed. (Wiley, New York, 1994).
[23] R. F. S. Job, J. Hatfield, P. Peploe, N. L. Carter, R. Taylor, and S. Morrell, ''General scales of community reaction to noise (dissatisfaction and perceived affectedness) are more reliable than scales of annoyance,'' J. Acoust. Soc. Am. **110**, 939–946 (2001).
[24] J. Hatfield, R. F. S. Job, P. Peploe, N. L. Carter, R. Taylor, and S. Morrell, ''The influence of psychological factors on the physiological and health effects of noise,'' Noise and Health **10**, 1–14 (2001).
[25] K. Zimmer and W. Ellermeier, ''Psychometric properties of four measures of noise sensitivity: A comparison,'' J Environ. Psychol. **19**, 295 (1999).
[26] G. Belojevic, E. Öhrström, and R. Rylander, ''Effects of noise on mental performance with regard to subjective noise sensitivity,'' Int. Arch. Occup. Environ. Health **64**, 293–301 (1992).
[27] G. Belojevic, B. Jakovljevic, and O. Oleksic, ''Subjective reactions to traffic noise with regard to some personality traits,'' Environ. Int. **23**, 221–226 (1997).
[28] B. Ekehammar and S. Dornic, ''Weinstein's noise-sensitivity scale: Reliability and construct validity,'' Percept. Mot. Skills **70**, 129–130 (1990).
[29] N. Moreira and M. Bryan, ''Noise annoyance susceptibility,'' J. Sound Vib. **21**, 449–462 (1972).
[30] S. M. Taylor, ''A path model of aircraft noise annoyance,'' J. Sound Vib. **96**, 243–260 (1984).
[31] N. D. Weinstein, ''Individual differences in reactions to noise: A longitudinal study in a college dormitory,'' J. Appl. Psychol. **63**, 458–566 (1978).
[32] N. D. Weinstein, ''Individual differences in critical tendencies and noise annoyance,'' J. Sound Vib. **68**, 241–248 (1980).
[33] J. M. Fields, ''The effect of numbers of noise events on people's reactions to noise: An analysis of existing survey data,'' J. Acoust. Soc. Am. **75**, 447–467 (1984).
[34] R. B. Bullen, A. J. Hede, and R. F. S. Job, ''Community reaction to noise from an artillery range,'' Noise Control Eng. **37**, 115–128 (1991).
[35] C. Levy-Leboyer and G. Moser, ''Individual differences in noise annoyance: Four explanations,'' in *Environmental Annoyance Characterisation, Measurement and Control*, edited by H.S. Koelega (Elsevier, Woudschoten, 1989).
[36] C. M. B. Anderson, ''The measurement of attitude to noise and noises,'' Acoustic report Ac. 52, National Physical Laboratory, Teddington (1971).
[37] F. J. Langdon, ''Noise nuisance caused by road traffic in residential areas,'' J. Sound Vib. **111**, 243–263 (1976).
[38] An intervening variable can be described (see Ref. 3) as a feature/attribute of a person that depends on the stimulus and also influences the response to this stimulus. In such a model each step adds to the prediction.
[39] An effect modifier is a feature/attribute of a person that changes the effect of a stimulus (e.g., noise) on the response (e.g., annoyance). The relation between noise and reaction is only valid for some values of noise sensitivity, but not for others.
[40] H. M. E. Miedema and H. Vos, ''Noise sensitivity and reactions to noise and other environmental conditions,'' J. Acoust. Soc. Am. **113**, 1492–1504 (2003).
[41] J. Hatfield, R. F. S. Job, A. J. Hede, P. Peploe, N. L. Carter, R. Taylor, and S. Morrell, ''The role of learned helplessness in human response to noise,'' Int. J. Behav. Med. **9**, 341–359 (2002).
[42] E. A. M. Franssen, C. M. A. G. van Wiechen, N. J. D. Nagelkerke, and E. Lebret, ''Aircraft noise around a large international airport and its impact on general health and medication use,'' Occup. Environ. Med. **61**, 405–413 (2004).
[43] M. M. Haines, S. A. Stansfeld, R. F. S. Job, B. Berglund, and J. Head, ''Chronic aircraft noise exposure, stress responses, mental health, and cognitive performance in school children,'' Psychol. Med. **31**, 265–277 (2001).
[44] A. J. Hede and R. B. Bullen, ''Human perception and reaction to noise,'' Architect. Sci. Rev. **3**, 58–64 (1981).
[45] R. B. Bullen, A. J. Hede, and E. Kyriacos, ''Reaction to aircraft noise in residential areas around Australian airports,'' J. Sound Vib. **108**, 199–225 (1986).
[46] A. L. Brown, ''Traffic noise annoyance along urban roadways: report on a survey in Brisbane, Sydney and Melbourne,'' Australian Road Research, Internal Report AIR 206-06 1978.
[47] G. J. Raw and I. D. Griffith, ''Individual differences in response to road traffic noise,'' J. Sound Vib. **121**, 463–471 (1988).

Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms

Bradford N. Gover[a] and John S. Bradley[b]
Institute for Research in Construction, National Research Council, 1200 Montreal Road, Ottawa, Ontario K1A 0R6, Canada

(Received 6 May 2004; revised 7 September 2004; accepted 7 September 2004)

Objective measures were investigated as predictors of the speech security of closed offices and rooms. A new signal-to-noise type measure is shown to be a superior indicator for security than existing measures such as the Articulation Index, the Speech Intelligibility Index, the ratio of the loudness of speech to that of noise, and the A-weighted level difference of speech and noise. This new measure is a weighted sum of clipped one-third-octave-band signal-to-noise ratios; various weightings and clipping levels are explored. Listening tests had 19 subjects rate the audibility and intelligibility of 500 English sentences, filtered to simulate transmission through various wall constructions, and presented along with background noise. The results of the tests indicate that the new measure is highly correlated with sentence intelligibility scores and also with three security thresholds: the threshold of intelligibility (below which speech is unintelligible), the threshold of cadence (below which the cadence of speech is inaudible), and the threshold of audibility (below which speech is inaudible). The ratio of the loudness of speech to that of noise, and simple A-weighted level differences are both shown to be well correlated with these latter two thresholds (cadence and audibility), but not well correlated with intelligibility. [DOI: 10.1121/1.1810300]

PACS numbers: 43.55.Hy, 43.71.Gv [NX] Pages: 3480–3490

I. INTRODUCTION

Offices and meeting rooms are often intended for confidential discussions where eavesdroppers and others outside the room should not be able to listen in to these discussions. Speech originating inside such a room being difficult to hear or understand in the adjoining spaces implies that the room provides good speech privacy. In cases where the degree of privacy is sufficiently high, one can speak of architectural speech ''security.'' Improved security would be provided, for instance, by a room constructed with boundaries having higher sound transmission loss. To assess the degree of security, it is necessary to measure whether the ''transmitted'' speech is audible or intelligible at positions outside the room. This paper reports the results of new subjective studies to evaluate measures of architectural speech security.

Defining the problem of architectural speech security in terms of the fraction of speech that can be understood makes it one of speech intelligibility. Investigations into speech privacy for open-plan office situations have used this same approach.[1] The distinction made by security designers is that ''privacy'' is a less stringent description of sound isolation than security. A condition where overheard speech is audible but only slightly intelligible corresponds to excellent privacy, but could be described as imperfect security. A high degree of speech security implies not only very low (or zero) speech intelligibility, but also minimal audibility of the transmitted speech sounds.

Speech intelligibility, speech privacy, and speech security are related to the level of the speech signal relative to the level of the noise at the listener position. The Articulation Index (AI) (Ref. 2) and the Speech Intelligibility Index (SII) (Ref. 3) are both frequency-weighted signal-to-noise ratio type measures, and can be calculated from the speech and noise spectra at the position of a listener. AI and SII are good indicators of the degree to which the speech is intelligible, but they do not necessarily represent zero intelligibility at their minimum values of zero. Furthermore, for conditions of zero intelligibility, where the rhythm of speech or some speech sounds may be audible, AI and SII provide no information. They are therefore seemingly ill-suited to the problem of assessing architectural speech security. A useful measure of security will indicate the degree to which the transmitted speech is intelligible or, in cases of zero intelligibility, audible.

The most significant published study of architectural speech privacy in buildings was by Cavanaugh *et al.* in 1962.[4] They presented a report on occupants' impressions of privacy in buildings, and how they related to AI. The main focus of the work was on privacy (i.e., freedom from distraction); however, there was a subcomponent of the study assessing so-called ''confidential privacy,'' which subjects were instructed corresponded to a situation where there was an ''assurance of not being overheard.'' One of their main results states that, ''the most critical 10% of the subjects began to feel a lack of [confidential] privacy when the articulation index reached 0.05.'' An important point is that the assessments were based on how private the subjects *felt* a situation was. The actual fraction of the speech they *could* understand was not measured. Nevertheless, this paper lays

[a]Electronic mail: brad.gover@nrc-cnrc.gc.ca
[b]Electronic mail: john.bradley@nrc-cnrc.gc.ca

the basis for relating privacy in offices to an objective measure.

In 1965, Young[5] published a revised computational procedure based on the data in Cavanaugh *et al.* He proposed a measure derivable from A-weighted levels of speech and noise, and single number indicators of transmission loss. This is a practical approach, easy to calculate, but at best no more accurate than the original Cavanaugh *et al.* method. This simplified method was not supported with additional subjective tests. The so-called ''Speech-Privacy Calculation'' has become accepted practice,[6] but, being based on the Articulation Index, is likely not appropriate for rating speech security, where zero intelligibility is expected.

This paper presents the derivation of a suitable objective measure, providing information regarding the degree of security both above and below the threshold of intelligibility.

II. EXPERIMENTAL DESIGN

Subjective listening tests were conducted, presenting subjects with sound fields simulating listening outside of a room in which a person was speaking. The spectra of the speech and of the noise at the listening position were separately measured, and were used to compute various objective measures. Relations were sought between these measures and the subjects' responses.

A. Listening tests

All listening tests were conducted in a sound-isolated test room located at the National Research Council in Ottawa. Anechoically recorded speech was filtered to simulate transmission through a wall and, along with background noise, was played into the room. Subjects in the room heard these sounds and responded to a test operator seated outside the room, who scored their response. Details of the tests and of the subjects follow. All tests were approved by NRC's Ethics Review Board.

1. Facility and hardware

The test room measured 9.2 m long by 4.7 m wide by 3.6 m high, and was constructed from concrete. The room is not connected to the building, and is resting on springs for vibration isolation. Sounds existing outside the room, therefore, are largely isolated from penetrating within. For the present study, the interior walls of the room were lined with 10-cm-thick absorbing foam, covered by curtains. There was a conventional T-bar ceiling with 25-mm-thick glass-fiber ceiling tiles installed, and the floor was covered with carpet. This interior treatment yielded a quite ''dead'' space. The measured background noise level in the room was 13.7 dBA.

Test speech was played over loudspeakers positioned at the front of the room. The background noise was played over another set of loudspeakers positioned above the ceiling, directly above the subject. Figure 1 shows a diagram of the setup. Having the noise and speech originate from different spatial locations is important for a realistic test. Listeners are better able to recognize speech in noise when the speech and noise arrive from different directions.[7]

A block diagram of the electroacoustic system used to produce the test sounds is shown in Fig. 2. The two blocks

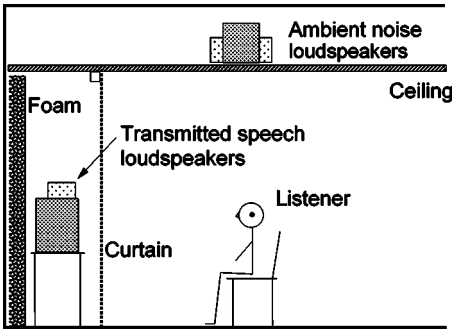


FIG. 1. Schematic of cross section through the test room, showing the location of the listener and the loudspeakers used to generate the test sound fields.

labeled ''DME32'' are Yamaha Digital Mixing Engines, which are highly flexible signal-processing boxes, able to perform the functions of many interconnected devices such as equalizers, filters, oscillators, etc. The outputs of the DME32s run through the power amplifiers into high-quality loudspeaker systems (Paradigm Compact Monitors, Paradigm PW subwoofers). One component in each DME32 was initially configured under computer control (via the RS232 interface) to equalize the playback path through the power amplifiers and loudspeakers to be flat at position of the listener's head (± 1 dB from 60 to 12 000 Hz).

The background noises for the test sound fields were generated by the internal noise generator of one of the DME32 units. This same unit shaped the spectrum and adjusted the level as desired, responding to control commands sent by the computer over the MIDI interface. One channel of the noise output was delayed by 300 ms relative to the other so as to avoid any unnatural perceptual effects caused by movement of the listener's head (as can be observed when listening to monaural material over a pair of loudspeakers).

The speech sounds were generated from playback of anechoically recorded source material stored on the computer in 16-bit, 44.1-kHz wave-file format. The output of the sound card ran into the second DME32, which performed the necessary equalization and level adjustment, again with com-

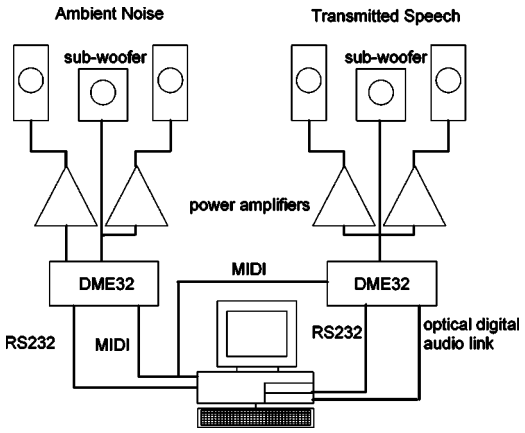


FIG. 2. Block diagram of the computer-controlled electroacoustic system used to create the test sounds. The speech, played from the computer, is processed by one of the Yamaha DME32 units to simulate transmission through a wall. The other DME32 unit generates spectrally shaped random background noise.

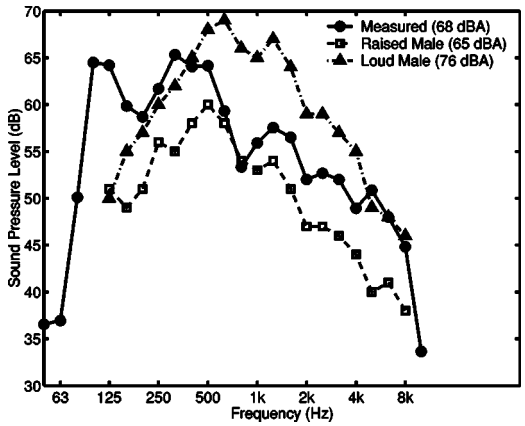


FIG. 3. Example measured spectrum of talker used in this work. Also shown are ''typical'' spectra for a male talker speaking with ''raised'' and ''loud'' effort as given in Ref. 11.

mands to switch settings received over the MIDI interface. Equalizer components in this DME32 simulated the various wall transmission loss curves.

2. Speech material

The test material used was phonetically balanced English sentences.[8] These sentences are of low predictability, and are more representative of conversation than isolated words. The response and scoring procedure had the subjects state out loud the words they thought they heard, and the test operator noted on a score sheet those that were correctly identified. In this way the fraction of words actually understood was scored. In intelligibility tests using rhyming words, the scoring can be done similarly, but subjects are able to guess correctly as often as 20% of the time,[9] which means they are particularly unsuitable in assessing security situations, where scores as low as 0% are expected. Furthermore, sentences can be more accurately scored than running speech, or so-called ''connected discourse,'' whereby the procedure usually involves allowing the subject to estimate the fraction of the words they are capable of understanding.[10] When the issue is security, rather than a sense of privacy, this distinction is important.

TABLE I. Wall descriptions and sound transmission class (STC) ratings for the walls simulated in the listening tests.

Wall Descriptor	Wall description	STC rating
Door	Solid core wood door, no seals.	20
Plenum	5/8 in. mineral fiber ceiling.	32
G13	89-mm wood stud wall with 13-mm gypsum board on both sides; cavity filled with glass-fiber batts.	34
G16	90-mm steel stud wall with 16-mm gypsum board on both sides; cavity filled with glass-fiber batts.	46

A total of nine recorded versions of the test sentences was acquired, spoken with different efforts by different talkers of both genders. The choice of recording used for the tests was the best quality recording available (16-bit, 44.1-kHz ''CD-quality'' digital), and was of a male talker speaking clearly. This speaking voice was most easily identified by subjects in pilot testing. That is, from a security point of view, this recording was the ''worst-case'' talker. The average of the spectral magnitude of four test sentences is shown in Fig. 3. The spectrum varies slightly from sentence to sentence; this curve is more representative of a ''long-term'' average. Also shown in Fig. 3 are ''typical'' speech spectra for a male talker speaking with two different speaking efforts, taken from Ref. 11. At this playback level (68 dBA), it can be seen that the speech corresponds to an effort somewhere between ''raised'' and ''loud.'' All speech spectral levels were measured over a 60-s period, looping the sentences continuously.

3. Walls

Walls to be simulated in the tests were selected from past measurements of actual wall samples in the wall testing facility at NRC. The transmission loss (TL) curves for the four walls used in the test are shown in Fig. 4; their descriptions are given in Table I. These four curves were selected since they are of different shapes, and representative of wall constructions typical of office environments.

The speech transmitted through any of these walls will be attenuated and spectrally distorted. Figure 5 shows typical

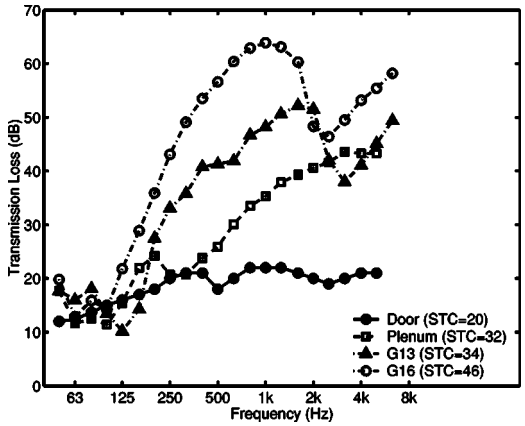


FIG. 4. Measured transmission loss curves for the walls selected to be simulated in the tests. The four wall descriptors are explained in Table I.

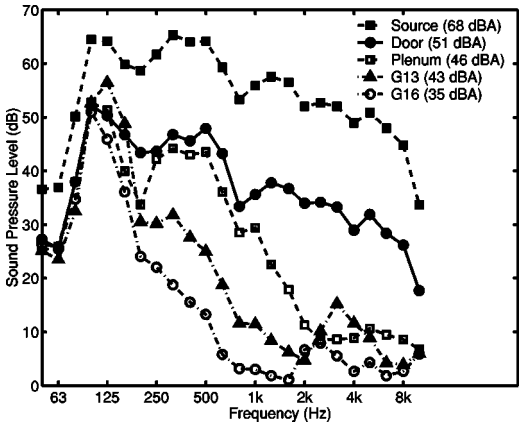


FIG. 5. ''Transmitted'' speech spectra, measured in the test room after filtering the source speech (top curve) through each simulated wall.

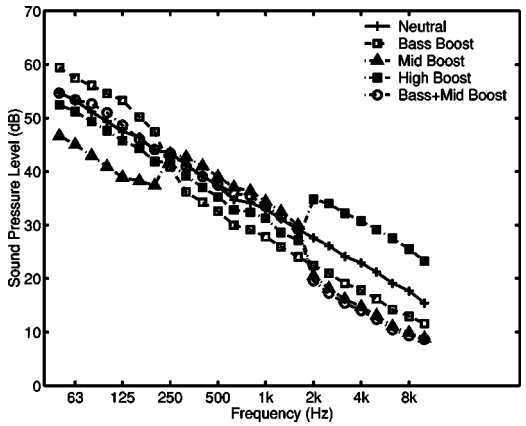


FIG. 6. Measured noise spectra, all corresponding to 45 dBA. The ''neutral'' spectrum has a −5-dB/octave roll-off. The other spectra were derived from this by boosting the bass frequencies (50–200 Hz), midfrequencies (250–1600 Hz), or high frequencies (2000–10 000 Hz) by 10 dB.

speech spectra measured in the test room after filtering by each TL curve.

4. Noise

All background noises used in the testing were spectrally shaped random noise. As discussed above, one of the Yamaha DME32 units was used to generate the noise internally, and equalizer components were used to shape the spectrum. Five different spectra were used, shown in Fig. 6.

The base case noise is the ''neutral'' spectrum, so-named under the RC naming convention.[12] It has a −5-dB/octave roll-off. The other spectra are derived from this by boosting ''low'' (50–200 Hz), ''mid'' (250–1600 Hz), or ''high'' (2000–10 000 Hz) frequency sections by 10 dB. These other noise types are used to systematically vary the spectral qualities of the interfering noise, representing cases that are more or less ''rumbly'' or ''hissy.''

B. Objective measures

From measurements of the one-third-octave spectral levels of the ''transmitted'' speech and noise, various numerical indicators can be calculated. The issue is to find a measure that is well-correlated with the subjective responses from the listening tests.

1. AI and SII

The Articulation Index (AI) (Ref. 2) and its more recent replacement, the Speech Intelligibility Index (SII) (Ref. 3), are measures calculated from the ratio of speech to noise in various frequency bands. The basic idea is that the contribution to the overall intelligibility from a particular frequency band is dependent on the ''effective'' ratio of signal to noise in that band, and on the importance of the band. The index itself is a weighted sum of the band contributions. Properties of the hearing system are built into the measure through the method in which the effective signal-to-noise ratio is determined from the signal and noise levels, and through the specification of the frequency band importance weights.

For the calculation of AI, the effective signal-to-noise level difference in a band is obtained by clipping the actual level difference, and by adjusting for the difference in level between the peak speech level and the rms speech level. The clipping is performed by setting all values below a specified minimum value $L=-12$ dB to be equal to -12 dB, and all values above a specified maximum value $U=+18$ dB equal to $+18$ dB. This clipped signal-to-noise is then shifted by adding 12 dB, to adjust for the difference between the peak and rms levels. The AI is given mathematically by

$$\mathrm{AI}=\frac{1}{30}\sum_b w_b\cdot[12+\min(\max(S_b-N_b,-12),18)], \tag{1}$$

where S_b is the speech level in decibels in frequency band b, N_b is the noise level in decibels in band b, and w_b are the frequency weightings. Here, $\max(x,L)$ is the larger of x or L; that is, L is the minimum clipped value of x. Values of x below L are set to L. Similarly, $\min(x,U)$ clips the upper value to U. The resulting value of AI is between 0 and 1. The calculation of SII is similar, but the determination of effective band signal-to-noise ratio is a little more sophisticated, taking into account masking. With regard to the practical minimum and maximum values for band signal-to-noise level differences, the SII calculation uses $L=-15$ dB, $U=+15$ dB.

2. Weighted signal-to-noise

A simplified approach that is similar to some of the steps involved in computing the AI or SII is to compute the actual (as opposed to effective) signal-to-noise level difference in each frequency band, and then simply perform a weighted sum across all bands. It does seem prudent, however, to specify a lower limit for the signal-to-noise level difference, clipping to this minimum value. In speech security situations, very high signal-to-noise level differences are unlikely to occur, so clipping will not be specified on the upper end. The expression for the resulting weighted signal-to-noise ratio $X_w(L)$ is

$$X_w(L)=\sum_b w_b\cdot\max(S_b-N_b,L), \tag{2}$$

where as above, S_b is the speech level in decibels in frequency band b, N_b is the noise level in decibels in band b, and w_b are the frequency weightings. Here, $\max(S_b-N_b,L)$ is the larger of L or the signal-to-noise level difference; that is, L is the minimum clipped value of the actual signal-to-noise level difference. Several weighting strategies for one-third-octave bands were considered; those resulting in the best-correlated indices are presented here: uniform weighting, AI-band importance weighting, and SII-band importance weighting. The weights, normalized so that they sum to unity, are given in Table II.

3. Loudness

Loudness (in sones) is a quantity that can be calculated from the spectrum of a signal, and has been shown to be related to the ''perceived loudness'' of sounds. The calcula-

TABLE II. 1/3-octave-band signal-to-noise weighting strategies considered in computing a weighted signal-to-noise ratio measure.

	Frequency (Hz)															
Weighting	160	200	250	315	400	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000
Uniform	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625	0.0625
AI	0.0000	0.0120	0.0300	0.0300	0.0420	0.0420	0.0601	0.0601	0.0721	0.0901	0.1111	0.1141	0.1021	0.1021	0.0721	0.0601
SII	0.0088	0.0101	0.0159	0.0306	0.0466	0.0612	0.0691	0.0752	0.0866	0.0893	0.0933	0.0950	0.0918	0.0893	0.0816	0.0558

tion of loudness is explained in Ref. 13. Measuring the speech loudness Λ_S and the noise loudness Λ_N separately, the loudness ratio r_Λ given by

$$r_\Lambda = \frac{\Lambda_S}{\Lambda_N}, \tag{3}$$

can be computed. This is a linear measure, so for instance if the speech is half as loud as the noise, the value of r_Λ will be 0.5.

4. A-weighted level difference

Weighted speech and noise levels can be computed from their spectra separately, and then the difference in these levels computed. The weighting scheme considered here is the familiar A-weighting, the result being the A-weighted level difference L_A, given by

$$L_A = 10\log\left(\sum_b w_b 10^{(S_b/10)}\right) - 10\log\left(\sum_b w_b 10^{(N_b/10)}\right), \tag{4}$$

where S_b, N_b, b, and w_b are as above. The terms on the right-hand side are the A-weighted signal and noise levels, respectively. The weights w_b are merely the nonlogarithmic A-weighting factors; $10\log(w_b)$ is the decibel correction factor for band b.

III. LISTENING TEST DESCRIPTIONS

A. Test 1: Intelligibility

The test conducted first consisted of a large number of sentences processed to simulate wide ranges of the conditions thought to be representative of actual offices and meeting rooms. Each of the four wall types and five background noise types was included. For each wall/noise combination, three signal-to-noise ratios were designed, corresponding to varying ''difficulty:'' easy, moderately difficult, and difficult. In the ''easy'' case it was judged that most listeners should be able to identify all words in a sentence. The ''difficult'' case was intended to be just above the threshold of intelligibility—some listeners could identify some (but not necessarily all) of the words. (It is not possible to design in advance the conditions that define the threshold we are seeking to find. Some of the difficult cases still had to correspond to nonzero intelligibility in order to find the threshold point.) In addition to these $4\times5\times3=60$ cases, 8 additional cases were constructed corresponding to 2 additional difficulties (''very difficult'' and ''very, very difficult'') for 1 noise type (neutral), for all 4 walls. These cases were constructed by reducing the speech level 3 and 9 dB below that used for the difficult case. In total, this yielded 68 physical conditions. Five different sentences were included for each condition, resulting in a test 340 sentences long. The range of levels for the ''source'' speech (notionally behind the simulated wall) resulted in measured ''transmitted'' speech levels of 28–49 dBA in the test room. The range of levels for the background noise was from 27–51 dBA.

The test was conducted in the following manner: the subject sat in the room with no test sounds playing. The noise was turned on, and a second or so later, the speech began. After the end of the sentence, the noise was switched off. At this point, the subject said out loud (into a talkback microphone) the words they thought they had heard. The subjects were encouraged to guess. The operator was outside the test room (monitoring the talkback microphone) scoring the responses, and would ask the subject to repeat an answer deemed ambiguous or incomprehensible. When the operator was satisfied, the computer was cued to play the next test sentence. The score for each sentence was computed as the percentage of words correctly identified; all words were counted, and no part scores were given.

Subjects completed the test over three different testing sessions, usually on different days. Each session consisted of two runs of about 57 sentences, separated by a brief break to avoid fatigue. The subjects listened to several practice sentences before each run. In total, one run of 57 sentences took about 20 min; the session for the day taking just under an hour, including breaks.

B. Test 2: Thresholds

A second test was conducted as a follow-up to the previous one, consisting of an additional 160 test sentences. These resulted from the 32 combinations of 2 wall types (G13 and G16)$\times$2 noise types (neutral and bass boost)$\times$8 signal-to-noise ratios, with 5 different sentences for each combination. These cases were distributed over the range of difficulty from all subjects able to understand at least one word, through to all subjects unable to detect the presence of speech at all. The range of notional source speech levels resulted in measured levels ranging from 20–45 dBA. The background noise levels ranged from 43–46 dBA. The best listeners from the previous test were used as subjects. (Here, ''best'' means these subjects correctly identified the most words, that is, they were the worst-case listeners from a security point of view.)

The test was conducted in a manner exactly like the previous one, except in how the subjects responded. They still spoke aloud, but were asked to respond to the following questions, written on an instruction sheet: *(1) Did you hear any speech sounds? (2) If yes, did you hear the rhythm or*

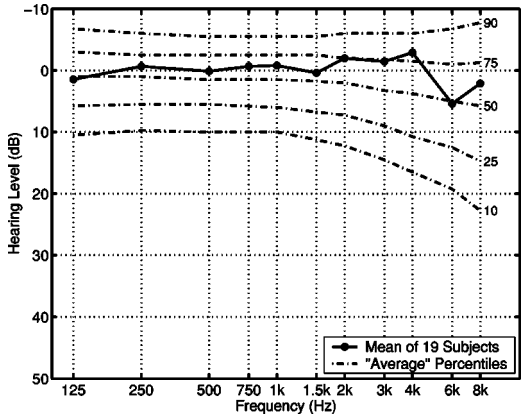


FIG. 7. Average HL curve for the 19 subjects shown with 10th, 25th, 50th, 75th, and 90th percentile HL curves for an ''average'' otologically normal listener. The percentiles give the percentage of the population with hearing loss worse than the curve. The ''average'' listener response is the mean of the 30-year-old and 40-year-old male and female responses given in Ref. 14.

cadence of the speech? (3) If yes, did you understand any of the words? (Tell the experimenter the words you were able to understand.)

C. Subjects

Subjects participating in the tests were volunteers; all were fluent English speakers, and none was compensated for participating. All respondents (54 in total) were given a standard audiometric hearing threshold test and a short trial of the intelligibility test, spanning all difficulty levels. The 36 subjects correctly identifying 62% or more of the words in the trial participated in the main test (test 1). The mean intelligibility score (across all 340 sentences) for each of these 36 subjects was computed and was used to split the group into the ''better'' subjects (the 19 scoring greater than 65%) and the ''worse'' subjects (the 17 scoring less than 65%). These better 19 subjects additionally participated in test 2. The analyses in this paper use only the scores for these 19 subjects, for both tests 1 and 2.

From the data provided in Ref. 14, an ''average'' otologically normal listener was constructed by averaging the hearing level (HL) data for 30- and 40-year-old males and females. The 10th, 25th, 50th, 75th, and 90th percentiles of these HL curves are shown in Fig. 7. Also shown in the figure is the average HL curve for the 19 better subjects from the tests. Twenty-five percent of the population of average 30/40-year-old male/female listeners can be expected to have hearing as good or better than this. We see therefore that our subjects had excellent hearing, but not ''unreasonably'' so.

Knowing what portion of the population was used as test subjects is important, but does not necessarily enable ''extrapolation'' of the test results to groups with less sensitive hearing. For certain, people with worse hearing will do more poorly on the listening tests (and therefore be less of a risk, from a security point of view). However, there are factors other than hearing loss that can cause reduced scores (for example, attention span, ability and desire to concentrate, native language).

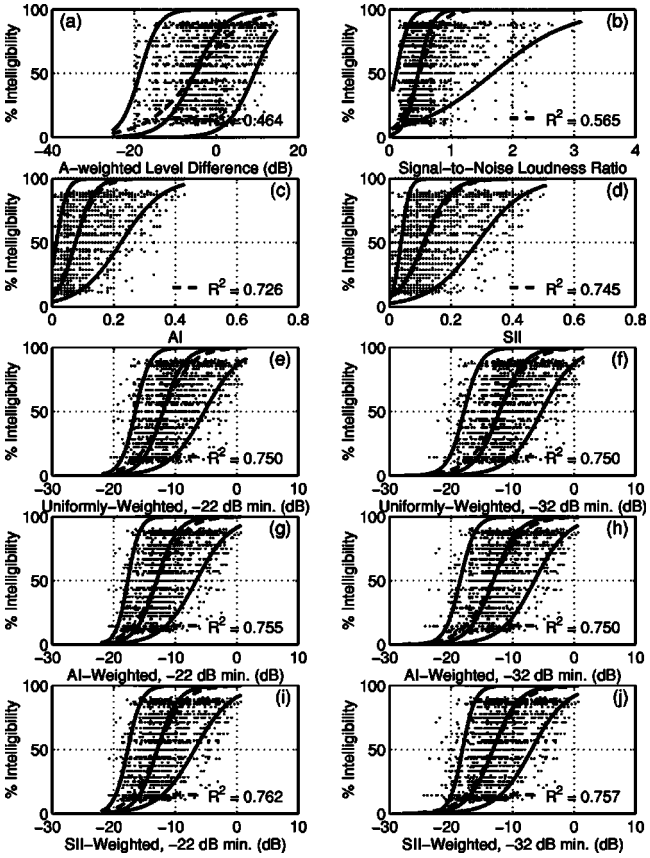


FIG. 8. Individual intelligibility scores for the 19 subjects from 500 sentences each (9500 points total), plotted versus the various indices. The dashed line is the least-squares Boltzmann function fit to the mean; the squared correlation coefficient (R^2) for this fit is shown. The three solid lines are the least-squares Boltzmann function fits to the 5th, 50th, and 95th percentiles (from bottom to top, respectively).

IV. RESULTS AND ANALYSIS

A. Intelligibility score

Figure 8 shows the intelligibility score results for all the sentences in both tests combined. Each panel of the figure shows the $19\times(340+160)=9500$ individual intelligibility scores for the 19 subjects for each sentence, as a percentage of the words correctly identified, plotted versus the various objective indicators computed from the speech and noise spectra (the dots). Overlaid on the plots are the least-squares best-fit Boltzmann curve to the mean of the data (the dashed curve), and fits to the 5th, 50th, and 95th percentile curves (solid curves, from bottom to top, respectively). The Nth percentile curve is defined so that $(100\text{-}N)\%$ of the recorded intelligibility scores exceeds the values on the line; that is, $(100\text{-}N)\%$ of the data points lies above the line, $N\%$ lies below. The square of the correlation coefficient R^2 (the ''coefficient of determination'') of the relationship between the scores and the best-fit curve is given. See Appendix A for the equation of the Boltzmann curve, and the fit parameters for each (Table III).

The squared correlation coefficients can be used to evaluate the goodness of the indicators as predictors of intelligibility. Values closer to unity indicate better relationships, closer to zero indicate worse ones. The R^2 for the

TABLE III. Values of the Boltzmann function parameters A and B from Eq. (A1) for the intelligibility score curves plotted in Fig. 8.

	Mean		5th percentile		50th percentile		95th percentile	
	A	B	A	B	A	B	A	B
A-weighted level difference	−5.1418	5.9696	9.3204	3.3068	−5.2999	3.6593	−18.6277	2.3182
Loudness ratio	0.4893	0.1693	1.6817	0.6347	0.4689	0.1120	0.1073	0.1047
AI	0.0767	0.0352	0.2240	0.0683	0.0756	0.0297	0.0138	0.0168
SII	0.1182	0.0459	0.2863	0.0776	0.1110	0.0428	0.0376	0.0132
Uniformly wtd, −22-dB clipped	−11.9266	2.3234	−5.2475	2.8743	−11.9736	1.7240	−16.4794	1.1815
AI-weighted, −22-dB clipped	−12.9251	2.2389	−6.5831	2.8026	−13.0123	1.7106	−17.7037	0.9459
SII-weighted, −22-dB clipped	−12.9106	2.1987	−6.8542	2.9735	−12.8981	1.6604	−17.7130	1.1150
Uniformly wtd, −32-dB clipped	−12.1875	2.5259	−5.2621	2.7047	−12.1053	1.9800	−17.8748	1.3434
AI-weighted, −32-dB clipped	−13.2084	2.5172	−6.4154	2.7317	−13.1395	1.9902	−18.5097	1.2247
SII-weighted, −32-dB clipped	−13.2092	2.4777	−6.8974	2.8654	−13.1850	2.0601	−18.1247	1.0317

A-weighted level difference (0.464) and the loudness ratio (0.565) are the lowest. AI and SII yield R^2 values of 0.726 and 0.745, respectively. The other six indices are weighted signal-to-noise ratios: three different weightings (uniform, AI, SII) for each of two different signal-to-noise clipping levels ($L=-22$ and -32 dB). The highest R^2 values are for the SII weighting, $R^2=0.762$ for -22-dB clipping, and $R^2=0.757$ for -32-dB clipping. The AI-weighted indices yield $R^2=0.755$ and $R^2=0.750$ for -22-dB and -32-dB clipping, respectively. The uniformly weighted indices result in $R^2=0.750$ for both -22- and -32-dB clipping. These clipping levels were selected after having assessed the relationships between the test results and the indices for clipping from -12 dB down to -32 dB, in 2-dB steps. These results are discussed in Appendix B.

All of the weighted signal-to-noise indices presented in Fig. 8 are well-correlated with the intelligibility scores. The differences among the R^2 values are statistically significant, but it is not clear that they are all practically significant. In general, the SII-weighted measures correlate slightly better than the others, the highest for -22-dB clipping of the signal-to-noise level difference.

Notice that AI and SII have the problem that at their minimum value of zero, subjects are still able to correctly identify words from the test sentences. They are therefore confirmed to be unsuitable as a security measure.

Figure 9 shows the curve fits to the 5th, 25th, 50th, 75th, and 95th percentiles of the intelligibility score data collected from both listening tests, plotted versus SII-weighted, -22-dB clipped signal-to-noise ratio. The parameters defining these Boltzmann curves are given in Table IV. Since the weights sum to unity, the minimum value of the measure is -22 dB, which is the endpoint of these curves. What the figure shows, for instance, is that at an index value of -15 dB, 75% of the subjects could identify about 10% of the overheard words, 50% of the subjects could identify 23% of the words, and only the best 5% of the subjects could identify 92% of the words. Dropping an additional 5 dB to an index value of -20 dB, only 25% of the subjects could identify more than 4% of the words, and only the best 5% of subjects could identify 11% of the words.

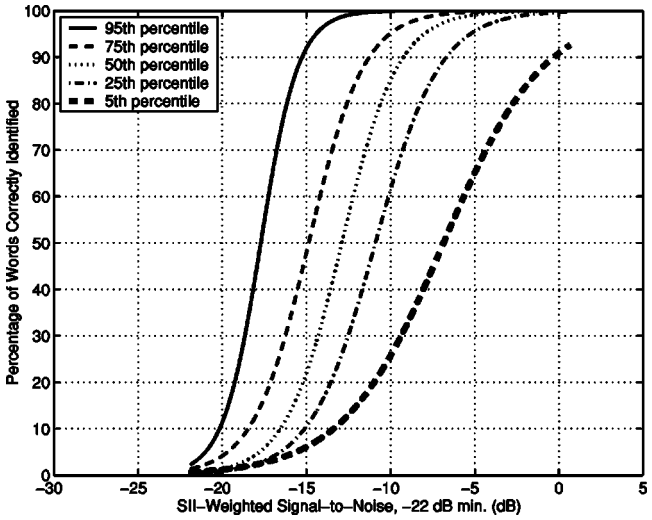


FIG. 9. Intelligibility score versus SII-weighted signal-to-noise. The curves from bottom to top represent the 5th, 25th, 50th, 75th, and 95th percentiles of the 9500 individual responses.

TABLE IV. Values of the Boltzmann function parameters A and B from Eq. (A1) for the intelligibility score percentile curves shown in Fig. 9.

	Intelligibility score percentiles	
	A	B
5th percentile	−6.8542	2.9735
25th percentile	−10.9010	1.9146
50th percentile	−12.8981	1.6604
75th percentile	−14.8641	1.6355
95th percentile	−17.7130	1.1150

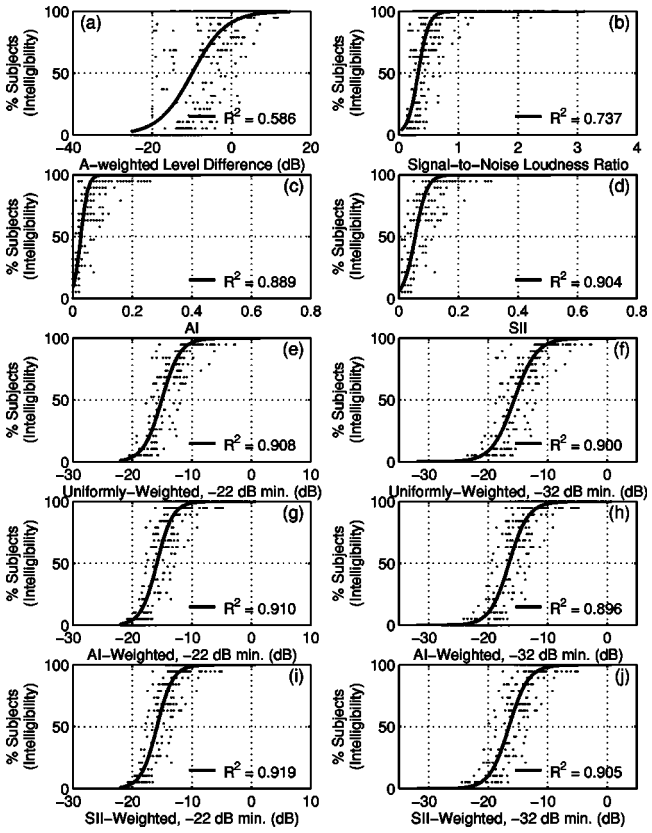


FIG. 10. Threshold of intelligibility: The dots are the percentage of the 19 subjects correctly identifying at least one word from each of the 340+160 sentences in both tests (500 dots total). The solid curve is the least-squares Boltzmann function fit to the mean. The squared correlation coefficient values (R^2) for the fits are shown.

B. Thresholds

For each of the 340+160=500 sentences in both tests, a tally was made of the percentage of the 19 subjects able to correctly identify at least one word. These data are plotted in Fig. 10 versus the objective indices (the dots). Also shown in each panel is the least-squares Boltzmann function fit, and associated R^2 value (function parameters are given in Table V). These are the threshold of intelligibility results. Appendix B discusses R^2 values for correlations with indices computed with other clipping levels.

Notice that as above for the word intelligibility scores, A-weighted level difference and loudness ratio were the poorest indicators ($R^2=0.586$ and 0.737, respectively). The R^2 values for AI and for SII were 0.889 and 0.904, respectively. The −22-dB clipped signal-to-noise ratios had the highest R^2 values (0.919, 0.910, and 0.908 for the SII, AI, and uniform weights, respectively). The R^2 for the −32-dB clipped schemes were 0.905 for SII weighting, 0.896 for AI weighting, and 0.900 for uniform weighted.

For the 160 sentences of test 2, a tally was made of the fraction of the 19 subjects able to: (1) identify the cadence or rhythm of the speech (including those identifying words), and (2) hear the presence of speech in the background noise (including those identifying cadence or words). These results are shown in Fig. 11, for the threshold of cadence, and in Fig. 12 for the threshold of audibility. Each panel shows the percentage of subjects (the dots), the least-squares Boltzmann function fit, and the R^2 value for the fit (see Table V for the fit parameters). Appendix B discusses R^2 values for correlations with indices computed with other clipping levels.

For the threshold of cadence analysis, R^2 was 0.918 for the A-weighted level difference, 0.956 for the loudness ratio. The R^2 values for AI and SII were 0.672 and 0.770, respectively. The −22-dB clipped measures had R^2 values of 0.912, 0.798, and 0.815 for uniform, AI, and SII weighting, respectively. The −32-dB clipped measures had R^2 values of 0.858, 0.686, and 0.691 for uniform, AI, and SII weighting, respectively.

For the threshold of audibility analysis, R^2 was 0.835 for the A-weighted level difference, 0.899 for the loudness ratio. The R^2 values for AI and SII were 0.389 and 0.566, respec-

TABLE V. Values of the Boltzmann function parameters *A* and *B* from Eq. (A1) for the threshold curves shown in Figs. 10–12.

	Threshold of intelligibility		Threshold of cadence		Threshold of audibility	
	A	*B*	*A*	*B*	*A*	*B*
A-weighted level difference	−9.8875	4.3067	−12.7405	2.2338	−16.6796	2.3335
Loudness ratio	0.3242	0.0882	0.2099	0.0413	0.1398	0.0288
AI	0.0271	0.0117	0.0033	0.0032	−0.0001	0.0020
SII	0.0547	0.0197	0.0150	0.0087	0.0047	0.0083
Uniformly wtd, −22-dB clipped	−15.0566	1.5448	−18.5891	0.8918	−19.9799	0.7142
AI-weighted, −22-dB clipped	−15.8685	1.3351	−18.7729	0.9573	−20.0918	1.0158
SII-weighted, −22-dB clipped	−15.8558	1.3363	−18.8710	0.8925	−20.1302	0.9513
Uniformly wtd, −32-dB clipped	−15.6432	1.8739	−20.0467	1.4037	−22.4119	1.8053
AI-weighted, −32-dB clipped	−16.4980	1.6114	−20.1055	2.0360	−22.9373	2.5032
SII-weighted, −32-dB clipped	−16.5059	1.6267	−20.3049	2.0092	−23.1323	2.4872

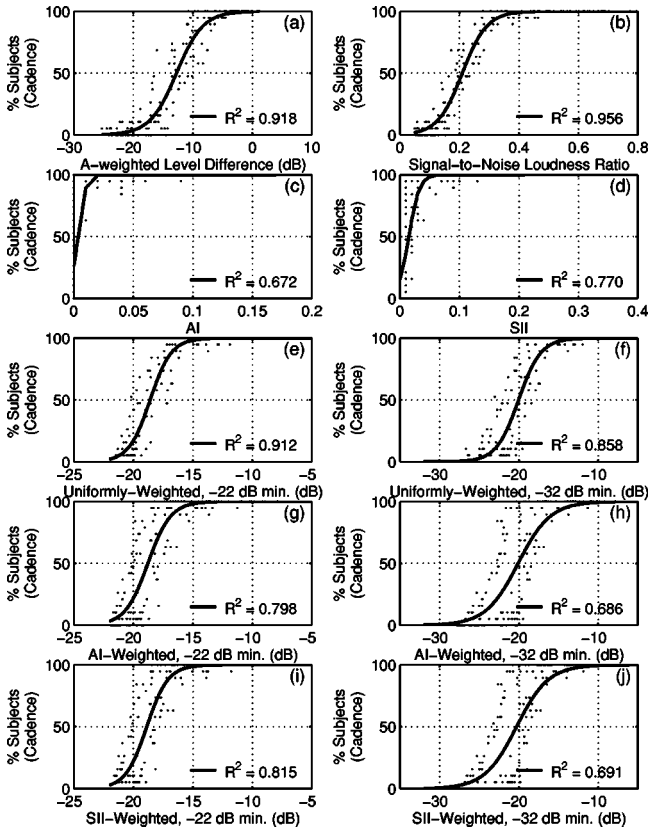


FIG. 11. Threshold of cadence: The dots are the percentage of the 19 subjects identifying the cadence of speech (including those correctly identifying some words) from each of the 160 sentences in the second test (160 dots total). The solid curve is the least-squares Boltzmann function fit to the mean. The squared correlation coefficient values (R^2) for the fits are shown.

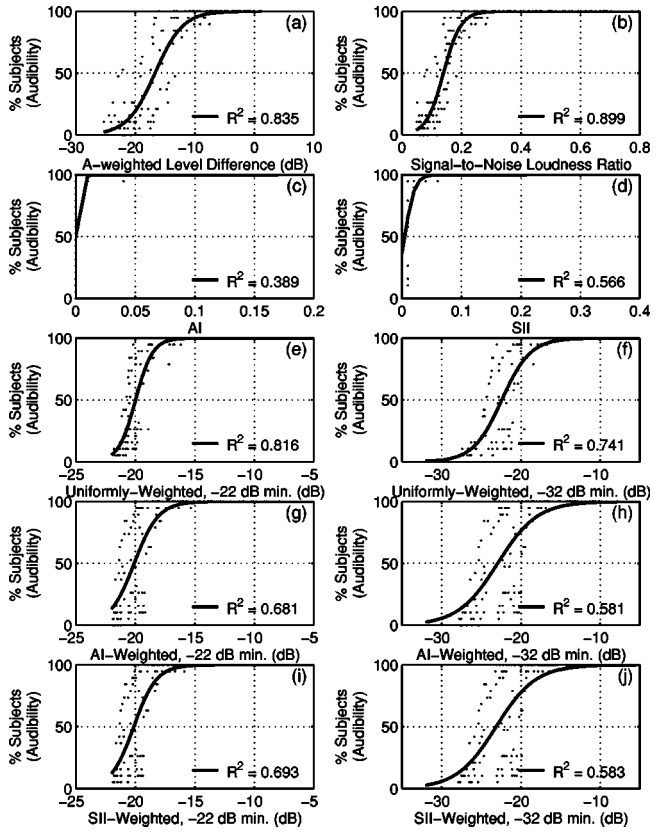


FIG. 12. Threshold of audibility: The dots are the percentage of the 19 subjects identifying the presence of speech sounds (including those identifying cadence and those identifying some words) from each of the 160 sentences in the second test (160 dots total). The solid curve is the least-squares Boltzmann function fit to the mean. The squared correlation coefficient values (R^2) for the fits are shown.

tively. The −22-dB clipped measures had R^2 values of 0.816, 0.681, and 0.693 for uniform, AI, and SII weighting, respectively. The −32-dB clipped measures had R^2 values of 0.741, 0.581, and 0.583 for uniform, AI, and SII weighting, respectively.

In general, for both cadence and audibility thresholds, the A-weighted level difference and loudness ratio are better indicators than the signal-to-noise ratio type measures, including AI and SII. The thresholds of cadence and of audibility involve the detection of sounds, not necessarily understanding of speech. They are more related to audibility and loudness than to intelligibility; the strong correlations with measures of such (loudness, A-weighted levels) indicate this. The −22-dB clipped measures were better correlated than the −32-dB clipped, but inspection of the figure indicates that the best-fit curves do not reach zero. This is analogous to the problem with AI and SII for intelligibility scores. For both thresholds, the uniformly weighted measures are better correlated than the AI and SII weighted ones. This possibly indicates the relative importance of the lower frequencies (below 800 Hz) for audibility, which are reduced by the intelligibility-derived weighting schemes, but are relatively important for transmitted speech sounds (see Fig. 5).

Figure 13 shows the curves for the threshold of cadence and the threshold of audibility: the percent of subjects able to detect the cadence or any speech sounds, versus A-weighted level difference. At −10 dB (i.e., speech level 10 dB *lower* than background noise level), 95% of the subjects could identify that there was speech, and 77% of the subjects could identify its cadence. Five decibels lower than this, at a level difference of −15 dB, 67% of subjects could hear the speech sounds, 27% identifying the cadence. The threshold of intel-

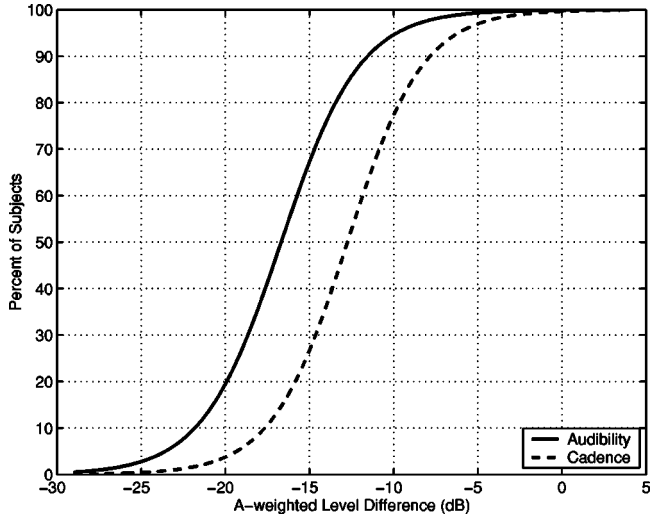


FIG. 13. Thresholds of cadence and audibility versus A-weighted level difference. The curves represent the percentage of subjects able to at least identify the cadence of speech (cadence) or to at least detect the presence of speech sounds (audibility).

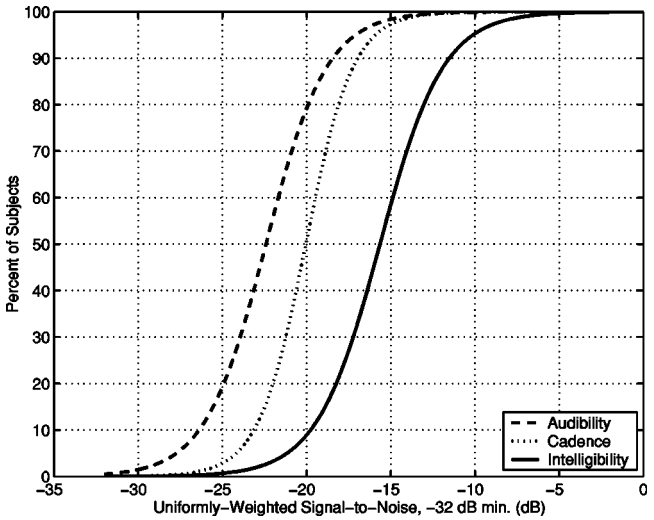


FIG. 14. Thresholds versus uniformly weighted, −32-dB clipped signal-to-noise index. Each curve represents the percentage of subjects able to correctly identify: at least one word from a sentence (intelligibility), the cadence of speech (cadence), or the presence of speech sounds (audibility).

ligibility curve was not included in Fig. 13, since the correlation with A-weighted level difference is so poor ($R^2 = 0.586$). A-weighted level difference is not a good measure for assessment of threshold of intelligibility.

A measure that is well-correlated with all three thresholds is the uniformly weighted −32-dB clipped signal-to-noise. This measure can be used to assess all thresholds, and indicate relative relations among them. The usefulness of uniformly weighted band levels has been explored by Tachibana *et al.* as an indicator of loudness.[15] This ties in well with the above-noted observation that detection of the thresholds of audibility and of cadence has to do with the loudness of the speech sounds. Figure 14 shows the threshold curves for the percentage of subjects able to correctly identify at least one word (intelligibility), able to identify at least the cadence (cadence), and able to at least hear some speech sounds (audibility) versus uniformly weighted, −32-dB clipped signal-to-noise index. For a measure value of −15 dB, 98% of the subjects could hear some speech sounds, and 60% could identify at least one word. Five decibels lower, at a value of −20 dB, only 8% of subjects could identify at least one word. At an index value of −25 dB, less than 1% of subjects could identify a word, and only 20% could hear speech sounds. If one considers 50% of the subjects as a threshold point (as is the norm), then the threshold of intelligibility could be said to be at about −15.5 dB, the threshold of audibility about 7 dB lower, at −22.5 dB.

V. CONCLUSIONS

Listening tests simulating speech transmission through a range of typical office wall constructions have been used to find objective measures of intelligibility and audibility suitable for architectural speech security situations. A speech-signal-to-noise ratio, restricting the 1/3-octave-band level differences to −22 dB and weighted using the band importance frequency weights from the SII calculation, was found to be a good measure of speech intelligibility. It has also been found that, in cases of zero intelligibility, both loudness ratio and simple A-weighted level difference are able to accurately predict audibility of the speech or its cadence. These measures (loudness ratio and A-weighted level difference) should not, however, be used to assess intelligibility. A uniformly weighted 1/3-octave-band signal-to-noise ratio clipped to −32 dB is a good indicator for all three thresholds.

The existing measures AI and SII are not suitable for evaluating speech security. While they are highly correlated with the listening test intelligibility scores, both fail to indicate zero intelligibility at their minimum values of zero. Furthermore, they provide no information regarding the thresholds of audibility and of cadence.

These results indicate that the Speech-Privacy Calculation,[5] which uses estimates of A-weighted signal-to-noise ratio to assess privacy relative to AI=0.05, is not ideal for assessing the threshold of intelligibility.

The relationships derived between intelligibility scores and thresholds and the objective measures are worst-case from a security point of view. They are therefore broadly applicable. The listeners were acute hearing and cued to expect to overhear speech; the speaking voice was strong and clear. Predicting a level of security from this work should err on the conservative side relative to listeners with less sensitive hearing and/or talkers with less clear voices.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of Ms. Kimberlee Cuthbert, who was responsible for scheduling subjects, running subject experiments, and assembling the measurement data, and Mr. Ketana Teav, who set up much of the equipment and wrote control software for the computer and hardware. We also would like to thank the subjects who each volunteered several hours of their time to help us do this work. Funding for this work was provided by Public Works and Government Services Canada (PWGSC) and the Royal Canadian Mounted Police (RCMP).

APPENDIX A: BOLTZMANN FUNCTION FITS

The sigmoidal curves fitted to the data from the listening tests are Boltzmann functions.[16] The functional form of this curve is given by

$$F(x) = 1 - \frac{1}{1 + \exp\left(\frac{x - A}{B}\right)}, \tag{A1}$$

where A and B are parameters defining the midpoint of the rise and the slope, respectively. The parameter A is the abscissa value corresponding to the 50% point of the curve: $F(A) = 0.5$. The curve has asymptotes of $+1$ at $x = \infty$ and 0 at $x = -\infty$. The values of A and B for the intelligibility score curves shown in Fig. 8 are given in Table III, those for the intelligibility score percentile curves shown in Fig. 9 are given in Table IV, and those for the threshold curves shown in Figs. 10–12 are given in Table V.

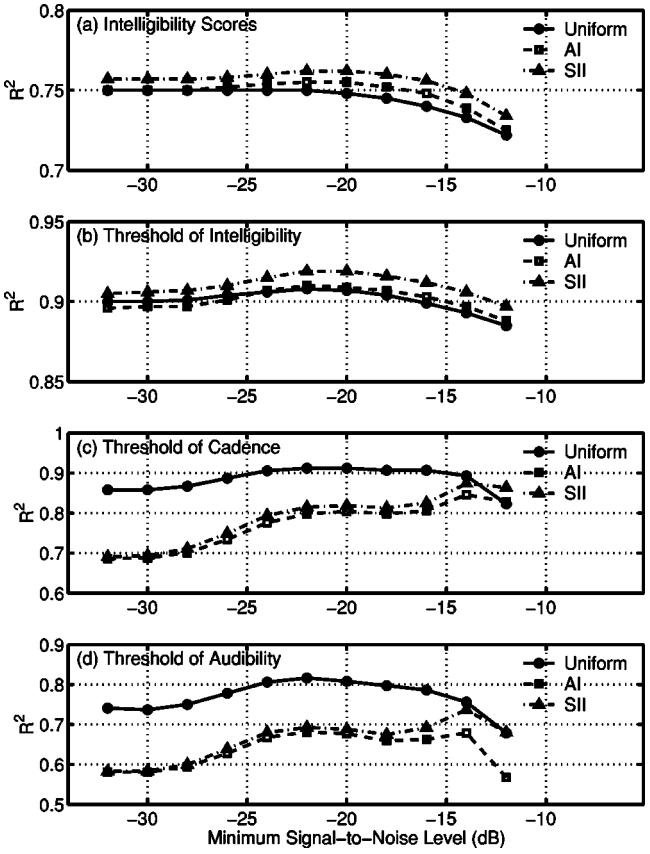


FIG. 15. Squared correlation coefficient (R^2) versus minimum band signal-to-noise clipping level. The correlation describes the relationship between the subjective test responses and the measure calculated using the corresponding level, for each of three weightings (uniform, AI, and SII) for (a) intelligibility scores; (b) threshold of intelligibility; (c) thresholds of cadence; and (d) threshold of audibility.

APPENDIX B: SIGNAL-TO-NOISE CLIPPING LEVEL

The parameter L in Eq. (2) defines the minimum value of the signal-to-noise level difference that is taken as contributing to the measure $X_w(L)$. The results of the listening tests were correlated with $X_w(L)$ for $L=-12$ dB to $L=-32$ dB, in 2-dB steps for the three weighting strategies (uniform, AI, and SII). Figure 15 shows the resulting R^2 values for the relationships of the measures with each of: (a) intelligibility scores; (b) threshold of intelligibility; (c) threshold of cadence; and (d) threshold of audibility. Clipping below -32 dB does not appreciably change the relationships; little of the collected data lies in that range.

For the intelligibility scores and the threshold of intelligibility, the correlations are much the same for all three weightings, the SII only slightly higher (notice the scale on the ordinate axis). There is a peak at around -22 dB, which is also about the point where the fit curve trends to zero intelligibility (see Figs. 8 and 10). Clipping below this, the R^2 trends to a constant value. Above this, not only does the correlation drop, but also the data start to clip before the intelligibility reaches zero, as happens with AI and SII.

The thresholds of cadence and audibility show that uniform weighting is better correlated than the AI and SII for all clipping levels. The R^2 actually drops with decreasing clipping level, but from inspection of Figs. 11 and 12, a level of -32 dB is necessary to ensure the data points are not clipped before audibility reaches zero.

[1] ASTM E1130-02, "Standard Test Method for Objective Measurement of Speech Privacy in Open Offices Using Articulation Index," (ASTM International, West Conshohocken, PA).
[2] ANSI S3.5-1969, "American National Standard Methods for the Calculation of the Articulation Index," (ANSI, New York, NY).
[3] ANSI S3.5-1997, "American National Standard Methods for Calculation of the Speech Intelligibility Index," (ANSI, New York, NY).
[4] W. J. Cavanaugh, W. R. Farrell, P. W. Hirtle, and B. G. Watters, "Speech privacy in buildings," J. Acoust. Soc. Am. **34**, 475–492 (1962).
[5] R. W. Young, "Re-vision of the speech-privacy calculation," J. Acoust. Soc. Am. **38**, 524–530 (1965).
[6] M. D. Egan, *Architectural Acoustics* (McGraw-Hill, New York, 1988.)
[7] F. Santon, "Intelligibilité de la parole masquée et rôle de la direction de la source masquante," "The Intelligibility of Speech and the role of the Masking Source Direction" Acustica **61**, 67–74 (1986).
[8] "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 227–246 (1969).
[9] A. C. C. Warnock, "Acoustical privacy in the landscaped office," J. Acoust. Soc. Am. **53**, 1535–1543 (1973).
[10] C. Speaks, B. Parker, C. Harris, and P. Kuhl, "Intelligibility of connected discourse," J. Speech Hear. Res. **15**, 590–602 (1972).
[11] W. O. Olsen, "Average speech levels and spectra in various speaking/listening conditions: A summary of the Pearson [*sic*], Bennett, & Fidell (1977) report," Am. J. Audiol. **7**, 1–5 (1998).
[12] ANSI S12.2-1995 (R1999), "American National Standard Criteria for Evaluating Room Noise," (ANSI, New York, NY).
[13] ISO 532-1975, "Acoustics—Method for calculating loudness level," (ISO, Geneva, Switzerland).
[14] ISO 7029-1984 (E), "Acoustics—Threshold of hearing by air conduction as a function of age and sex for otologically normal persons," (ISO, Geneva, Switzerland).
[15] H. Tachibana, Y. Sonoda, K. Iwamoto, S. Kuwano, and S. Namba, "Validity of arithmetic average of sound pressure levels in octave bands as a loudness index," J. Acoust. Soc. Jpn. (E) **14**, 197–204 (1993).
[16] "ORIGIN User's Manual Version 5," Microcal Software Inc., 1997.

Robustness of spatial average equalization: A statistical reverberation model approach[a)]

Sunil Bharitkar,[b)] Philip Hilmes, and Chris Kyriakakis
Department of Electrical Engineering—Systems, University of Southern California, Los Angeles, California 90089-2564

(Received 28 December 2003; revised 28 June 2004; accepted 4 September 2004)

Traditionally, multiple listener room equalization is performed to improve sound quality at all listeners, during audio playback, in a multiple listener environment (e.g., movie theaters, automobiles, etc.). A typical way of doing multiple listener equalization is through spatial averaging, where the room responses are averaged spatially between positions and an inverse equalization filter is found from the spatially averaged result. However, the equalization performance, will be affected if there is a mismatch between the position of the microphones (which are used for measuring the room responses for designing the equalization filter) and the actual center of listener head position (during playback). In this paper, we will present results on the effects of microphone-listener mismatch on spatial average equalization performance. The results indicate that, for the analyzed rectangular configuration, the region of effective equalization depends on (i) the distance of a listener from the source, (ii) the amount of mismatch between the responses, and (iii) the frequency of the audio signal. We also present some convergence analysis to interpret the results. [DOI: 10.1121/1.1810238]

PACS numbers: 43.55.Br, 43.60.Cg, 43.55.Jz [NX] Pages: 3491–3497

I. INTRODUCTION

A typical room is an acoustic enclosure that can be modeled as a linear system whose behavior at a particular listening position is characterized by an impulse response. The impulse response yields a complete description of the changes a sound signal undergoes when it travels from a source to a receiver (microphone/listener). The signal at the receiver consists of direct path components, discrete reflections that arrive a few milliseconds after the direct sound, as well as a reverberant field component. In addition, it is well established that room responses change with source and receiver locations in a room.[1,2]

Specifically, the time of arrival of the direct and multipath reflections and the energy of the reverberant component will vary from position to position. In other words; a room response at position i, $p_{f,i}$, can be expressed as $p_{f,i}=p_{f,d,i}+p_{f,\mathrm{rev},i}$; whereas the room response at position j, $p_{f,j}$, can be expressed as $p_{f,j}=p_{f,d,j}+p_{f,\mathrm{rev},j}$ where $p_{f,d}$ is the frequency response for the direct path component, and $p_{f,\mathrm{rev}}$ is the response for the multipath component. An example of time domain responses at two positions, displaced a few feet apart from each other, in a room with a reverberation time of about 0.25 s, is shown in Fig. 1 along with the direct component, early reflections, and late reverberant components. Figure 2 shows the corresponding frequency response from 20 Hz to 20 kHz.

One of the goals in equalization is to minimize the spectral deviations (viz., correcting the peaks and dips) found in the magnitude response through an equalization filter. This correction of the room response significantly improves the quality of sound played back through a loudspeaker system. In essence, the resulting system formed from the combination of the equalization filter and the room response should have a perceptually flat frequency response.

One of the important considerations is that the equalization filter has to be designed such that the spectral deviations in the magnitude response (e.g., Fig. 2) are minimized *simultaneously* for all listeners in the environment. Simultaneous equalization is an important consideration since listening has evolved into a group experience (e.g., as in-home theaters, movie theaters, and concert halls). An example of performing only a single position equalization (by designing an inverse filter for position 1) is shown in Fig. 3. The top plot shows the equalization result at position 1 (which shows a flat response under an ideal filter design).[3] However, the equalization performance is degraded at position 2 with the use of this single position filter, as can be seen in the lower plot. For example, comparing Figs. 2 and 3, it can be seen that the response around 50 Hz at position 2, after single position equalization, is at least 7 dB below the response before equalization.

One method for providing simultaneous multiple listener equalization is by spatially averaging the measured room responses at different positions, for a given loudspeaker, and stably inverting the result. The microphones are positioned, during measurements, at the expected center of a listener's head. An example of performing spatial average equalization is shown in Fig. 4. Clearly, the spectral deviations are significantly minimized for both positions through the spatial average equalization filter.[4–6]

Although spatial average equalization is aimed at achieving uniform frequency response coverage for all listeners, its performance is often limited due to (i) mismatch

[a)]Submitted in December 2003 to the Journal of the Acoustical Society of America, revised June 2004.
[b)]Current address: 2461 Coolidge Ave., Ste. 1, Los Angeles, California 90064; electronic mail: sunil@audyssey.com

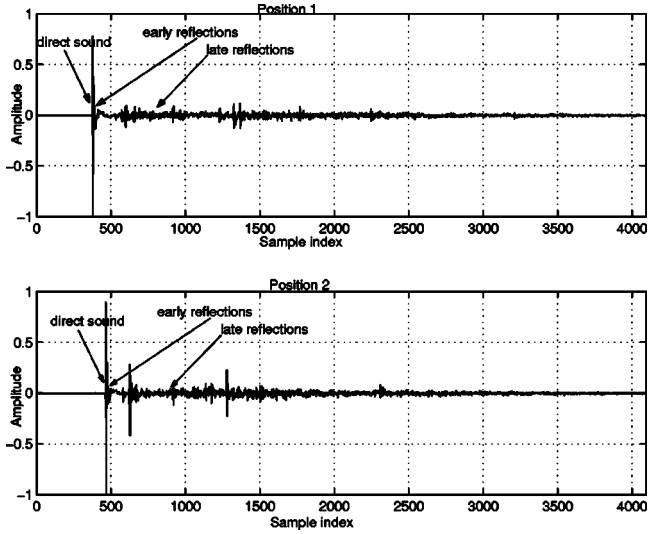


FIG. 1. Examples of room acoustical responses, having the direct and reverberant components, measured at two positions a few feet in a room.

between the microphone measurement location and the actual location for the center of the listener head, or (ii) variations in listener locations (e.g., head movements).

In this paper, we present a method for evaluating the robustness of spatial averaging based equalization, due to the introduction of variations in room responses [generated either through (i) or (ii)], for a rectangular listener arrangements relative to a fixed sound source. The proposed approach uses a statistical description for the reverberant field in the responses (viz., via the normalized correlation functions) in a rectangular listener configuration for a rectangular room.[7] A similar approach is followed in Ref. 8 for determining variations in performance. However, this was done with a single position equalization in mind and is focused for microphone array applications (e.g., sound source localization). Talantzis *et al.*[9] used a similar analysis for understanding the effect of source displacements, but this analysis was also presented for a microphone array setup without spatial average equalization.

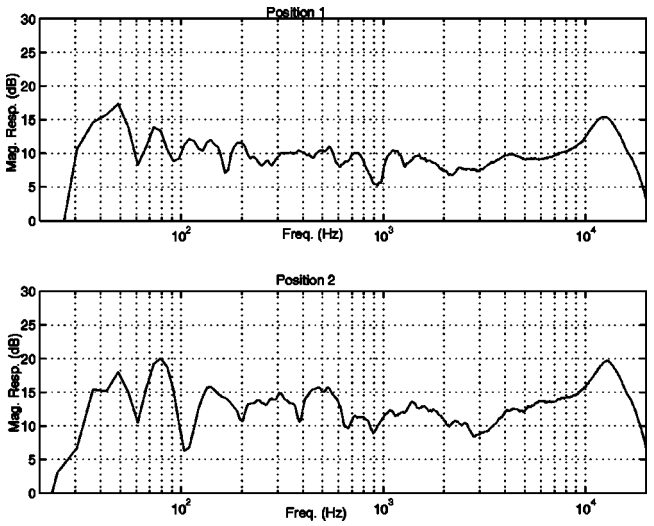


FIG. 2. Magnitude responses of room responses of Fig. 1 showing different spectral deviations (from flat) at the two listener positions.

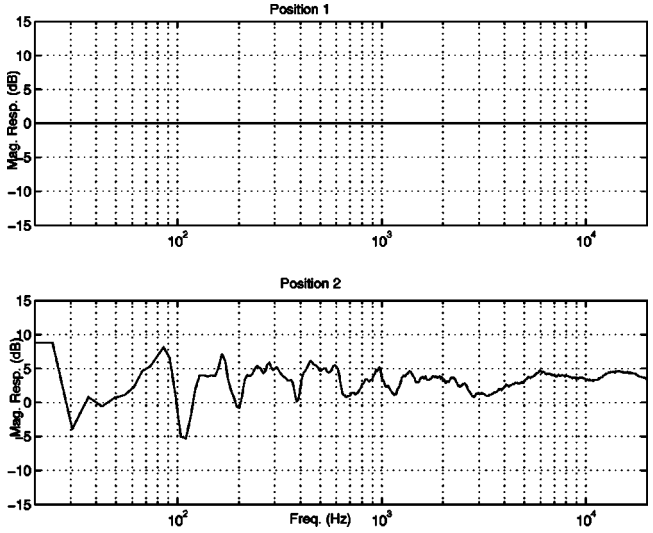


FIG. 3. Magnitude responses, upon single position equalization, of responses of Fig. 2. Specifically, the equalization filter is designed to correct for deviations at position 1, but the equalized response at position 2 is degraded.

The advantage of the proposed approach is that (i) it is based on an established theory of the statistical nature of reverberant sound fields;[13] (ii) it can be applied to a large frequency range above the Schroeder frequency, for typical size rooms, unlike modal equations that are valid for low frequencies having wavelengths greater than $1/3 \min[L_x, L_y, L_z]$;[11] and (iii) the computational complexity, due to the approximations, is low.

In Sec. II, we introduce a background necessary for the development of the robustness analysis. Specifically, an introduction is provided for the deterministic direct component, and the statistical reverberant field correlations. In Sec. III we present the mismatch measure for analyzing the effects of mismatch between the microphone (during measurement of room responses) and listener position (during playback) with a spatial average equalizer.[10] Additionally, a convergence analysis of the equalization mismatch error, for

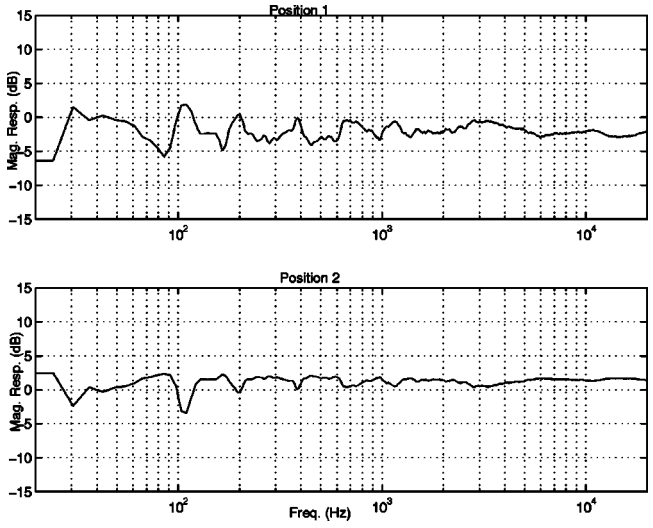


FIG. 4. Magnitude responses, upon spatial average equalization, of responses of Fig. 2. Specifically, the equalization filter is designed to correct for deviations, on an average, at positions 1 and 2.

spatial average equalization, is presented at the end of the section. In Sec. IV we present results based on simulations for typical rectangular listener arrangements relative to a fixed source. A rectangular configuration has been considered, as this is common in large environments (e.g., movie theaters, concert halls) as well as in typical home-theater setups. The analysis can be extended to arbitrary listening configurations. In Sec. V we conclude the paper.

II. ROOM ACOUSTICS FOR SIMPLE SOURCES

The sound pressure $p_{f,i}$ at location i and frequency f can be expressed as a sum of direct field component, $p_{f,d,i}$, and a reverberant field component, $p_{f,\mathrm{rev},i}$, as given by

$$p_{f,i}=p_{f,d,i}+p_{f,\mathrm{rev},i}. \tag{1}$$

The direct field component for sound pressure, $p_{f,d,i}$, of a plane wave, at farfield listener location i for a sound source of frequency f located at i_0, can be expressed as[11]

$$\begin{aligned} p_{f,d,i}&=-jk\rho cS_f g_f(i|i_0)e^{-j\omega t},\\ g_f(i|i_0)&=\frac{1}{4\pi R}e^{jkR},\\ R^2&=|i-i_0|^2, \end{aligned} \tag{2}$$

where $p_{f,d}(i|i_0)$ is the direct component sound pressure amplitude, S_f is the source strength, $k=2\pi/\lambda$ is the wave number, $c=\lambda f$ is the speed of sound (343 m/s) and ρ is the density of the medium (1.25 kg/m^3 at sea level).

The normalized correlation function[12] which expresses a statistical relation between sound pressures, of reverberant components, at separate locations i and j, is given by

$$\frac{E\{p_{f,\mathrm{rev},i}p^*_{f,\mathrm{rev},i}\}}{\sqrt{E\{p_{f,\mathrm{rev},i}p^*_{f,\mathrm{rev},i}\}}\sqrt{E\{p_{f,\mathrm{rev},j}p^*_{f,\mathrm{rev},j}\}}}=\frac{\sin kR_{ij}}{kR_{ij}}, \tag{3}$$

where R_{ij} is the separation between the two locations i and j relative to an origin, and $E\{\cdot\}$ is the expectation operator.

The reverberant-field mean square pressure is defined as

$$E\{p_{f,\mathrm{rev},i}p^*_{f,\mathrm{rev},i}\}=\frac{4c\rho\Pi_a(1-\bar{\alpha})}{S\bar{\alpha}}, \tag{4}$$

where Π_a is the power of the acoustic source, $\bar{\alpha}$ is the average absorption coefficient of the surfaces in the room, and S is the surface area of the room.

The assumption of a statistical description [as given in (3), (4)] for reverberant fields in rooms is justified if the following conditions are fulfilled:[13] (1) Linear dimensions of the room must be large relative to the wavelength. (2) Average spacing of the resonance frequencies must be smaller than one-third of their bandwidth (this condition is fulfilled in rectangular rooms at frequencies above the Schroeder frequency, $f_s=2000\sqrt{T_{60}/V}$ Hz (T_{60} is the reverberation time in seconds and V is the volume in m^3). (3) Both source and microphone are in the interior of the room, at least a half-wavelength away from the walls.

Furthermore, under the conditions in Ref. 13, the direct and reverberant sound pressure are uncorrelated.

III. MISMATCH ANALYSIS FOR SPATIAL AVERAGE EQUALIZATION

A. Analytic expression for mismatch performance function

A performance function, W_f, that is used for analyzing the effects of mismatch, for spatial average equalization, of room responses is given as

$$\begin{aligned} \bar{W}_f&=\frac{1}{N}\sum_{i=1}^{N}\epsilon_{f,i}(r),\\ \epsilon_{f,i}(r)&=E\{|\tilde{p}_f(r)\bar{p}_f^{-1}-p_{f,i}\bar{p}_f^{-1}|^2\}. \end{aligned} \tag{5}$$

In (5), $\epsilon_{f,i}(r)$ represents the equalization error in the r neighborhood of the equalized location i having response $p_{f,i}$ (the r neighborhood is defined as all points at a distance of r from location i). The neighboring response, at a distance r from location i, is denoted by $\tilde{p}_f(r)$, whereas the spatial average equalization response is denoted by $\bar{p}_f$. Thus, response $\tilde{p}_f(r)$ is the response corresponding to the displaced center of the head position of the listener (viz., with a displacement of r). To get an intermediate equalization error measure, $\epsilon_{f,i}(r)$, the expectation is performed over all neighboring locations at a distance r from the equalized location i. Furthermore, the final performance function $\bar{W}_f$ is the average of all the equalization errors, $\epsilon_{f,i}(r)$, in the vicinity of the N equalized locations. In essence, the displacement (distance) r can be interpreted as a "mismatch parameter," since a room response measured at displacement r will be different than the response measured at a nominal location i.

For simplicity, in our analysis, we assume variations in responses due to displacements (or mismatch) in a horizontal plane (x-y plane). The analysis, presented in this paper, can be extended to include displacements on a spherical surface. Thus, (5) can be simplified to yield

$$\epsilon_{f,i}(r)=E\left\{\left|\frac{\tilde{p}_f(r)N}{\Sigma_{j=1}^{N}p_{f,j}}-\frac{p_{f,i}N}{\Sigma_{j=1}^{N}p_{f,j}}\right|^2\right\}. \tag{6}$$

An approximate simplification for (5) can be done by using the Taylor series expansion.[14] Accordingly, if g is a function of random variables, x_i, with average values $E\{x_i\}=\bar{x}_i$, then $g(x_1,x_2,\ldots,x_n)=g(x)$ can be expressed as $g(x)=g(\bar{x})+\Sigma_{i=1}^{n}g_i'(\bar{x})(x_i-\bar{x}_i)+g(\hat{x})$, where $g(\hat{x})$ is a function of order 2 (i.e., all its partial derivatives up to the first order vanish at $(\bar{x}_1,\bar{x}_2,\ldots,\bar{x}_n)$. Thus, to a zeroth order of approximation, $E\{g(x)\}\approx g(\bar{x})$.

Hence, an approximation for (6) is given as

$$\epsilon_{f,i}(r)\approx N^2\frac{E\{\tilde{p}_f(r)\tilde{p}_f(r)^*-\tilde{p}_f(r)p^*_{f,i}-\tilde{p}_f(r)^*p_{f,i}+p_{f,i}p^*_{f,i}\}}{\Sigma_j\Sigma_k E\{p_{f,j}p^*_{f,k}\}}. \tag{7}$$

We use the following identities for determining the denominator of (7):

$$E\{p_{f,j}p^*_{f,k}\}=E\{p_{f,d,j}p^*_{f,d,k}+p_{f,\mathrm{rev},j}p^*_{f,\mathrm{rev},k}\}, \tag{8}$$

$$|kc\rho S_f|^2=4\pi\Pi_a c\rho, \tag{9}$$

$$E\{p_{f,d,j}p^*_{f,d,k}\}=\frac{\Pi_a c\rho}{4\pi R_j R_k}e^{jk(R_j-R_k)}, \tag{10}$$

$$E\{p_{f,\mathrm{rev},j}p^*_{f,\mathrm{rev},k}\}=\frac{4c\rho\Pi_a(1-\overline{\alpha})}{S\overline{\alpha}}\frac{\sin kR_{jk}}{kR_{jk}}, \tag{11}$$

$$R_{jk}=\sqrt{R_j^2+R_k^2-2R_jR_k\cos\theta'_{jk}}. \tag{12}$$

In summary, (8) is obtained by using (1) and knowing that the reverberant and direct field components of sound pressure are uncorrelated, (9) is derived in Ref. 11 (p. 311), (10) is determined by using (2) and (9), and (11) is determined from (3) and (4). In (12), which is the cosine law, θ_{jk} is the angle, subtended at the source at i_0, between locations j and k.

Thus, the denominator term in (7) is

$$\sum_j\sum_k E\{p_{f,j}p^*_{f,k}\}=\sum_j\sum_k\left(\frac{\Pi_a c\rho}{4\pi R_jR_k}e^{jk(R_j-R_k)}+\frac{4c\rho\Pi_a(1-\overline{\alpha})}{S\overline{\alpha}}\frac{\sin kR_{jk}}{kR_{jk}}\right). \tag{13}$$

Now, the first numerator term in (7) is

$$\begin{aligned}E\{\tilde{p}_f(r)\tilde{p}_f(r)^*\}&=E\{\tilde{p}_{f,d}(r)\tilde{p}_{f,d}(r)^*+\tilde{p}_{f,\mathrm{rev}}(r)\tilde{p}_{f,\mathrm{rev}}(r)^*\},\\ E\{\tilde{p}_{f,d}\tilde{p}^*_{f,d}\}&=|k\rho cS_f|^2E\{g_f(\tilde{i}|i_0)g_f^*(\tilde{i}|i_0)\}\\ &=|k\rho cS_f|^2E\left\{\frac{1}{(4\pi)^2|\tilde{R}|^2}\right\},\end{aligned} \tag{14}$$

where, $\tilde{R}$ is the distance from a source at i_0 relative to spatial average equalized location i, and is determined by using cosine law (viz., $\tilde{R}=\sqrt{R_i^2+r^2-2R_ir\cos\theta_i}$, where θ_i is the angle subtended at the source between location i and the location in the r neighborhood of location i). The result from applying the expectation can be found by averaging over all locations in a circle in the r neighborhood of location i (since for simplicity we have assumed mismatch in the horizontal or x-y plane). Thus,

$$E\left\{\frac{1}{|4\pi\tilde{R}|^2}\right\}=\frac{1}{2}\frac{1}{(4\pi)^2}\int_{-1}^{1}\frac{d(\cos\theta_i)}{R_i^2+r^2-2R_ir\cos\theta_i}. \tag{15}$$

Simplifying (15) and substituting the result in (14) gives

$$E\{\tilde{p}_{f,d}(r)\tilde{p}_{f,d}(r)^*\}=\frac{|k\rho cS_f|^2}{2(4\pi)^2R_ir}\log\left|\frac{R_i+r}{R_i-r}\right|=\frac{\Pi_a\rho c}{8R_ir\pi}\log\left|\frac{R_i+r}{R_i-r}\right| \tag{16}$$

$$E\{\tilde{p}_{f,\mathrm{rev}}(r)\tilde{p}_{f,\mathrm{rev}}(r)^*\}=\frac{4c\rho\Pi_a(1-\overline{\alpha})}{S\overline{\alpha}}. \tag{17}$$

The result in (16) is obtained by using (9), whereas (17) is a restatement of (4). Thus,

$$E\{\tilde{p}_f(r)\tilde{p}_f(r)^*\}=\frac{\Pi_a\rho c}{8R_ir\pi}\log\left|\frac{R_i+r}{R_i-r}\right|+\frac{4c\rho\Pi_a(1-\overline{\alpha})}{S\overline{\alpha}}. \tag{18}$$

The correlation, $E\{\tilde{p}_{f,d}(r)p_{f,d,i}(r)^*\}$, in the direct-field component for the second term in the numerator of (7) is

$$|k\rho cS_f|^2\frac{1}{2(4\pi)^2}\int_{-1}^{1}\frac{e^{jk(\sqrt{R_i^2+r^2-2R_ir\cos\theta_i}-R_i)}d\cos\theta_i}{R_i\sqrt{R_i^2+r^2-2R_ir\cos\theta_i}}=\frac{\Pi_a\rho c}{4\pi R_i^2}\frac{1}{(4\pi R_i)^2}\frac{\sin kr}{kr}. \tag{19}$$

The reverberant field correlation for the second term in the numerator of (7) can be found using (3), and is

$$E\{\tilde{p}_{f,\mathrm{rev}}(r)p^*_{f,\mathrm{rev},i}\}=\frac{4c\rho\Pi_a(1-\overline{\alpha})}{S\overline{\alpha}}\frac{\sin kr}{kr}. \tag{20}$$

The third numerator term in (7) can be found in a similar manner as compared to the derivation for (19) and (20).

The last term in the numerator of (7) is computed to yield

$$E\{p_{f,i}p^*_{f,i}\}=\frac{\Pi_a\rho c}{4\pi R_i^2}+\frac{4\rho c\Pi_a(1-\overline{\alpha})}{S\overline{\alpha}}. \tag{21}$$

Equation (21) can be obtained by substituting $j=k=i$ in (10) and (11) respectively. Substituting the computed results into (7), and simplifying by canceling certain common terms in the numerator and the denominator, the resulting equalization error due to displacements (viz., mismatch in responses) is

$$\begin{aligned}\epsilon_{f,i}(r)&\approx\frac{N^2}{\psi_1}\left[\frac{1}{8R_ir\pi}\log\left|\frac{R_i+r}{R_i-r}\right|+2\psi_2+\frac{1}{2\psi_3}-\left(\frac{1}{\psi_3}+2\psi_2\right)\frac{\sin kr}{kr}\right]\\ \psi_1&=\sum_j\sum_l\left(\frac{1}{4\pi R_jR_l}e^{jk(R_j-R_l)}+\psi_2\frac{\sin kR_{jl}}{kR_{jl}}\right)\\ \psi_2&=\frac{4(1-\overline{\alpha})}{S\overline{\alpha}},\\ \psi_3&=2\pi R_i^2,\\ R_{jl}&=\sqrt{R_j^2+R_l^2-2R_jR_l\cos\theta_{jl}}.\end{aligned} \tag{22}$$

Finally, substituting (22) into (5) yields the necessary equation for $\overline{W}_f$.

B. Analysis of equalization error

In this section, we present an analysis for the behavior of the equalization error at each listener. This analysis helps in understanding, theoretically, the degradation (from a "steady-state" perspective) of equalization performance at different listener positions and at different frequencies.

Throughout the analysis we assume that $r<R_i$, for small mismatch between the microphone position and the center of the listener head position relative to the distance between the microphone and the source. Thus, in (22), $\log|(R_i+r)/(R_i-r)|\to 0$.

Now, for $r/\lambda>1$, the equalization error (22) converges to a steady-state value, $\epsilon^{SS}_{f,i}(r)$:

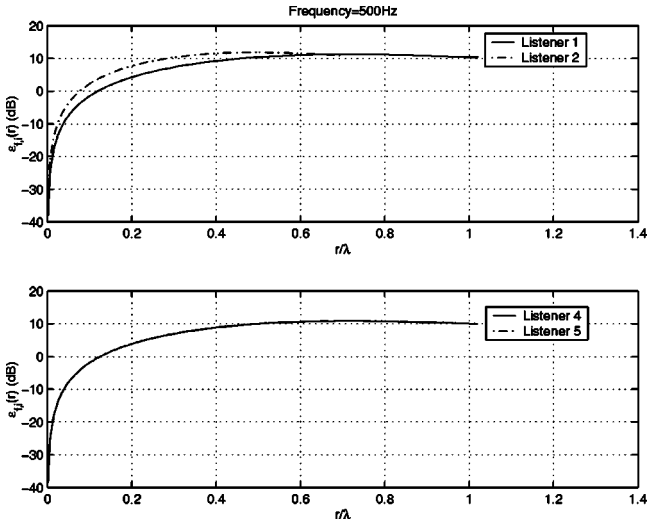


FIG. 5. $\epsilon_{f,i}(r)$ for the listeners at different distance from the source; $r/\lambda = 0$ corresponds to the optimal position, $f = 500$ Hz.

$$\epsilon_{f,i}^{ss}(r) \approx \frac{N^2}{\psi_1}\left[2\psi_2 + \frac{1}{2\psi_3}\right] = k_1\left(k_2 + \frac{1}{4\pi R_i^2}\right) \propto \frac{1}{R_i^2}; \quad (23)$$

since $\sin kr/kr \to 0$. This implies that listeners at larger distances will have lower steady-state equalization errors than listeners closer to the source for a given wavelength of sound. Primarily, the inverse relationship between $\epsilon_{f,i}^{ss}(r)$ and R_i, at steady state in (23) is due to the direct path sound field correlations [viz., $1/2\psi_3$ obtained from (21)] at position i.

IV. RESULTS

We simulated (22) for frequencies above the Schroeder frequency $f_s = 77$ Hz (i.e., $T_{60} = 0.7$ s, $V = 8\ \text{m} \times 8\ \text{m} \times 8\ \text{m}$).

In this setup, we simulated a rectangular arrangement of six microphones, with a source in the front of the arrangement. Specifically, microphones 1 and 3 were at a distance of 3 m from the source, microphone 2 was at 2.121 m, microphones 4 and 6 were at 4.743 m, and microphone 5 was at 4.242 m. The angles θ_{1k} in (12) were (45,90,18.5,45,71.62) degrees for $(k = 2,...,6)$, respectively. Thus, the distances of the listeners from the source are such that $R_6 = R_4 > R_5 > R_1 = R_3 > R_2$.

The equalization error, $\epsilon_{f,i}(r)$, results are depicted for different listeners in Figs. 5–8 for four frequencies ($f = 500$ Hz, $f = 1$ kHz, $f = 5$ kHz, and $f = 10$ kHz) as a function of r/λ, where the mismatch parameter $0 \leqslant r \leqslant 0.7$ m (r/λ corresponds to no mismatch condition). Specifically, only the results for listeners 1 and 2 are shown in the top panels since listener 3 is symmetric relative to source/listener 2 (hence the results of listener 1 and 3 are identical). Similarly, only the results for listeners 4 and 5 are shown in the bottom panels.

We observe the following.

(1) It can be seen that the steady-state equalization error at listener 2 is higher than that at listener 1 (top panel). This follows from (23) (since $R_1 = R_3 > R_2$). Similar results

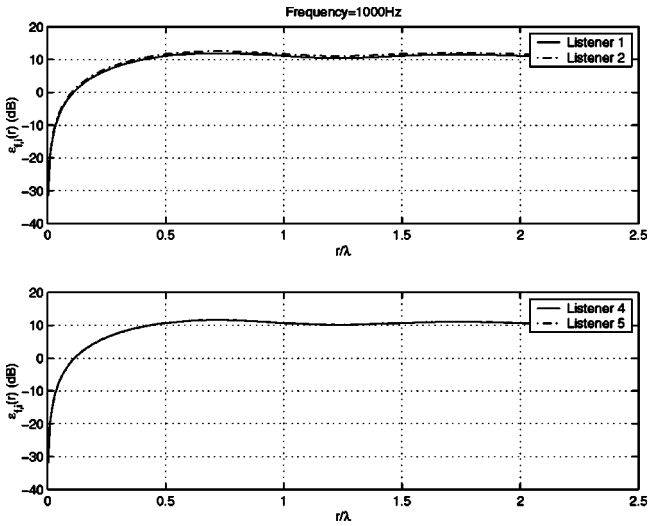


FIG. 6. $\epsilon_{f,i}(r)$ for the listeners at a different distance from the source, $f = 1$ kHz.

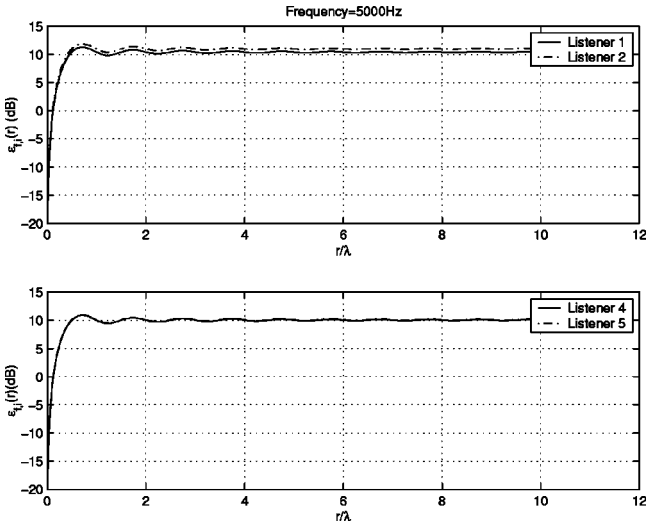


FIG. 7. $\epsilon_{f,i}(r)$ for the listeners at different distance from the source, $f = 5$ kHz.

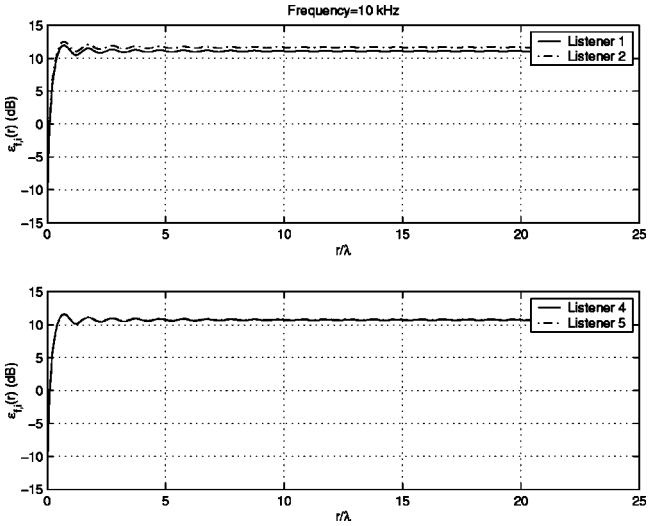


FIG. 8. $\epsilon_{f,i}(r)$ for the listeners at different distance from the source, $f = 10$ kHz.

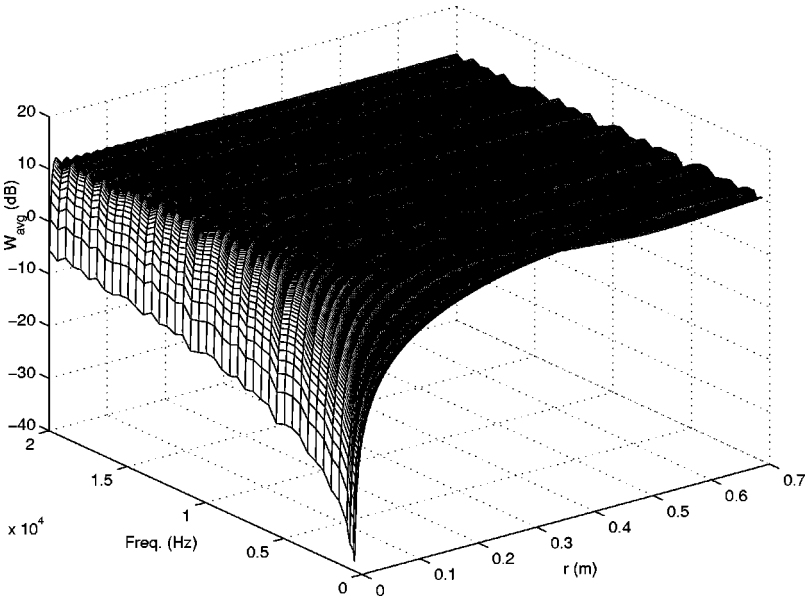


FIG. 9. $\overline{W}$ for various mismatch parameters and frequencies between between 20 Hz and 20 kHz.

can be predicted for the equalization errors for listeners 4 and 5 (this is not immediately obvious in the bottom panels, since R_4 is close to R_5).

(2) Furthermore, the non-steady state equalization region, for a given equalization error, is larger (better) for listeners farther from the source. For example, the equalization region is a circle of a radius 0.025λ for listener 2, whereas it is 0.04λ for listener 1 at $\epsilon_{f,i}(r) = -10$ dB and $f = 500$ Hz. This effect is dominant at lower frequencies but not easily noticeable at higher frequencies (as can be seen from the initial rise of the error toward a peak value before reaching a steady-state value).

(3) The equalization error shows a $\mathrm{sinc}(2r/\lambda)$ dependance after the initial peak (as emphasized in Fig. 6). This dependence arises from the finite correlation of the reverberant field before it reaches a negligible value at steady state.

Finally, Fig. 9 summarizes the average equalization error plot [i.e., W_{avg} of (5)], over all listeners, for frequencies beyond f_s and mismatch parameter r ranging from 0 to 0.7 m. This measure is a composite measure weighing the equalization error at all positions equally, and shows that the performance degrades for all frequencies with increasing mismatch or displacement in meters. Also, the degradation, for small displacement r (of the order of 0.1 m), is larger for higher frequencies. For example, it can be seen that the slope of the W_{avg} curves in the frequency region around 200 Hz is lower than the slopes of the curves for frequencies around 10 kHz. Alternate measures with nonuniform weighting, depending on the ''importance'' of a listening position, may be instead used. Thus, such a measure could potentially be used, to give an overall picture, during comparisons to other approaches of multiple listener equalization.

V. CONCLUSIONS

In this paper, we analyzed the performance of spatial averaging equalization, in a multiple listener environment, used during sound playback. As is well known, room equalization at multiple positions, allows for high quality sound playback in the room. However, as is typically the case in room equalization, the microphone positions during measurement of the room response will not necessarily correspond to the center of the head of the listener, leading to a frequency-dependent degradation due to mismatch between the measured response and the actual response corresponding to the center of listener head during playback. Several interesting observations can be made from the results, including (i) the influence of frequency and distance on the size of equalization region, (ii) the steady-state equalization error being dependent on the distance of the listener from the source, and (iii) the dependence of the reverberant field correlation on the equalization error. Future goals will be directed to using the proposed method for comparing different multiple listener equalization techniques in terms of their robustness to response mismatch.

ACKNOWLEDGMENTS

This research has been supported in part by the National Science Foundation (NSFERC, Cooperative Agreement No. EEC 9529152) and the United States Department of Army (Contract No. DAAD 19-99-D-0046).

[1] H. Kuttruff, *Room Acoustics*, 3rd ed. (Elsevier Applied Science, New York, 1991).

[2] J. Mourjopoulos, ''On the variation and invertibility of room impulse response functions,'' J. Sound Vib. **102**, 217–228 (1985).

[3] In practice, a low pass filter with a large cutoff frequency (e.g., 10 kHz), depending on the direct to reverberant energy, is applied to the equalization filter to prevent audio from sounding bright.

[4] The filter was a finite impulse response (FIR) filter of duration 8192 samples, and the design of lower-order filters and analysis for such designs will not be considered in this paper. Interested readers are requested to see, Refs. 5 and 6.

[5] S. Bharitkar and C. Kyriakakis, ''A cluster centroid method for room response equalization at multiple locations,'' *Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '01)*, New York, October 2001.

[6] S. Bharitkar and C. Kyriakakis, ''Perceptual multiple location equalization with clustering,'' *Proceedings of the 36th IEEE, Asilomar Conference on*

Signals, Systems, & Computers, Pacific Grove, California, November 2002.
[7] A rectangular room is considered since the assumptions for a statistical reverberant sound field have been verified for this shape room and in practice rectangular shaped rooms are found commonplace.
[8] B. D. Radlovic, R. C. Williamson, and R. A. Kennedy, ''Equalization in acoustic reverberant environment:robustness results,'' IEEE Trans. Speech Audio Process. **8**, 311–319 (2000).
[9] F. Talantzis and D. B. Ward, ''Robustness of multichannel equalization in an acoustic reverberant environment,'' J. Acoust. Soc. Am. **I**, 247250 (2002).
[10] S. Bharitkar, P. Hilmes, and C. Kyriakakis, ''Robustness of averaging equalizers to spatial variations: A statistical approach,'' in Ref. 6.
[11] P. M. Morse and K. Uno Ingard, *Theoretical Acoustics* (Princeton University Press, Princeton, NJ, 1986).
[12] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson, ''Measurement of correlation coefficients in reverberant sound fields,'' J. Acoust. Soc. Am. **27**, 1072–1077 (1955).
[13] M. R. Schroeder, ''Frequency correlation functions of frequency responses in rooms,'' J. Acoust. Soc. Am. **34**, 1819–1823 (1962).
[14] M. Kendall and A. Stuart, *The Advanced Theory of Statistics*, 4th ed. (Griffin Press, London, 1976), p. 246.

Measurement of in-duct acoustic properties by using a single microphone with fixed position

Y. S. Choy and Lixi Huang[a)]
Department of Mechanical Engineering, The Hong Kong Polytechnic University, Hunghom, Kowloon, Hong Kong

(Received 13 April 2004; revised 12 August 2004; accepted 10 September 2004)

Acoustic properties of sound absorption materials and other acoustic structures can be measured in an impedance tube using the well-established two-microphone method to resolve the two traveling wave components of a standing wave pattern. The accuracy of such measurements depends crucially on the calibration of the two microphones placed in close proximity. To eliminate such calibration, the one-microphone method [Chu, J. Acoust. Soc. Am. **80**, 555–560 (1986)] uses the same microphone to probe at two positions sequentially using the voltage driving the loudspeaker as a reference signal. A variant of this method is introduced in this study in which the microphone is fixed at one position while a rigid end plate moves between two positions to resolve the standing wave. The sound source is installed as a side branch, and its driving signal is also used as a reference in the two-step measurement. Close agreement is found with the established two-microphone method, and factors which might affect the accuracy of the new technique are discussed. As a demonstration of the robustness of the method, a low-budget electret microphone is used and the result also matches well with those obtained by the two-microphone method with high-quality condenser type microphones. [DOI: 10.1121/1.1811476]

PACS numbers: 43.55.Ev, 43.85.Bh [DKW] Pages: 3498–3504

I. INTRODUCTION

There are four methods to determine the in-duct acoustic properties, such as the reflection coefficient and absorption coefficient. They are the standing wave method using a probe microphone, multi-point, two-microphone, and one-microphone methods. Their strengths and weaknesses have been discussed by Jones and Parrott (1989).

The so-called standing wave method (SWR) (Lippert, 1953) uses a probe microphone to measure the ratio of the successive maxima and minima of a standing wave pattern in order to find out the acoustic impedance, from which the reflection and absorption coefficients are deduced. This is a tedious and time-consuming process. The multi-point method uses one microphone to take measurements at multiple points, and the least-square method is used to curve-fit the measured pattern and deduce the acoustic properties. Two to six pressure measurements per half-wavelength are required (Jones and Parrott, 1989). Again, the measurement points can be varied by using a traversing probe microphone. The main weakness of both SWR and multi-point methods is the requirement of the hardware movement of microphone, which is time consuming. To overcome this problem, the two-microphone method was introduced by Seybert and Ross (1977). Two microphones are placed at two different positions with a certain separation distance. The signals from the two microphones are acquired at the same time in order to find out the transfer function between them, from which the forward and backward waves in a standing wave pattern are resolved. Random excitation is used to find out the auto-spectral and cross-spectral densities of the two locations so that the whole spectrum of acoustic properties can be found at once. This method was further developed by Chung and Blaser analytically and experimentally (1980a, b). Their method involved the decomposition of the waveform into the incident and reflected components using a simple transfer function of the pressures at two locations. The excitation signal can be random or harmonic. The comparison between the transfer function method and SWR with different selected points was investigated in detail by Chu (1988). The absorption and impedance found by the two-microphone method with varied microphone separation distance is better than those with fixed separation and SWR method. A similar approach for different distances between two measurement points was also taken by Fahy (1984). Chu (1986a) also extended the transfer function studies by adding the effect of tube attenuation.

Compared with the SWR and multiple-point methods, the two-microphone method saves time but requires the knowledge of accurate amplitude and phase relationships between the two microphones, so the calibration is necessary before taking a measurement. The calibration is carried out by swapping the pair of microphones which are flush mounted on the duct with absorptive termination according to ISO 10534-2 (1998). To eliminate the calibration, Chu (1986b) introduced the so-called one-microphone method in which a single microphone is used to measure sound at the two measurement points of an otherwise two-microphone rig. In this sense, the methodology is similar to the two-microphone method. Signals taken from the single microphone at different positions are referenced to the analog signal which drives the loudspeaker. The result is good compared with the SWR method although there are some

[a)]Author to whom correspondence should be addressed. Electronic mail: mmlhuang@polyu.edu.hk

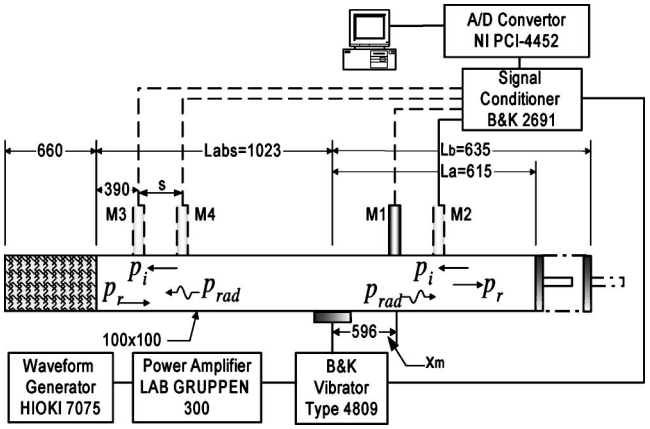


FIG. 1. Experimental setup for the new one-microphone method with provision for the two-microphone method employed for validation purpose.

deviations at low frequencies (<200 Hz). The one-microphone method can eliminate the error of phase mismatch between two microphones. In both two- and one-microphone methods, the two measurements become one when the separation distance between the two measurement positions is equal to an integer multiple of the half-wavelength. In other words, there are frequency blind spots in these methods. To overcome this problem, a third measurement position might be needed. In this study, an alternative configuration for the single microphone method is introduced. In the new design, the loudspeaker is flush mounted to the duct wall as a side branch, and only one fixed microphone insertion hole is provided. A rigid end plate provides the rigid wall boundary and the design is such that the position of the plate can be adjusted easily without the provision of any extra microphone insertion hole. Measurements taken with two positions for the rigid end plate resolve the two traveling wave components. As is shown below, the present one-microphone method is based on an acoustic arrangement different from all previous impedance tube rigs, and the new rig is also expected to bring some convenience in terms of implementation. In what follows, expressions are derived for the reflection and absorption coefficients of the acoustic specimen as a function of the readings of the two-step measurement. Experimental validation is carried out by the two-microphone method. Factors affecting the accuracy of the new single microphone method are discussed towards the end.

II. THEORY

The theoretical model is shown in Fig. 1, which also serves as an illustration of the experimental rig to be described in the next section. The geometry resembles a standard impedance tube. The specimen to be tested is put at one end of the tube at a distance L_{abs} from the center of the flush mounted piston ($x=0$) driven by a vibrator. A movable rigid plate is installed on the right-hand side at $x=L_a$, L_b during two measurements, respectively. The measurement microphone (labeled M1) is located at $x=x_m$, while three other microphones illustrated by broken lines indicate the use of three extra microphones for the validation purpose in this study.

The acoustic field in the duct can be described by the superposition of acoustic radiation from the piston, indicated by p_{rad}, and the standing wave pattern formed by the reflecting ends of the duct, labeled as p_i and p_r for the left- and right-traveling components, respectively. Note that a near field exists around the piston before the radiated waves evolve into plane waves when the frequency is below the cut-on frequency of the duct. In what follows, p_{rad} denotes the complex amplitude of the plane traveling wave and the near field is ignored in the following derivation where measurement points are sufficiently far away from the piston center. Note also that p_i and p_r are the constant amplitudes for the traveling waves throughout the duct length. The pressure at $x=x_m$ (microphone 1) is written as follows:

$$p_1=(p_{rad}+p_r)e^{-ikx_m}+p_ie^{ikx_m}. \tag{1}$$

Note that the time dependence of $\exp(i\omega t)$ is left out of all formulations, and all pressure amplitudes are divided by the complex amplitude of the voltage signal which drives the vibrator–piston assembly. In other words, the pressures measured can be regarded as a result of vibration of a unit amplitude for all measurements.

The normal particle velocity vanishes at the acoustically rigid plate at $x=L_a$, so the relationship between the wave components is

$$p_i=(p_{rad}+p_r)e^{-2ikL_a} \quad \text{or} \quad p_r=p_ie^{2ikL_a}-p_{rad}. \tag{2}$$

Combining Eqs. (1) and (2),

$$p_i=\frac{p_1}{e^{ikx_m}+e^{ik(2L_a-x_m)}}. \tag{3}$$

When the rigid plate is moved to $x=L_b$, the wave components p_i and p_r are changed but the radiation pressure p_{rad} remains unchanged and independent of the reflections at the tube ends. It is implicitly assumed here that the change in the input mechanical impedance from the air in the duct on the piston is negligible when the rigid end plate moves from one position to the next during the two-step measurement. This assumption is validated in the experiment, and the small variation of piston response is analyzed in the next section. The reflection wave components for the rigid end plate position at $x=L_b$ are denoted by p_i' and p_r'. Similar to Eq. (3),

$$p_i'=\frac{p_1'}{e^{ikx_m}+e^{ik(2L_b-x_m)}}. \tag{4}$$

In addition to these equations, the impedance and reflection coefficient at the interface of the absorption material at $x=-L_{abs}$ is assumed to be the same for the two measurements. The reflection coefficient is written as

$$\frac{R}{I}=\frac{p_r}{p_{rad}+p_i}e^{2ikL_{abs}}=\frac{p_r'}{p_{rad}+p_i'}e^{2ikL_{abs}}, \tag{5}$$

where I is the complex amplitude of the total traveling wave incident on the specimen, and R is the complex amplitude of the reflected wave, cf. Fig. 1.

Solving these two equations for p_{rad} and R/I,

$$\frac{R}{I}=\left(\frac{p_r-p_r'}{p_i-p_i'}\right)e^{2ikL_{abs}}. \tag{6}$$

Within Eq. (6), p_i and p_i' can be found from Eqs. (3) and (4), while p_r and p_r' are found from the second equation in (2) when the rigid plate is placed at $x=L_a$ and $x=L_b$, respectively. Hence,

$$p_r-p_r'=p_ie^{2ikL_a}-p_i'e^{2ikL_b}.$$

Substitution of these results into Eq. (6) gives the complex reflection coefficient R/I as well as the energy coefficients of reflection (β) and absorption (α) defined as follows:

$$\frac{R}{I}=\frac{(p_1/p_1')\cos[k(L_b-x_m)]e^{2ikL_a}-e^{ik(L_a+L_b)}\cos[k(L_a-x_m)]}{(p_1/p_1')\cos[k(L_b-x_m)]-e^{ik(L_a-L_b)}\cos[k(L_a-x_m)]}e^{2ikL_{abs}},\quad \beta=\left|\frac{R}{I}\right|^2,\quad \alpha=1-\beta. \tag{7}$$

When the distance between the measurement microphone and the rigid end plate is a quarter wavelength, say for the second measurement, $k(L_b-x_m)=\pi/2$, the pressure at the measuring point vanishes, $p_1'=0$, and Eq. (7) becomes the type of 0/0. This is similar to what happens in the two-microphone measurement when the distance between the two microphones is a multiple of a half-wavelength. Theoretically, the problem can be resolved by taking the ratio of the derivative with respect to k for $\cos[k(L_b-x_m)]/p_1' \rightarrow -1/(\partial p_1'/\partial k)$, which requires the use of data for the current as well as the neighboring frequencies. Having said that, one must be aware that such a derivative may not be numerically reliable in an experiment where broadband excitation is used, leading to small ripples in the measured data. If, however, harmonic excitation is used, one can always skip this particular frequency during the test. If the result at the particular frequency is definitely desired, one can always move the rigid end plate to a slightly different position. Such a move would be easier to implement in the current device than drilling another microphone insertion hole in the two-microphone method. In the present experimental study, this singularity is not encountered as the operational frequency is chosen to be below 1000 Hz while the quarter-wavelength frequency for $L_b-x_m=40$ mm is 2144 Hz.

III. EXPERIMENTAL VALIDATION

Figure 1 also shows the schema for the experimental rig with lengths labeled in mm. The duct cross section was 100 $\times$100 mm^2. The duct wall was made of 15-mm-thick acrylic, which is believed to be acoustically rigid. The first cut-on frequency of the duct was 1700 Hz. For the purpose of cross-checking, two pairs of B&K $\frac{1}{2}$-in. condenser-type microphones (type number 4189) were installed flush with the duct walls. Microphone 1 (M1) was regarded as the measurement microphone in the present one-microphone method. The additional microphone (M2) was used to couple with M1 to conduct the two-microphone measurement for the reflection coefficient at the surface of the movable rigid plate. Two more microphones (M3,M4) were installed at the left-hand side to measure the acoustic properties of the specimen at $x=-L_{abs}$ for comparison with the results of the one-microphone method. Note that only one microphone, M1, is needed when the device is used in practice. A wide separation distance of 80 mm was used for the microphone pairs in order to have a good measurement accuracy at lower frequencies. The distance between the rigid plate and M1 was 20 mm in the first measurement and 40 mm in the second measurement when the rigid plate was moved by $L_b-L_a=20$ mm. The movement was increased to 40 mm for frequencies lower than 200 Hz. The movable plate was tightened after the adjustment of its location. The microphones were supported by B&K's Nexus four-channel conditioning amplifier (type 2691), and the signals were acquired through the National Instruments' AD conversion card type PCI-4452. The voltage signal fed to the vibrator (B&K type 4809), which drove the flush-mounted piston, was also captured together with the microphone signals. A harmonic analog signal was generated by a function generator (Hioki 7050), amplified by B&K's power amplifier (LAB Gruppen 300) before it was fed to the vibrator. The frequencies ranged from 100 to 1000 Hz with an interval of 20 Hz. The advantage of the pure tone tests is its better signal-to-noise ratio than broadband excitation, which is good for the comparison between the two-microphone method and the present method. The piston had a surface area of 50$\times$50 mm^2, flush mounted in the middle of the duct. In order to allow free oscillation of the piston, there was a 2-mm gap between the edges of the duct and the piston. To avoid noise leaking through the gap, a 0.6-mm-thick flexible rubber was used to seal the gap. Fiberglass of density 40 kg/m^3 and flow resistivity of about 9000 kg/m^{3}s were used as the acoustic specimen to be tested in the left-hand side tube. Tests were conducted with the test section filled up by fiberglass to three different depths: 60, 185, and 660 mm. Since the duct was not very short, the acoustic attenuation had to be taken into account when deducing the reflection coefficient. This is discussed in Sec. IV.

Figure 2 shows the comparison of the absorption coefficient α of the glass fiber measured by the present one-microphone method (open circles) and the two-microphone

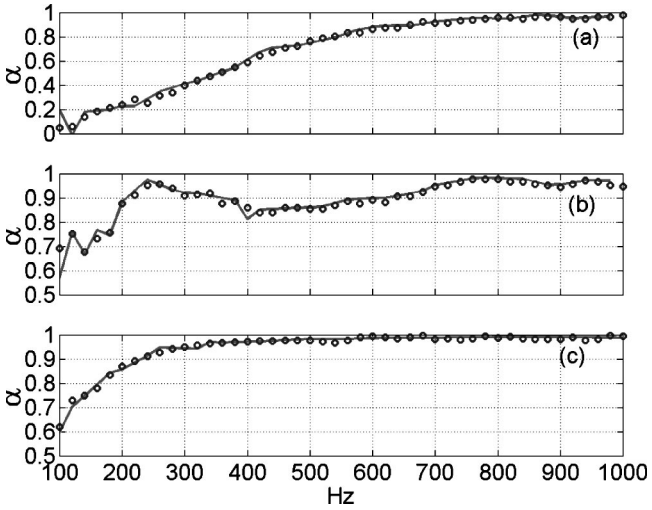


FIG. 2. Comparison of the absorption coefficient between the present method (○) and the two-microphone method (———). (a) is for the absorption material of 60 mm in depth, (b) 185 mm, and (c) 660 mm.

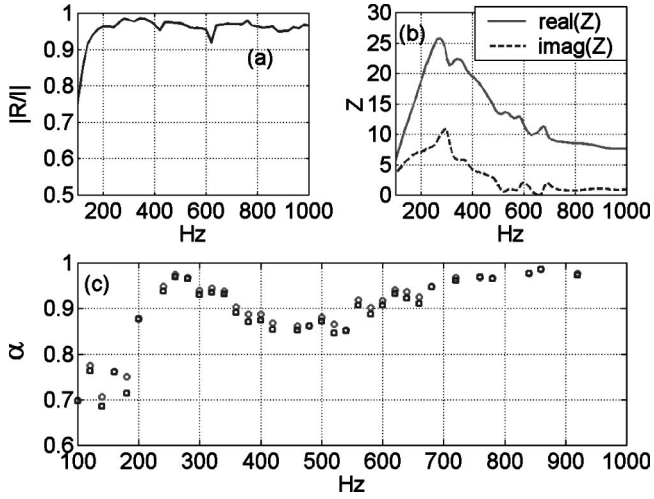


FIG. 3. Effect of the rigidity of the movable plate. (a) is the reflection coefficient at the surface of the movable plate measured by the two-microphone method. (b) is the real (———) and imaginary (– –) parts of the normalized impedance at the surface of the movable plate. (c) is the comparison of the absorption coefficient calculated by Eq. (7) (○) and Eq. (8) (□).

method (solid line). Figures 2(a)–(c) are the results for the fiberglass filling of depth 60, 85, and 660 mm, respectively. The agreement between the two methods for all three tests is good except a few points below about 150 Hz. Above 150 Hz, the mean deviations between the two methods are 1.4%, 1.1%, and 0.6% for the three tests, respectively. Similar results can be obtained by using random signal excitation to the piston.

IV. MEASUREMENT ACCURACY

This section is devoted to the investigation of the effects of various special factors which might affect the accuracy of the current measurement technique. These factors include the vibration of the so-called rigid plate, the effect of the end plate position on the piston response, noise leaking in the piston gap, and other noise attenuation mechanisms present in the system. The attenuation would be significant if the distance between the measurement microphone M1 and the acoustic specimen is very large. For the errors caused by the finite size of the microphone sensing membrane, calibration error of phase mismatch between two microphones, spectral analysis for signals with noisy environment, etc., the readers are referred to the earlier works of Seybert and Soenarko (1981) as well as Bodén and Åborn (1986).

A. Rigidity of the movable plate

If the rigid plate is not entirely rigid, sound reflection on its surface would be less than 100%. The extent to which the rigid plate assumption was correct for the current rig was assessed by the reflection coefficient by the two-microphone method using microphones M3 and M4 shown in Fig. 1. The ratio of linear amplitudes, $|R/I|$, is shown in Fig. 3(a). On average, $|R/I|$ was about 0.96. However, below 100 Hz, it could be lower than 0.9. The impedance of the so-called rigid plate is normalized by that of air, $Z=p'/(\rho_0 c_0 u)$, and the result is shown in Fig. 3(b). When the finite impedance is used for the end plate, the absorption coefficient for the acoustic specimen is recalculated as

$$\alpha = 1 - \left| \frac{\left[\frac{Z+1}{Z-1}\right]\left[\frac{p_1}{p_1'}\right]\left[\frac{(Z-1)e^{ikx_m}+(Z+1)e^{i(2kL_b-x_m)}}{(Z-1)e^{ikx_m}+(Z+1)e^{i(2kL_a-x_m)}}\right]e^{2ikL_a}-e^{2ikL_b}}{\left[\frac{p_1}{p_1'}\right]\left[\frac{(Z-1)e^{ikx_m}+(Z+1)e^{i(2kL_b-x_m)}}{(Z-1)e^{ikx_m}+(Z+1)e^{i(2kL_a-x_m)}}\right]-1} \right|^2 . \quad (8)$$

Figure 3(c) compares α deduced from the rigid-plate assumption (open circles) and that from the above formula (open squares). It is found that the correction of the end plate vibration amounts to about 1.5% in terms of α except for 180 Hz where a deviation of 4.0% is found. The conclusion is that the end plate vibration is not significant.

B. Effect of sound attenuation in the duct system

A long duct of over 1.6 m is used in the present setup in Fig. 1. This may cause excessive sound attenuation due to the frictional loss, heat transfer, etc. There may also be a certain amount of noise leaking through the clearance between the piston and the duct wall. These losses can be cal-

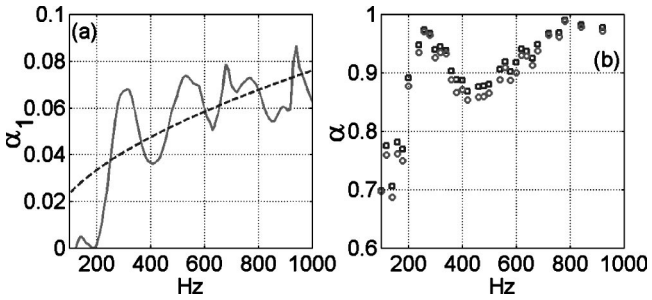


FIG. 4. The effect of sound energy dissipation along a rigid duct. (a) compares the measured energy dissipation coefficient of the straight duct segment in the present rig (———) with prediction of such losses based on a traveling wave theory (Pierce, 1991) (––). (b) is the comparison of the absorption coefficient with (○) and without (□) correction for the sound energy loss along the rigid duct segments.

culated by the net sound energy flux into a control volume which encloses the long duct segment. As shown in Fig. 1, the control volume can be considered to be the duct segment from the center of microphones M3 and M4 to that of M1 and M2. Denote the intensity of sound traveling from the position of microphone *m* towards microphone *n* as I_{mn}, the total sound energy dissipated inside the control volume is found as $\delta I = I_{34} - I_{43} + I_{21} - I_{12}$, and an energy dissipation coefficient is defined as $\alpha_1 = \delta I / \bar{I}$ where $\bar{I}$ is the average of the four intensity magnitudes. To find α_1, the rig shown in Fig. 1 was revised as follows. The test specimen on the left-hand side duct was removed, and a loudspeaker was installed on the left-hand side to serve as the sound source. By using two microphones on the left-hand side and another two on the right-hand side, all the wave components can be found, hence α_1. The vibrator-piston assembly was installed in the middle segment of the duct which had a length of 60 cm. When the middle segment of the duct was removed, the total duct length became 1.0 m between the centers of the two pairs of microphones. The energy dissipation coefficient is shown in Fig. 4(a) as the solid line. The energy dissipation in this case is expected to come mainly from the friction between sound and the duct walls, for which the analytical prediction (see Pierce, 1991, Eq. 10-5.8) for a single traveling wave is also shown in Fig. 4(a) as a dashed line. The two lines are close to each other in general, validating the general assumption for the dissipation mechanism. The wavy behavior of the measured dissipation coefficient probably accounts for the standing wave effect. Since such standing wave composition is not known *a priori* in the eventual one-microphone measurement, the analytical prediction of attenuation for a traveling wave is used as a correction factor for the decay of each traveling wave along the duct. The effect of this correction is shown in Fig. 4(b), where the open squares are the results without the correction and open circles are those with attenuation correction. The mean correction is about 1.2% for α. Generally speaking, the correction is not large and such a correction can be further reduced or even ignored if the distance between the measurement microphone and the specimen is minimized. The current test rig had the provision for the two-microphone measurement for the validation purpose, and a reduction of duct length is anticipated when such provision is removed for practical use of the one-microphone method.

C. Piston response

The working principle of the present one-microphone method is based on the assumption that the radiated wave from the piston for a given voltage input, p_{rad}, is not affected by the location of the rigid end plate. Factors that uphold and undermine this assumption are briefly discussed here before a quantitative experimental examination of the assumption is reported. The actual amplitude of the piston vibration is influenced by the input mechanical impedance, say Z_{in}, derived from the fluid loading on the piston inside the duct (cf. Kinsler *et al.*, 2000). This impedance diverges at a duct resonance frequency when the measurement sample is a hard wall and all acoustic attenuations are ignored. Under this extreme condition, the constant p_{rad} assumption is unlikely to be upheld and the current method would fail around such resonance frequency. However, it is anticipated that the actual magnitude of Z_{in} is rather limited when the normal friction is accounted for. It is also limited by the finite dissipative element which normally exists in any specimen to be measured. The variation of Z_{in} at a nonresonant frequency is even more limited. Having said that, a loudspeaker made of paper cone is unlikely to possess sufficient mechanical impedance such that Z_{in} is deemed negligible. The device used in the current test is a B&K vibration exciter type 4809 with a force rating of 45 N. A piston of mass 294 g is attached as the payload. Note that a heavier piston would reduce the vibration amplitude but the system impedance can be further increased, leading to a better satisfaction of the constant p_{rad} assumption. The extent to which the piston vibration was affected by the swapping of the end plate position was investigated quantitatively. The response of the piston was measured by a B&K accelerometer (type 4374 with Nexus conditioning amplifier type 2692) placed on top of the piston. The voltage input to the vibrometer, v_{in}, and the acceleration output from the accelerometer, v_{out}, were sampled simultaneously by the computer. The amplitude ratio of the two signals is shown in Fig. 5 for the range of frequencies from 60 to 1000 Hz. The solid line is the result for the first position of the end plate, $L_a = 615$ mm, and open circles for the second, $L_b = L_a + 40$ mm. The two results are hardly distinguishable in the figure. The average deviation is 0.8%, and the effect of such deviation on the results of one-microphone measurement is found to be negligible.

D. Inaccuracy caused by the calibration error

Whether the deviation between the one- and two-microphone measurements is large or not can be judged by the range of uncertainty involved in the latter method due to microphone calibration. Assuming that the transfer function between the actual sound pressure *p* and the digital reading *A* from a microphone input channel is J_1, J_2 for the two microphones, respectively, and that a standing wave pattern is formed by an incident wave Ie^{-ikx} and a reflected wave Re^{ikx}, where $x = 0$ is located at the center of the two micro-

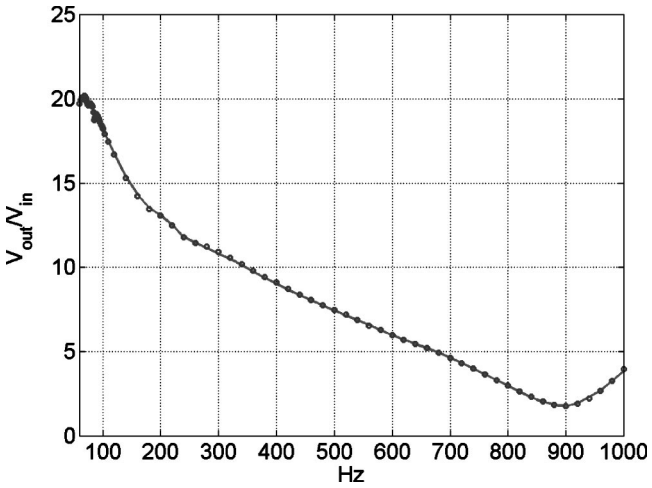


FIG. 5. Comparison of the amplitude ratio of signals from the piston acceleration and the voltage input to the vibrator when the end plate is located at $L_a = 615$ mm (———) and $L_b = L_a + 40$ mm (○).

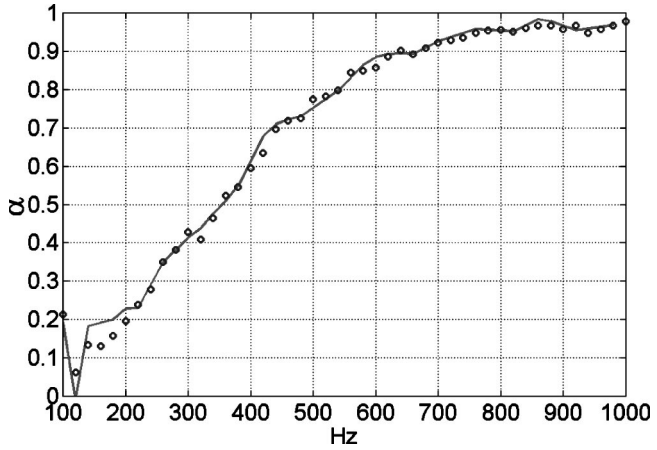


FIG. 7. The absorption coefficient measured by an electret microphone in the present rig (○) compared with the two-microphone measurement using two B&K microphones (———).

phones separated by a distance of s, the wave components I and R are resolved in one session of two-microphone measurement as follows:

$$I = \frac{J_1 A_1 e^{-ik(L_1+s)} - J_2 A_2 e^{-ikL_1}}{e^{-iks} - e^{iks}} \quad \text{and}$$

$$R = \frac{-J_1 A_1 e^{ik(L_1+s)} + A_2 J_2 e^{ikL_1}}{e^{-iks} - e^{iks}}.$$

The linear reflection coefficient R/I is

$$\frac{R}{I} = e^{2ikL_1}\left(\frac{-A_1 e^{iks} + A_2 J}{A_1 e^{-iks} - JA_2}\right), \tag{9}$$

where $J = J_2/J_1$ is the complex ratio of the microphone responses. When the complex ratio J is subject to an uncertainty of δJ, the resultant energy coefficients of reflection and absorption can be found out easily. Figures 6(a) and (b) show the amplitude and phase of calibration factor J, respectively. Two typical sets of data are shown for two calibration exercises, and the difference between the two gives an estimate for the magnitude ratio $|\delta J|$ and phase angle difference $\delta\theta$. Typically, the difference in amplitude was about 0.002 for frequencies greater than 300 Hz. However, much higher deviation was found at lower frequencies. The effect of δJ on the measured absorption coefficient α for the 60-mm-deep fiberglass filling in the present rig is shown in Figs. 6(c) and (d). Figure 6(c) shows the maximum (dashed line) and minimum (solid line) level of the absorption coefficient when $|J|$ changes from 0.99 to 1.01 and phase angle θ changes from -0.005 to 0.015. Figure 6(d) shows the range of absorption coefficient $\delta\alpha$. It is found that $\delta\alpha$ associated with such δJ can be greater than the deviation between the one-microphone and two-microphone measurements as shown in Fig. 2. The uncertainty is also larger than the frictional losses shown in Fig. 4(a). The comparison between the two-microphone method and Chu's (1986b) one-microphone method shows that the deviation for the reflection coefficients at very low frequencies (<250 Hz) is about 0.1, which is slightly poorer than what is achieved here (cf. Fig. 3 of Chu, 1986b).

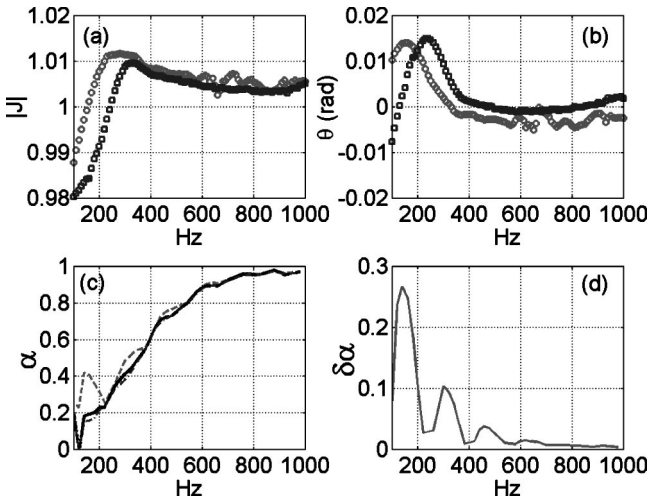


FIG. 6. The uncertainty of the two-microphone method. (a) shows the magnitude of the calibration factor J, and (b) is the phase of J during two calibration trials (□,○). (c) is the maximum (––) and minimum level (———) of absorption coefficient caused by the variation of J. (d) is the uncertainty of the absorption coefficient caused by such variation.

E. Use of an electret microphone

The accuracy of the one-microphone measurement depends on the consistency of the transfer function between the measurement microphone and the sound source, and it eliminates the need of calibration between two microphones. To demonstrate the full advantage of this method, a very low-budget microphone was used instead of B&K's $\frac{1}{2}$-in. condenser-type microphones which have a flat frequency response down to at least 20 Hz. A miniature (0.1-in.) electret microphone produced by Tibbetts Industries (151 series) was tested. It had a flat frequency response from 300 to 3000 Hz, and its response at 100 Hz was 4 dB below that at 300 Hz. Figure 7 shows the comparison of α measured by using such

an electret microphone in the present method with that of the two-microphone method. The medium deviation was about 1.9%, slightly higher than 1.4% from B&K microphones.

V. CONCLUSIONS

A new rig of one-microphone measurement method for in-duct standing waves is proposed and experimentally validated against the well-established two-microphone method. The new rig involves the use of a side-branch sound source and a movable rigid plate which provides the necessary configuration variation to achieve two independent measurements. The main advantage of the rig is that the measurement microphone is installed at its fixed position during the measurements. Results are summarized as follows.

(1) The use of a fixed microphone eliminates the need to drill many microphone insertion holes as required by the two-microphone method and the existing one-microphone method.
(2) The agreement of the sound reflection and absorption coefficients between the present one-microphone method and the two-microphone method is very satisfactory.
(3) The complete flexibility for the rigid plate to vary its position eliminates the blind spots in spectrum where measurements are rendered impossible when the separation distance between the two microphone positions coincides with the half wavelength.
(4) Sound energy losses and other factors of uncertainties are discussed, and the errors are found to be generally acceptable when compared with the uncertainties in the two-microphone measurement arising from the microphone calibration.
(5) A very low budget electret microphone is also found to yield satisfactory results in the present one-microphone method.

ACKNOWLEDGMENTS

The authors thank the Research Grants Council of the Hong Kong SAR Government for its support of the project (PolyU 5169/02E). Very careful checking by an anonymous reviewer also improved the manuscript.

Bodén, H., and Åborn, M. (**1986**). ''Influence of errors on the two-microphone method for measuring acoustic properties in ducts,'' J. Acoust. Soc. Am. **79**, 541–549.
Chu, W. T. (**1986a**). ''Extension of the two-microphone transfer function method for impedance tube measurement,'' J. Acoust. Soc. Am. **80**, 347–348.
Chu, W. T. (**1986b**). ''Transfer function technique for impedance and absorption measurements in an impedance tube using a single microphone,'' J. Acoust. Soc. Am. **80**, 555–560.
Chu, W. T. (**1988**). ''Further experimental studies on the transfer-function technique for impedance tube measurements,'' J. Acoust. Soc. Am. **83**, 2255–2260.
Chung, J. Y., and Blaser, D. A. (**1980a**). ''Transfer function method of measuring in-duct acoustic properties. I. Theory,'' J. Acoust. Soc. Am. **68**, 907–913.
Chung, J. Y., and Blaser, D. A. (**1980b**). ''Transfer function method of measuring in-duct acoustic properties. II. Experiment,'' J. Acoust. Soc. Am. **68**, 914–921.
Fahy, F. J. (**1984**). ''Rapid method for the measurement of sample acoustic impedance in a standing wave tube,'' J. Sound Vib. **97**, 168–170.
ISO10534-2:1998(E) (**1998**). Acoustic-Determination of sound absorption coefficient and impedance in impedance tubes (International Standard Organization).
Jones, M. G., and Parrott, T. L. (**1989**). ''Evaluation of a multi-point for determining acoustic impedance,'' Mech. Syst. Signal Process. **3**, 15–35.
Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**2000**). *Fundamentals of Acoustics* (Wiley, New York).
Lippert, W. K. R. (**1953**). ''The practical representation of standing waves in an acoustic impedance tube,'' Acustica **3**, 153–160.
Pierce, A. D. (**1991**). *Acoustics: An Introduction to its Physical Principles and Applications* (Acoustical Society of America, New York).
Seybert, A. F., and Ross, D. F. (**1977**). ''Experimental determination of acoustic properties using a two-microphone random-excitation technique,'' J. Acoust. Soc. Am. **61**, 1362–1370.
Seybert, A. F., and Soenarko, B. (**1981**). ''Error analysis of spectral estimates with application to the measurement of acoustic parameters using ransom sound fields in ducts,'' J. Acoust. Soc. Am. **69**, 1190–1199.

Investigation of the validity of radiosity for sound-field prediction in cubic rooms

Eva-Marie Nosal[a)]
Department of Mathematics, University of British Columbia, 1984 Mathematics Road, Vancouver, BC, V6T 1Z2 Canada

Murray Hodgson
School of Occupational and Environmental Hygiene and Department of Mechanical Engineering, University of British Columbia, 2206 East Mall, Vancouver, BC, V6T 1Z3 Canada

Ian Ashdown
Heart Consultants Limited, 620 Ballantree Road, West Vancouver, BC, V7S 1W3 Canada

(Received 3 October 2003; revised 2 June 2004; accepted 12 September 2004)

This paper explores acoustical (or time-dependent) radiosity using predictions made in four cubic enclosures. The methods and algorithms used are those presented in a previous paper by the same authors [Nosal, Hodgson, and Ashdown, J. Acoust. Soc. Am. **116**(2), 970–980 (2004)]. First, the algorithm, methods, and conditions for convergence are investigated by comparison of numerous predictions for the four cubic enclosures. Here, variables and parameters used in the predictions are varied to explore the effect of absorption distribution, the necessary conditions for convergence of the numerical solution to the analytical solution, form-factor prediction methods, and the computational requirements. The predictions are also used to investigate the effect of absorption distribution on sound fields in cubic enclosures with diffusely reflecting boundaries. Acoustical radiosity is then compared to predictions made in the four enclosures by a ray-tracing model that can account for diffuse reflection. Comparisons are made of echograms, room-acoustical parameters, and discretized echograms. [DOI: 10.1121/1.1811473]

PACS numbers: 43.55.Ka, 43.55.Br [MK] Pages: 3505–3514

I. INTRODUCTION

In the initial phase[1] of the current research, algorithms and methods were developed for acoustical radiosity for arbitrary polyhedral enclosures, based on a numerical solution to the integral equation. These were programed in MATLAB (Ref. 2) and validated by comparison to analytical solutions for a spherical enclosure.

Further research is required to understand the necessary and sufficient conditions for convergence of the solution found by the algorithms and methods to the analytical solution. The numerical solution has been applied several times in the literature[3–7] but, apart from Schroeder and Hackman's[8] investigations for a two-dimensional enclosure, there has been little or no systematic investigation of the effect of various prediction parameters, such as the space and time resolutions. Questions that need to be addressed include: how many patches and/or elements are needed?; what time resolution is fine enough?; how far (in time) does the prediction need to be made?; what part of this should be exact (without the averaging outlined in Nosal *et al.*[1])?; and, what is the error introduced by the use of configuration factors to approximate form factors[1]? This research explores acoustical radiosity thoroughly for one case—that of a cubic enclosure—and leads to a better understanding of the factors involved in acoustical radiosity predictions.

Similarly, little effort has previously been put into the comparison of acoustical radiosity to other prediction methods. Work by Le Bot and Bocquillet,[9] in which radiosity and ray tracing are compared, focuses only on steady-state sound-pressure levels. Further comparison of the two prediction techniques is another objective of this research. To do so, predictions are made for the four cubic enclosures using RAYCUB, a ray-tracing model that can account for diffuse reflection, and are compared to the predictions made by acoustical radiosity. Acoustical radiosity and ray tracing with diffuse reflection are expected to give identical predictions in limiting cases.[9]

II. PARAMETERS, ECHOGRAMS, AND DISCRETIZED ECHOGRAMS

Listener perception of a sound field is affected by two main attributes of the field. The first is the arrival time and strength of the direct signal and the first-order reflections. Differences between sound fields with respect to these reflections can be seen by comparing echograms—perhaps the most fundamental and detailed method of comparison. This comparison reveals information about the prominent reflections and other fine details about the distribution of energy in time.

The other perceptible attribute of the field is the temporal distribution of energy. It is difficult to get information about this distribution by comparison of echograms, so other

[a)]Current affiliation: School of Ocean and Earth Sciences and Technology, Department of Geology and Geophysics, University of Hawaii at Manoa, 1680 East-West Road POST 813, Honolulu, HI, 96822. Electronic mail: nosal@hawaii.edu

TABLE I. Distribution of absorption in the cubic rooms. In all cases, $\bar{\alpha} = 1/6$.

Cube	Distribution of absorption
1	$\alpha = \frac{1}{6}$ over all walls
2	$\alpha = 1$ on the floor, 0 on the other walls
3	$\alpha = \frac{1}{2}$ on the floor and the ceiling, 0 on the other walls
4	$\alpha = \frac{1}{2}$ on the floor and the front wall, 0 on the other walls

methods of comparison must be used. One method is to compare room-acoustical parameters, most of which are measures of energy distribution (for example, clarity and definition[10] are defined by ratios of early and late energies). Parameters used for comparisons in this paper are sound-pressure level (L_p), center time (TS), early decay time (EDT), reverberation time (RT), clarity (C_{80}), and definition (D_{50}).[10] They have been well correlated with listener perception.[11] Nevertheless, much information about the exact distribution of energy is lost to the parameters, since they are not highly sensitive to subtle changes in energy distribution.

To enable comparison of such changes, ''echogram discretization'' is used in this research as a way to compare total energy levels within small time steps. The idea is very simple: time is discretized, and all energy arriving between points of discretization is summed to give the total energy for the corresponding time step (a similar method of comparison has previously been used by Suh *et al.*[12]). Obviously, the resolution used in the echogram discretization must be coarser than the time resolution used in the radiosity algorithm (or, in the case of measurements, the inverse of the sampling frequency). Note that using a longer time interval in the radiosity algorithm is not equivalent to using a small time interval in radiosity and then using the longer interval in the discretization of the echogram. We have chosen to use discretized echograms with 50-ms time steps. The resolution is somewhat arbitrary (not corresponding to perceptual issues or otherwise), and was chosen since it provided a substantial, but not overwhelming, amount of information. In the cases considered, the use of smaller or larger time steps gave similar results and led to similar conclusions.

III. RADIOSITY PREDICTIONS

A. Preliminary remarks

The radiosity algorithm was run numerous times for four cubic rooms. Each cube had walls 8 m long, and an average absorption coefficient of 1/6. The absorption was distributed as indicated in Table I.

Predictions were made, varying: (1) the number of patches (only a single-level hierarchy was applied); (2) the resolution of the time discretization; (3) the length of time for which exact and approximate predictions are made; and (4) the form-factor prediction method. The effect of the varying distribution of absorption (among the four cubes) was investigated using the results, and comparisons were made with predictions by ray tracing. For each prediction, the source was at the center of the cube and had power of 0.005 W. The receiver was located inside the enclosure 2 m from the floor, the front wall, and a sidewall, respectively (by symmetry of the configurations, which sidewall does not matter). Air absorption was neglected throughout—in the way that it is incorporated, it has no effect on convergence.

B. Patch size

1. Specifications

The numerical solution is based on a discretization of the boundary. This results in discretization error, which should be reduced as the resolution of the meshing is increased. In particular, as the resolution increases, the numerical solution approaches the analytical solution. To minimize computational requirements, it is desirable to minimize the number of patches used in the discretization.

For each cube, the radiosity algorithm was run ten times with an increasing number of patches. The first prediction for each cube had each wall as a single patch, giving a total of six patches. For the second prediction, each wall was divided into four equally sized, rectangular patches, giving 24 patches in total. The third prediction had nine patches on each wall for a total of 54 patches, the fourth had 16 giving 96 in total, and so on (in general, the nth prediction had n^2 patches on each wall and $6n^2$ patches in total).

The predictions were made with a discretization period of $\Delta t = 1/24\,000$ s. This was chosen based on initial predictions that indicated that there is almost no variation in room-acoustical parameter predictions if time is discretized more finely than $\Delta t = 1/18\,000$ s. Form factors were found using the analytical formulas given by Gross *et al.*[13]

In the methods used here,[1] predictions are made using the exact radiosity algorithm for time steps up to t_{exact}. For time steps between t_{exact} and t_{final}, which is the last time step for which predictions are made, radiation densities are estimated using an averaging technique.[1]

t_{exact} and t_{final} were set to 0.8 and 1.0 s, respectively, based on a criterion developed by Nosal.[14] The criterion guarantees that the reverberation time predicted from the truncated impulse response is within ϵ of the reverberation time that would be obtained from the full impulse response, provided that truncation of the full impulse response occurs at a time greater than $t_{\min}(\epsilon, \text{RT})$, where

$$t_{\min}(\epsilon, \text{RT}) = \frac{\text{RT}}{6\ln(10)} \times \ln\left[1 + \frac{(10^{-3}e^{[3\ln(10)\epsilon]/\text{RT}}) - 1}{10^{-35/10} - 10^{-5/10}(10^{-3}e^{[3\ln(10)\epsilon]/\text{RT}})}\right], \quad (1)$$

where RT is the reverberation time for the full response.

To apply the criterion, exponential decay was assumed. Miles[6] showed that, in general, the decay curves predicted using the integral equation are strictly exponential after some time. Consequently, even though the first part of the sound decay is not exponential, the assumption of exponential decay along the entire curve may not be unreasonable in this application. Then, according to Eying's classical formula for reverberation time[10]

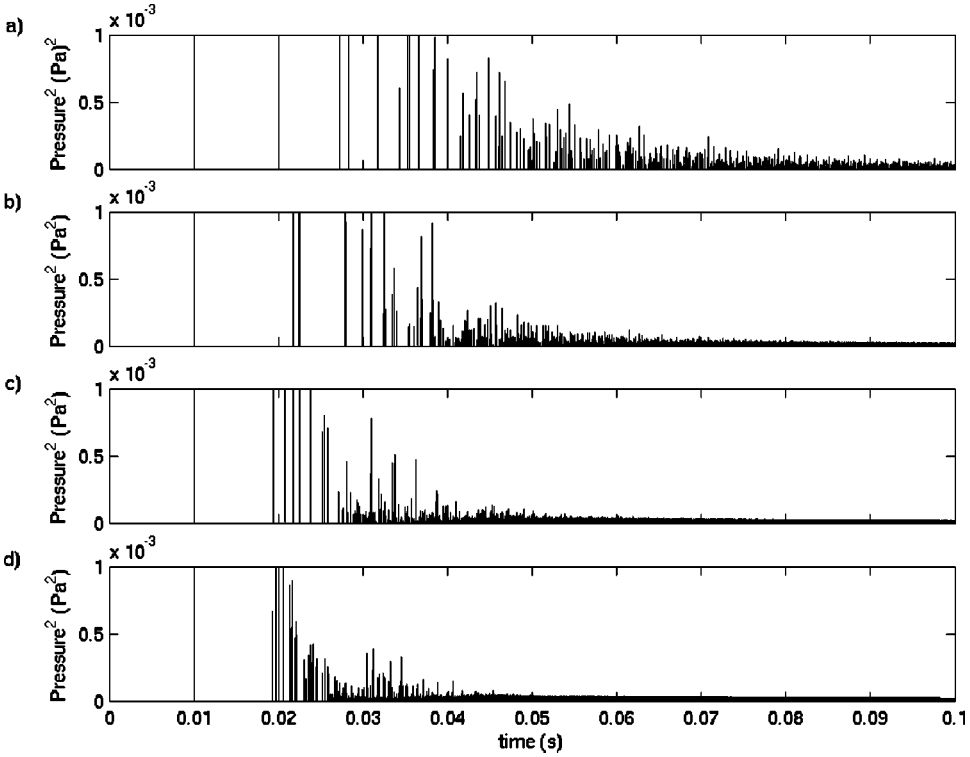


FIG. 1. Echograms for cube 1 from radiosity predictions with (a) 24; (b) 54; (c) 150; and (d) 600 patches ($\Delta t = 1/24\,000$, $t_{\text{exact}} = 0.8$ s, and $t_{\text{final}} = 1.0$ s).

$$\text{RT} = \frac{24V\ln(10)}{c(4mV - S\ln(1-\overline{\alpha}))} \approx 0.161\frac{V}{4mV - S\ln(1-\overline{\alpha})}. \quad (2)$$

The RTs in the cubic rooms should be 1.1782 s [in Eq. (2), V, S, and $\overline{\alpha}$ are room volume, surface area, and average absorption coefficient, respectively, c is the speed of sound, and m is the air absorption coefficient]. Although this reverberation time may not be perfectly accurate (due to the limiting diffuse-field assumption made by Eyring[15]), it can be taken as an initial estimate for use in Eq. (1). If it is desired to predict RT to within 0.01 s of the full impulse response RT (which represents very high accuracy) then, according to this equation, the required t_{final} is given by

$$t_{\text{final}} \geqslant t_{\min}(0.01, 1.1782) = 0.9264 \text{ s}. \quad (3)$$

$t_{\text{final}} = 1.0$ was chosen correspondingly, and t_{exact} was chosen to be quite close to t_{final} (since the effects of varying t_{exact} were unknown).

2. Convergence

Figure 1 shows the echograms for cube 1 obtained from predictions with 24, 54, 150, and 600 patches, respectively. As the meshing is refined, the energy arriving at the receiver is spread out in time, with fewer distinct reflections. Indeed, the echogram for 600 patches has none of the prominent, large peaks that the echogram for 24 patches has (apart from the direct signal). An explanation for this observation is that, with a finer meshing, the energy leaving a patch is divided into more parts to propagate to the other patches, so that large units of energy are quickly dissipated. This is what is expected in a room with diffusely reflecting boundaries.

In a real room, with partially specularly reflecting boundaries, some distinct reflections would be expected, suggesting that fewer patches may yield echograms that look more similar to measured echograms. However, this should not be interpreted as meaning that predictions with fewer patches yield more realistic predictions. The arrival times and amplitudes of the distinct reflections in such predictions are unrealistic and do not correspond to times and amplitudes that would result from specular reflection.

Figure 2 shows a discretized echogram for cube 1 with time intervals of 0.05 s. For each time interval, the pressure-squared sums for predictions with 6, 54, 150, and 600 patches, respectively, are plotted. From the graphs, it is evident that more patches tend to result in more energy in the very early time steps, with levels becoming more uniform at about 0.15 s. This trend reverses itself from about 0.2 to 0.35 s, after which energy levels are very similar. The most noticeable changes in the echograms and discretized echograms occur for predictions with less than 150 patches. For more than 150 patches, there is less change, with predictions for 150 patches similar to those for 600 patches. This indicates that 150 patches are sufficient for accurate predictions.

In Table II, the predicted values of the various parameters are given for varying numbers of patches. These values are for cube 1; the trends are very similar for all four cubes. Each of the parameters considered (L_p, C_{80}, D_{50}, EDT, RT, and TS) converges to a finite value as the number of patches increases. The percent difference between consecutive predictions decreases with increasing number of patches [here, consecutive means between the nth and the $(n+1)$th prediction, with $6n^2$ and $6(n+1)^2$ patches, respectively]. This suggests convergence by the ratio test.[16] The validation of the numerical solution for the sphere[1] provides confidence that the numerical solution does indeed converge (with increasing number of patches) to the analytical solution to the integral equation.

C_{80} is the slowest to converge of all the parameters.

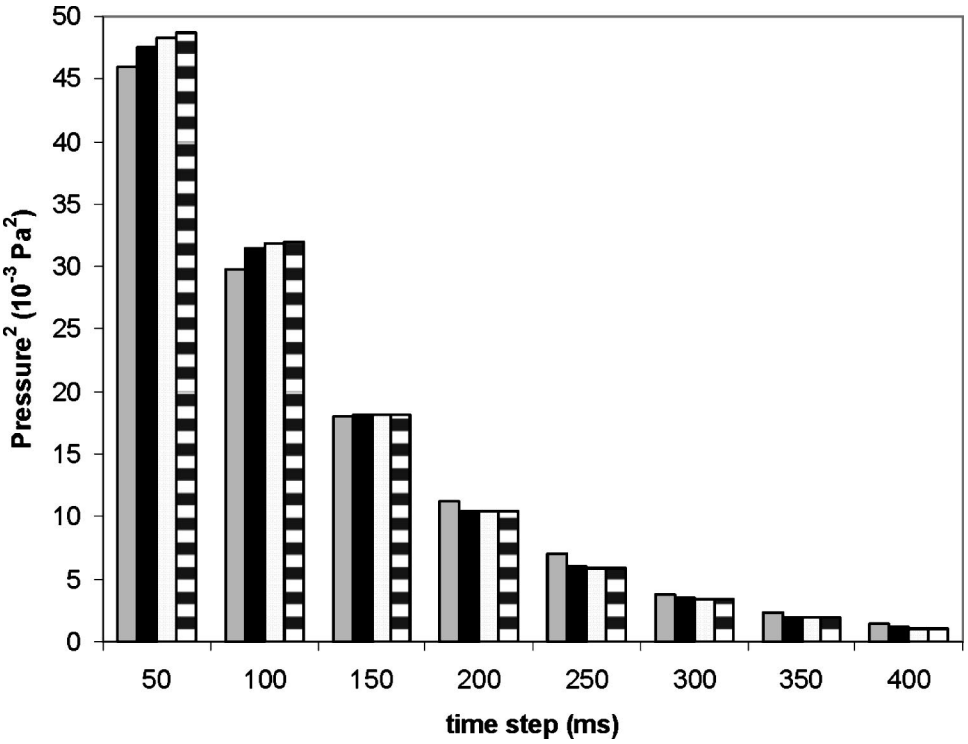


FIG. 2. Discretized echogram for cube 1 with 6 patches; 54 patches; 150 patches; 600 patches ($\Delta t = 1/24\,000$, $t_{\text{exact}} = 0.8$ s, and $t_{\text{final}} = 1.0$ s).

However, even in the case where it is slowest to converge (cube 1), the differences between the fifth and sixth predictions of C_{80} are less than 1%. In fact, all of the other parameters (in all cubes) have differences less than 1% between the third and fourth predictions. Indeed, a fine subdivision of the enclosure is not necessary in these trials, and 150 patches are sufficient for good predictions, as in the echograms. Worth noting is that the differences between the L_p's for the first and second prediction is less than 1% (in all cubes). This suggests that (in a cubic room) radiosity can be used with a very coarse subdivision (one patch per wall in this case) to predict steady-state sound-pressure levels in rooms with diffusely reflecting boundaries, even with nonuniformly distributed absorption.

It is important to realize that required patch size does not depend on the frequency of the sound that is being considered. This is due to the fact that phase is not accounted for in radiosity. Indeed, the integral equation is the same for any frequency, except for the values of absorption coefficients and air absorption exponents; consequently, convergence of the numerical solution to the analytical solution is not frequency dependent (apart from changes in convergence with varying absorption). Convergence of the numerical solution to the analytical solution is not to be confused with convergence to the true solution (the solution that would be obtained in a real room). It is possible that the analytical solution to the integral equation (which is the converged numerical solution) is closer to the true solution at higher frequencies, because phase effects are less significant at these frequencies. The phase independence of radiosity might also help to explain the adequacy of a coarse subdivision of the enclosure surfaces. Still, it is important to keep the relationship between patch size and the wavelength of sound being considered in mind when working with acoustical radiosity.

TABLE II. Parameter predictions with varying number of patches for cube 1 ($\Delta t = 1/24\,000$, $t_{\text{exact}} = 0.8$ s, and $t_{\text{final}} = 1.0$ s).

Number of patches	L_p (dB)	C_{80} (dB)	D_{50} (%)	EDT (s)	RT (s)	TS (ms)
6	84.83	1.60	43.29	1.35	1.42	102.93
24	84.82	1.85	45.04	1.27	1.23	97.16
54	84.83	1.92	45.43	1.24	1.21	95.95
96	84.85	2.97	45.74	1.23	1.20	95.02
150	84.86	2.00	45.87	1.23	1.20	94.67
216	84.86	2.02	46.98	1.22	1.19	94.42
294	84.86	2.04	46.04	1.22	1.19	94.28
384	84.86	2.05	46.14	1.22	1.19	94.17
486	84.87	2.06	46.18	1.22	1.19	94.11
600	84.87	2.06	46.20	1.22	1.19	94.07

3. Absorption distribution

It is of interest to use predictions made for the four different cubes to investigate the effect of different distributions of absorption. As mentioned above, the average absorption coefficient in each cube was 1/6. Refer to Table III for the predicted parameters (from the run with 150 patches).

RT is longest in cube 1, with uniform absorption, and shortest in cube 2, which has absorption on the floor. The

TABLE III. Parameter predictions for the four cubic rooms (150 patches, $\Delta t = 1/24\,000$, $t_{\text{exact}} = 0.8$ s, and $t_{\text{final}} = 1.0$ s).

Cube	L_p (dB)	C_{80} (dB)	D_{50} (%)	EDT (s)	RT (s)	TS (ms)
1	84.86	2.00	45.87	1.23	1.20	94.67
2	84.99	2.85	49.88	1.08	1.06	84.5
3	84.63	2.39	47.83	1.17	1.16	89.7
4	84.67	2.35	47.58	1.17	1.16	90.2

13% difference in RT in these cubes indicates that absorption distribution has a significant effect on RT. Following a line of reasoning set forward by Schroeder and Hackman,[8] cube 2 has the lowest RT because the absorbent surface ''sees'' only reflective surfaces. In the other cubes, the absorbing surfaces see surfaces that absorb some of the sound, so they receive lower sound energy than they would from reflecting surfaces. In cube 1, no surfaces are fully reflecting, so that absorbing surfaces see only other absorbing surfaces, resulting in a lower decay rate and higher RT. Since the absorbing surfaces in cubes 3 and 4 see both reflecting and absorbing surfaces, the RTs are between those in cube 1 and cube 4.

L_p's are very similar (within 0.4 dB) for all four enclosures. Predicted values of C_{80} and D_{50}, which are lower for cube 1 and higher for cube 4, are consistent with the predictions of RT. For higher RTs, lower C_{80} and D_{50} values are expected, because a slower decay rate means more late energy, giving lower early-to-late energy ratios. Similarly, higher and lower TSs for cubes 1 and 4, respectively, are consistent with the RT results.

4. Computational efficiency

Experimental data suggest that radiosity, in a single-level hierarchy, is associated with a computational time complexity of order n^2, where n is the number of patches. This can be explained theoretically as follows. At each time step, and for each element, $(n-1)$ elements must be considered in the calculations. Since there are n elements in total, this gives $n\cdot(n-1)$ calculations for each time step, giving a time complexity of order n^2. For a two-level hierarchy with n elements and m patches, we expect a complexity of order $m\cdot n$.

Also suggested by experimental data is a memory complexity of order n; the memory requirement increases linearly with the number of patches. Once again, this makes sense theoretically because, for each element, radiation densities must be computed for all time steps, so that there are always $n\cdot$(number of time steps) values to be stored.

Refer to Fig. 3 for plots of elapsed time versus number of patches, and for memory requirement versus number of patches. The run times and memory requirements are for predictions for cube 2 (and are similar for the other cubes). A Pentium IV 2.2-GHz PC was used for these predictions.

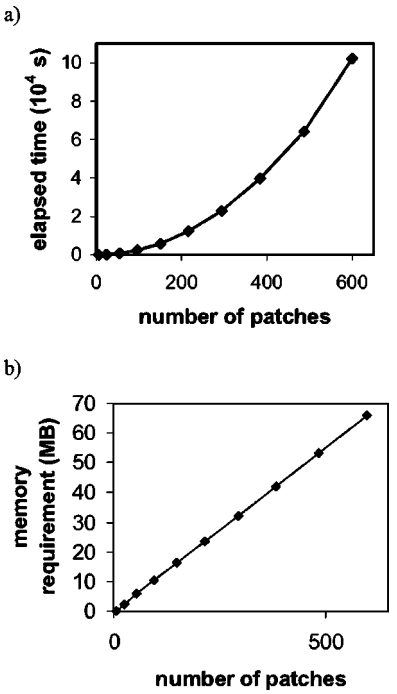


FIG. 3. Variation of (a) elapsed time; and (b) memory requirement with number of patches for cube 2 ($\Delta t = 1/24\,000$, $t_{\text{exact}} = 0.8$ s, and $t_{\text{final}} = 1.0$ s).

C. Time resolution

First, it should be recognized that the nature of time resolution in the radiosity algorithm is quite different from sampling periods used in digital systems and in experimentation. The radiosity algorithm does not simply sample the response every Δt seconds, where Δt is the discretization period. Rather, energy within a time interval is summed to give the energy level for the corresponding time step. Also, phase is neglected in radiosity, so signal frequency does not affect the algorithm (as discussed above). Previous authors have referred to Nyquist's and Shannon's theorems, claiming that, by these theorems, a discretization frequency $(1/\Delta t)$ of at least twice the maximum signal frequency is sufficient. Because it is not clear if the application of these theorems is appropriate, they are not used here.

On the other hand, the Courant number criterion is applicable here. The Courant number[17] can be expressed as $c\Delta t/\Delta x_{\min}$, where $\Delta x_{\min}$ is the minimum distance between patches. To satisfy the Courant number criterion, this number must be less than 1; if the Courant number is too large, then the time resolution must be refined to satisfy the criterion. This ensures that the distance traveled in one time step is not more than the distance between patches. In all investigations performed in this research, the Courant number criterion is satisfied by default (time resolution is fine enough).

To investigate the effect of resolution, further predictions were made for all cubes, with discretization periods of $\Delta t = 1/16\,000$, 1/12 000, 1/8000/, 1/4000, 1/2000, and 1/1000 s (recall that the initial predictions had $\Delta t = 1/24\,000$). For these predictions, 600 patches were used to mesh the cubes. Once again, t_{exact} and t_{final} were set to 0.8 and 1.0 s, respectively.

Several echograms for cube 1 are given in Fig. 4. Echograms for the different resolutions look quite similar. Already for $\Delta t = 1/2000$, the distinct early signals are in the same place, and of the same amplitude, as the ones for $\Delta t = 1/24\,000$. The overall amplitudes of the signals in the later parts of the echogram are lower for a finer time resolution, as expected—smaller time steps result in more time steps, each with less energy. From this, it might be concluded that discretization periods of $\Delta t = 1/4000$ are sufficient, and that already at $\Delta t = 1/2000$, reasonable results are obtained.

Discretized echograms for the different time resolutions are almost identical. There tends to be slightly more energy in the early time steps for finer resolutions. This tendency

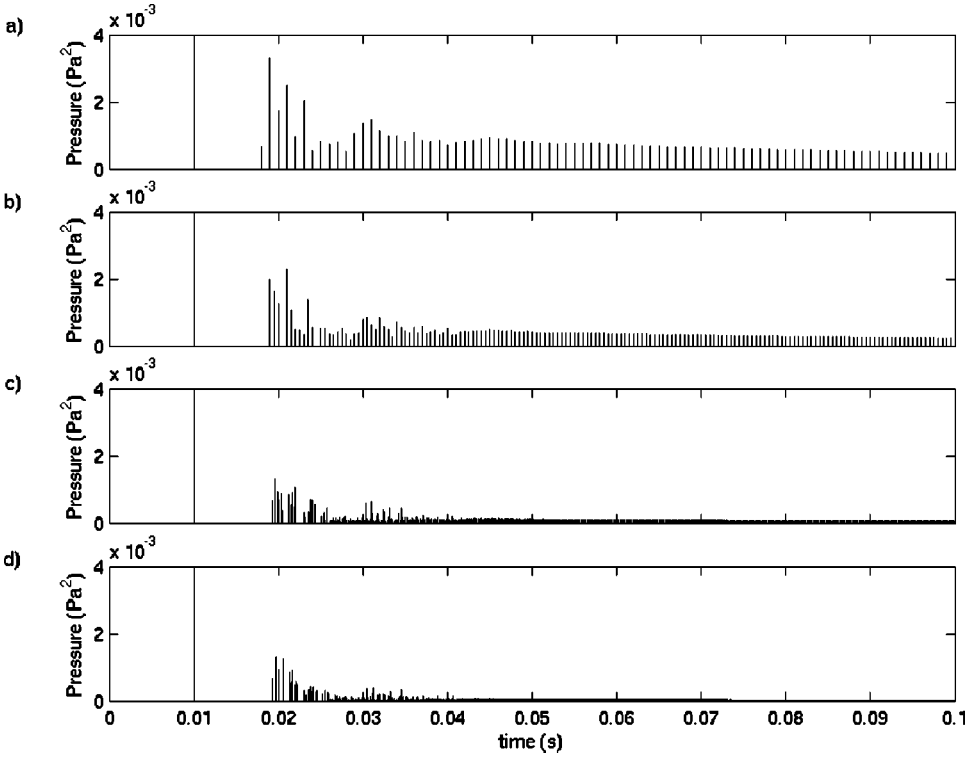


FIG. 4. Echograms for cube 1 with four time resolutions Δt: (a) $\Delta t = 1/1000$ s; (b) $\Delta t = 1/2000$ s; (c) $\Delta t = 1/8000$ s; (d) $\Delta t = 1/24\,000$ s (600 patches, $t_{\text{exact}} = 0.8$ s, and $t_{\text{final}} = 1.0$ s, range reduced to cut off the direct signal, which is the same for all predictions).

reverses for later time steps (after about 0.2 s). Interestingly, this is a similar trend to that for varying patch discretization. Nevertheless, the differences for the varying time resolutions are, for all practical purposes, negligible.

Values for the various parameters versus $1/\Delta t$ are given in Table IV. The values are for cube 4, which has the most significant differences in parameters for different resolutions. Trends for the other cubes are very similar. As is the case for varying patch sizes, the parameter predictions converge with finer resolution. L_p is unaffected by discretization period. This is expected, because different time resolutions do not alter the total energy in the system; they only distribute it differently in time. Values of C_{80}, D_{50}, and TS change minimally for the different discretizations (differences between predictions with $\Delta t = 1/1000$ and $\Delta t = 1/24\,000$ are less than 1%). The most significant changes are for EDT and RT. Even for these parameters, however, predictions for $\Delta t = 1/2000$ are within 1% of predictions for $\Delta t = 1/4000$, and the differences for predictions for $\Delta t = 1/4000$ and $\Delta t = 1/24\,000$ are less than 1%. As with the echograms, this suggests that discretization periods of $\Delta t = 1/4000$ are sufficient for accurate predictions.

TABLE IV. Parameter predictions with varying $1/\Delta t$ for cube 4 (150 patches, $t_{\text{exact}} = 0.8$ s, and $t_{\text{final}} = 1.0$ s).

$1/\Delta t$ (s^{-1})	L_p (dB)	C_{80} (dB)	D_{50} (%)	EDT (s)	RT (s)	TS (ms)
1 000	84.67	2.31	47.86	1.20	1.20	91.21
2 000	84.67	2.32	47.66	1.19	1.19	90.69
4 000	84.67	2.33	47.57	1.18	1.18	90.45
8 000	84.67	2.33	47.53	1.18	1.17	90.30
12 000	84.67	2.35	47.62	1.18	1.16	90.20
16 000	84.67	2.35	47.58	1.17	1.16	90.20
24 000	84.67	2.35	47.58	1.17	1.16	90.20

Both time complexity and memory requirements are of the order $1/\Delta t$, as expected. Figure 5 shows elapsed time versus $1/\Delta t$, and memory requirement versus $1/\Delta t$, for the predictions for cube 1. Once again, the computer was a Pentium IV 2.2-GHz PC.

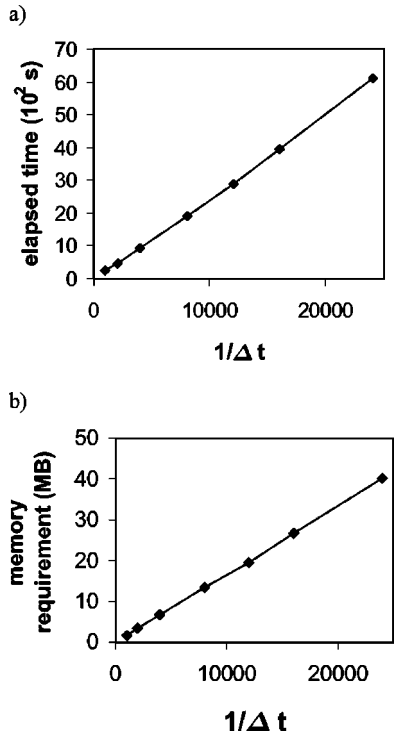


FIG. 5. Variation of (a) elapsed time; and (b) memory requirement with discretization frequency $1/\Delta t$ for cube 1 (150 patches, $t_{\text{exact}} = 0.8$ s, and $t_{\text{final}} = 1.0$ s).

TABLE V. Parameter predictions with varying t_{exact} (s) for cube 1 (t_{final} = 2 s, 150 patches, $\Delta t = 1/12\,000$).

t_{exact} (s)	L_p (dB)	C_{80} (dB)	D_{50} (%)	EDT (s)	RT (s)	TS (ms)
0.05	84.85	1.88	45.89	1.21	1.16	93.75
0.1	84.85	2.00	45.88	1.22	1.16	94.13
0.2	84.86	2.00	45.87	1.23	1.17	94.49
0.3	84.86	2.00	45.87	1.23	1.18	94.61
0.4	84.86	2.00	45.87	1.23	1.19	94.65
0.6	84.86	2.00	45.87	1.23	1.20	94.67
0.8	84.86	2.00	45.87	1.23	1.20	94.67
1.0	84.86	2.00	45.87	1.23	1.20	94.69
1.2	84.86	2.00	45.87	1.23	1.21	94.69
1.4	84.86	2.00	45.87	1.23	1.21	94.69

D. Time limits

It was shown [Eq. (3)] that, to get accurate predictions of RT, the impulse response must be predicted to at least $t_{final} = 0.93$ s. Although this gives an idea of what value of t_{final} to use in predictions, it is still not known what time t_{exact} is required. It is also of interest to investigate the effect of changing t_{final} in the predictions to verify that $t_{final_{min}} = 0.93$ s. To do so, predictions were made for the cubes, with varying t_{final} and t_{exact}. In this investigation, 150 patches were used with $\Delta t = 1/12\,000$.

First, as an "exact" case, a prediction was made using $t_{final} = 2$ s and $t_{exact} = 1.4$ s. This was confirmed as the limit to which predictions converge with increasing time limits; there was no change (up to four decimal places) in parameter predictions between this case and predictions with $t_{final} > 2$ s and/or $t_{exact} > 1.4$ s.

For the first set of predictions, t_{final} was held constant at 2 s while t_{exact} was varied (between 0.05 and 1.4 s). In Table V, the parameter predictions for cube 1 are given for varying t_{exact}. From these predictions, it is clear that RT is the parameter most sensitive to changes in t_{exact}. Still, in all cubes, the RT prediction with $t_{exact} = 0.6$ s is different from the exact answer by less than 1%. For all other parameters, $t_{exact} = 0.1$ s is sufficient for the same accuracy (at this exact time limit, RT predictions differ by less that 5%).

To investigate the effect of t_{final}, another set of predictions held t_{exact} constant and varied t_{final}. Table VI gives predictions of the parameters with $t_{exact} = 0.6$ s and varying t_{final}. Predictions do not change significantly as t_{final} is increased beyond 1 s. This is consistent with Eq. (1). Indeed, the prediction with $t_{exact} = 0.6$ s and $t_{final} = 1.0$ s puts all parameter predictions within 1% of the values for the exact case, for all cubes.

TABLE VI. Parameters predictions with varying t_{final} (s) for cube 1 (t_{exact} = 0.6 s, 150 patches, $\Delta t = 1/12\,000$).

t_{final} (s)	L_p (dB)	C_{80} (dB)	D_{50} (%)	EDT (s)	RT (s)	TS (ms)
0.7	84.85	2.00	45.92	1.23	1.16	94.46
0.8	84.86	2.00	45.91	1.23	1.19	94.60
0.9	84.86	2.00	45.91	1.23	1.20	94.66
1.0	84.86	2.00	45.91	1.23	1.20	94.69
1.1	84.86	2.00	45.91	1.23	1.20	94.69
1.2	84.86	2.00	45.90	1.23	1.20	94.69
1.3	84.86	2.00	45.90	1.23	1.20	94.69
1.5	84.86	2.00	45.90	1.23	1.21	94.69
1.6	84.86	2.00	45.90	1.23	1.21	94.69

E. Form factors

As outlined in Nosal *et al.*,[1] HELIOSFF (Ref. 18) was used to find form factors for arbitrary polyhedral rooms. HELIOSFF approximates the form factors by configuration factors; integrals are taken from one point on a patch, rather than from all points on the patch, to the second patch. For this reason, it is necessary to check that the use of these form factors does not significantly affect predictions. To do so, predictions were made for all four cubes under the same conditions as those outlined in Sec. B, using form factors given by HELIOSFF instead of the analytical values. In all cases, the parameters other than C_{80} are within 1% of those found using analytical form factors (in Sec. B) for the meshing with 96 patches. C_{80} is within 1% of the analytical form factor predictions for 150 patches. Echograms and discretized echograms are also very similar to those predicted in Sec. B. It was concluded that approximations made in the determination of form factors by HELIOSFF do not significantly affect predictions, particularly when a reasonably fine (150 patches or more, in this case) meshing of the enclosure is used.

IV. RAY TRACING

A. Diffuse ray tracing with RAYCUB

Having established some understanding of, and confidence in, the acoustical radiosity prediction results for the cubic enclosures, predictions can be compared to ray-tracing predictions for the case of diffuse surface reflection, as assumed by radiosity. The ray-tracing model used for this comparison (RAYCUB) was developed by Ondet and Barbry;[19] modifications were made to predict echograms and acoustical parameters, and to account for diffuse reflections as described by Kuttruff.[20] RAYCUB allows the user to input relevant room characteristics (room dimensions, source/receiver positions, absorption coefficients, air-absorption exponents). Also input are the size of the receiver (a cube of finite volume), the number of rays to be traced, the number of ray trajectories, and the sampling rate (fixing the echogram time-bin size), as well as a seed number (to initialize the random number generator). To account for diffuse reflection, two random numbers are selected to determine the direction of the reflected ray. The first number is chosen in the interval $(-\pi,\pi)$ to define the azimuthal angle of the reflected ray. The polar angle of the reflected ray is defined by the arccosine of the square root of the second number, chosen in the interval (0,1). Details about this method can be found in Kuttruff.[20]

B. Comparisons

The ray-tracing model was used to predict the echograms for the four cubic rooms. To ensure high preci-

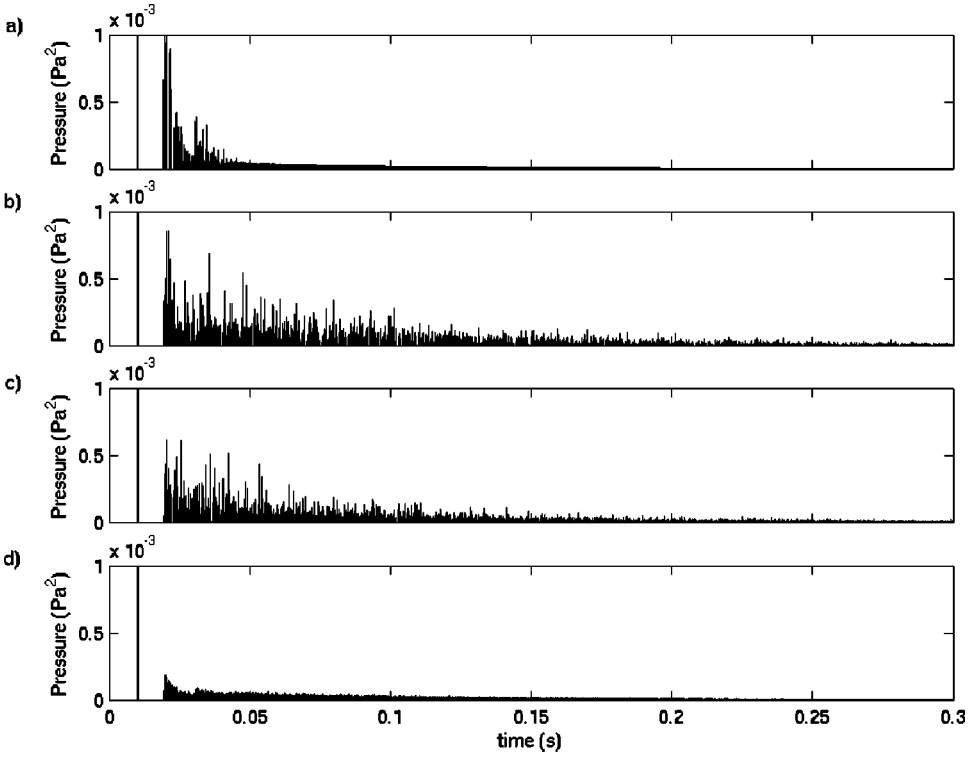


FIG. 6. Echograms for cube 1 predicted by (a) radiosity; (b) ray tracing with 1 million rays; (c) ray tracing with 50 million rays; (d) ray tracing with 1 million rays averaged over 64 runs. Radiosity: 600 patches, $\Delta t = 1/24\,000$, $t_{\text{exact}} = 0.8$ s, and $t_{\text{final}} = 1.0$ s. Ray tracing: diffuse reflection, 500 reflections, 24 kHz, $(0.1\text{ m})^3$ receiver. Range reduced to cut off the direct signal to show detail.

sion, a million rays were traced for 500 reflections at a sampling frequency of 24 kHz. The receiver was a cubic cell with side lengths of 0.1 m.

Echograms obtained by radiosity (with 600 patches and $\Delta t = 1/24\,000$) and ray tracing for cube 1 are shown in Figs. 6(a) and 6(b) (echograms for the other cubes are similar). The echograms are clearly quite different; the echogram for ray tracing has more peaks than the radiosity echogram. It seems that radiosity smears the energy in time more than ray tracing with diffuse reflection.

When the discretized echograms are compared, small differences in the distribution of energy is apparent for all cubes (Fig. 7 shows the discretized echogram for cube 3—the other cubes showed similar trends). In general, it can be observed that total energy levels are higher for radiosity in very early time steps (before 200 ms), and that levels become more similar in later time steps. This suggests, as did observation of the echograms, that radiosity tends to smear (or diffuse) energy faster than ray tracing with diffuse reflection.

Values of the parameter predictions for the cubes are given in Table VII(a). In general, ray tracing with diffuse reflection gives predictions quite similar to predictions by radiosity (see Table III). The parameter that exhibits the greatest difference between predictions is C_{80}; differences for these parameters never exceed 16%. RT, EDT, TS, and D_{50} predictions are all within 8%. Radiosity predictions for L_p are all within 2% of ray-tracing predictions. Note that, for the cubic enclosures, ray-tracing predictions of L_p are consistently lower than predictions by radiosity.

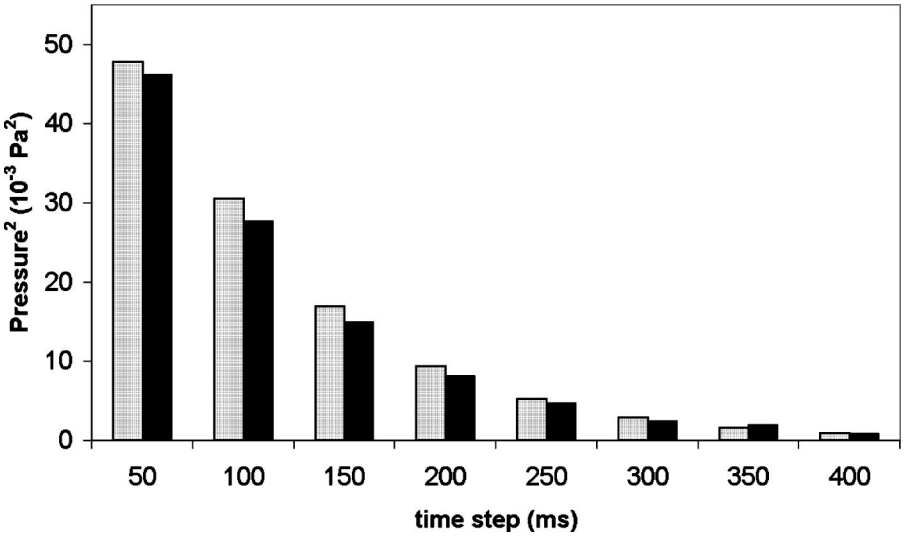


FIG. 7. Discretized echograms with time resolution of 0.05 s for cube 3 predicted by radiosity (600 patches, $\Delta t = 1/24\,000$, $t_{\text{exact}} = 0.8$ s, and $t_{\text{final}} = 1.0$ s) and ray tracing (1 million rays).

TABLE VII. Parameter predictions by ray tracing for the four cubic rooms with (a) 1 million rays; (b) 50 million rays; (c) average over 64 runs with 1 million rays [diffuse reflection, 500 reflections, 24 kHz, $(0.1\ \mathrm{m})^3$ receiver].

	Cube	L_p (dB)	C_{80} (dB)	D_{50} (%)	EDT (s)	RT (s)	TS (ms)
(a)	1	84.58	2.15	46.23	1.21	1.24	92.89
	2	83.29	2.45	46.41	1.03	1.09	86.41
	3	83.95	2.24	47.97	1.26	1.18	92.91
	4	83.86	2.31	46.46	1.15	1.13	91.82
(b)	1	84.57	1.98	46.89	1.22	1.22	93.71
(c)	1	84.57	2.02	45.87	1.22	1.23	94.50
	2	83.45	2.83	49.69	1.13	0.98	84.95
	3	84.31	2.31	46.24	1.18	1.16	91.38
	4	83.84	2.24	46.80	1.18	1.18	91.31

C. Discussion

The differences in the predictions of the parameters, echograms, and discretized echograms for radiosity and ray tracing with diffuse reflection are likely a consequence of the different manner in which diffuse reflection is incorporated in the models. Indeed, a diffusely reflected ray in RAYCUB is not scattered in all directions (as it is in radiosity) but only in one random direction. Consequently, the energy of the ray is not broken into many parts. Rays with high amounts of energy retain the energy, resulting in peaks in the echogram and less smearing of energy. In radiosity, all energy is immediately scattered, so that no larger ''bundles'' of energy remain intact to propagate through the system to cause peaks in the echogram. Also, because of this scattering in radiosity, energy is quickly smeared in time, giving different discretized echograms and parameter predictions.

Still, in the limiting cases (with infinitely many rays traced indefinitely with a point receiver in ray tracing, and with infinitely many patches and continuous time in radiosity) the two methods should be equivalent.[9] It was, therefore, surprising to find such big differences, particularly in the echograms (where the difference is most pronounced). Several attempts were made to investigate this, including performing further predictions with finer meshing of the room surfaces in radiosity, and further predictions with ray tracing. Increasing the number of patches did not improve the comparison but, as the number of rays increased in ray tracing, the echograms became more similar. Echograms from a simulation for cube 1 with 50 million rays is shown in Fig. 6(c), and parameter predictions for cube 1 are given in Table VII(b).

To check for noise artifacts in ray tracing, which could be the source of the problem, the number of rays hitting the receiver in the ray-tracing simulations was recorded. It was approximately 18 000 for 1 million rays, 178 860 for a run with 10 million rays, and 902 014 for 50 million rays. Indeed, the number of hits increased linearly with the number of rays. When the ray-tracing model was run again with different initialization numbers, the echograms were different, indicating that the solution may not have converged. Certainly, the echogram for the case of 50 million rays looks more similar to the radiosity echogram than that for the case of 1 million rays—some of the most predominant peaks were reduced in amplitude. Still, it cannot be concluded that the two methods give similar echograms. Predicted discretized echograms also changed with increasing numbers of rays, but those for ray tracing with 50 million rays were not obviously closer to radiosity than those with fewer rays. Errors in parameter values did not improve with increasing number of rays.

It would be of interest to run RAYCUB for an even larger number of rays. This could not be done because of time constraints (the run with 10 million rays took 15 h on a Pentium III 500-MHz PC; the one with 50 million rays took 80 h on a Pentium 4 2.2-GHz PC). There is a possibility that, with many more rays, ray tracing would converge to the radiosity solution. If this were the case, then it could be concluded that radiosity presents a more efficient way to predict the details of echograms for rooms with diffuse reflection (the run for radiosity took about 5 h, and only seconds for changing receiver positions once the enclosure was rendered). It may also be the case, however, that, with a much finer meshing, radiosity echogram predictions more closely resemble ray-tracing predictions. Very long run times prohibit an exhaustive investigation of the issue at this time. A possible conclusion is that the presented results may not represent converged solutions, but that they do represent practical limits in the algorithms used and in current computing capabilities.

A second idea was to average many ray-tracing runs, each with a different random initialization number. This was done for all cubes with 1 million rays averaged over 64 runs. The resulting echogram for cube 1 is shown in Fig. 6(d), and parameter predictions are given in Table VII(c). The tail (after about 0.05 s) of the echogram looks more similar to the radiosity echogram than the nonaveraged echogram, but the early part is missing some of the more prominent arrivals that are seen in the radiosity echogram—they have been lost in the averaging process. Parameter predictions are, in general, much closer to radiosity predictions than before. Once again, L_p were consistently lower than predictions by radiosity.

An alternative approach would be to use a ray tracer that incorporates diffuse reflection in a different way. It could, for instance, spawn a hundred or more rays for reflection from each diffuse surface, and assign a portion of the incident ray's energy (according to Lambert's cosine law) to each spawned ray. This approach would likely reduce the noise of the stochastic solution and may give results more similar to the radiosity predictions.

V. CONCLUSION

Several things can be learned from the predictions made for the four cubic ($8\times8\times8$-m) enclosures using the acoustical radiosity algorithms and methods developed in Nosal *et al.*[1] First, the predictions suggest that 150 patches, discretization periods of 1/4000 s, and impulse responses predicted exactly for slightly over half of the total predicted length of the response, where cutoff times are determined by Eq. (1), are sufficient for accurate prediction in these enclosures. Second, different distributions of absorption have a significant effect on the sound fields in enclosures with diffusely reflecting boundaries—more uniform distributions

lead to slower decays (hence longer reverberation times, higher center times, and lower clarities and definitions). Moreover, computational time complexity increases as the square of the number of patches, while memory complexity increases linearly with the number of patches. Both time and memory complexity are inversely proportional to time resolution. Furthermore, estimation of form factors by way of configuration factors (such as those predicted by HELIOSFF) does not significantly affect prediction accuracy. In general, radiosity smears energy in time, and eliminates any strong reflections from the echogram. This may make it insufficient for realistic rendering (auralization) of sound fields, in which strong signals have considerable effect on listener perception.

Although not expected, significant differences between radiosity and diffuse ray-tracing echogram predictions were found. There is reason to believe that the ray-tracing predictions had not converged for 1 million rays, or even for 50 million rays, and that echograms for radiosity and ray tracing would look more similar for converged predictions. Ray-tracing predictions averaged over 64 runs had more similarities in the late decay and in parameter predictions. Limitations of the algorithms, and of current computing capabilities, may be the restrictive factor for a perfect comparison.

Despite the differences in the detailed echograms, the discretized echograms predicted by radiosity are similar to those predicted by ray tracing, as are parameter predictions. In the case of the cubic enclosures considered in this research, these are within 16% for C_{80}, 8% for RT, EDT, TS, and D_{50}, and 2% for L_p.

The question of why radiosity and ray tracing predict different echograms is of immediate interest to future research in acoustical radiosity. Ideally, further simulations would be run, for a sufficiently large number of rays in ray tracing and, possibly, for a very fine meshing of the enclosure for radiosity. In addition, it would be interesting to look further into the conditions for convergence of acoustical radiosity and to do further comparisons to ray tracing—possibly for noncubic, and even nonrectangular enclosures. Beyond this, comparisons of radiosity predictions (and those by ray tracing for various surface-reflection laws) to measurements made in real rooms are needed to validate the methods.

ACKNOWLEDGMENTS

We gratefully acknowledge the Natural Sciences and Engineering Research Council of Canada for financial support, and the Institute of Applied Mathematics at UBC for facilitating collaboration between the authors. Also, many warm thanks to Sander Calisal and Wolfgang Heidrich for their valuable suggestions and input to this work.

[1] E.-M. Nosal, M. R. Hodgson, and I. Ashdown, "Improved algorithms and methods for room sound-field prediction by acoustical radiosity in arbitrary polyhedral rooms," J. Acoust. Soc. Am. **116**(2), 970–980 (2004).
[2] MATLAB is a registered trademark of *The Math Works, Inc.*
[3] J. Kang, "Reverberation in rectangular long enclosures with diffusely reflecting boundaries," Acust. Acta Acust. **88**, 77–87 (2002).
[4] H. Kuttruff, "A simple iteration scheme for the computation of decay constants in enclosures with diffusely reflecting boundaries," J. Acoust. Soc. Am. **98**(1), 288–293 (1995).
[5] H. Kuttruff, "Energetic sound propagation in rooms," Acust. Acta Acust. **83**, 622–628 (1997).
[6] R. N. Miles, "Sound field in a rectangular enclosure with diffusely reflecting boundaries," J. Sound Vib. **92**(2), 203–226 (1984).
[7] J. Shi, A. Zhang, J. Encarnação, and M. Göbel, "A modified radiosity algorithm for integrated visual and auditory rendering," Comput. Graph. **17**(6), 633–642 (1993).
[8] M. R. Schroeder and D. Hackman, "Iterative calculation of reverberation time," Acustica **45**, 269–273 (1980).
[9] A. LeBot and A. Bocquillet, "Comparison of an integral equation on energy and the ray-tracing technique in room acoustics," J. Acoust. Soc. Am. **108**(4), 1732–1740 (2000).
[10] H. Kuttruff, *Room Acoustics*, 4th ed. (Spon, London, 2000).
[11] A. G. Sotiropoulou and D. B. Fleming, "Concert hall acoustic evaluations by ordinary concert-goers. II. Physical room acoustics criteria subjectively significant," Acustica **81**(1), 10–19 (1995).
[12] J. S. Suh and P. A. Nelson, "Measurement of transient response f rooms and comparison with geometrical acoustic models," J. Acoust. Soc. Am. **105**(4), 2304–2317 (1999).
[13] U. Gross, K. Spindler, and E. Hahne, "Shape-factor equations for radiation heat transfer between plane rectangular surfaces of arbitrary position and size with rectangular surfaces of arbitrary position and size with rectangular boundaries," Lett. Heat Mass Transfer **8**, 219–227 (1981).
[14] E.-M. Nosal, "Criteria for impulse response length for accurate reverberation time prediction," J. Acoust. Soc. Am. (submitted).
[15] M. R. Hodgson, "When is diffuse-field theory applicable," Appl. Acoust. **49**(3), 197–207 (1996).
[16] G. J. Etgen, *Salas and Hille's Calculus—Several Variables*, 7th ed. (Wiley, New York, 1995).
[17] J. H. Ferziger, *Numerical Methods for Engineering Application*, 2nd ed. (Wiley Interscience, New York, 1998).
[18] HELIOSFF is a trademark of *by Heart Consultants Limited* ⟨www.helios32.com⟩.
[19] A. M. Ondet and J. L. Barbry, "Modeling of sound propagation in fitted workshops using ray tracing," J. Acoust. Soc. Am. **85**(2), 787–796 (1989).
[20] H. Kuttruff, "Simulierte Nachhallkurven in Rechteckräumen mit diffusem Schallfeld," [Simulated reverberation curves in rectangular rooms with diffuse sound fields], Acustica **25**(6), 333–342 (1971).

On the kinematics of broadband multipath scintillation and the approach to saturation

John A. Colosi
Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543

Arthur B. Baggeroer
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 3 November 2003; revised 5 August 2004; accepted 5 August 2004)

Utilizing a simple model in which the acoustic wave function is a sum of independent Gaussian wave packets, the relative intensity variance or scintillation index (SI) is analytically calculated. The model has an unspecified probability density function (PDF) for wave packet amplitudes and Gaussian PDFs for travel-time-induced and non-travel-time-induced phase shifts; amplitudes and both phase shifts are assumed to be mutually uncorrelated. It is shown that a proper treatment of the mean field is required to obtain the saturation value, SI=1, in the limit of a large number of interfering wave packets. The analytic formulas for SI allow identification of important wave packet parameters in the approach to saturation. Criteria are identified for both broadband and narrow-band cases for which the approach to saturation is from above and below 1. It is demonstrated that the broadband approach to saturation is much slower than the narrow-band cases, since wave packets separated in time by more than an inverse bandwidth do not strongly contribute to interference. This effect is quantified by the time–bandwidth product. The model is also used to obtain an analytic expression for pulse time spread; it is shown that multipath conditions which favor a rapid approach to saturation do not favor large pulse spread.
[DOI: 10.1121/1.1798356]

PACS numbers: 43.60.Cg, 43.30.Ft, 43.20.Fn [WLS] Pages: 3515–3522

I. INTRODUCTION

Long-range ocean acoustic propagation experiments in both deep and shallow water are characterized by an ubiquitous broadband multipath arrival structure which requires interpretation (Duda *et al.* 1992; Colosi *et al.*, 2001; Chiu *et al.*, 2004; Fredricks *et al.*, 2003). Of fundamental importance to ocean remote sensing and signal processing in this multipath environment is the development of strong wave scattering and the approach to saturation where the real and imaginary parts of the wave-field envelope become independent Gaussian random variables. Understanding the approach to saturation is an exercise in understanding how independent contributions from the scattered wave field produce interference so that the sum of all these contributions leads (via the central limit theorem) to Gaussian noise.[1] A fully saturated wave field has an intensity probability density function (PDF) that is the two degree of freedom chi-square distribution (i.e., negative exponential)

$$P(I)=\frac{1}{\langle I\rangle}\exp\left[-\frac{I}{\langle I\rangle}\right], \tag{1}$$

so that the moments of the field become

$$\langle I^n\rangle = n!\langle I\rangle^n. \tag{2}$$

Also, the phase of the wave field is uniformly distributed between $-\pi$ and π. A useful measure of saturation is the relative intensity variance or scintillation index given by

$$\mathrm{SI}=\frac{\langle I^2\rangle}{\langle I\rangle^2}-1, \tag{3}$$

and, using Eq. (2) it is seen that the saturation value of SI is 1. Figure 1 shows a sketch of two possible routes to saturation. In one case, which is typically a high-frequency or geometrical optics case, SI overshoots 1, forming the partially saturated or strong focusing regime (Flatté *et al.*, 1979) and saturation is approached from above 1. In the other case, which is typically a low-frequency or diffractive case, SI simply grows monotonically to approach 1 from below. In this paper, kinematical conditions are obtained which describe the approach to saturation from either above or below 1.

In studying the approach to saturation, analytic results are difficult to obtain. Dashen (1979), for example, has computed first-order corrections to Eq. (2), in the partially saturated regime, using the path integral formalism so that

$$\langle I^n\rangle = n!\langle I\rangle^n[1+\tfrac{1}{2}n(n-1)\alpha], \tag{4}$$

where α can be computed explicitly using a spectrum of index of refraction variations.[2] In this case SI$=1+2\alpha$, and α is always positive. Later, Dashen *et al.* (1993) improved upon the series expansion to get faster convergence in the case of a constant background wave speed and homogeneous isotropic turbulence with a pure power law. Comparing to Monte Carlo numerical simulations, they were able to accurately predict moments up to $\langle I^4\rangle$ from the saturated region where SI$=1$ to near the partially saturated region where SI can have a maximum. However, these approaches are limited in that there is no waveguide, the Markov approximation is made, the approach to saturation is only from above 1, and there is no account of broadband effects. Numerical approaches have been very useful for atmospheric optics appli-

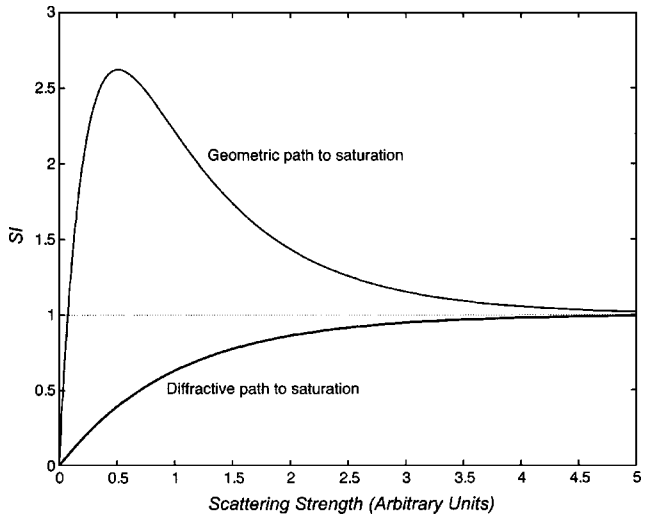


FIG. 1. Sketch of two different routes to saturation. The *y* axis is scintillation index and the *x* axis is a generic scattering strength index which could be propagation range, variance of the index of refraction, intensity variance computed by weak fluctuation theory, etc.

cations (Martin and Flatté, 1988, 1990; Flatté, Bracher, and Wang, 1994) and for some ocean cases (Ewart and Percival, 1986; Ewart, 1989), but it is hard to isolate the important physics, and again broadband effects have not been treated.

In this paper a simple analytic form of SI is put forth which is based on the interference of independent random phasors modeled as Gaussian wave packets. Explicitly, a sum of Gaussian wave packets (Gasiorowicz, 1974) at time t can be written

$$p(t)=\sum_{k=1}^{N} a_k \mathrm{e}^{-(t-t_k^{gr})^2/(2\Delta T^2)} \cos[\omega_0(t-t_k^{ph})-\theta_k], \quad (5)$$

where ω_0 is the center frequency, a_k is a real amplitude, ΔT is the time width of the Gaussian wave packet ($\Delta\omega \equiv 1/\Delta T$ is the frequency width of the Gaussian), t_k^{gr} and t_k^{ph} are the group and phase delays, and θ_k is a frequency-independent phase shift. It must be noted that the amplitudes a_k, and the delays t_k^{gr} and t_k^{ph} are all slow functions of geophysical time, but this dependence is not explicitly noted because we are considering independent realizations of the field. This model is quite general and may equally describe the interference along a single temporally resolved wavefront (given adequate bandwidth) or it may describe the interference over the entire arrival pattern as in single-frequency cases. Acoustic waves are nondispersive, so $t^{gr}=t^{ph}$, and the phase shift θ_k may be associated with, for example, the number of caustics that the wave packet has traversed. Utilizing the nondispersive property and demodulating Eq. (5) leads to ψ, the complex envelope of p

$$\psi(t)=\sum_{k=1}^{N} a_k \mathrm{e}^{-(t-t_k)^2/(2\Delta T^2)} \mathrm{e}^{-i\omega t_k} \mathrm{e}^{-i\theta_k}, \quad (6)$$

where we denote $t_k=t^{gr}=t^{ph}$. For the forthcoming analysis it is useful to write the demodulated field at the mean travel time of the multipaths t_m so that

$$\psi(t_m)=\mathrm{e}^{-i\omega t_m}\sum_{k=1}^{N} a_k \mathrm{e}^{-\delta t_k^2/(2\Delta T^2)} \mathrm{e}^{-i\omega\delta t_k} \mathrm{e}^{-i\theta_k}$$

$$=\mathrm{e}^{-i\omega t_m}\sum_{k=1}^{N} a_k \mathrm{e}^{-Q^2\phi_k^2/2} \mathrm{e}^{-i\phi_k} \mathrm{e}^{-i\theta_k}, \quad (7)$$

where $Q\equiv\Delta\omega/\omega_0$, and $\phi_k\equiv\omega_0\delta t_k=\omega_0(t_m-t_k)$ is the phase change associated with the travel time of the pulse.[3] In Eq. (7) we have assumed that the mean travel times of the multipaths are the same, and thus this term will have no effect on the statistics of intensity. The multipaths are assumed uncorrelated and for an individual multipath the amplitude, group delay, and phase shifts are all assumed to be uncorrelated, so that

$$\begin{aligned}
&\langle a_k\phi_j\rangle=\langle a_k\theta_j\rangle=\langle\phi_k\theta_j\rangle=0,\\
&\langle a_k a_j\rangle=\langle a\rangle^2 (k\neq j), \quad \langle a_k a_k\rangle=\langle a^2\rangle\\
&\langle\phi_k\phi_j\rangle=\Phi^2\delta_{k,j}, \quad \text{and } \langle\theta_k\theta_j\rangle=\Theta^2\delta_{k,j}.
\end{aligned} \quad (8)$$

The phases ϕ_k and θ_k are assumed to be zero-mean Gaussian random variables with variances Φ^2 and Θ^2, but the PDF of the amplitude a is unconstrained.[4] N may be considered the number of independent multipaths, which in general will be different from the total number of multipaths. In real random media problems N is a random variable,[5] but in this study we will treat it as a constant. If N is independent of the other parameters in this problem, the results of this paper could be used with a PDF of N to further ensemble average the intensity moments.

Loosely, the physical basis of the model [Eq. (7)] is ray theory, where the group delay ϕ_k is associated with travel-time fluctuations, the phase θ_k is associated with the number of caustics that the ray has passed through, and the amplitude variation is associated with fluctuations of the ray tube. This model is useful since recent ray theory results (Brown *et al.*, 2003; Wolfson and Tomsovic, 2001; Beron-Vera *et al.*, 2003) have been able to describe several observables from long-range acoustic transmission experiments. The ray theory also provides analytic models for both amplitude and phase variability. However, the analogy is not perfect as the phase, θ_k, in ray theory is not a continuous random variable, but is quantized in units of $\pi/2$.[6] Also, Eq. (7) can be interpreted in mode language where variations of the group and phase delays (both continuous random variables) are associated with mode coupling. However, in this analogy the assumption of a symmetric Gaussian wave packet is not appropriate to the dispersive nature of modal wave packets.

A noteworthy omission in the present model is that correlations between multipaths are not treated. This effect is known to be crucial to the partially saturated regime where N is small and strong coherent caustic formation dominates the field (Flatté *et al.*, 1979; Mikhalevsky, 1982). Further, it has been suggested from path integral results (Flatté, 1983) and numerical simulations [Bracher and Flatté, 1993; Flatté, Bracher and Wang, 1994] in an atmospheric optics context (power-law random media) that the approach to saturation is controlled by the disintegration of multipath correlations, where the small-scale index of refraction fluctuations contributes to the creation of independent multipaths, while the

large-scale modulations create correlations between these multipaths. This physical picture is the basis of so-called convolution PDFs like the log-normal convolved with exponential intensity PDF that has had some success in describing numerical simulations through turbulence spectra [Bracher and Flatté, 1993; Flatté, Bracher, and Wang, 1994]. The present model takes a different, more kinematical view, and exposes constraints on parameters like the source Q function, number of interferers N, phase variances Φ^2 and Θ^2, and amplitude variations of the interfering wave as the saturation limit is reached. This approach should be valid in some neigborhood of full saturation once multipath correlations have weakened.

Using the present model [Eqs. (7) and (8)] it is found that a proper treatment of the mean field is essential to obtaining a correct understanding of the approach to saturation. This point has been appreciated in the optical propagation literature and is the basis of the Rice–Nakagami amplitude distribution (Rice, 1944; Strohbehn, 1978, and references therein). Several important results are found from defining a proper mean centered wave function. First, under both narrow-band and broadband conditions the scintillation index can approach 1 from either above or below 1, yet for the broadband case it is much more difficult to achieve an approach from below 1. Further, it is found that broadband signals approach saturation more slowly because the effective number of interfering waves is reduced by a factor proportional to the time–bandwidth product. That is to say, wave packets more than an inverse bandwidth away from each other do not strongly interfere and thus do not contribute to saturation. The ratio of $\langle a^4\rangle/\langle a^2\rangle^2$, effectively the kurtosis of the multipath amplitude PDF, has significance for both broadband and narrow-band cases and competes against N to drive the system to saturation. Further, it is shown that for broadband cases, conditions which favor a rapid approach to saturation do not favor a large pulse time spread.

The outline of this paper is as follows. In Sec. II analytic calculations of intensity moments are presented and used in the analysis of SI. In Sec. III the approach to saturation is discussed and a model based on ray theory is put forward to demonstrate the approach to saturation as a function of propagation range. Section IV presents results for pulse time spread as saturation is approached. A summary is given in Sec. V.

II. A STATISTICAL MODEL OF MULTIPATH SCINTILLATION

Details of the model calculations are now given. The wave function, ψ, is described by Eq. (7) and the assumed multipath statistics stated in Eq. (8) are used. The mean intensity is

$$\langle I\rangle=\langle\psi\psi^*\rangle=\sum_{k=1}^{N}\sum_{j=1}^{N}\langle a_k a_j\rangle\langle e^{i(\theta_k-\theta_j)}\rangle \times\langle e^{-Q^2(\phi_k^2+\phi_j^2)/2}e^{i(\phi_k-\phi_j)}\rangle. \quad (9)$$

Making use of the assumption of Gaussian phases, the following basic results can be easily obtained:

$$\langle e^{-mQ^2\phi^2/2}e^{in\ \phi}\rangle=\frac{1}{(1+mQ^2\Phi^2)^{1/2}}\times\exp\left[-\frac{n^2}{2}\frac{\Phi^2}{(1+mQ^2\Phi^2)}\right], \quad (10)$$

$$\langle e^{-mQ^2\phi^2/2}\rangle=\frac{1}{(1+mQ^2\Phi^2)^{1/2}}, \quad (11)$$

and m is required to be greater than zero. In Eq. (9) there are N, $j=k$ terms in the double sum and they are evaluated using Eq. (11), while there are $N(N-1)$ terms with $j\neq k$ and these are evaluated using Eq. (10). Combining both results gives

$$\langle I\rangle=\frac{N\langle a^2\rangle}{(1+2Q^2\Phi^2)^{1/2}}+\frac{N(N-1)\langle a\rangle^2}{1+Q^2\Phi^2}F_1\gamma^2, \quad (12)$$

where

$$F_m(\Phi,Q)=\exp\left[-\frac{\Phi^2}{1+mQ^2\Phi^2}\right], \quad \text{and}$$

$$\gamma^2=\exp[-\Theta^2]. \quad (13)$$

The mean intensity is seen to be the sum of two terms, the first of which comes from the correlation of two identical wave packets, and the second which comes from the assumed uncorrelated nonidentical pairs.

The calculation for $\langle I^2\rangle$ proceeds along similar lines

$$\langle I^2\rangle=\langle\psi\psi^*\psi\psi^*\rangle=\sum_{k=1}^{N}\sum_{j=1}^{N}\sum_{m=1}^{N}\sum_{n=1}^{N}\langle a_k a_j a_m a_n\rangle \times\langle e^{i(\theta_k-\theta_j+\theta_m-\theta_n)}\rangle \times\langle e^{-Q^2(\phi_k^2+\phi_j^2+\phi_m^2+\phi_n^2)/2}e^{i(\phi_k-\phi_j+\phi_m-\phi_n)}\rangle, \quad (14)$$

except the combinatorics is more difficult. A cumbersome but straightforward calculation utilizing Eqs. (10) and (11) leads to the mean-square intensity

$$\langle I^2\rangle=\frac{N\langle a^4\rangle}{(1+4Q^2\Phi^2)^{1/2}}+\frac{N(N-1)\langle a^2\rangle^2}{(1+2Q^2\Phi^2)}(F_2^4\gamma^8+2) +\frac{4N(N-1)\langle a^3\rangle\langle a\rangle(F_3F_1)^{1/2}\gamma^2}{((1+3Q^2\Phi^2)(1+Q^2\Phi^2))^{1/2}} +\frac{2N(N-1)(N-2)\langle a^2\rangle\langle a\rangle^2F_1\gamma^2}{(1+Q^2\Phi^2)(1+2Q^2\Phi^2)^{1/2}}(F_2^2\gamma^4+2) +\frac{N(N-1)(N-2)(N-3)\langle a\rangle^4F_1^2\gamma^4}{(1+Q^2\Phi^2)^2}. \quad (15)$$

The mean-square intensity is seen to be the sum of five terms which come from the five different ways of combining four wave packets. The first and last terms are from combinations of all identical and all different wave packets, while the third term is from three identical wave packets and one different. The second and fourth terms are, respectively, from combinations of two pairs of identical wave packets, and one pair of identical and one pair of different wave packets. An im-

portant parameter in these equations gauging the effects of bandwidth is the "time–bandwidth" product $Q\Phi = (\Delta\omega/\omega_0)(\omega_0\langle(\delta t^{gr})^2\rangle^{1/2}) = \Delta\omega\langle(\delta t^{gr})^2\rangle^{1/2}$ (Munk and Wunsch, 1983). However, in the limit of very large N (fixing the other variables), one finds that $\langle I^2\rangle \to \langle I\rangle^2$ and thus SI $\propto N^{-1}$. It is seen that as N becomes large, SI$\to$0 so that a proper saturation value of SI=1 is not obtained. In this case the moments are dominated by the contributions from products of independent wave packets of which there are a large number [for example, the second term in Eq. (12) which scales like $N(N-1)$ as opposed to the first term which scales like N]. These independent wave packet contributions are associated with a background energy, or mean field which is not directly related to scintillation. Thus, to examine this issue in more detail we compute the mean field. Taking the expectation value of Eq. (7) gives the mean field

$$\langle\psi\rangle = \frac{N\langle a\rangle}{(1+Q^2\Phi^2)^{1/2}} F_1^{1/2}\gamma. \tag{16}$$

For $Q=0$ the mean field reduces to the familiar expression $N\langle a\rangle\exp[-(\Phi^2+\Theta^2)/2]$. It is seen that in a multipath environment, the mean field may attain a significant value since increasing N can offset the effects of increasing Φ or Θ: In ray theory, N can increase exponentially. Further, the dependence of the mean field on Φ goes from exponential ($\exp[-\Phi/2]$) for a narrow-band signal to a weaker $1/(Q\Phi)$ for a broadband signal.

To rectify the mean-field problem, a new mean centered wave function, $\psi_r = \psi - \langle\psi\rangle$ is defined which leads to the result that

$$\langle I_r\rangle = \langle I\rangle - \langle\psi\rangle^2 = \frac{N\langle a^2\rangle}{(1+2Q^2\Phi^2)^{1/2}} - \frac{N\langle a\rangle^2}{1+Q^2\Phi^2} F_1\gamma^2. \tag{17}$$

Note here that after removing the mean field from the mean intensity, the dependence on N has dropped from quadratic to linear. Another cumbersome calculation leads to the result

$$\begin{aligned}\langle I_r^2\rangle &= \langle I^2\rangle + 2\langle\psi\rangle^2(2\langle I\rangle + \langle\psi^2\rangle) - 3\langle\psi\rangle^4 - 4\langle\psi\rangle \\ &\quad\times\langle\psi\psi^*\psi\rangle \qquad (18)\\ &= \frac{N\langle a^4\rangle}{(1+4Q^2\Phi^2)^{1/2}} + \frac{N(N-1)\langle a^2\rangle^2}{(1+2Q^2\Phi^2)}(F_2^4\gamma^8+2) \\ &\quad - \frac{4N\langle a^3\rangle\langle a\rangle(F_3F_1)^{1/2}\gamma^2}{((1+3Q^2\Phi^2)(1+Q^2\Phi^2))^{1/2}} \\ &\quad - \frac{2N(N-2)\langle a^2\rangle\langle a\rangle^2F_1\gamma^2}{(1+Q^2\Phi^2)(1+2Q^2\Phi^2)^{1/2}}(F_2^2\gamma^4+2) \\ &\quad + \frac{3N(N-2)\langle a\rangle^4F_1^2\gamma^4}{(1+Q^2\Phi^2)^2}. \qquad (19)\end{aligned}$$

Equations (17) and (19) are used to analytically compute the scintillation index SI, and it is found that Gaussian statistics are indeed obtained in the limit of large N. These equations will guide the interpretation of the approach to saturation, which is discussed next.

III. APPROACH TO SATURATION

Equations (17) and (19) are still a bit unwieldy, so it is instructive to look at two distinct cases of (1) single frequency ($Q=0$) and large phase variance (Φ, $\Theta \gg 1$) and (2) large time–bandwidth product, ($Q\Phi \gg 1$). In this discussion, a narrow-band signal ($Q \ll 1$) is associated with the single-frequency case. In the narrow-band and broadband cases, the scintillation index is found to be[7]

$$\begin{aligned}\mathrm{SI} &\simeq 1 + \frac{1}{N}\left(\frac{\langle a^4\rangle}{\langle a^2\rangle^2} - 2\right) \\ &\quad + \mathrm{O}(\gamma^2, F_0(\Phi,0)) \quad \text{narrow band,} \qquad (20)\end{aligned}$$

$$\begin{aligned}\mathrm{SI} &\simeq 1 + \frac{1}{N}\left(Q\Phi\frac{\langle a^4\rangle}{\langle a^2\rangle^2} - 2\right) \\ &\quad + \mathrm{O}(\gamma^2, 1/Q\Phi), \quad \text{broadband.} \qquad (21)\end{aligned}$$

Note that because of the complex dependency of Eqs. (17) and (19) on Q, the asymptotic narrow-band and broadband cases [Eqs. (20) and (21)] cannot be related to one another as a smooth function of Q. Several points are noteworthy from Eqs. (20) and (21).

A. Approach to saturation from above or below 1

First, SI will approach 1 if the second terms in Eqs. (20) and (21) go to zero. This limit will be approached from either above or below 1 depending on the sign of the terms multiplying $1/N$. In the narrow-band case the transition from an approach from above 1 to below 1 is at $\langle a^4\rangle/\langle a^2\rangle^2 = 2$, while the broadband case requires $Q\Phi\ \langle a^4\rangle/\langle a^2\rangle^2 = 2$, with the additional requirement that $Q\Phi \gg 1$. The magnitude of $\langle a^4\rangle/\langle a^2\rangle^2$ is a measure of the importance of the tails in the PDF of the amplitude, and can be considered a kurtosis measure. In subsequent discussion we will refer to this parameter as the kurtosis of the amplitude distribution. Therefore, in a broadband context, the PDF of amplitude fluctuations from the interferers must fall off very rapidly in the tails to get an approach to saturation from below 1.

To get a feeling for the orders of magnitude, some sample PDFs for the multipath amplitudes (a) can be examined. For the idealized case of no amplitude fluctuations (i.e., where $\langle a^4\rangle/\langle a^2\rangle^2 = 1$), the broadband case yields $Q\Phi = 2$ for the transition. The narrow-band case always approaches 1 from below (Dyer, 1970). A Rayleigh PDF for a (exponential in intensity) has been obtained for mode amplitudes (Dozier and Tappert, 1978; Colosi and Flatté, 1996). For the narrow-band case with Rayleigh amplitudes, a vanishing first-order correction term is obtained since $\langle a^4\rangle/\langle a^2\rangle^2 = 2$ [Eq. (2)] so that higher order terms are need to describe the approach to saturation. However, for the broadband case and Rayleigh amplitudes, one finds the transition between SI approaching 1 from above or below at $Q\Phi = 1$. As a final example, ray theory results indicate that the PDF of a is close to log normal (Wolfson and Tomsovic, 2002; Brown *et al.*, 2003; Beron-Vera *et al.*, 2003), and for a log-normal PDF (variance σ^2 and mean μ) the moments of a are

$$\langle a^n\rangle = \exp\left[\frac{\sigma^2 n^2}{2} + \mu n\right], \quad \text{so that} \quad \frac{\langle a^4\rangle}{\langle a^2\rangle^2} = \exp[4\sigma^2]. \tag{22}$$

Therefore, the condition for $\langle a^4\rangle/\langle a^2\rangle^2 = 2$ is $\sigma = (\ln 2)^{1/2}/2 \rightarrow 1.8\,\text{dB}$, which is a very modest value for the amplitude variation, and is in fact much smaller than simulated multipath amplitude variations (see Sec. III C.) Thus, it is seen that the conditions to get an approach to saturation from below 1 impose a strong restriction on the PDF of multipath amplitude fluctuations.

B. Corrections to Rayleigh statistics

In both broadband and narrow-band situations, the correction to SI=1 near saturation is controlled by the factor

$$\beta_{\text{SI}} = \frac{1}{N_{\text{eff}}}\frac{\langle a^4\rangle}{\langle a^2\rangle^2}\begin{cases} N_{\text{eff}} = N & \text{for } Q = 0, \\ N_{\text{eff}} = N/Q\Phi & \text{for } Q \neq 0, \end{cases} \tag{23}$$

where N_{eff} is the effective number of interfering waves. In the narrow-band case $N_{\text{eff}} = N$, but in the broadband case $N_{\text{eff}} = N/(Q\Phi)$. Thus, the time–bandwidth product $Q\Phi$ increases the correction to SI=1, so that the system is further from saturation. It must be noted that in the narrow-band case the phase variances Φ^2 and Θ^2 are on equal footing as both terms are associated with exponentially decaying terms (i.e., F_0 and γ^2). In the narrow-band case these travel-time-induced and non-travel-time-induced phases contribute to the approach to saturation only to the extent of wiping out all but a few terms in SI. The broadband case is different because the travel-time-induced phase shifts are now tied to the shifting of the pulse envelope, and thus we see that the β_{SI} parameter is a function of Φ^2 and not a function of Θ^2 (and thus γ^2). In this broadband case, it is seen that group delays have a larger role in the approach to saturation.

As an acoustic field propagates out in range, it is expected that N, Φ, and $\langle a^4\rangle/\langle a^2\rangle^2$ will all increase. The approach to saturation as a function of range will depend on the competition of these factors as related by the β_{SI} expression [Eq. (23)]. Increasing N speeds the approach to saturation, but increasing $\langle a^4\rangle/\langle a^2\rangle^2$ can counteract the N effect and slow down the approach to saturation. Further, for the broadband case the counterintuitive result is obtained that increasing phase variance Φ actually slows the approach to saturation by reducing N_{eff}. This means physically that pulses separated by more than an inverse bandwidth will not significantly interfere and thus contribute to saturation.

C. A ray chaos example

It is instructive to use Eqs. (17) and (19) in an example to understand potential range variation of SI; the approximate forms for SI [Eqs. (20) and (21)] are also used to demonstrate their regime of validity. A ray theory example is treated since there are theoretical models for all parameters. However, we will ignore the influence of the non-travel-time-induced phase fluctuations θ, since we have seen theoretically that at best they are on equal footing with the travel-time-induced phase fluctuations in the narrow-band case, and in the broadband case they are much less significant than the travel-time-induced phase fluctuations. For ray theory the PDF of ray amplitude is log normal, and the variance σ^2 grows linearly with range (Wolfson and Tomsovic, 2001; Beron-Vera *et al.*, 2003). Further, the ray theory model has $N = \exp(\nu R)$ (Tappert and Tang, 1996) and gives the exponential increase in multipaths. A reasonable model is to assume Φ^2 to grow diffusively with range. So, for this example, parametrizations of the form

$$N(R) = \exp(R\nu), \quad \Phi^2(R) = \Phi_0^2 R/R_0,$$

$$\langle a^n\rangle(R) = \exp[\tfrac{1}{2}\sigma_0^2 R n^2/R_0], \tag{24}$$

are adopted. Numerical simulations of ray propagation through internal wave sound-speed perturbations for a long-range propagation experiment suggest $\Phi_0^2 \sim 4\pi^2$, $\sigma_0^2 \sim 5$, $R_0 = 3000\,\text{km}$, and $\nu \sim 1/100$ km (Beron-Vera *et al.*, 2003). Note that in this model the correction to Rayleigh statistics parameter

$$\beta_{\text{SI}} = \frac{1}{N}\frac{\langle a^4\rangle}{\langle a^2\rangle^2} = \frac{\exp[4\sigma_0^2 R/R_0]}{\exp[\nu R]}$$
$$= \exp[-R(\nu - 4\sigma_0^2/R_0)], \tag{25}$$

is governed by two competing exponentials, and so the saturation limit SI=1 can only be obtained if $\nu > 4\sigma_0^2/R_0$. Equations (17) and (19) as well as Eqs. (20) and (21) are used to compute SI, and Fig. 2 shows the anticipated range variation. The modeled SI values are above 1 in the first few 100 km, and then saturate very close to 1 at about 2000 km. The broadband case is seen to reach saturation at a longer range.

IV. SATURATION AND THE MEAN PULSE

In broadband propagation the approach to saturation is accompanied by a distortion of the pulse from the transmitted form (Flatté *et al.*, 1979). In the previous section it was shown that the approach to saturation is slowed by increasing $Q\Phi$, since pulses separated in time by more than an inverse bandwidth will not significantly interfere. In this section it will be shown how this kinematical constraint imposes a limit on pulse time spread. Pulse distortion can be quantified by examining the mean pulse and its time spread (Dashen, Flatté, and Reynolds, 1985). The calculation of the mean pulse begins with Eq. (6), which can be written

$$\psi(t) = \mathrm{e}^{-i\omega t_m}\sum_{k=1}^{N} a_k \mathrm{e}^{-(t-t_m-\delta t_k)^2/(2\Delta T^2)}\mathrm{e}^{-i\omega\delta t_k}\mathrm{e}^{-i\theta_k}$$
$$= \mathrm{e}^{-i\omega t_m}\sum_{k=1}^{N} a_k \mathrm{e}^{-Q^2(\phi(t)-\phi_k)^2/2}\mathrm{e}^{-i\phi_k}\mathrm{e}^{-i\theta_k},$$

$$\phi(t) = \omega_0(t - t_m). \tag{26}$$

Utilizing the same assumptions as before [Eq. (8)], the mean pulse is then simply written as

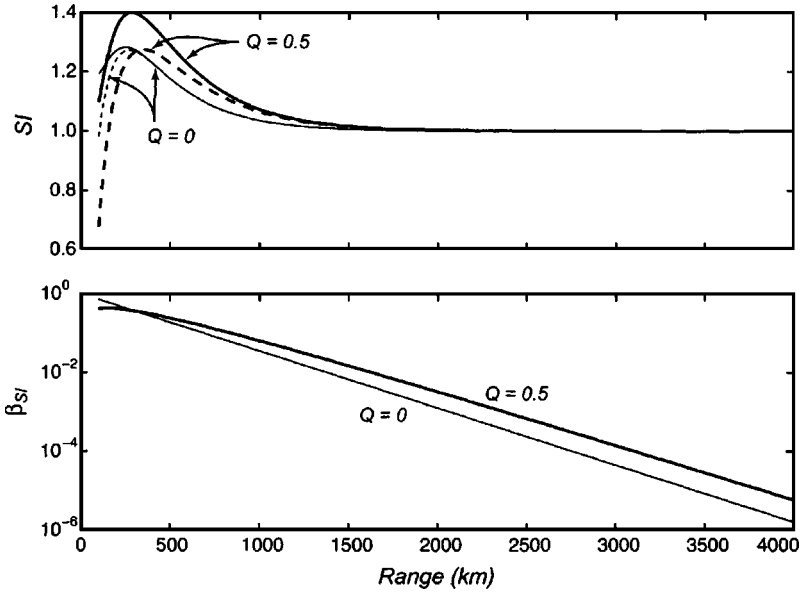


FIG. 2. Upper panel: Range variation of SI using a ray theory model as discussed in the text. The heavy curves are for $Q=0.5$ and the light curves are for $Q=0$. Dashed curves are for asymptotic estimates of SI using Eqs. (21) (heavy) and (20) (light). Lower panel: Range evolution of β_{SI} for $Q=0.5$ and $Q=0$.

$$\langle I(\phi(t))\rangle=\langle\psi(t)\psi^*(t)\rangle=\sum_{k=1}^{N}\sum_{j=1}^{N}\langle a_k a_j\rangle\langle e^{i(\theta_k-\theta_j)}\rangle \times\langle e^{-Q^2[(\phi(t)-\phi_k)^2+(\phi(t)-\phi_j)^2]/2}e^{i(\phi_k-\phi_j)}\rangle. \tag{27}$$

The expectation values in the mean pulse involve basic Gaussian integrals which can be solved by traditional means to obtain

$$\langle I(\phi(t))\rangle=I_0(t)\left(\frac{N\langle a^2\rangle}{(1+2Q^2\Phi^2)^{1/2}}\exp\left[\frac{2Q^2\Phi^2Q^2\phi^2(t)}{1+2Q^2\Phi^2}\right]+\frac{N(N-1)\langle a\rangle^2F_1\gamma^2}{1+Q^2\Phi^2}\exp\left[\frac{Q^2\Phi^2Q^2\phi^2(t)}{1+Q^2\Phi^2}\right]\right), \tag{28}$$

where $I_0(t)=\exp(-Q^2\phi^2(t))$ is the unperturbed pulse. Here, $\langle I(t=t_m)\rangle$ is the same as the mean intensity in Eq. (12). The reader will note that the expression for the mean pulse does not have the mean field removed [though the modification is trivial via Eq. (17)]: Here, the issue of the mean field is not important and we are interested in the time spread of the total field.

The pulse width is defined as the full width at half maximum intensity and so, for the unperturbed pulse, the width is simply $2(\ln 2)^{1/2}/\Delta\omega$. The width of the perturbed pulse τ_0 is investigated in the limit of large time–bandwidth product $Q\Phi\gg1$. In this case τ_0 is obtained from the solution of a quadratic equation yielding

$$y=-\zeta+(\zeta^2+\zeta+1/2)^{1/2},\quad \zeta=\beta_\tau\exp(Q^{-2})/\sqrt{8}, \tag{29}$$

$$\beta_\tau=\frac{\langle a^2\rangle}{\langle a\rangle^2}\frac{Q\Phi}{N-1}\frac{1}{\gamma^2}, \tag{30}$$

$$\tau_0^2=4\tau^2\ln y^{-2},\quad \tau=\Phi/\omega_0. \tag{31}$$

It proves useful to define the pulse spread relative to the unperturbed pulse and measured in units of inverse bandwidth $\tau_b=1/\Delta\omega$, so the spread is

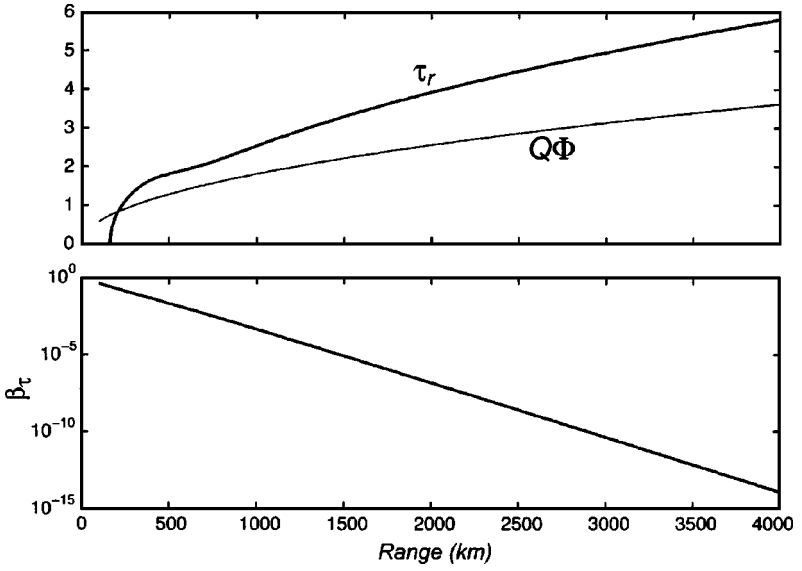


FIG. 3. Upper panel: Range variation of the relative pulse time spread (τ_r, measured in inverse bandwidths) using a ray theory model as discussed in the text. The time spread is computed using Eq. (32), with $Q=0.5$ (heavy line). Also plotted is the range dependence of $Q\Phi$ (light line). Lower panel: Range evolution of β_τ.

$$\tau_r^2 = 4(Q^2\Phi^2 \ln(y^{-2}(\beta_\tau, Q) - \ln 2). \quad (32)$$

The pulse spread is controlled by a parameter much like the scintillation index parameter β_{SI}; this parameter is called the spread parameter β_τ. For SI to approach 1, β_{SI} must approach zero. Pulse spread grows in two ways which can occur simultaneously; one in which β_τ increases and one in which Φ increases. In the first case when β_τ is large then $y \to 0$, making τ_0 large; in this case pulse spread, unlike SI, is affected by the frequency-independent phase shift factor γ^2, which under strong scattering conditions ($\gamma \to 0$) will contribute to increasing the spread. Thus, as seen before, conditions which favor an approach to saturation ($\beta_{\mathrm{SI}} \ll 1$) do not favor large pulse spread ($\beta_\tau \gg 1$). To demonstrate these effects, τ_0 is computed for the previously described ray theory model. Figure 3 shows the evolution in range of τ_r and it is seen that the pulse spread does not grow rapidly, but increases like $\sqrt{R}$ in accord with the increase in Φ.

V. SUMMARY

A kinematic model of broadband multipath scintillation is put forward to describe scintillation index SI and pulse timespread τ_0 dependencies on source Q, number of interferers N, travel-time and non-travel-time-induced phase delay variances Φ and Θ, and interferer's amplitude variability. This model assumes independent multipaths, and thus is strictly valid only in a neighborhood where SI is near 1, though the extent of this neighborhood is not known. It is found that SI can approach 1 from either above or below, and this behavior is critically dependent on the kurtosis of the multipath amplitude PDF ($\langle a^4\rangle/\langle a^2\rangle^2$), and the time–bandwidth product ($Q\Phi$). For narrow-band and broadband signals, the conditions to approach 1 from above are

$$\frac{\langle a^4\rangle}{\langle a^2\rangle^2} > 2, \quad \text{narrow band}, \quad (33)$$

$$\frac{\langle a^4\rangle}{\langle a^2\rangle^2} > \frac{2}{Q\Phi}, \quad \text{broadband}. \quad (34)$$

These relations have important implications for the partially saturated regime, since an approach to saturation from above 1 is associated with a strong focusing regime and a local maximum in SI. Bandwidth is seen to place strong constraints on the problem, and only multipath amplitude PDFs with relatively small kurtosis will result in an approach to saturation from below 1. The corrections to Rayleigh statistics at SI=1 are governed by the parameter β_{SI} [Eq. (23)] and here it is seen that increasing multipath amplitude kurtosis and time–bandwidth product competes with increasing N to drive the system to saturation. In this case one sees the counterintuitive result that increasing phase variance Φ^2 actually slows the approach to saturation. Further, it is shown that group delays (i.e., phase variations associated with travel-time changes, Φ^2) are the important phase variations leading to saturation.

Regarding pulse time spread τ_0, two mechanisms are shown to lead to pulse broadening. In one case, a pulse spread parameter $\beta_\tau \gg 1$ [Eq. (30)] leads to large pulse spread, and in the other increasing Φ leads to spread. The spread parameter β_τ is closely related to the saturation parameter β_{SI}, so that the condition $\beta_{\mathrm{SI}} \ll 1$ (which implies $\beta_\tau \ll 1$) for saturation is not conducive to large pulse spread through the first mechanism.

The results of this paper are kinematic, not dynamic, and it is the hope of the authors that these results will help guide the interpretation of numerical simulations, and further theory on the difficult topic of the approach to saturation.

ACKNOWLEDGMENTS

The inspiration for this work evolved out of several enlightening discussions with Fredrick Tappert over the applicability of ray methods to the saturation limit. Fred considered the ''Dead Horse'' relation [Eq. (20)] as the nail in the coffin of ray theory, since exponential increases in N and the amplitude moments do not necessarily lead to $\beta_{\mathrm{SI}} \to 0$. This is Woods Hole Oceanographic Institution Contribution Number 11030.

[1]Lossy media pose some extra difficulties; See for example, Creamer, 1996.
[2]This correction, however, breaks down for large n of order $(2/\alpha)^{1/2}$.
[3]The definition of Q is unconventional, as the reciprocal relation is often used (i.e., $\omega_0/\Delta\omega$). The present definition of Q provides for a cleaner notation in this study.
[4]Phase PDFs with variances greater than $(2\pi)^2$ can be problematical since in practice wrap-around effects would make it difficult, if not impossible, to distinguish amoung competing PDFs.
[5]In analyzing the K distribution, Jakeman and Pusey (1978) consider the cases where N has Poisson and negative binomial distributions.
[6]In ray theory the phase delay is related to the Maslov index μ by the relation $\theta = \mu\pi/2$. Computations by Beron-Vera *et al.* (2003) for long-range ray propagation through internal waves show that the PDF of Maslov index is very close to Gaussian.
[7]In Eq. (21) a term of the form $(N-1)\exp(-2Q^{-2})\gamma^8/N$ has been neglected since in the limit $Q\Phi \gg 1$ it is also expected that $\gamma \exp(-\Theta^2) \ll 1$. Numerical simulations of ray propagtion through internal waves by Beron-Vera *et al.* (2003) shows that at long range the variance of Maslov index justifies the neglect of γ.

Beron-Vera, F. J., Brown, M. G., Colosi, J. A., Tomsovic, S., Virovlyansky, A. L., Wolfson, M. A., and Zaslavsky, G. M. (**2003**). ''Ray dynamics in a long-range acoustic propagation experiment,'' J. Acoust. Soc. Am. **114**(3), 1226–1242.

Brown, M. G., Colosi, J. A., Tomsovic, S., Virovlyansky, A. L., Wolfson, M. A., and Zaslavsky, G. M. (**2003**). ''Ray dynamics in ocean acoustics,'' J. Acoust. Soc. Am. **113**(5), 2533–2547.

Chiu, C. S., Ramp, S. R., Miller, C. W., Lynch, J. F., Duda, T. F., and Yang, T. Y. (**2004**). ''Acoustic intensity fluctuations induced by South China Sea internal tides and solitons,'' IEEE J. Ocean. Eng. (in press).

Colosi, J. A., and Flatté, S. M. (**1996**). ''Mode coupling by internal waves for multimegameter acoustic propagation in the ocean,'' J. Acoust. Soc. Am. **100**(6), 3607–3620.

Colosi, J. A., Tappert, F., and Dzieciuch, M. (**2001**). ''Further analysis of intensity fluctuations from a 3252-km acoustic propagation experiment in the eastern North Pacific Ocean,'' J. Acoust. Soc. Am. **110**(1), 163–169.

Creamer, D. (**1996**). ''Scintillating shallow water waveguides,'' J. Acoust. Soc. Am. **99**(5), 2825–2838.

Dashen, R. (**1979**). ''Path integrals for waves in random media,'' J. Math. Phys. **20**, 894–920.

Dashen, R., Flatté, S. M., and Reynolds, S. (**1985**). ''Path-integral treatment of acoustic mutual coherence functions for rays in a sound channel,'' J. Acoust. Soc. Am. **77**(5), 1716–1722.

Dashen, R., Wang, G. Y., Flatté, S. M., and Bracher, C. (**1993**). ''Moments of intensity and log intensity: new asymptotic results for waves in power-law media,'' J. Opt. Soc. Am. **10**(6), 1233–1242.

Dozier, L. B., and Tappert, F. D. (**1978**). ''Statistics of normal mode amplitudes in a random ocean. I. Theory,'' J. Acoust. Soc. Am. **63**, 353–365.

Duda, T. F., Flatté, S. M., Colosi, J. A., Cornuelle, B. D., Hildebrand, J. A., Hodgekiss, W. S., Worcester, P. F., Howe, B. M., Mercer, J. A., and Spindel, R. C. (**1992**). ''Measured wavefront fluctuations in 1000-km pulse propagation in the Pacific Ocean,'' J. Acoust. Soc. Am. **92**(2), 939–955.
Dyer, I. (**1970**). ''Statistics of sound propagation in the ocean,'' J. Acoust. Soc. Am. **48**(1), 337–345.
Ewart, T. (**1989**). ''A model of the intensity probability distribution for wave propagation in random media,'' J. Acoust. Soc. Am. **86**(4), 1490–1498.
Ewart, T., and Percival, D. B. (**1986**). ''Forward scattered waves in random media—The probability distribution function of intensity,'' J. Acoust. Soc. Am. **80**(6), 1745–1753.
Flatté, S. (**1983**). ''Wave propagation through random media: Contributions from ocean acoustics,'' Proc. IEEE **71**(11), 1267–1294.
Flatté, S., Bracher, C., and Wang, G. Y. (**1994**). ''Probability-density functions of irradiance for waves in atmospheric turbulence calculated by numerical simulation,'' J. Opt. Soc. Am. **11**(7), 2080–2092.
Flatté, S., Dashen, R., Munk, W., Watson, K., and Zachariasen, F. (**1979**). *Sound Transmission through a Fluctuating Ocean* (Cambridge University Press, Cambridge).
Fredricks, A., Colosi, J. A., Lynch, J. F., Gawarkeiwicz, G., Chiu, C. S., and Abbot, P. ''Analysis of multipath scintillations observed during the summer 1996 New England shelfbreak PRIMER study,'' J. Acoust. Soc. Am. (submitted).
Gasiorowicz, S. (**1974**). *Quantum Physics* (Wiley, New York).
Jakeman, E., and Pusey, P. N. (**1978**). ''Significance of *K* distributions in scattering experiments,'' Phys. Rev. Lett. **40**(9), 546–550.
Martin, J., and Flatté, S. M. (**1988**). ''Intensity images and statistics from numerical simulation of wave propagation in 3-D random media,'' Appl. Opt. **27**(11), 2111–2126.
Martin, J., and Flatté, S. M. (**1990**). ''Simulation of point-source scintillations through three-dimensional random media,'' J. Opt. Soc. Am. **7**(5), 838–847.
Mikhalevsky, P. N. (**1982**). ''Envelope statistics of partially saturated processes,'' J. Acoust. Soc. Am. **72**(1), 151–158.
Munk, W., and Wunsch, C. (**1983**). ''Ocean acoustic tomography: Rays and modes,'' Rev. Geophys. Space Phys. **21**, 777–793.
Rice, S. O. (**1944**). ''Mathematical analysis of random noise,'' Bell Syst. Tech. J. **23**, 282.
Strohbehn, J. W. (**1978**). ''Modern Theories in the Propagation of Optical Waves in a Turbulent Medium,'' in Topics in Applied Physics: Laser Beam Propagation in the Atmosphere, 45–106.
Tappert, F. D., and Tang, X. (**1996**). ''Ray chaos and eigenrays,'' J. Acoust. Soc. Am. **99**(1), 185–195.
Wolfson, M. A., and Tomsovic, S. (**2001**). ''On the stability of long-range sound propagation through a structured ocean,'' J. Acoust. Soc. Am. **109**(6), 2693–2703.

Development of a laser photothermoacoustic frequency-swept system for subsurface imaging: Theory and experiment

Ying Fan and Andreas Mandelis
Center for Advanced Diffusion-Wave Technologies, Department of Mechanical and Industrial Engineering, 5 King's College Road, University of Toronto, Toronto M5S 3G8, Canada

Gloria Spirou
Department of Medical Biophysics, University of Toronto and Ontario Cancer Institute/Princess Margaret Hospital, 610 University Ave, Toronto M5G 2M9, Canada

I. Alex Vitkin
Department of Radiation Oncology, University of Toronto and Ontario Cancer Institute/Princess Margaret Hospital/University Health Network, 610 University Ave, Toronto M5G 2M9, Canada

(Received 23 April 2004; revised 28 September 2004; accepted 29 September 2004)

In conventional biomedical photoacoustic imaging systems, a pulsed laser is used to generate time-of-flight acoustic information of the subsurface features. This paper reports the theoretical and experimental development of a new frequency-domain (FD) photo-thermo-acoustic (PTA) principle featuring frequency sweep (chirp) and heterodyne modulation and lock-in detection of a continuous-wave laser source at 1064 nm wavelength. PTA imaging is a promising new technique which is being developed to detect tumor masses in turbid biological tissue. Owing to the linear relationship between the depth of acoustic signal generation and the delay time of signal arrival to the transducer, information specific to a particular depth can be associated with a particular frequency in the chirp signal. Scanning laser modulation with a linear frequency sweep method preserves the depth-to-delay time linearity and recovers FD-PTA signals from a range of depths. Preliminary results performed on rubber samples and solid tissue phantoms indicate that the FD-PTA technique has the potential to be a reliable tool for biomedical depth-profilometric imaging. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1819393]

PACS numbers: 43.60.Lq, 43.80.Vj, 43.80.Qf [YHB] Pages: 3523–3533

I. INTRODUCTION

The photoacoustic (PA), or more precisely, photo-thermo-acoustic (PTA) effect, is the process of light being absorbed by a material, creating a temperature change followed by a localized volume expansion leading to the generation of acoustic waves. In addition to obvious applications in the area of subsurface depth profilometry of defects in materials,[1] there have been many advances in applying photoacoustic monitoring to soft tissue imaging, cancerous lesion detection, and subdermal depth profilometry in the last decade. In recent years, the field of laser optoacoustic (or photoacoustic) applications to biomedical imaging has enjoyed very rapid development,[2–4] becoming the object of broader attention by the biophotonics community.[5] This is so because PA detection has shown concrete promise of depth profilometric imaging in turbid media at depths significantly larger than accessible by purely optical methodologies.[6] In state-of-the-art laser PA instrumentation and measurement systems in turbid media, as developed by some of the major research groups in this field,[6–10] pulsed lasers have always been the PA signal source of choice. The major reasons for this choice are twofold: (a) Following optical absorption of a short laser pulse by turbid tissue, optical-to-thermal energy conversion and localized PTA volume expansion detection, the acoustic propagation received within approximately 1 μs after the end of the laser pulse is essentially thermally adiabatic: it carries information about the thermal shape of the absorber, which substantially coincides with its geometric shape before any significant heat conduction can deform the image at later times;[11] (b) In pulsed photoacoustics a large amount of the available energy is imparted to the Fourier spectral components of the PTA signal response, which correspond to the early times (or high frequencies) after the arrival of the acoustic pulse at the transducer, thus yielding acceptable signal-to-noise ratios under coadded transient pulse detection.[12] Pulsed PTA detection, however, presents disadvantages in terms of laser jitter noise, acoustic and thermal noise within the wide bandwidth of the transducer, easy-to-miss nascent factors of contrast-generating subsurface features amidst the series of stronger acoustic reflections for other neighboring structural inhomogeneities and boundaries, as well as strong background signals from sound-scattering tissue. These mechanisms tend to limit system detectivity and signal-to-noise ratio (SNR) and construe important limitations because they may seriously compromise the contributions to the signal of contrast-generating subsurface features and thus limit the ability to monitor nascent and small-size tumors by the PTA process. In addition, very large pulsed-laser peak irradiances incident on living tissue may have detrimental effects, and for this reason very low average pulse energies (<5 mJ) are used at the expense of SNR. Besides, it is difficult to construct linear, low-noise detection systems for a wide range of pulsed amplitudes, a common requirement for patient-specific diagnostics. Nor-

mally, acoustic responses from turbid media are time gated and Fourier transformed into the frequency domain in order to determine and match the peak response of the transducer with the frequency contents of the PTA signal.[13] Quantitatively, tissue inhomogeneity parameter measurements are derived from the peak of the frequency spectrum of the transformed signal.

Frequency-domain (FD) PTA methodologies can offer alternative detection and imaging schemes with concrete advantages over pulsed laser photoacoustics. These advantages include: (a) Low fluence of the harmonic or frequency-swept (chirped) laser modulation, with the concomitant advantage of a much higher tissue damage threshold. A combination of harmonically modulated and chirped detection methodologies[12] can overcome the possible disadvantage of lower signal levels under single-frequency harmonic modulation at high, thermally adiabatic frequencies ($\sim$ MHz), while retaining the speed and wide temporal range of pulsed laser responses. The superior signal-to-noise ratio of the ultranarrow lock-in amplifier bandpass filter can offset much of the SNR deterioration at MHz frequencies. Frequency chirps may also recover the strength of the high-frequency Fourier components through fast-Fourier transformation of the frequency-domain transfer function to time-domain impulse-response, thus matching the major advantage of pulsed-laser excitation; (b) Depth profilometry over a very wide range of frequencies. The depth range in turbid media depends on the acoustic velocity and the optical extinction coefficient at the probe wavelength. Therefore, through precise time gating by use of optimal boundary (minimum and maximum) frequencies in a laser chirp, depth localization can be effected in a much more controllable fashion than with a pulsed laser. The chirp can be further tailored to center around a particular feature of interest, thus offering a high degree of depth controllability and efficient rejection of undesirable acoustic reflections outside the time gate, thus minimizing unwanted interference effects. (c) Possible parallel multichannel lock-in signal processing and image generation in quasireal time;[14,15] (d) A substantially wider signal dynamic range through use of lock-in filtering; and (e) A simple instrumental normalization procedure through division with a reference signal in the frequency-domain, as opposed to nontrivial deconvolution in the time domain, especially with highly nonlinear (e.g., resonant) ultrasonic transducers. Yet, FD and/or hybrid methods have not been pursued historically in biomedical PTA imaging.

In this paper we report the theoretical and experimental development of a new frequency-swept lock-in PTA depth profilometric imaging methodology for ultimate use with biomedical turbid media. The problem of linear PTA pressure wave generation from a solid immersed in a liquid is addressed and a mathematical model for a three-layered liquid–solid–liquid system is formulated. A modulated cw laser beam incident on the top surface of an absorbing solid is considered as the source of PTA waves inside the solid. Coupled small-amplitude (linear) wave equations are solved in the solid and fluid, and the resulting predictions are compared to experimental data obtained using a laser operating within the relatively nonabsorbing range of human tissue (1064 nm) and a commercial acoustic transducer as the receiver. The present one-dimensional model also accounts for acoustic attenuation and optical absorption effects. The single-frequency model was further extended to include linearly frequency-swept laser amplitude modulation via chirping the laser source, and theoretical data were compared to experimental results on a rubber sample immersed in water. The results indicate that the frequency-domain PTA field is closely associated with the location and dimensions of a subsurface object, its acoustic impedance, acoustic damping property, and optical absorption coefficient. By scanning across a sample surface area of interest, 2D and/or 3D images can be obtained by combining the depth profilometric scans obtained at each point and at a fixed acoustic delay time. Preliminary experimental results using artificial subsurface defects in rubber and solid tissue phantoms are also presented. This method allows one to increase subsurface feature resolution by time gating and imaging only selected depth slices to the exclusion of other overlying or underlying features.

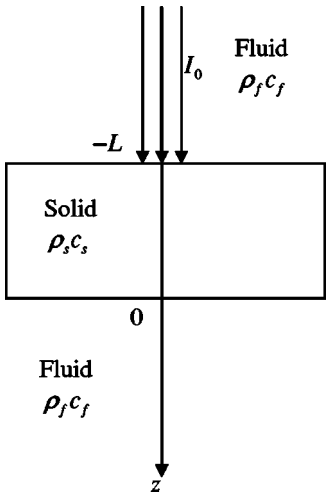


FIG. 1. Geometry used for formulating the frequency-swept lock-in PTA problem.

II. THEORY OF PHOTO-THERMO-ACOUSTIC WAVE GENERATION IN LAYERED STRUCTURES

A. Coupled PTA boundary value problems

Figure 1 shows the geometry used for the one-dimensional mathematical model. The configuration closely corresponds to the experimental geometry. It contains three coupled layers: the top and bottom layers are composed of a fluid, with the middle layer composed of a solid. The top layer is assumed to be semi-infinite fluid and occupies the spatial region $-\infty < z \leqslant -L$. It has density ρ_f, and speed of sound c_f. The solid layer has thickness L, density ρ_s, speed of sound c_s, specific heat at constant pressure C_{Ps}, optical absorption coefficient at the laser wavelength μ_a, bulk modulus K_s, and isobaric volume expansion coefficient β_s. For the quantitative measurements on nonturbid media presented in this work, the scattering coefficient of the medium and the contributions and complications of the diffuse photon-density wave replacing the purely absorbing ther-

moacoustic signal source are ignored for simplicity. However, the mathematical formalism to deal with these issues has been developed[16,17] and will be applied to quantitative measurements of light-scattering phantoms, such as those used qualitatively in Sec. III, in a future publication. Since the thermal diffusion length, $\mu(\omega)=\sqrt{2\alpha_s/\omega}$, is negligible compared to the diameter of the heat source ($\sim$3 mm), the mathematical model can be considered in the 1D regime. The bottom layer extends from $0\leqslant z<\infty$. We do not consider the finite thickness of the bottom layer because in our experiments no reflections from the fluid–container interface were observed.

An analytical solution of the coupled PTA problem in the form of spectral integrals can be obtained by converting the time-domain equations to their frequency-domain counterparts using Fourier transforms (FTs).[18,19] The spectral Fourier component of the heat conduction equation in the solid is the thermal-wave equation

$$\frac{\partial^2}{\partial z^2}\theta_s(z,\omega)-\left(\frac{i\omega}{\alpha_s}\right)\theta_s(z,\omega)=-\frac{1}{\lambda_s}H(z,\omega),\quad -L\leqslant z\leqslant 0, \tag{1}$$

where $\theta_s(z,\omega)$ is the temperature rise above ambient. α_s and λ_s are, respectively, the thermal diffusivity and conductivity of the solid medium. The spectral component $H(z,\omega)$ at any angular frequency ω of the thermal source is given by

$$H(z,\omega)=\mu_a I_0 e^{-\mu_a(L+z)+i\omega t}. \tag{2}$$

Here, I_0 is the laser fluence and $e^{i\omega t}$ is the modulation factor. The term $e^{i\omega t}$ is implied in each term of the solutions and will be omitted hereafter. Since the thermal diffusion length, $\mu(\omega)=\sqrt{2\alpha_s/\omega}$, is negligible compared to the sample thickness, the solid layer can be considered as semi-infinite. The general solution of Eq. (1) is

$$\theta_s(z,\omega)=Ge^{-\sigma_s(z+L)}+Ee^{-\mu_a(z+L)}, \tag{3}$$

where $\sigma_s^2=i\omega/\alpha_s$ is the complex thermal-wave diffusion coefficient in the solid. G is an unknown coefficient that needs to be determined and E is the source term defined as

$$E=\frac{\mu_a I_0}{\lambda_s(\sigma_s^2-\mu_a^2)}. \tag{4}$$

Assuming a nonviscous fluid with no direct internal PTA source, the FT component of the heat conduction equation inside such a fluid can be written as

$$\frac{\partial^2}{\partial z^2}\theta_f(z,\omega)-\left(\frac{i\omega}{\alpha_f}\right)\theta_f(z,\omega)=0,\quad -\infty<z\leqslant -L,\quad 0\leqslant z<\infty, \tag{5}$$

where $\theta_f(z,\omega)$ is the temperature rise and α_f is the thermal diffusivity of the fluid. The general solution for Eq. (5) can be written as

$$\theta_f(z,\omega)=Fe^{\sigma_f(z+L)}. \tag{6}$$

Here, F is an unknown coefficient and $\sigma_f^2=i\omega/\alpha_f$.

By using the thermal boundary conditions at the fluid/solid interface

$$\theta_f(-L,\omega)=\theta_s(-L,\omega),$$

$$\lambda_s\frac{\partial}{\partial z}\theta_s=\lambda_f\frac{\partial}{\partial z}\theta_f, \tag{7}$$

one can determine the constants F and G as

$$F=\frac{E(\lambda_s\sigma_s-\lambda_s\mu_a)}{(\lambda_f\sigma_f+\lambda_s\sigma_s)}, \tag{8}$$

$$G=-Eb_{sf}, \tag{9}$$

where

$$b_{sf}=\frac{(\lambda_s\mu_a+\lambda_f\sigma_f)}{(\lambda_f\sigma_f+\lambda_s\sigma_s)} \tag{10}$$

By introducing in the solid a particle/molecule displacement potential, $\phi_s(z,\omega)$, the coupled wave equations in the solid and fluid can be easily solved. The displacement potential is related to the displacement vector, $U_s(z,\omega)$, as

$$U_s(z,\omega)=\frac{\partial}{\partial z}\phi_s(z,\omega),\quad -L\leqslant z\leqslant 0. \tag{11}$$

Due to laser PTA excitation by a large spot-size laser beam, only longitudinal waves are assumed to propagate in an isotropic solid. This assumption allows the use of the Helmholtz equation, which is satisfied by the displacement potential, ϕ_s

$$\frac{\partial^2}{\partial z^2}\phi_s(z,\omega)+k_s^2\phi_s(z,\omega)=\left(\frac{K_s\beta_s}{\rho_s c_s^2}\right)\theta_s(z,\omega), \tag{12}$$

where $k_s=\omega/c_s$ is the acoustic wave number in the solid for small-amplitude acoustic perturbations. The general solution to this equation is

$$\phi_s(z,\omega)=Ae^{ik_s z}+Be^{-ik_s z}+Ce^{-\sigma_s(z+L)}+De^{-\mu_a(z+L)}. \tag{13}$$

Here, A and B are integration constants to be determined. Constants C and D are found to be

$$C=\frac{-\mu_a I_0 K_s\beta_s b_{sf}}{\rho_s c_s^2\lambda_s(\sigma_s^2-\mu_a^2)(\sigma_s^2+k_s^2)},$$

$$D=\frac{\mu_a I_0 K_s\beta_s}{\rho_s c_s^2\lambda_s(\sigma_s^2-\mu_a^2)(\alpha^2+k_s^2)}. \tag{14}$$

Inside the fluid, since wave sources have a potential nature, liquid motion will be potential motion. By introducing a scalar potential of the velocity field

$$v(z,\omega)=\frac{\partial}{\partial z}\psi_{fi}(z,\omega),\quad -\infty<z\leqslant -L,\quad 0\leqslant z<\infty, \tag{15}$$

where the subscript $i=1,2$ indicates the top and bottom fluid, respectively, one can obtain the photo-thermo-acoustic wave equation [Eq. (16)] for a nonviscous fluid[19]

$$\frac{\partial^2}{\partial z^2}\psi_{fi}(z,\omega)+k_f^2\psi_{fi}(z,\omega)=0, \tag{16}$$

where $k_f=\omega/c_f$ is the wave number for small-amplitude acoustic perturbations in the fluid.

The small-amplitude pressure change in the fluid is related to the velocity potential, Ψ_{fi}, by

$$P(z,\omega) = -i\omega\rho_f \psi_{fi}(z,\omega). \tag{17}$$

The general solutions to Eq. (16) can be written as

$$\psi_{f1}(z,\omega) = C_1 e^{ik_f(z+L)}, \quad -\infty < z \leqslant -L,$$

$$\psi_{f2}(z,\omega) = C_2 e^{-ik_f z}, \quad 0 \leqslant z < \infty. \tag{18}$$

The constants (A,B,C_1,C_2) in Eqs. (13) and (18) can be determined through the boundary conditions of stress and velocity continuity at the two interfaces $z=0,-L$[19]

$$\rho_s c_s^2 \frac{\partial^2}{\partial z^2}\phi(0,\omega) - K_s\beta_s\theta_s(0,\omega) = -P(0,\omega) = i\omega\rho_f\psi_{f2}(0,\omega),$$

$$\rho_s c_s^2 \frac{\partial^2}{\partial z^2}\phi(-L,\omega) - K_s\beta_s\theta_s(-L,\omega) = P(-L,\omega) = i\omega\rho_f\psi_{f1}(-L,\omega),$$

$$i\omega\frac{\partial}{\partial z}\phi_s(0,\omega) = \frac{\partial}{\partial z}\psi_{f2}(0,\omega),$$

$$i\omega\frac{\partial}{\partial z}\phi_s(-L,\omega) = \frac{\partial}{\partial z}\psi_{f1}(-L,\omega). \tag{19}$$

Using the boundary conditions (19), coefficient C_1 can be determined as

$$C_1(\omega) = \frac{\mu_a I_0 K_s\beta_s c_f [A_n(\omega)e^{ik_s L} + B_n(\omega)e^{-ik_s L} + 2i\omega c_f b_{sf}(\mu_a^2+k_s^2)(i\sigma_s\rho_f + k_f\rho_s)e^{-\sigma_s L} + C_n(\omega)]}{c_s\omega\lambda_s(\sigma_s^2-\mu_a^2)(\mu_a^2+k_s^2)(\sigma_s^2+k_s^2)[(\rho_s c_s+\rho_f c_f)^2 e^{ik_s L} - (\rho_s c_s - \rho_f c_f)^2 e^{-ik_s L}]},$$

where

$$A_n(\omega) = (\rho_s c_s + \rho_f c_f)[ic_s k_s^2(\sigma_s^2 + k_s^2 - \mu_a^2 b_{sf} - k_s^2 b_{sf}) - \omega(\mu_a\sigma_s^2 + \mu_a k_s^2 - \sigma_s\mu_a^2 b_{sf} - \sigma_s k_s^2 b_{sf})],$$

$$B_n(\omega) = (\rho_s c_s - \rho_f c_f)[\omega(\mu_a\sigma_s^2 + \mu_a k_s^2 - \sigma_s\mu_a^2 b_{sf} - \sigma_s k_s^2 b_{sf}) + ic_s k_s^2(\sigma_s^2 + k_s^2 - \mu_a^2 b_{sf} - k_s^2 b_{sf})],$$

and

$$C_n(\omega) = 2i\omega c_f(\sigma_s^2 + k_s^2)(-k_f\rho_s - i\mu_a\rho_f)e^{-\mu_a L}. \tag{20}$$

Finally, the spectral component of the experimentally measurable pressure wave in the front fluid can be written as

$$P(z,\omega) = -i\omega\rho_f C_1(\omega)e^{ik_f(z+L)}. \tag{21}$$

B. Acoustic attenuation effects

Equation (12) for the FT of the one-dimensional displacement potential can be modified to allow for acoustic attenuation effects in the solid. The resulting Helmholtz equation can be written as

$$\frac{\partial^2}{\partial z^2}\phi_s(z,\omega) + k_{\gamma s}^2\phi_s(z,\omega) = \left(\frac{K_s\beta_s}{\rho_s c_s^2}\right)\theta_s(z,\omega),$$

where

$$k_{\gamma s} = \frac{\omega}{c_s} - i\frac{\gamma\omega}{2\pi}. \tag{22}$$

Here, γ is the acoustic attenuation coefficient with dimensions of Nepers/m. By assuming the wave number, $k_{\gamma s}$, to be complex, its imaginary part would account for the acoustic absorptive properties of the material. Therefore, the effects of acoustic attenuation can be incorporated in the PTA pressure wave equations by simply substituting $k_{\gamma s}$ for k_s in all relevant spectral equations described above.

C. Special case: Optically opaque solid

For an optically opaque solid, the exponential term, $e^{-\mu_a L}$, can be ignored due to the vanishingly small optical absorption length, μ_a^{-1}. Therefore, by using the boundary conditions [Eq. (19)], coefficient C_1 can be relatively simplified

$$C_1(\omega) = \frac{\mu_a I_0 K_s\beta_s c_f [A_n(\omega)e^{ik_s L} + B_n(\omega)e^{-ik_s L} + 2i\omega c_f b_{sf}(\mu_a^2+k_s^2)(i\sigma_s\rho_f + k_f\rho_s)e^{-\sigma_s L}]}{c_s\omega\lambda_s(\sigma_s^2-\mu_a^2)(\mu_a^2+k_s^2)(\sigma_s^2+k_s^2)[(\rho_s c_s+\rho_f c_f)^2 e^{ik_s L} - (\rho_s c_s - \rho_f c_f)^2 e^{-ik_s L}]},$$

where

$$A_n(\omega) = (\rho_s c_s + \rho_f c_f)[ic_s k_s^2(\sigma_s^2 + k_s^2 - \mu_a^2 b_{sf} - k_s^2 b_{sf}) - \omega(\mu_a\sigma_s^2 + \mu_a k_s^2 - \sigma_s\mu_a^2 b_{sf} - \sigma_s k_s^2 b_{sf})],$$

$$B_n(\omega) = (\rho_s c_s - \rho_f c_f)[\omega(\mu_a\sigma_s^2 + \mu_a k_s^2 - \sigma_s\mu_a^2 b_{sf} - \sigma_s k_s^2 b_{sf}) + ic_s k_s^2(\sigma_s^2 + k_s^2 - \mu_a^2 b_{sf} - k_s^2 b_{sf})]. \tag{23}$$

Again, by substituting coefficient, $C_1(\omega)$, into Eq. (21), one can obtain the spectral component of the pressure wave in the top fluid layer, which is a complex number. Both the amplitude and phase of the pressure field will be plotted and compared to the experimental frequency-scan results in Sec. III.

The frequency-domain numerical results will be further converted into time domain using inverse Fourier transfor-

TABLE I. Optical and elastic properties used as input parameters for the mathematical model.

μ_s (1/cm)	K_s (N/m^2)	β_s (1/C)	c_s (m/s)	c_f (m/s)	ρ_s (kg/m^3)
100	1.5×10^9	6×10^{-4}	1600	1474	1320
ρ_f (kg/m^3)	λ_s (W/mK)	λ_f (W/mK)	α_s (m^2/s)	α_f (m^2/s)	γ (Nepers/m)
998	0.13	0.61	0.06×10^{-6}	0.1×10^{-6}	55×10^{-6}

mation (IFT) and used as theoretical fits to the experimental data in frequency-swept measurements (Sec. IV).

III. PTA FREQUENCY-SCAN SIMULATIONS, INSTRUMENTATION, AND EXPERIMENTAL RESULTS

Theoretical simulations were performed for the simple case of a black rubber sample immersed in water. There are two reasons to use rubber samples in this project. First is to verify the mathematical model. The numerically simulated results show the difference between PTA signals generated by an opaque and a translucent layer. The rubber sample is an example of such an opaque solid. Second, the rubber samples can provide us with preliminary results to test and optimize the instrumental aspects of the imaging system while we are still trying to achieve a better understanding of wave generation from scattering phantoms. Two input parameters, the thickness of the solid and its optical absorption coefficient, were changed independently for each simulation to illustrate the frequency-domain PTA signal generation through the developed theory. Table I (Refs. 20–22) presents the optical and elastic properties used as input parameters for the mathematical model. For a fully opaque solid, Eqs. (23) and (21) were used to calculate the laser-induced acoustic field within the actual experimental frequency range. Figures 2(a) and (b) display the theoretical amplitude and phase of the pressure field obtained from a 2.5-mm-thick rubber piece. Figures 3(a) and (b) are the numerically simulated results obtained from 5-mm-thick rubber using the developed theory. The frequency-domain amplitude plots contain a series of oscillations, superposed on a decaying envelope. Due to the intermediate step of optical energy conversion into thermal waves, the spectral energy content decreases with increasing laser modulation frequency. This well-known thermal-wave effect controls the further conversion of thermal energy into acoustic waves, thus turning the PTA signal envelope into a low-pass-equivalent filter. The amplitude oscillations are due to standing acoustic waves between the top and bottom surfaces of the rubber. For an opaque solid, laser light is only absorbed at the top surface of the solid, where thermoelastic waves are then generated. In the one-dimensional model, the acoustic waves propagate in two directions, upward into the fluid and downward into the solid material. In a 3D model, this longitudinal character of the photo-thermo-acoustic wave would still prevail as the signal is dominated by boundary reflections. The wave components reflected from the bottom surface will lag in phase compared to those generated at the top surface of the solid. Under harmonic excitation, a standing wave will be formed, so that at frequencies

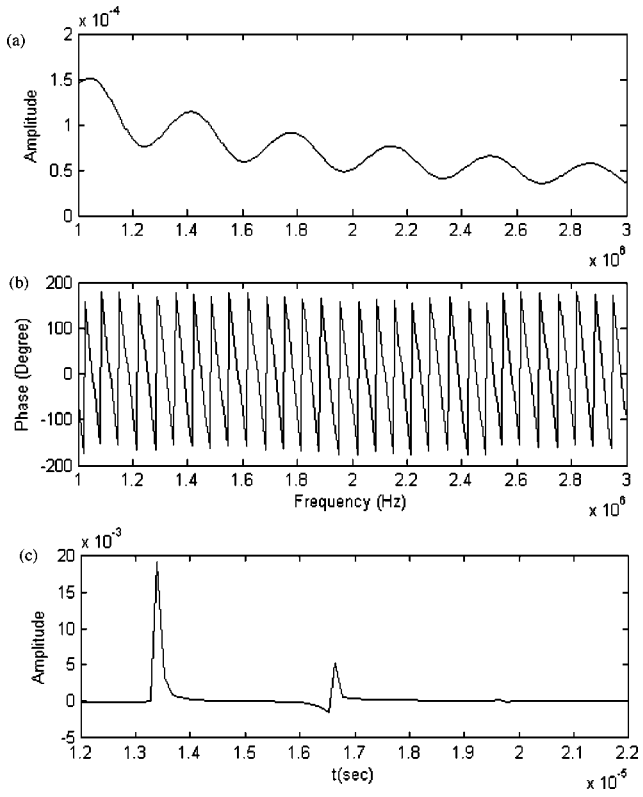


FIG. 2. Frequency-domain (amplitude and phase) and time-domain (amplitude) simulation of the PTA pressure field obtained using a 2.5-mm-thick rubber piece ($\mu_a = 100\ cm^{-1}$).

$$f_n = \frac{(2n+1)c_s}{4L}, \quad n = 0,1,2,3..., \tag{24}$$

and

$$f_n' = \frac{nc_s}{2L}, \quad n = 1,2,3..., \tag{25}$$

there will be destructive interference minima and constructive interference maxima, respectively, as shown in the amplitude plots of Figs. 2(a) and 3(a). It is clear that the frequency values, f_n and f_n' are related to the thickness of the opaque solid.

The PTA phase patterns of Figs. 2(b) and 3(b) are related to sample thickness, the acoustic impedance at the sample–water interface, and the location of the receiver. Despite the appearance of these curves, the phase of the pressure field is continuous as a function of frequency. Since phase is measured within the range of $\pm\pi$, a 360-deg discontinuity occurs as the value of phase reaches $+\pi$ or $-\pi$, both experimentally and computationally.

Figures 2(c) and 3(c) are the time-domain amplitude plots of the simulated pressure field, corresponding to Figs. 2(a) and 3(a), respectively. These time-domain plots were converted from their frequency-domain counterparts using the inverse Fourier transformation (IFT). The two peaks in each plot indicate the acoustic waves generated at the top surface and reflected from the bottom surface of the solid. The delay time between the two peaks is equal to twice the thickness of the solid, divided by the speed of sound in the medium.

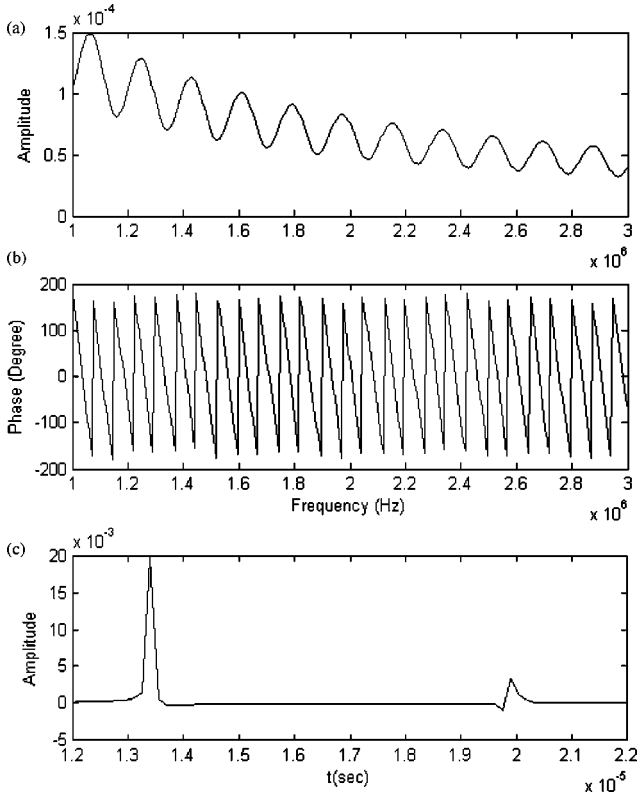


FIG. 3. Frequency-domain (amplitude and phase) and time-domain (amplitude) simulation of the PTA pressure field obtained using a 5-mm-thick rubber piece ($\mu_a = 100\ \mathrm{cm}^{-1}$).

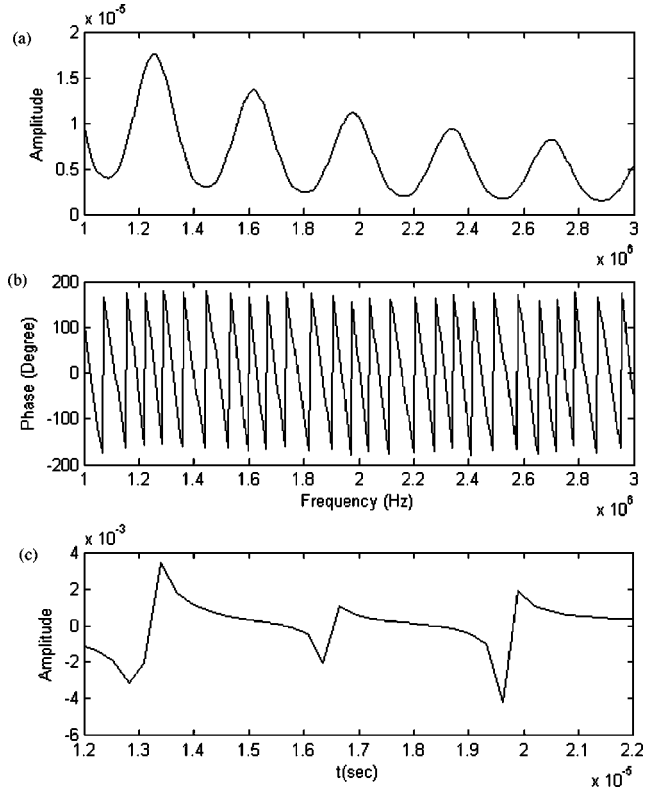


FIG. 4. Frequency-domain (amplitude and phase) and time-domain (amplitude) simulation of the PTA pressure field obtained from a 5-mm-thick translucent solid with the same thickness of rubber in Fig. 3 and optical absorption coefficient $\mu_a = 4\ \mathrm{cm}^{-1}$.

To illustrate the effects of varying optical absorption (penetration) depth on the PTA signal, Fig. 4 features the same rubber material as Fig. 3, but with a smaller optical absorption coefficient, representing a translucent solid. For such a solid, laser light can penetrate deeper into the bulk and create a distributed light/heat source. Due to the distributed heat sources, thermoelastic waves are generated by local compressions and expansions. All these acoustic waves created throughout the thickness of the translucent solid contribute to the total pressure field. Although the thicknesses of the solids in Figs. 3 and 4 are the same, different interference patterns are observed due to the differences in absorption depth. As for the phase plots, no significant changes can be observed between Figs. 3(b) and 4(b), owing to the very large range of phases involved. Figure 4(c) is the IFT time-domain amplitude plot corresponding to Fig. 4(a), and it contains three major peaks. The first two peaks are caused by the acoustic waves generated at the top and bottom liquid–solid interfaces. The third peak is caused by the first reflection of acoustic waves from the bottom surface. Comparing Fig. 4(c) to Fig. 3(c), which is obtained from an opaque solid with the same thickness and similar thermoelastic properties, the delay time between the first two peaks in Fig. 4(c) is half of that in Fig. 3(c). This is so because the acoustic waves created by the subsurface-distributed light/heat sources at the back interface of the translucent solid propagate backwards and are detected without requiring the forward wave to undergo reflections from the bottom surface. Another important feature caused by the decrease of the optical absorption coefficient is the decrease in signal amplitude and broadening of PTA response, which is also shown in Fig. 4(c). This feature is caused by the deeper penetration of light and a more even distribution of heat source inside the material.

A block diagram of the experimental setup for the frequency scan measurements is shown in Fig. 5. The optical source used to generate PTA pressure waves was an ytterbium fiber laser (IPG Photonics, 1064 nm). Using a function generator (Stanford Research Systems, DS345) to drive the acousto-optic modulator (Neos Technologies, N15180-1.06-Gap) chosen for its optimal modulation depth, the intensity of the laser beam was modulated harmonically at selected frequencies. The laser beam was focused onto the specimen. An acoustic mirror (Panametrics, F102) with a small hole to allow passage of the incident light beam was used to reflect

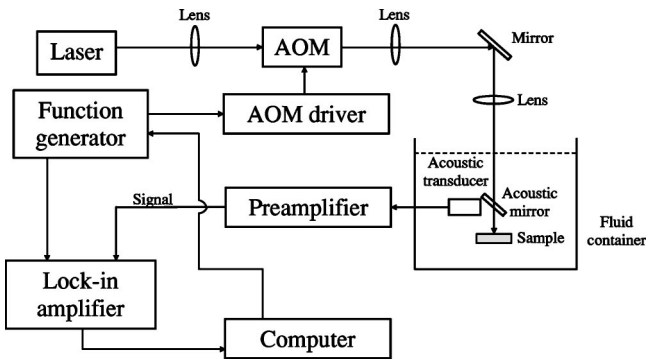


FIG. 5. Block diagram of the experimental PTA setup used for frequency scans.

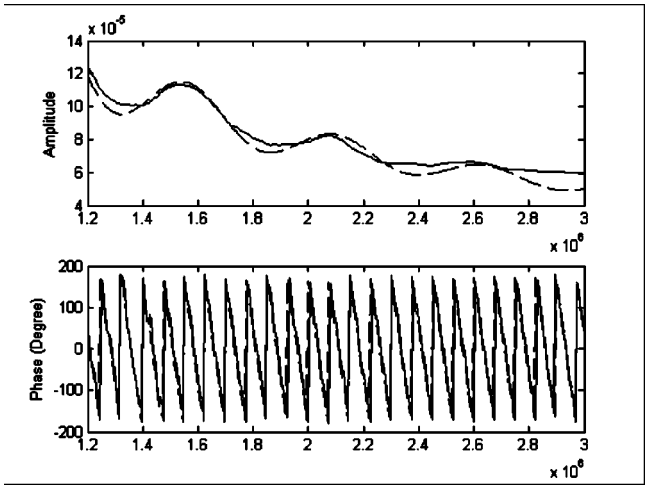


FIG. 6. Frequency-domain amplitude and phase of the PTA pressure field obtained from a 1.8-mm-thick rubber piece. Solid line: experimental results. Dashed line: best-fitted numerical results.

the sound waves to a commercial acoustic transducer (Panametrics, V382). The received signal was amplified by a preamplifier (Panametrics, 5676) and sent to a high-frequency lock-in amplifier (Stanford Research Systems, SR844). A computer program (MATLAB 6.0) was used to control the frequency of the function generator and acquire the data from the lock-in amplifier.

The specimen used for the frequency scan measurements was a 1.8-mm-thick piece of rubber immersed in water. Frequency scans were performed at several lateral positions. At each position, the scan was repeated five times within the frequency range of 1.2–3 MHz. The frequency step size was 50 kHz. The results at each lateral position were coadded and averaged to eliminate white noise. The frequency spectra were normalized using the transfer function curve of the acoustic transducer supplied by the manufacturer.

The experimental and best-fitted frequency-scan results using the system of Fig. 5 and Eq. (23) for an opaque solid are shown in Fig. 6. The thickness and density of the rubber sample were measured to be 1.8 mm and 1320 kg/m^3, respectively. The measured room temperature was 24 °C, from which the density and the speed of sound in water were calculated[23] to be 998 kg/m^3 and 1489 m/s, respectively. To obtain the best fits to the entire frequency record of the pressure responses, the exact values of the bulk modulus, K_s, isobaric volume expansion coefficient, β_s, acoustic attenuation coefficient, γ, thermal diffusivities, $\alpha_{f,s}$, and thermal conductivities, $\lambda_{f,s}$, were not as important as the speed of sound, and the optical absorption coefficient of the rubber sample. For these secondary parameters, the listed values in Table I were used for the numerical simulations. The important parameters used for the theoretical fits were the speed of sound, $c_s = 1800$ m/s, and the optical absorption coefficient, $\mu_a = 100$ cm^{-1}. The value of the optical absorption coefficient is close to the upper limit of the resolvable absorption coefficient range. Although this number is not excessively high, it is well-known that PA saturation sets in at around 5×10^3 cm^{-1}.[24] Above this limit, no significant changes were observed for the amplitude and phase plots, and the solid is considered as a black absorber. A good agreement is observed in Fig. 6, between the measured and calculated pressure field against frequency. The slight discrepancy in the amplitude plot is within the estimated range of uncertainties of the measurement system, which was mainly caused by nonuniformity of the specimen and system noise of the electrical circuit.

IV. PTA FREQUENCY SWEEPS AND HETERODYNED SUBSURFACE IMAGING

A rapid frequency-scanning (chirp generation) system was designed and implemented to allow for fast PTA depth profilometry of subsurface inhomogeneities through a wide range of depths at a fixed spatial coordinate. Besides their advantage of speed for full frequency-spectrum acquisition compared to point-to-point frequency scans, another major feature of frequency-swept thermoacoustic signals is their ability to be Fourier transformed into a time-delay domain inverse spectrum which carries the depth profilometric information in a series of time sequences equivalent to the impulse response of the acoustic system.[25,26] Therefore, frequency-swept PTA signals are hybrid between time- and frequency domains,[12] a unique property which can be used to enhance signal-to-noise ratio significantly through heterodyning and lock-in noise filtering. The block diagram of the frequency-sweep heterodyne subsystem used to modify the PTA signal generation and detection part of Fig. 5 is shown in Fig. 7. A linear frequency sweep (chirp) signal was generated by a function generator (FG1) to modulate the intensity of the ytterbium laser. The chirp signal of FG1 was triggered by a delay-pulse generator (Stanford Research Systems, DG 535), which was also used to trigger the second function generator (FG2). The output of FG2 was mixed with the output of the PTA ultrasonic transducer using a mixer (Mini-circuits, ZAD-3). The output of the mixer was further sent to a low-pass filter (LPF1, Stanford Research Systems, SR 640). This signal was then mixed with a single harmonic frequency, ω_0 using a second mixer. The single-frequency signal was generated by the internal oscillator of the high-frequency lock-in amplifier (LIA; Stanford Research Systems SR 844). The output signal of mixer 2 was filtered using the second low-pass filter (LPF2) and then sent to the LIA as the input signal. The amplitude and phase of the LIA output were stored in a computer for display and analysis.

The PTA signal generation flow chart associated with the circuit of Fig. 7 is described in Fig. 8. The chirp signal generated by FG1 can be written as $\cos[(a+bt)t]$, where $a = 1$ MHz is the starting frequency and $b = 4$ MHz/ms is the sweep rate. This chirp signal was triggered by the DG 535 delay-pulse generator, which also triggered FG2. The output of FG2, delayed by a controlled time, τ, through the delay pulse generator can be written as $\cos\{[a+b(t-\tau)]t\}$. The intensity of the laser beam was modulated by the acousto-optic modulator according to the chirp signal generated by FG1. At the acoustic transducer, the received signal can be written as $\cos\{[a+b(t-(z/c))]t\}$, where z represents the depth at which the acoustic signal originates, and c is the speed of sound in the probed medium. Due to the linear relationship between the depth and the delay time when the transducer

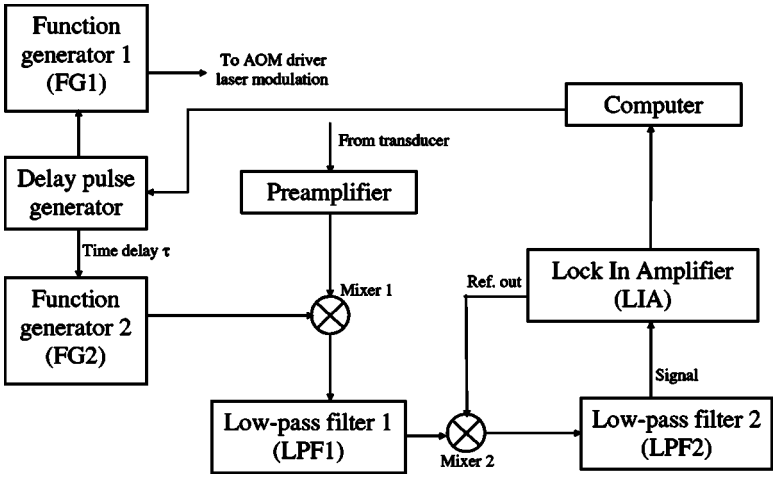


FIG. 7. Block diagram of the circuit used in the PTA system of Fig. 5 for frequency-swept (chirp) heterodyned depth profilometric scanning imaging.

receives the signal, this expression shows that the information at a certain depth can be related with the frequency components of the chirp signal. At mixer 1, the frequency components from the two input channels underwent addition and subtraction. The resulting high-frequency part was then removed by low-pass filter 1. The remaining low-frequency part of the signal was downshifted and contained a wide spectrum of frequencies, each with time-delayed information. Since both the input signals of mixer 1 contained the same starting frequency, a, this component was canceled out after subtraction. The output signal from LPF1 can be written as $\cos\{[b(\tau-(z/c))]t\}$. To perform PTA depth profilometry, mixing this output (second mixer) with a single harmonic fixed frequency, ω_0, generated by the internal oscillator of the LIA, and low-pass filtering the two sidebands of the mixed signal, yielded an output which can be represented by $\cos\{[\omega_0-b(\tau-(z/c))]t\}$. The frequency downshifted output was detected by the LIA. By scanning the chirp delay time, τ, a nonzero LIA signal output was expected at ω_0 only when $\tau=z/c$. Therefore, τ scans at a fixed spatial coordinate are equivalent to depth coordinate scans and can yield information from different probe depths in the sample at a fixed lateral coordinate point. Scanning over a predetermined two-dimensional area of the sample will generate a subsurface 3D image.

The solid line in Fig. 9 is the result of a single (x,y) point scan using the PTA imaging system, while the dashed line is the simulated result. The chirp signal covered a frequency range from 1 to 5 MHz and the step size of the delay time used for the scan was 0.1 μs. The specimen was a 5.4-mm-thick rubber piece, which had the same material properties as the one used for frequency-scan tests. The frequency-domain simulated result was calculated by substituting the thickness, material properties, and the chirp frequency range into Eqs. (23) and (21). The corresponding time-domain pressure field was obtained by applying inverse Fourier transformation to the frequency-domain result. The peaks were generated by solid–liquid interface reflections of acoustic waves. Good agreement was obtained between the numerical and experimental results. Even though the laser beam was focused, thus creating in principle a 3D PTA pro-

FG1: $\cos((a+bt)t)$;

a – starting freq.; b – sweep rate

Acoustic transducer: $\cos[(a+b(t-\frac{z}{c}))t]$

FG2: $\cos[(a+b(t-\tau))t]$

Mixer1+LPF1: $\cos[b(\tau-\frac{z}{c})t]$

LIA's internal oscillator: $\cos[\omega_0 t]$

Mixer2+LPF2: $\cos[(\omega_0-b(\tau-\frac{z}{c}))t]$

LIA registers a signal only when : $\tau = \frac{z}{c}$

FG —— Function generator
LPF —— Low pass filter
LIA —— Lock-in amplifier

FIG. 8. Frequency-swept and heterodyned PTA signal generation flow chart.

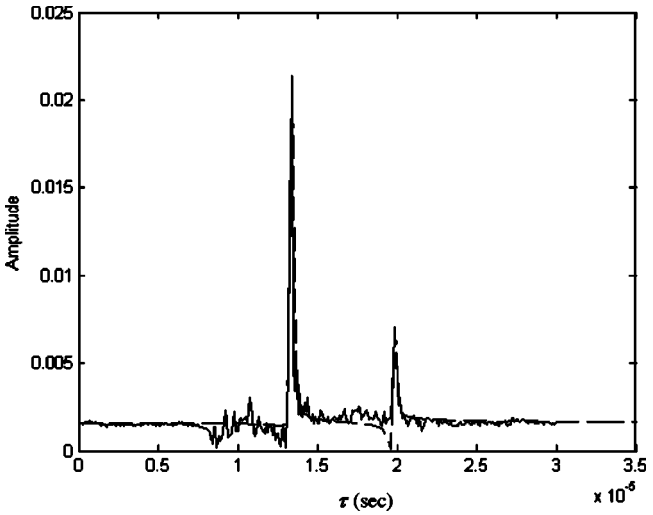


FIG. 9. A single-point scan of a 5.4-mm-thick rubber piece using the PTA chirp heterodyne scanning system. Solid line: experimental data; dashed line: numerical fits to the theory. The horizontal axis indicates time delay $\tau=z/c$.

file, the reflection spectrum from a short distance between interfaces is expected to be very similar to the 1D situation and to be adequately interpreted by our 1D theory, as observed. An interesting feature here is the increase of ''background'' noise in the measured pressure field after 8 μs. Eight μs corresponds to the delay time required by the acoustic transducer to receive the first echo generated at the acoustic mirror, Fig. 5. The background noise is due to the multiple reflections of sound waves within the probed medium.

Although both the single-frequency scan system (Fig. 5) and the chirp scan system (Fig. 7) can perform depth profilometric measurements, the PTA chirp scan system has significant advantages of precise depth control, fast scanning speed, and high signal-to-noise ratio. The modified chirp scan system shown in Fig. 7 performs an inverse Fourier transformation to convert the frequency-domain measured pressure field into time-domain signals, which carry the depth profilometric information in a series of time sequences equivalent to the impulse response of an acoustic system. This important step makes it possible to construct an operator-controlled subsurface slice-by-slice image in 3D space. In order to obtain a frequency-scan plot, as shown in Fig. 6, one has to perform point-by-point frequency-scans using the single frequency scan system, followed by signal averaging and normalization, which significantly hampers the operating speed. On the other hand, the chirp scan system has the capability to perform full frequency-spectrum acquisition in $<$ms duration. In conclusion, through heterodyning and lock-in noise filtering, the PTA chirp scan system can generate time-domain equivalent signals with a high signal-to-noise ratio, using a fraction of operating time.

Figure 10 presents the depth profilometric amplitude and phase images of a 2D scan over another rubber sample, which had a curved top surface. The meaning of a 2D scan is actually a surface line scan containing a full depth scan at each surface coordinate point. The thickness of this sample was 7 mm at the center and 3 mm around the edges. The scan was performed from left to right with a spatial step size of around 1 mm. The step size of the delay time was 0.1 μs, which corresponds to a distance of around 150 μm in water. The images clearly show the top surface and the bottom surface at two edges of the solid. However, at the center part of the cross section, the bottom surface is not visible, due to excessive acoustic attenuation in the thicker body of the material. From this result it was determined that the detectable thickness of this opaque rubber sample was around 3 mm.

Preliminary depth profilometric scans using solid phantoms with optical and acoustical properties resembling those of blood-deficient (normal) and blood-rich (vascularized) tissue were also performed. The solid phantoms are homogeneous, 3-mm-thick, monolithic, and made of polyacrylamide with different absorption coefficients. The reason we show this set of results is to partially verify the validity of the theory and compare the PTA signals obtained from an opaque vs a translucent solid without scattering. The second reason we present these results is to show potential applications of this technique to subsurface tissue imaging. The optical absorption coefficients at 1064 nm wavelength are

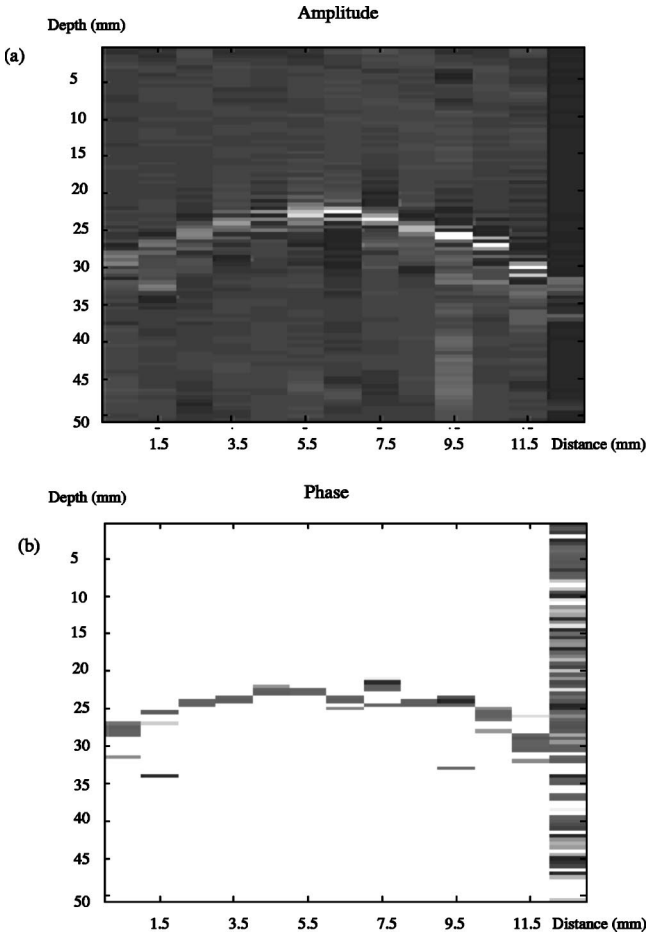


FIG. 10. Depth profilometric images of a rubber sample with curved top surface obtained using the PTA chirp heterodyne scanning system. (a) Amplitude (b) Phase.

around 0.1 and 10 cm^{-1} for blood-deficient dermis and oxygenated blood, respectively.[27] The effective scattering coefficient for breast tissue is around 1.2 cm^{-1}.[6] To illustrate the effects of varying optical absorption coefficient (optical penetration depth) on the PTA signal, depth profilometric scans were performed on a set of single-layered solid phantoms made of polyacrylamide, diluted with different percentages of India ink to closely mimic the absorption and scattering properties of human tissue. The thickness of the solid phantoms was 3 mm. Their optical absorption coefficients, μ_a, and effective optical scattering coefficients, μ_s', ranged from 3 to 15 cm^{-1} and 0.15 to 3 cm^{-1}, respectively. These optical properties were obtained from the literature.[28]

Figure 11 shows three sets of 1D scan results (amplitude and phase) obtained from samples with different μ_a and μ_s'. A direct quantitative comparison between the experimental and numerical results is not practical, since our model does not include the scattering effects of the solid. Following the developed formalism,[16,17] these issues will be addressed in forthcoming publications. However, certain features observed from the experimental results still agree with the predictions of the present mathematical model. (A) A decrease in optical absorption coefficient always results in diminished signal amplitude and broadening of the PTA response. (B) The peaks and troughs corresponding to the top and bottom

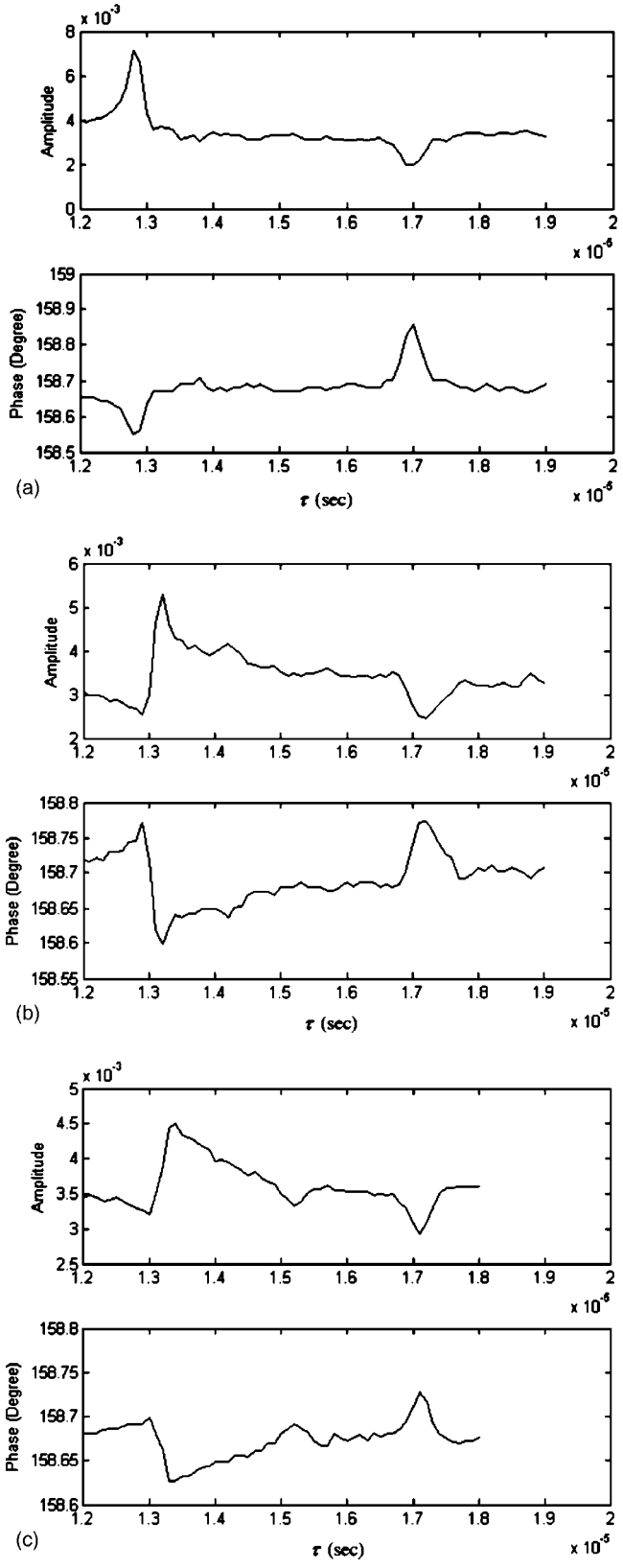


FIG. 11. Single-point scans of 3-mm-thick solid phantoms using the PTA chirp heterodyne scanning system. The optical absorption and scattering coefficients of the phantoms are (a) $\mu_a=9\ \mathrm{cm}^{-1}$ and $\mu_s'=2\ \mathrm{cm}^{-1}$; (b) $\mu_a=7\ \mathrm{cm}^{-1}$ and $\mu_s'=1.2\ \mathrm{cm}^{-1}$; (c) $\mu_a=3\ \mathrm{cm}^{-1}$ and $\mu_s'=0.15\ \mathrm{cm}^{-1}$.

surface reflections are clearly visible in both amplitude and phase traces. The generation of peaks or troughs is determined by the acoustic impedance at the fluid–solid boundary. For the solid phantoms, the density was measured to be 1160 kg/m^3. The sound velocity, c_s, was 1550 m/s, which was measured using the time-of-flight technique. (C) For samples with optical absorption depths commensurate with their thickness, as shown in Fig. 11(c), an additional trough is generated halfway between reflections of the incident acoustic wave at the front and back surface, corresponding to direct absorption at the back interface and launching of an acoustic pulse traveling toward the front surface (transducer location).

V. CONCLUSIONS

A linear frequency-domain PTA theory has been developed for a composite liquid–solid–liquid one-dimensional geometry which includes acoustic attenuation effects and natural mixed (rigid and free) boundary conditions at the solid–liquid interface.[19] Numerical simulations have been carried out to study the effects of varying sample thickness and optical absorption coefficients on the PTA field. The theory was then applied to the case of a 1.8-mm-thick rubber sample immersed in water. The simulated PTA field was compared with the experimental results obtained from a frequency scan system. The results display the following trends: (1) The amplitude of the pressure field contains acoustic standing-wave patterns. (2) For an opaque solid, the interference pattern is mainly determined by the thickness of the solid. For a translucent solid, it is strongly dependent on the optical absorption coefficient. (3) The frequency values of the minima or maxima contained in the measured frequency scan results can be used to measure the acoustic speed and the optical absorption coefficient of the experimental system. For the tested black-rubber/water system, the measured acoustic speed is 1800 m/s, which is within the range reported in the literature.[22]

A heterodyned lock-in PTA imaging technique using an intensity-modulated frequency-swept (chirped) laser beam was also developed as an extension of the single-frequency PTA setup. Compared to the single-frequency scan system, the PTA chirp scan method has significant advantages of precise depth localization, fast scanning speed, and high signal-to-noise ratio. This lock-in system can be used to perform subsurface slice-by-slice two-channel (amplitude and phase) imaging from operator determined, precisely controlled depths, thus improving the reliability and signal-to-noise ratio (i.e., dynamic range) of photoacoustic imaging over conventional pulsed laser detection. Good agreement was obtained between the numerical and experimental results on a flat, 5.4-mm-thick rubber sample and a rubber specimen with curved top surface. Experiments were also carried out on solid tissue phantoms with different optical absorption and scattering coefficients, indicating that the PTA depth profilometry system using a continuous-wave laser source and frequency-swept, heterodyned detection may be applicable in a biomedical imaging setting.

ACKNOWLEDGMENTS

The authors wish to acknowledge a CHRP Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC) which made this research possible. We also wish to thank Professor William Whelan, Ryerson University, for his valuable advice on this project.

[1] A. C. Boccara and D. Fournier, ''Principles and perspectives of photothermal and photoacoustic phenomena—historical perspectives,'' in *Progress in Photothermal and Photoacoustic Science and Technology*, edited by A. Mandelis (North-Holland, New York, 1992), 4–30.

[2] K. P. Koestli, M. Frenz, H. P. Weber, G. Paltauf, and H. Schmidt-Kloiber, ''Pulsed optoaoustic tomography of soft tissue with a piezoelectric ring sensor,'' *Biomedical Optoacoustics*, edited by A. A. Oraevsky, Proc. SPIE Vol. 3916, 67-74 (Bellingham, WA, 2000).

[3] A. A. Oraevsky, A. A. Karabutov, S. V. Solomatin, E. V. Savateeva, V. G. Andreev, Z. Gatalica, H. Singh, and R. D. Fleming, ''Laser optoacoustic imaging of breast cancer *in vivo*,'' *Biomedical Optoacoustics II*, Proc. SPIE Vol. 4256, 6–15 (A. A. Oraevsky, Ed., Bellingham, WA, 2001).

[4] M. Eghtedari, M. Motamedi, V. L. Popov, N. A. Kotov, and A. A. Oraevsky ''Optoacoustic imaging of gold nanoparticles targeted to breast cancer cells,'' *Biomedical Optoacoustics*, Proc. SPIE Vol. 5320, 21–28 (A. A. Oraevsky, Ed., Bellingham, WA, 2004).

[5] H. Hogan, ''Photonics sounds off,'' in Biophotonics International, Vol. 7, 40–45 (September/October 2000).

[6] V. G. Andreev, A. A. Karabutov, S. V. Solomatin, E. V. Savateeva, V. Aleynikov, Y. Z. Zhulina, R. D. Fleming, and A. A. Oraevsky, ''Optoacoustic tomography of breast cancer with arc-array transducers,'' in *Biomedical Optoacoustics*, Proc. SPIE Vol. 3916, 36–47 (A. A. Oraevsky, Ed., Bellingham, WA, 2000).

[7] C. G. A. Hoelen, R. G. M. Kolkman, M. Letteboer, R. Berendsen, and F. F. de Mul, ''Photoacoustic tissue scanning (PATS),'' in *Optical Tomography and Spectroscopy of Tissue III*, Proc. SPIE Vol. 3597, 336–343 (1999).

[8] P. C. Beard and T. N. Mills, ''An optical detection system for biomedical photoacoustic imaging,'' *Biomedical Optoacoustics*, Proc. SPIE Vol. 3916, 100–109 (A. A. Oraevsky, Ed., Bellingham, WA, 2000).

[9] P. M. Henrichs, J. Meador, K. Mehta, T. Miller, A. Yee, and A. A. Oraevsky, ''Phantoms for development of LOIS as a modality for diagnostic imaging of breast cancer,'' *Biomedical Optoacoustics*, Proc. SPIE Vol. 5320, 8–15 (A. A. Oraevsky, Ed., Bellingham, WA, 2004).

[10] G. Ku, X. Wang, X. Xie, G. Stoica, and L. H. Wang, ''Photoacoustic tomography of rat brain *in vivo* using multibandwidth ultrasonic detection,'' *Biomedical Optoacoustics*, Proc. SPIE Vol. 5320, 172–178 (A. A. Oraevsky, Ed., Bellingham, WA, 2004).

[11] A. A. Karabutov, N. B. Podymova, and V. S. Letokhov, ''Time-resolved optoacoustic measurement of absorption of light by inhomogeneous media,'' Appl. Opt. **34**, 1484–1487 (1995).

[12] A. Mandelis, ''Signal-to-noise ratios in lock-in amplifier synchronous detection: A generalized communications systems approach with application to frequency-, time-, and hybrid (rate-window) photothermal measurements,'' Rev. Sci. Instrum. **65**, 3309–3323 (1994).

[13] A. Oraevsky and A. Karabutov, ''Ultimate sensitivity of time-resolved opto-acoustic detection,'' *Biomedical Optoacoustics*, Proc. SPIE Vol. 3916, 228–239 (A. A. Oraevsky, Ed., Bellingham, WA, 2000).

[14] D. Fournier, F. Charbonnier, and A. C. Boccara, ''Method and device for multichannel analysis detection,'' French Patent 2666 (1993).

[15] J. Selb, S. Leveque-Fort, L. Pottier, and C. Boccara, ''Setup for simultaneous imaging of optical and acoustic contrasts in biological tissues,'' *Biomedical Optoacoustics II*, Proc. SPIE Vol. 4256, 200–207 (A. A. Oraevsky, Ed., Bellingham, WA, 2001).

[16] A. Mandelis and C. Feng, ''Frequency-domain theory of laser infrared photothermal radiometric detection of thermal waves generated by diffuse-photon-density wave fields in turbid media,'' Phys. Rev. E **65**, 021909–021909 (2002).

[17] L. Nicolaides, Y. Chen, A. Mandelis, and I. A. Vitkin, ''Theoretical, experimental, and computational aspects of optical property determination of turbid media using frequency-domain laser infrared photothermal radiometry,'' J. Opt. Soc. Am. A **18(10)**, 2548–2556 (2001).

[18] A. Karabutov and V. Gusev, *Laser Optoacoustics* (AIP, New York, 1993).

[19] A. Mandelis, N. Baddour, Y. Cai, and R. Walmsley, ''Laser induced linear photo-thermo-acoustic pressure-wave pulses in water,'' *J. Opt. Soc. Am.* (submitted).

[20] D. R. Lide, *CRC Handbook of Chemistry and Physics* (CRC, Boca Raton, FL), pp. 14–38.

[21] D. P. Almond and P. M. Patel, *Photothermal Science and Techniques* (Chapman and Hall, London, 1996), pp. 16–17.

[22] J. Krautkramer and H. Krautkramer, *Ultrasonic Testing of Materials*, 3rd ed. (Springer, Berlin, 1983).

[23] N. Bilaniuk and G. S. K. Wong, ''Speed of sound in pure water as a function of temperature,'' J. Acoust. Soc. Am. **93**, 1609–1612 (1993).

[24] A. Rosencwaig and A. Gersho, ''The photoacoustic effect with solids,'' J. Appl. Phys. **47**, 64–69 (1976).

[25] R. C. Hayser, ''Acoustical measurements by time delay spectrometry,'' J. Audio Eng. Soc. **15**, 370–379 (1967).

[26] R. C. Hayser, ''Determination of loudspeaker signal arrival times I,'' J. Audio Eng. Soc. **19**, 734–741 (1971).

[27] M. C. Pilatou, R. G. M. Kolkman, E. Hondebrink, R. Berendsen, F. F. M. DeMul, ''Photo-acoustic Imaging of Microvascular Structure in Tissue,'' *Biomedical Optoacoustics*, Proc. SPIE Vol. 3916, 48–54 (A. A. Oraevsky, Ed., Bellingham, WA, 2000).

[28] D. D. Royston, ''Optical properties of scattering and absorption materials used in the development of optical phantoms at 1064 nm,'' J. Biomed. Opt. **1**, 110–116 (1996).

The temporal representation of speech in a nonlinear model of the guinea pig cochlea

Stephen D. Holmes[a)]
School of Psychology, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom

Christian J. Sumner
MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD, United Kingdom

Lowel P. O'Mard and Ray Meddis
Centre for the Neural Basis of Hearing at Essex, Department of Psychology, University of Essex, Colchester, CO4 3SQ, United Kingdom

(Received 20 April 2004; revised 15 September 2004; accepted 15 September 2004)

The temporal representation of speechlike stimuli in the auditory-nerve output of a guinea pig cochlea model is described. The model consists of a bank of dual resonance nonlinear filters that simulate the vibratory response of the basilar membrane followed by a model of the inner hair cell/auditory nerve complex. The model is evaluated by comparing its output with published physiological auditory nerve data in response to single and double vowels. The evaluation includes analyses of individual fibers, as well as ensemble responses over a wide range of best frequencies. In all cases the model response closely follows the patterns in the physiological data, particularly the tendency for the temporal firing pattern of each fiber to represent the frequency of a nearby formant of the speech sound. In the model this behavior is largely a consequence of filter shapes; nonlinear filtering has only a small contribution at low frequencies. The guinea pig cochlear model produces a useful simulation of the measured physiological response to simple speech sounds and is therefore suitable for use in more advanced applications including attempts to generalize these principles to the response of human auditory system, both normal and impaired. [DOI: 10.1121/1.1815111]

PACS numbers: 43.64.Bt, 43.66.Ba [BLM] Pages: 3534–3545

I. INTRODUCTION

Speech is the most important sound that the human auditory system has to process. Physiological recordings from the mammalian auditory nerve (AN) in response to speechlike stimuli such as synthesized vowels and syllables (e.g., Deng and Geisler, 1987a; Miller and Sachs, 1983; Miller *et al.*, 1997; Sachs and Young, 1979; Young and Sachs, 1979) have determined the limits of peripheral representation, and posed clear tasks for processing by the central nervous system. Although such animal models have provided valuable insights, it is difficult to infer the processing mechanism from the measured responses alone. Furthermore, the nonlinear nature of the system makes it difficult to predict the responses to new complex stimuli, often requiring more data collection. Animal experiments are complex and time consuming, even using efficient measurement techniques such as the Spectral Manipulation Procedure (SMP) (LePrell *et al.*, 1996). A nonlinear computational model able to simulate the physiological data would be very useful in assessing how this coding would apply to a much wider range of speech stimuli.

Recently, Sumner *et al.* (2003b) presented a complete nonlinear filterbank model of the guinea-pig auditory nerve and evaluated it against AN responses to pure tones. The model was able to reproduce the pure tone tuning across a wide range of levels and best frequencies (BFs), the compressive cochlear nonlinearities, and rate-level functions associated with different AN fiber types at a given BF. A dual resonance nonlinear (DRNL) filter design (Meddis *et al.*, 2001) was used to simulate the response of the basilar membrane (BM). Transduction of the BM response by inner hair cells was simulated with a revised version of the Meddis inner hair-cell (IHC) (Meddis, 1986; Sumner *et al.*, 2002, 2003a). This allowed accurate reproduction of phase locking, adaptation and the characteristics associated with AN fibers of different spontaneous rate. In addition to superior response to pure tones, the model's nonlinear properties may introduce different responses to complex stimuli from that obtained from models that include linear filters, such as the gammatone filter (Johanesma, 1972; Patterson *et al.*, 1988).

Here, the cochlear model[1] of Sumner *et al.* (2003) is evaluated quantitatively in its ability to reproduce AN responses to complex speechlike stimuli. This is an important step in the refinement of the model. It helps to strengthen its value as a computational tool, advancing toward the goal of being able to predict the response to arbitrary stimuli. A variety of nonlinear filter models have been proposed previously (e.g., Deng and Geisler, 1987b; Goldstein, 1995; Irino and Patterson, 2001; Zhang *et al.*, 2001). While speechlike stimuli have been tested in some of the models (Deng and Geisler, 1987b; Sachs *et al.*, 2002), only Bruce *et al.* (2003) have attempted a systematic comparison with physiological data. Contrasts will be made with the responses of the cur-

a)Electronic mail: s.d.holmes@bham.ac.uk

rent model and those of the model of Bruce *et al.*

This study focuses on the temporal aspects of AN coding of steady-state vowels. Sachs and Young (1979) first demonstrated that a place representation based upon AN spike rates was insufficient to effectively code vowel spectra, especially in the presence of background noise (Sachs *et al.*, 1983). However, timing information, in the form of phase locking by AN fibers to prominent spectral features, robustly codes the vowel spectra across signal level in cats (Young and Sachs, 1979) even in the presence of background noise (Sachs *et al.*, 1983). This is important for speech representation because it provides level-invariant coding which is missing for a code based on spike rate. Young and Sachs proposed the average localized synchronized rate (ALSR) as a suitable representation of this information. This code is based on temporal information, but also includes rate information.

Palmer *et al.* (1986) and Palmer (1990) recorded the responses to single and double vowels from a large population of AN fibers across a wide range of BFs in the guinea-pig. From the raw spike times, they calculated several measures of temporal coding of vowels at different signal levels. Period histograms locked to the fundamental period of the vowels (or the combined period for double vowels synthesized with different pitches) were collected and these were used to generate synchronization spectra using Fourier analysis. The most prominent component in this FFT spectra was obtained for every fiber, and the fiber BF was plotted against the frequency of this largest component to produce a dominant component analysis. ALSRs were also calculated from the FFT spectra for the population of fibers. All of these analyses highlighted the way that prominent spectral features, such as formant harmonics, were strongly represented in the pattern of AN phase locking and were consistent across signal levels.

These two studies of Palmer and colleagues provide the physiological data for the evaluation of the guinea pig cochlea model presented here. The model parameters used here differ only slightly from previously published values. These changes were necessary to simulate more accurately the new data set, and are described with the model in Sec. II. The responses of the new model are shown in Sec. III. The new parameters still reproduce the previously modeled data. Thus, this study adds significantly to the value of the model by expanding the modeled data set to include complex sounds that are of biological significance for the human auditory system.

Nonlinear filtering is thought to play an important role in the global (cross-fiber) representation of speech sounds in the auditory nerve. Groups of fibers across wide BF ranges show coherent patterns of temporal responses phase locked to the major formant frequencies of the vowel (Secker-Walker and Searle, 1990). Clear discontinuities across the best-frequency axis corresponding to a shift in the dominant response from one formant to another are suggestive of significant nonlinear processing. This tendency of individual AN fibers to phase-lock to nearby formant peaks rather than local harmonic frequencies may be important to our understanding of the role of formant frequencies in identifying and differentiating vowel sounds.

Section IV investigates the features of the model response that are important to vowel formant representation, and also considers vowel representation in a human cochlea. The response of the revised nonlinear filterbank is compared to the previous generation of models based on linear gammatone filters, and also a version of the DRNL that is linear, but has unchanged pure-tone thresholds. In this way we test the hypothesis that nonlinear filtering contributes to the temporal coding of speech stimuli. These comparisons show that in fact linear filters are capable of highlighting formants across fiber populations if they are matched to the correct bandwidth. However, they do not match detailed patterns in the physiological data as closely as the full nonlinear model. Removal of the IHC/AN complex also shows that this contributes to the vowel formant representation. Further, a linear filterbank with human bandwidths does not produce an adequate temporal representation of the formants.

Formant phase-locking appears to be enhanced by a process of synchrony suppression. In physiological recordings from cats (e.g., Miller *et al.*, 1997; Young and Sachs, 1979), one harmonic shows greater phase locking relative to the other harmonics in the AN output than in the input signal spectrum. Bruce *et al.* (2003) demonstrated through selective ''impairment'' of their cat auditory model that this synchrony suppression is largely generated on the BM, with a small additional contribution from the IHC/AN complex. Synchrony suppression in the context of the current guinea pig model is investigated in Sec. IV, and appears to be somewhat different, showing more suppression in the IHC/AN than the BM. The current poor understanding of low frequency hearing does mean we must be careful to conclude that BM nonlinearity does not contribute significantly to speech representation. However, this study highlights the importance of the continuing and more accurate measurement of the auditory periphery if we are to understand the basis of speech perception. It also demonstrates the value of computational models in the investigation of important issues, such as the value of nonlinear filtering in auditory processing.

II. MODEL DESCRIPTION

A. Middle ear filtering

Middle ear filtering is modeled by two cascaded linear band-pass Butterworth filters. One filter is second-order, with lower and upper cutoffs of 4 and 25 kHz. The second is third-order and has lower and upper cutoffs of 0.55 and 30 kHz. A scaling factor of 1.41×10^{-10} m/s/uPa produces realistic stapes velocities.

B. Mechanical BM filtering

A bank of dual-resonance nonlinear (DRNL) filters is used to simulate the mechanical bandpass filtering of the BM. The DRNL filter has been described in detail elsewhere (Meddis *et al.*, 2001; Sumner *et al.*, 2003b). Figure 1 shows the architecture of a single filter, representing a single BM place. Each DRNL filter consists of two pathways, one linear and the other nonlinear. These represent the passive BM vibration, and the active process within the cochlea, respectively. Each pathway consists of a cascade of first-order gam-

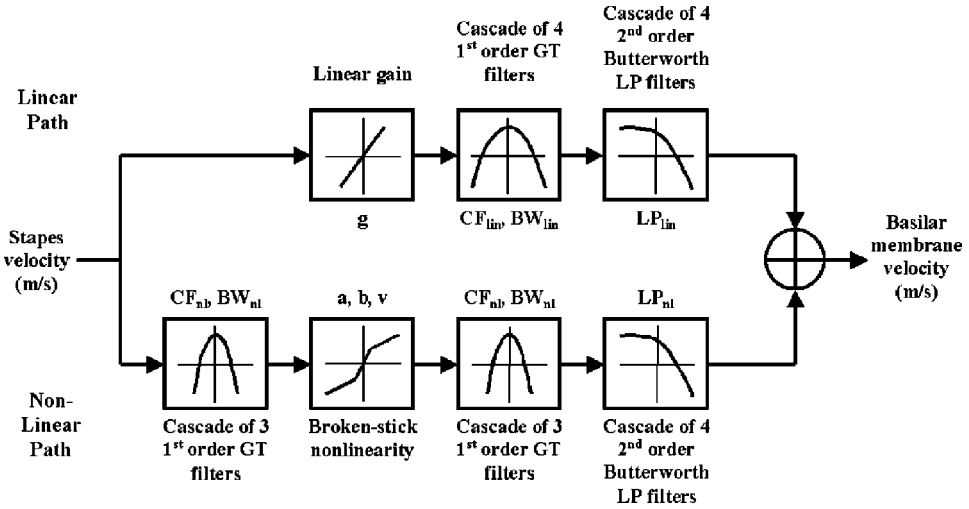


FIG. 1. Schematic diagram of the DRNL filter structure, representing a single BM location.

matone filters (Patterson *et al.*, 1988) and a cascade of second-order Butterworth low-pass filters. In the nonlinear pathway a compression function is sandwiched in the middle of the gammatone filter cascade. The output of the linear and nonlinear pathways are summed to produce the filter output. The compression function in the nonlinear pathway is defined by

$$y[t] = \mathrm{SIGN}(x[t]) \times \mathrm{MIN}(a|x[t]|, b|x[t]|^{v}), \quad (1)$$

where $x[t]$ is the input, $y[t]$ is the output, and a, b, and v are parameters.

Sumner *et al.* (2003b) fitted the DRNL to the pure tone responses of guinea-pig AN fibers across a wide range of BFs. To implement a complete filterbank, several of the model parameters vary as a function of the nonlinear filter BF. The relationship of parameter value and BF_{nl} is defined by

$$\text{Parameter} = 10^{\rho_0 + m \log_{10} BF_{nl}}, \quad (2)$$

where ρ_0 and m are different for each parameter that varies with BF. Values for these variable parameters are shown in Table I. The compression exponent, v, is set to 0.1 dB/dB for all filter BFs. Initially the number of filters in each cascade was also fixed across BF. In the present study, it was found that improved frequency threshold curves (FTCs) were produced by varying the linear path cascade with BF. This is detailed in Sec. II D. Other parameters fixed across BF are shown in Fig. 1. Here, the entire filterbank comprises 100 filters, with BFs from 0.1 to 5 kHz, evenly spaced on a log scale.

TABLE I. Basilar membrane filtering (DRNL) parameters that vary as a function of nonlinear BF where p_0 and m refer to Eq. (2) in the text. Where the parameters have changed from Sumner *et al.* (2003b), this is highlighted by inclusion of the earlier parameter values in the final column. Dashes indicate no change.

		Current values		Sumner *et al.* (2003b) values	
Parameter	Description	p_0	m	p_0	m
CF_{lin}	Best frequency of linear path gammatone filters	0.339	0.895	⋯	⋯
BW_{Lin}	Bandwidth of linear path gammatone filters	1.42	0.5	1.3	0.53
g	Linear path gain	5.68	−0.97	⋯	⋯
BW_{nl}	Bandwidth of nonlinear path gammatone filters	0.8	0.58	⋯	⋯
a	Low level gain in nonlinear path	1.87	0.45	⋯	⋯
b	Compressive region gain in nonlinear path	−5.65	0.875	⋯	⋯

C. IHC transduction

The process of mechanical to electrical transduction is simulated by a model of inner hair cell/auditory nerve processes (Sumner *et al.*, 2002, 2003a). This model includes simulations of cilia transduction and receptor potential response, presynaptic calcium processes, neurotransmitter dynamics at the synapse, and AN refractory effects. The parameters used in all of the simulations described here are those for the HSR fiber in Sumner *et al.* (2003b, Table I). In the simulations that follow, spikes were generated for 100 HSR AN fibers at each BF.

D. Model development

It was found that few changes to the model parameters from Sumner *et al.* (2003b) were required in order to match the new data closely, while retaining the correct response to Sumner *et al.*'s simulations. Table I shows the DRNL parameter values. The linear path bandwidth was increased at lower BFs and the gammatone cascade of the linear pathway was altered. Previously a cascade of three gammatone filters was used at all BFs. This produced too much tail response at low BFs and too little at high BFs. Sumner *et al.* fitted FTCs using data from Evans (1972). Better fits were obtained here with a cascade of 4 at BFs below 8 kHz and 2 at higher BFs. The only other change was to reduce the lower cutoff of the second middle ear filter from 700 to 550 Hz, as it was noted that thresholds were slightly too high at very low BFs. Otherwise the model was unchanged from Sumner *et al.*

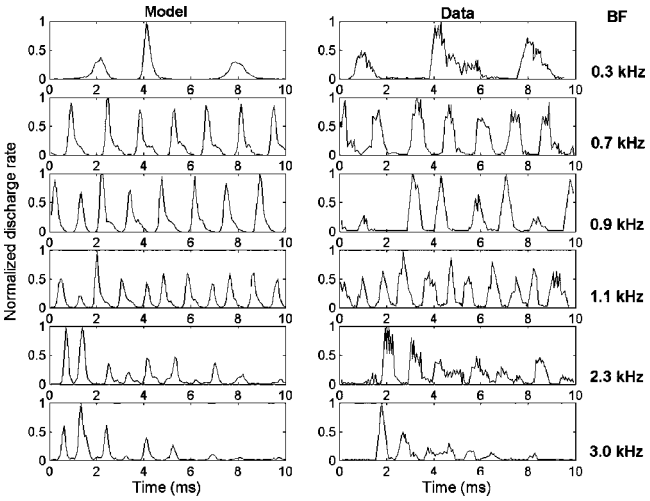


FIG. 2. Period histograms locked to the vowel period (10 ms) showing data from Palmer *et al.* (1986) and model simulation of the response to the vowel /ɑ/ presented at 80 dB SPL, of six AN fibers with BFs as indicated. The firing rate has been normalized to the value of the largest histogram bin.

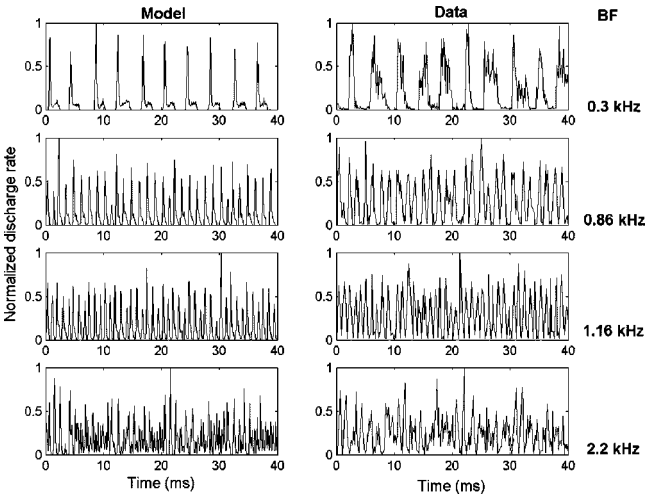


FIG. 4. Period histograms locked to the stimulus period (40 ms) showing data from Palmer (1990; right column) and model simulation (left column) of the response to the double vowel /ɑ(100 Hz)/+/i(125 Hz)/ presented at 85 dB SPL, of four AN fibers with BFs as indicated. The firing rate has been normalized to the value of the largest histogram bin.

III. MODEL EVALUATION

A. Individual fiber period histograms and FFT magnitude spectra

Period histograms (PHs) were accumulated from model AN fibers in response to vowel sounds identical to those used by Palmer *et al.* (1986) and Palmer (1990). Figure 2 shows the data from Palmer *et al.* (right column) and the model response (left column) for six fibers in response to /ɑ/ at 80 dB SPL, lasting 400 ms. Conditions for the model were as close as possible to the data. BFs were set equal to those of the guinea-pig AN fibers. The period histograms were set to the fundamental period (10 ms). The first 40 ms and the last 20 ms were excluded to avoid onset and offset effects. The model replicates the fine structure of the response for each fiber quite closely. The FFT magnitude spectra calculated from each PH are shown in Fig. 3. They show the degree of phase locking to each stimulus harmonic. The model responses are similar to the data, and show peaks at similar frequencies for all BFs, although sometimes with a somewhat broader range of harmonics picked out.

A further test of the model is its response to double vowels, where twice the number of formants are more closely packed in frequency. PHs recorded from four guinea-pig fibers in response to the double vowel /ɑ/+/i/ presented at 85 dB SPL for 500 ms (Palmer, 1990), together with the output of the model AN at the same BFs, are shown in Fig. 4. The fundamental frequencies of the vowels /ɑ/ and /i/ were 100 and 125 Hz, respectively. Therefore these PHs have a 40-ms period, in order to include an integer number of stimulus cycles for both vowels. At each of the four BFs, the model and data correspond closely. This is confirmed in the FFT magnitude spectra for the corresponding PHs, shown in Fig. 5. The responses of all four channels show almost the same pattern of phase locking to stimulus components as the

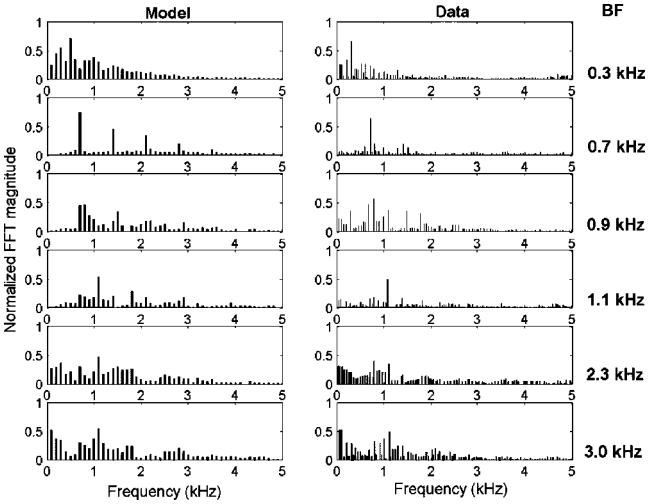


FIG. 3. FFT magnitude spectra obtained from the period histograms in Fig. 2. Data from Palmer *et al.* (1986; right column) and model (left column) simulations are shown. The firing rate has been normalized to the value of the first FFT bin.

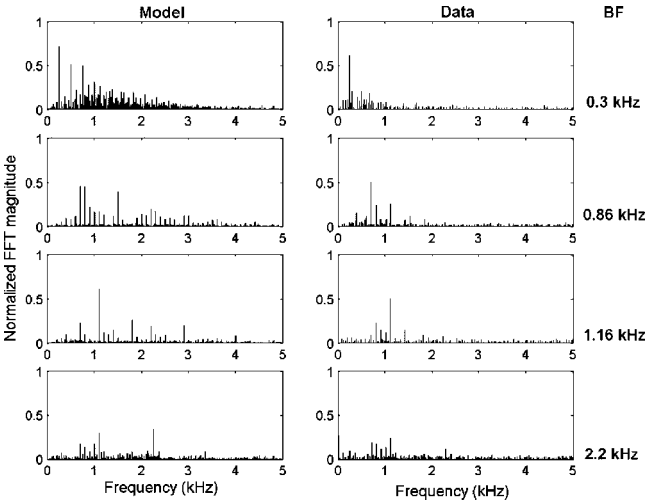


FIG. 5. FFT magnitude spectra obtained from the period histograms in Fig. 4. Data from Palmer (1990; right column) and model simulations (left column) are shown. The firing rate has been normalized to the value of the first FFT bin.

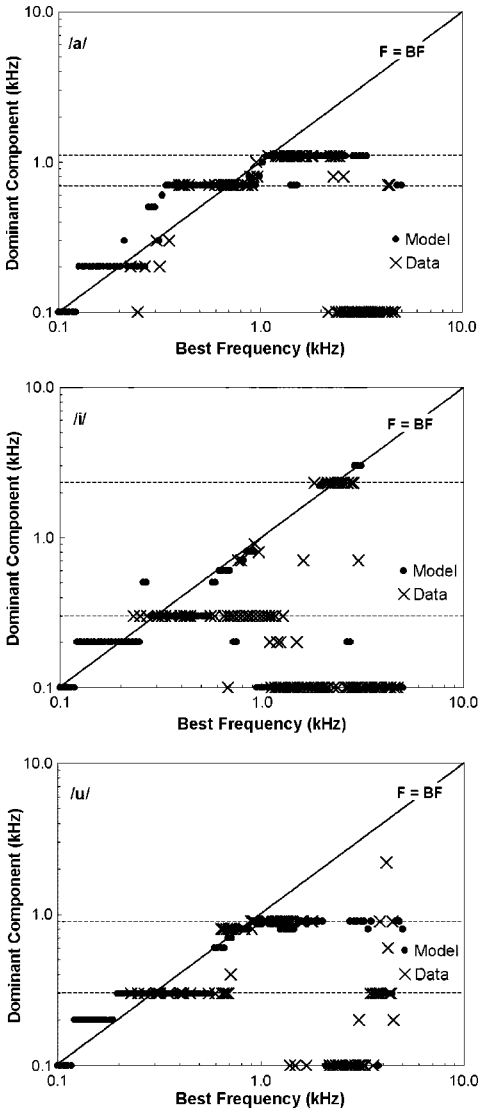


FIG. 6. Dominant component versus best frequency for the synthesized vowels /ɑ/, /i/, and /u/ presented at 80 dB SPL. The model response is shown as solid points, data by Palmer *et al.* (1986) as crosses. Horizontal broken lines indicate frequencies of the first three vowel formants. The diagonal line shows where the dominant component frequency is equal to the fiber BF.

data. The one difference occurs in the 2.2-kHz model channel, which shows more response to the second formant (F2) of /i/ at 2.29 kHz. This suggests the model may not have quite wide enough filters at this frequency. However, as will be seen in the next section, only a small number of model channels show this response.

B. Dominant component analysis

From the FFT magnitude spectra obtained for model channels/AN fibers, it is possible to conduct a dominant component analysis across the entire frequency range. For each channel/fiber the stimulus component with the largest value is selected. The channel/fiber BF is then plotted against the frequency of the dominant component.

For the single vowels /ɑ/, /i/ and /u/ that Palmer *et al.* (1986) analyzed, the dominant component analysis is shown in Fig. 6 (crosses). Also shown is the model response (dots).

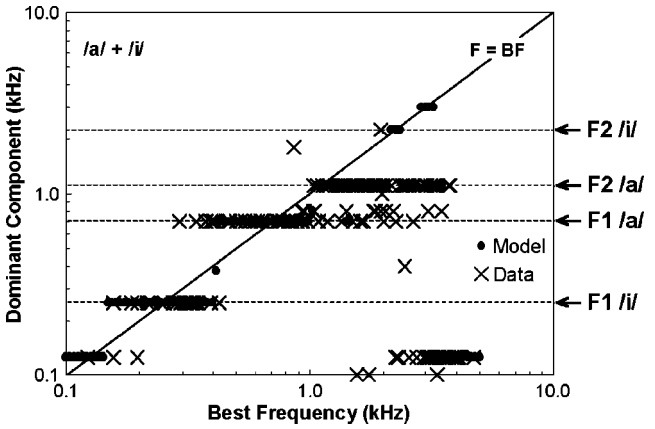


FIG. 7. Dominant component versus best frequency for the synthesized double vowel /ɑ(100 Hz)/+/i(125 Hz)/ presented at 85 dB SPL. The model response is shown as solid points, data by Palmer *et al.* (1986) as crosses. Horizontal broken lines indicate frequencies of the vowel formants, the details of which are indicated to the right of the figure. The diagonal line shows where the dominant component frequency is equal to the fiber BF.

The vowel formants are indicated by horizontal broken lines on the figures. Formants are highly salient in the pattern of dominant component phase locking by AN fibers, with a range of BFs showing synchronization to each formant frequency. The figures show that at certain BFs, the animal data can show more than one pattern of phase locking. At higher BFs, the fiber response may be dominated by either a high frequency formant or the vowel fundamental. These differences at the same BF are likely to be due to differences between individual animals. The only major difference between animal and model data is seen in the analysis for /i/, where F1 is under-represented between 0.5 and 1 kHz in the model. The differences are small though, in comparison to the strong overall correspondence of model and data across all three vowels.

The same comparison is made for the /ɑ/+/i/ double vowel response recorded by Palmer (1990; crosses) and the model output (dots) in Fig. 7. Again, the model captures the typical phase locking pattern shown by the data, although it is unable to show the split in dominant component in the data around 1–2 kHz for reasons discussed above. There is also too much phase locking to the high frequency formants (F2,F3) of /i/. However, the single set of model parameters used does produce phase locking to the components around which the data is most densely packed at most BFs and so reproduces the key features.

C. Average localized synchronized rate profiles

The average localized synchronized rate (ALSR) was first introduced by Young and Sachs (1979) as a way to combine both rate-place and phase-locked temporal information in a single representation. It is a measure of the synchronization to each stimulus harmonic, averaged across AN fibers whose BF is close to the frequency of each harmonic. In order to calculate the ALSR, an FFT is calculated from the PSTH of each fiber. From this the synchronization index (R_k) is obtained for each stimulus harmonic. The ALSR for the kth harmonic of the vowel is then defined as

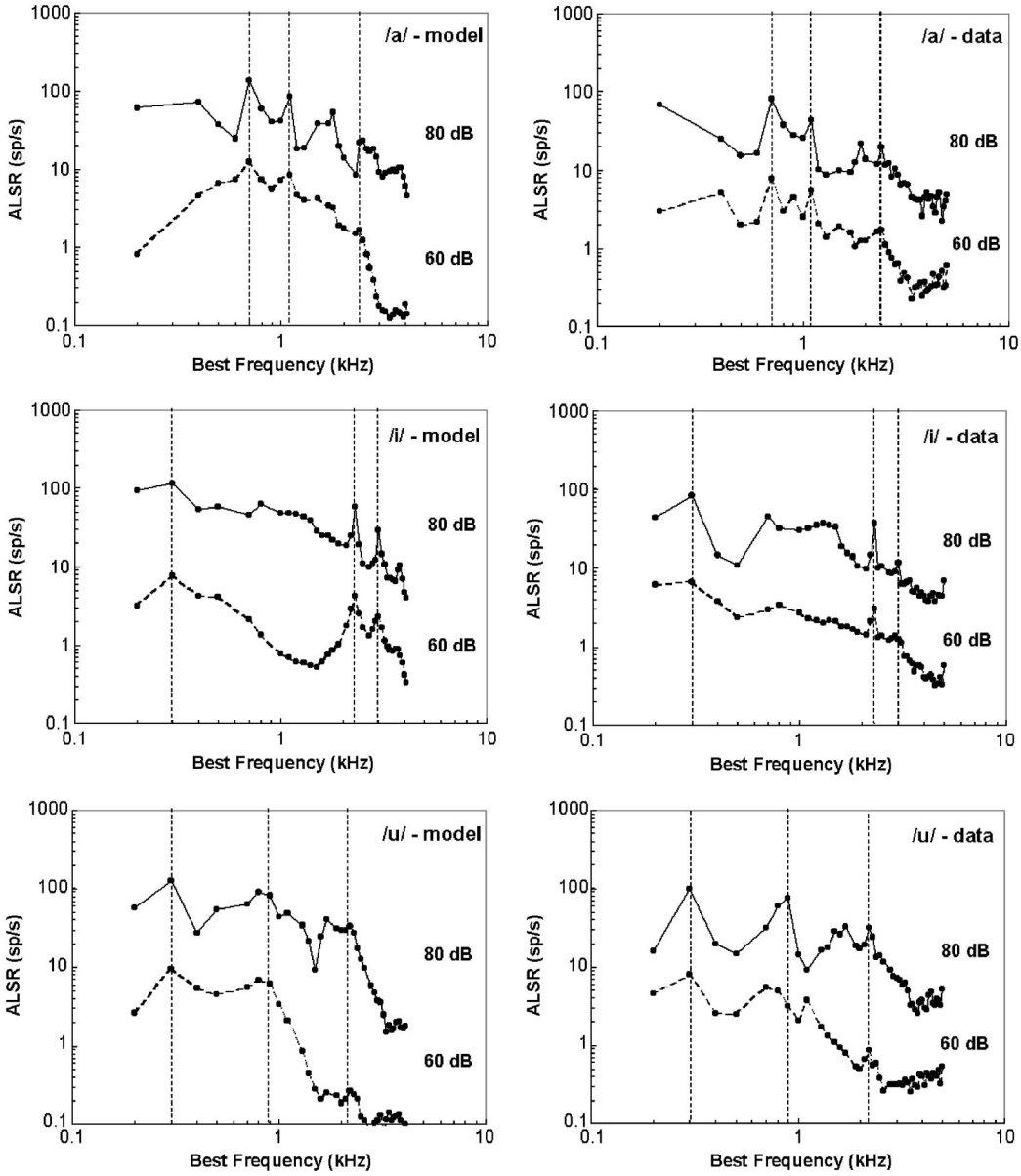


FIG. 8. ALSR functions on a logarithmic frequency scale calculated from the response to the synthesized vowels /ɑ/, /i/, and /u/ presented at 60 and 80 dB SPL. Data for 60 dB SPL is displaced one order of magnitude down. The model response is shown on the left, data from Palmer *et al.* (1986) on the right. Vowel formant frequencies are indicated by vertical lines.

$$\mathrm{ALSR}(k) = \frac{1}{M_k} \sum_{l \in C_k} R_{kl}, \tag{3}$$

where R_{kl} is the Fourier magnitude of the kth component of the response of the lth fiber expressed in spikes/second, C_k is the set of fibers with CFs within ± 0.25 oct of kf_0, M_k is the number of fibers in C_k, and f_0 is the fundamental frequency of the vowel. Data points are excluded from plots where the frequencies correspond to harmonics of F1 or F2, or sum or difference tones of the same.

Studies have shown that the ALSR captures the important spectral features of a stimulus well and is robust across large changes in signal level (Young and Sachs, 1979) and background noise (Sachs *et al.*, 1983). Palmer *et al.* (1986) and Palmer (1990) used ALSRs to display the response to vowels in their studies. ALSRs for the three vowels they used, at signal levels of 60 and 80 dB SPL, together with the response of the model to the same synthesized stimuli are shown in Fig. 8. The harmonics closest to F1–F3 of each vowel are indicated by broken vertical lines. Overall, there is very close correspondence between the model and the data. Formant peaks are apparent in both, and where the peaks are small, or peaks unrelated to formants are present in the data, the same peaks are generally present in the model. Changes with level are reproduced reasonably well. Synchronization to F2 and F3 of /i/ is, however, too strong in the model. Nevertheless the relatively high synchronization to harmonics in the trough between F1 and F2 of /i/ are reproduced at 80 dB, if less so at 60 dB.

Palmer (1990) recovered separate ALSRs for the constituent vowels of the double vowel /ɑ/+/i/. The two vowels had different fundamental frequencies, one of 100 Hz, the second of 125 Hz. Thus ALSR values were calculated for

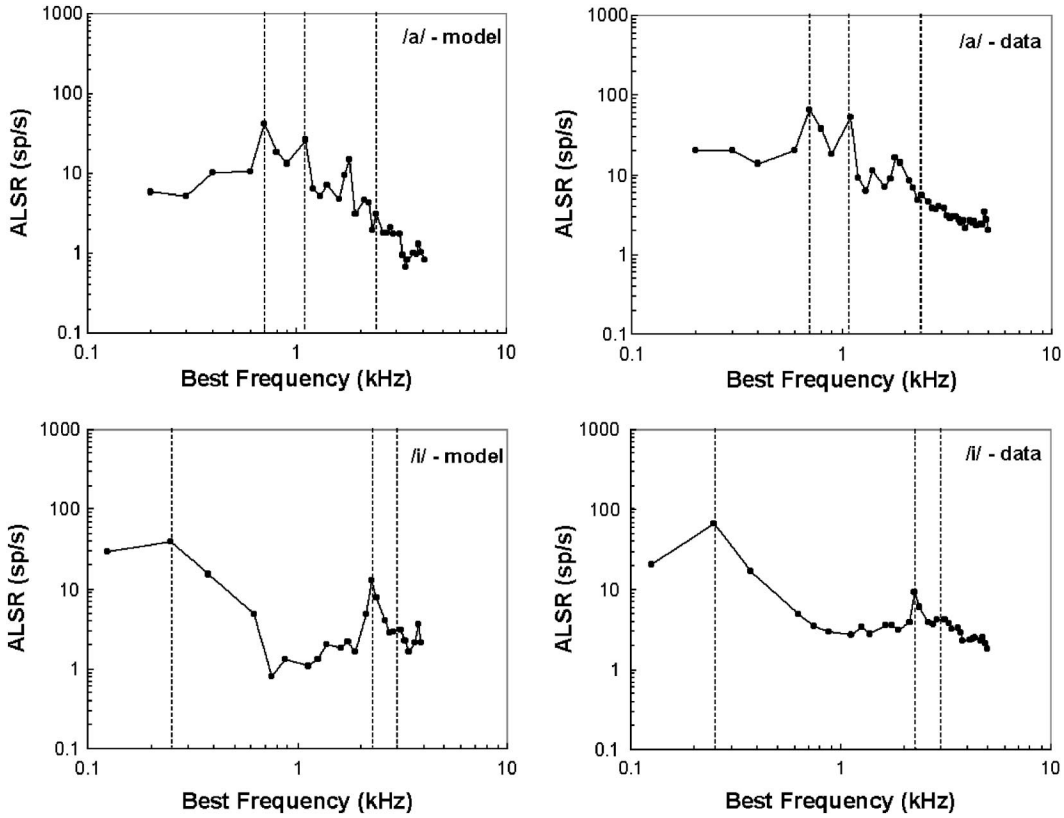


FIG. 9. ALSR functions on a logarithmic frequency scale calculated from the response to the synthesized double vowel /ɑ(100 Hz)/+/i(125 Hz)/ presented at 85 dB SPL. The model response is shown on the left, data from Palmer (1990) on the right. Plots are shifted vertically by one order of magnitude for clarity. The scales refers to 80 dB SPL. Vowel formant frequencies are indicated by vertical lines.

frequencies spaced 25 Hz apart. The ALSRs for the two vowels were recovered by sampling the complete ALSR at multiples of their respective fundamental frequencies. Palmer found that the recovered ALSRs showed almost the same pattern of peaks and troughs as those for the vowels presented singly. The ALSRs for /ɑ(100 Hz)/ and /i(125 Hz)/ based on Palmer's data, and the corresponding model ALSRs obtained in the same manner, are shown in Fig. 9. Shared harmonics of the two vowels are excluded from these plots. The model ALSRs correspond well with the data. They also closely match the ALSRs derived from single vowels in Fig. 8. There is greater synchronization to F2 of /i/, as per the results of the dominant component analysis in Fig. 6. Other than this difference, the ALSRs recovered from the combined activity show a close correspondence.

In order to determine how robust the ALSRs are across a wider range of signal levels, Fig. 10 shows ALSRs for the single vowel /ɑ/ at levels of 27 to 77 dB SPL. No comparable guinea pig data has been published for this simulation, however Young and Sachs (1979) show similar analyses recorded from the cat. Figure 10 includes their data for the vowel /ɑ/, chosen since this was the vowel with the closest match in formant frequencies to the vowels used here. Direct comparisons must be made with caution due to the different species, but Fig. 10 shows that the formant peaks are retained down to the lowest signal levels, much like Young and Sachs' data. The guinea pig ALSRs do show less prominent formant peaks than the cat data, but this is also shown in Palmer *et al.*'s guinea pig data at 60 and 80 dB SPL (Fig. 8). This may be due to differences in filter bandwidths, or synchrony suppression which is investigated in the next section.

IV. THE ROLE OF TUNING AND NONLINEARITY

In this section the properties of the model which affect the quality of representations are investigated, in particular the contributions of BM filter shape and synchrony suppression. Figure 11 shows the responses of various filterbank models to the vowel /ɑ/, presented at 80 dB SPL, represented as a dominant component analysis. The responses to other vowels showed similar differences between the models.

Figure 11(a) shows the response of a gammatone filterbank with the output fed through the older version of the IHC model (Meddis, 1986). The filters have been tuned to give the same 10-dB bandwidths as the DRNL using the same method described by Sumner *et al.* (2003b). The middle-ear model was modified slightly from that described in Sec. II A to produce the same audiogram as the DRNL model.[2] This filterbank shows a quite different response at low BFs to the nonlinear DRNL (cf. Fig. 6). At low BFs F1 is over represented. This is due to the high frequency slope of the gammatone filters, which is too shallow compared to guinea pig auditory filters. Gammatone filters are symmetric about BF on a linear frequency scale, and so alter the representation, which in animals is based upon more asymmetric filter shapes at high input intensities. At BFs from 1 to 2 kHz F0 is overrepresented. At 80 dB SPL input level, the gammatone filters are too narrow to pick out the formants, since

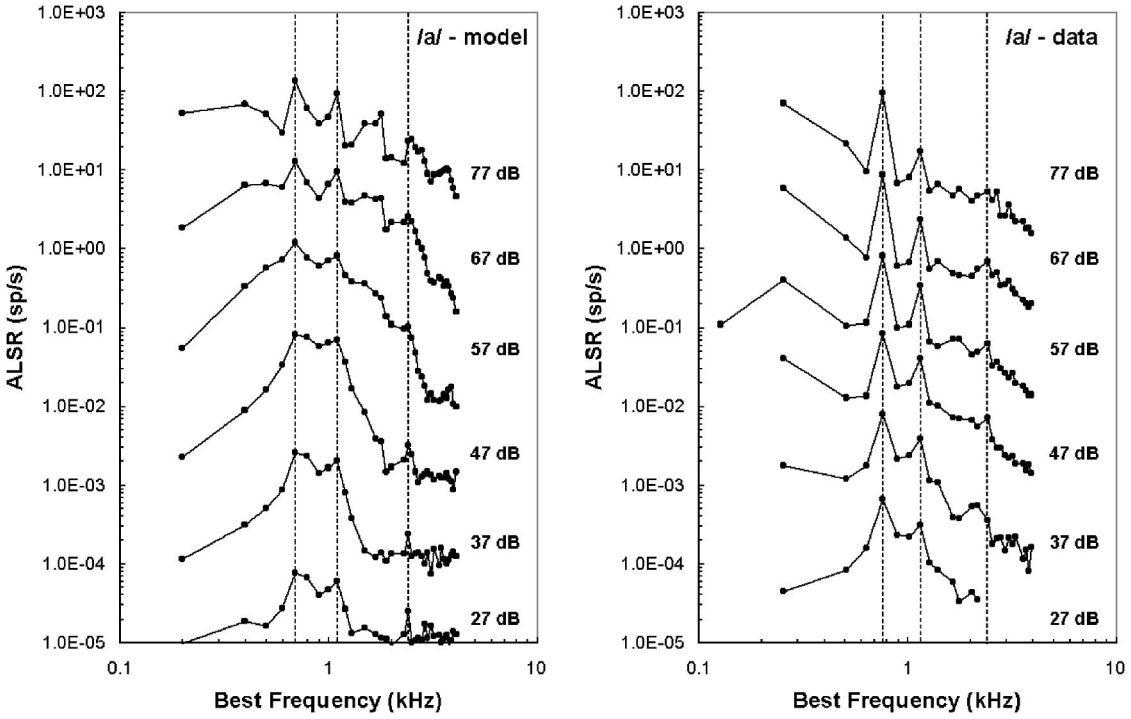


FIG. 10. ALSR functions on a logarithmic frequency scale calculated from the model response to the synthesized vowel /ɑ/ (left column) and corresponding cat data (data from Young and Sachs, 1979) (right column) presented at intensities from 27 to 77 dB SPL. Plots are shifted vertically by one order of magnitude for clarity. The scales refers to 77 dB SPL. Vowel formant frequencies are indicated by vertical lines.

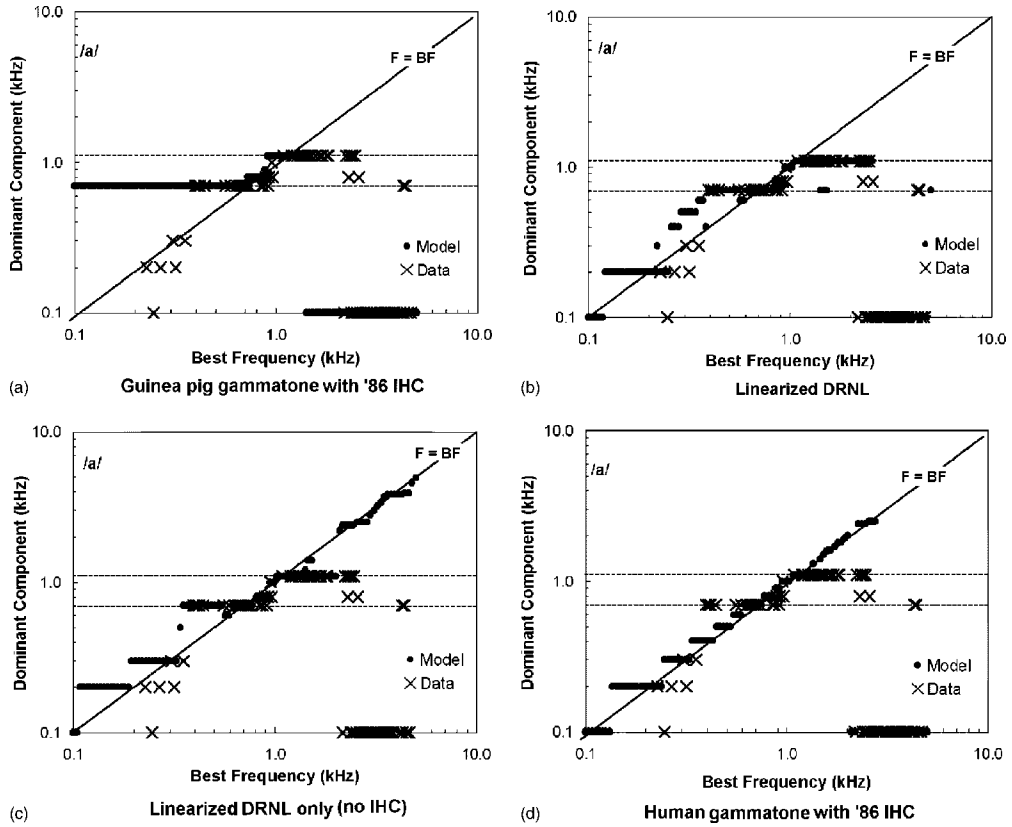


FIG. 11. Dominant component analysis of the reponse to /ɑ/ for four auditory nerve models. (a) Linear gammatone filterbank with guinea-pig bandwidths, and Meddis 86 IHC. (b) Linearized version of the DRNL filterbank model. (c) Linearized DRNL filterbank model only (without the IHC/AN stage). (d) Linear gammatone filterbank with human bandwidths, and Meddis 86 IHC.

the filter bandwidths do not vary with level. The broad low frequency tail characteristic of the physiological response at high intensities is absent.

Figure 11(b) shows the response of the DRNL filterbank presented in previous sections, but with the compression removed ($v=1$, $b(\rho_0)=1e5$) so the output of the normally nonlinear pathway is allowed to grow linearly at all input levels. The pure tone thresholds for this model are identical to the full nonlinear model. The response is more similar to that of the nonlinear DRNL (cf. Fig. 6) than the gammatone. Detail differences exist in the representation of low-frequency components, and there is less response to F2 (1100 Hz) in the linearized DRNL. These comparisons suggest that BM nonlinearity is not crucial for reproducing the data, and introduces only small differences at low BFs. The lack of a major nonlinear effect is not surprising since the pure tone response of the model is mostly linear at low BFs. At high BFs nonlinearity is more significant, and the differences would likely be larger.

Figure 11(c) shows the linearized DRNL filterbank output alone, with no IHC/AN complex. As a consequence of phase-locking limits—shown by the DRNL response to high frequency components in (c) which is lacking in (b)—the fit to the data is slightly degraded when the AN response is omitted. Otherwise, this simulation shows that the IHC/AN complex does contribute much to the vowel representation.

Finally, Fig. 11(d) shows the response of a gammatone filterbank with the old (Meddis, 1986) haircell, with bandwidths set to match human ERBs. The representation here is remarkably different, with BF components dominating at all BFs, and no widespread response to the formants.

Figures 11(b) and (c) suggest that in this model BM nonlinearity makes a small contribution to vowel formant representation, while the IHC/AN complex mostly effects higher stimulus frequencies. However, this conclusion is drawn from a dominant component analysis, which may obscure more detailed changes through the discarding of all but the largest response in each model channel. Bruce *et al.* (2003) have shown that synchrony suppression, particularly on the BM, but also within the IHC, contributes to the dominance of vowel formants in temporal representations of vowels in the cat AN. The current model is now compared to the only available guinea pig data in order to understand whether this is also true in the guinea pig. Palmer (1990) derived FFTs from AN spikes in a guinea pig AN fiber with a BF of 2.01 kHz when the double vowel /ɑ/+/i/ was presented at signal levels of 55–95 dB SPL. Figure 12 shows the model response and the corresponding data (b), derived in the same way as those in Sec. III A. The model response includes both the full model (a) and the model BM output only (c).

The data [Fig. 12(b)] shows the way in which phase locking shifts to different vowel harmonics as the compression and filter bandwidth characteristics vary with level. The response shifts from formant harmonics close to BF at low levels, when the filter is sharply tuned, to more intense but remote formant harmonics at high intensities with broader BM tuning. The full model shows a similar shift with level to the data, although there are detail differences. At higher intensities, the model is dominated by the 1100-Hz harmonic, whereas the data shows 700 Hz as the largest component. Figure 7 shows that in the data there are fibers showing 700, 800, and 1100 Hz as dominant components around 2 kHz BF. Therefore the model does match some fibers, but not the single illustrative fiber chosen by Palmer.

Of more interest is the lack of synchrony suppression in the data. Multiple harmonics are strongly represented, even at high levels. This contrasts strongly with the pattern shown in the cat (Bruce *et al.*, 2003, Fig. 8), where one component almost completely dominates the output at high intensities. The guinea pig data looks more like an impaired cat fiber (assumed to have lost BM nonlinearity and some IHC sensitivity) with little synchrony suppression. This difference explains the contrast in the formant peaks in the ALSRs for the model and cat data in Fig. 10. The full model (a) does show synchrony suppression at 85 and 95 dB SPL, when the 1100-Hz component dominates strongly. The BM-only model (c) shows much less synchrony suppression than the full model even at 95 dB. The prominent 2.25-kHz peak at lower signal levels in the BM response is a consequence of the narrow filters at this level, not suppression. The decline in phase locking in guinea pig AN fibers above 1.5–2 kHz explains why this prominent harmonic on the BM is not as strongly represented in the full model output. The suppression that exists in the model is generated more within the IHC/AN complex, not on the BM. In fact the current model probably shows too much suppression within the IHC at high input levels.

V. DISCUSSION

Simulations of the guinea pig cochlea have been described in Sec. III. These have produced results that match physiological data closely in terms of phase locking to stimulus components. The simulated AN response of individual model channels to both single- and double-vowels captures the important features of phase locking in the data. When all of these channels are combined to give the population response, in both the dominant component analysis and the ALSR, a close correspondence is found. The ALSR response across a wide range of signal levels also shows a similarly robust formant representation to the available cat data.

The filterbank was initially based upon parameters derived from the response of the guinea pig AN to pure tones (Sumner *et al.*, 2003b). Very little modification to those filterbank parameters was necessary in order to produce close matches to the response to vowels. The modifications simply improved the lower BF thresholds and FTC shapes. The change from the parameters used in the earlier simulations may indicate that the data previously modeled did not sufficiently constrain the low-BF responses. The focus in Sumner *et al.* (2003b) was upon higher BF fibers, which may have led to less optimal parameters for lower-BF fibers. In order to determine whether the parameters adopted in this study retained the correct responses to pure tones, the simulations in Sumner *et al.* (2003b) were repeated. Results were not seriously affected, and the good match to data was repeated. Therefore the current parameters can be viewed as a refinement of the previous ones.

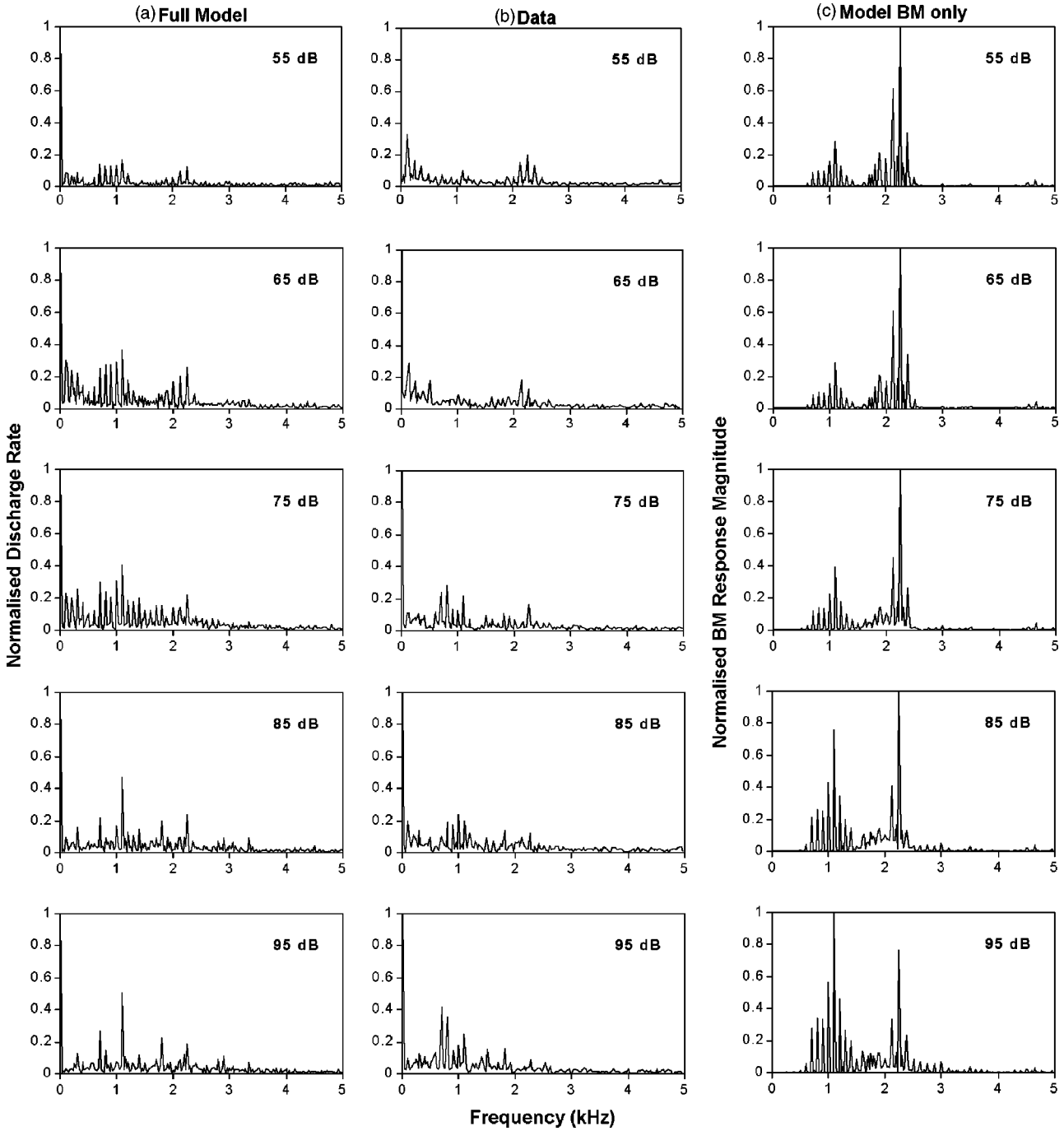


FIG. 12. FFT magnitude spectra obtained from period histograms locked to the stimulus period for a fiber with BF of 2.01 kHz, showing (a) full model AN spike rate, (b) AN spike data from Palmer (1990) and (c) BM model response magnitude. Stimulus is the double vowel /ɑ(100 Hz)/+/i(125 Hz)/ presented at levels of 55–95 dB SPL. The spike rate/BM response has been normalized to the value of the largest histogram bin.

The results in Fig. 11 compare different linear and nonlinear models and highlight the usefulness of computational models in investigating the role of nonlinearity in speech representation. They suggest, perhaps surprisingly, that filter shape and not nonlinearity is crucial for the observed representations in the guinea pig. The current results contrast with those of Bruce *et al.* (2003) and Deng and Geisler (1987b). Deng and Geisler demonstrated that when a linear and a nonlinear model were matched for frequency threshold curve (FTC) shape, the nonlinear model produced a better match to the animal data. The response of their linearized model resembled our gammatone simulation of human auditory filters, locking to harmonics rather than formants. The removal of BM nonlinearity had a much larger effect in Deng and Geisler' model than in the current one. This is due to differences in the degree of nonlinearity at low BFs. Deng and Geisler assumed significant nonlinearity across all BFs. Sumner *et al.* (2003b) fitted the current model to data from Cooper and Yates (1994) which showed less nonlinearity at a 1.8-kHz place than at higher BFs. Therefore, although parameters for BFs below 1.8 kHz are extrapolated using the logarithmic function in Eq. (2), they are more constrained than Deng and Geisler's parameters. The parameters do produce a good match to the data, although more physiological

measurements are required at low BFs in order to confirm the degree of nonlinearity at low frequencies in the guinea pig.

In our guinea pig cochlear model, the broad frequency selectivity allows intense harmonics below BF to be picked out. The asymmetrical shape of the DRNL filters, with a steep slope on the high frequency side and a shallow slope below BF, causes harmonics with frequencies greater than the filter BF to be rejected. Thus harmonics lower in frequency than the first formant are apparent in the dominant component analysis of even the linearized DRNL, while the more symmetrical gammatone filters do not show this property. Even if the bandwidths of the gammatone filters were adjusted manually according to the expected input level, they would not show the correct formant representation due to their symmetry about BF.

Section IV also suggests that the spread of temporal coding of formants along the cochlear partition might be less in humans than some other species. It is true that human cochlear filters are sharper than many animals (Shera *et al.*, 2002) and may be sufficient to support a reasonable spectral code. However, any conclusions drawn from our simulations about the importance of a temporal code for speech must be made with caution. Emerging evidence suggests that low frequency hearing in humans is in fact quite nonlinear (Lopez-Poveda *et al.*, 2003; Nelson and Schroder, 2004). Therefore the gammatone filterbank may not provide a realistic representation of responses to broadband signals. The human auditory nerve may behave more like the model of Deng and Geisler (1987b), which assumed significant nonlinearity at all BFs. Given suitable human compression and tuning data, the DRNL model architecture should be able to predict the human temporal response to speech. This highlights the need for further psychophysical characterization of the human auditory periphery to more tightly constrain computational models.

The results presented in Figs. 11 and 12 suggest that BM nonlinearity has only a small effect on speech representation, at least in guinea pigs. However, most of the spectral energy of vowel stimuli falls at frequencies well below those at which guinea pigs are maximally sensitive. At higher BFs nonlinearity is far more significant and the differences between a nonlinear and linear model would probably be much larger. Attempts have been made to generate vowel sounds that are normalized to account for cochlear differences between species (Recio *et al.*, 2002). Unfortunately such stimuli are not suitable for assessing temporal responses since their formant frequencies fall well outside the range of phase locking in guinea pigs. Therefore no attempt was made to study responses to normalized vowels at higher BFs. The reduced nonlinearity of low-BF places may explain the relatively weak BM synchrony suppression shown in Fig. 12 for the current model compared to the cat model of Bruce *et al.* (2003). Our results may appear contradictory to their demonstration of the importance of BM-generated synchrony suppression. However we feel that the current simulations are not opposed to their findings, but may illustrate a species difference. Of course the relative lack of guinea pig data makes these conclusions tentative.

The guinea pig cochlea model presented here can effectively reproduce an average AN population response to vowels. As described here, it is a suitable model for further investigation of speech coding in the AN. A future goal is the transfer of principles extracted from animal studies to models of human hearing and hearing impairment in particular. For example, it has been shown that one kind of human hearing loss may be simulated in our model by a single change to the DRNL compression function parameter *a* (Lopez-Poveda and Meddis, 2000). Similarly, Sachs *et al.* (2002) and Bruce *et al.* (2003) showed that auditory trauma in the cat could be modeled through a small set of parameter changes. A fuller appreciation of these issues should lead to improvements in speech coding algorithms in digital hearing aids. Another future application is the use of peripheral models as inputs to more ambitious models of processing in the central nervous system itself. In such cases, any errors in modeling the periphery will propagate throughout the system. As a consequence, detailed and rigorous evaluation of the peripheral model against available physiological data will assume even more importance.

ACKNOWLEDGMENTS

The work was supported by an Essex University studentship and a grant from the Wellcome Foundation. The authors would like to thank the two anonymous reviewers for their helpful comments and suggestions on an earlier version of this paper.

[1]The software used for the simulations in this study is available for download in two forms. DSAM (Development System for Auditory Modelling) is a C library containing an extensive collection of auditory model components, and support routines. AMS (Auditory Model Simulator) is a cross-platform application, providing a flexible GUI interface for all the models supported by DSAM. These can be downloaded from www.essex.ac.uk/psychology/hearinglab/dsam, together with the simulation scripts required to configure the model correctly.

[2]The order of the second bandpass is raised to 3; lower cut-offs for the two filters are 100 Hz and 6 kHz, respectively; overall gain is adjusted to a scalar value of 5.62×10^{5}; a first-order low-pass-filter with gain of 1.68×10^{-2} and a cutoff of 10 Hz acts as an integrator to convert stapes velocity to displacement.

Bruce, I. C., Sachs, M. B., and Young, E. D. (**2003**). ''An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses,'' J. Acoust. Soc. Am. **113**, 369–388.

Cooper, N. P., and Yates, G. K. (**1994**). ''Non-linear input-output functions derived from the responses of guinea-pig cochlear nerve fibers: Variations with characteristic frequency,'' Hear. Res. **78**, 225–243.

Deng, L., and Geisler, C. D. (**1987a**). ''Responses of auditory-nerve fibers to nasal consonant-vowel syllables,'' J. Acoust. Soc. Am. **82**, 1977–1988.

Deng, L., and Geisler, C. D. (**1987b**). ''A composite auditory model for processing speech sounds,'' J. Acoust. Soc. Am. **82**, 2001–2012.

Evans, E. F. (**1972**). ''The frequency response and other properties of single fibers in the guinea-pig cochlear nerve,'' J. Physiol. (London) **226**, 263–287.

Goldstein, J. L. (**1995**). ''Relations among compression, suppression, and combination tones in mechanical responses of the basilar membrane: data and MBPNL model,'' Hear. Res. **89**, 52–68.

Irino, T., and Patterson, R. D. (**2001**). ''A compressive gammachirp auditory filter for both physiological and psychophysical data,'' J. Acoust. Soc. Am. **109**, 2008–2022.

Johannesma, P. I. M. (**1972**). ''The pre-response stimulus ensemble of neurons in the cochlear nucleus,'' in *Proceedings of the Symposium on Hearing Theory* (IPO Eindhoven, The Netherlands), pp. 58–69.

LePrell, G., Sachs, M. B., and May, B. (**1996**). ''Representation of vowel-like spectra by discharge rate responses of individual auditory-nerve fibers,'' Aud. Neurosci. **2**, 275–288.

Lopez-Poveda, E. A., and Meddis, R. (**2000**). ''A computational model for simulating basilar membrane nonlinearity in subjects with normal and impaired hearing,'' International Hearing Aid Research Conference, PA9, Lake Tahoe, CA.

Lopez-Poveda, E. A., and Meddis, R. (**2001**). ''A human nonlinear cochlear filterbank,'' J. Acoust. Soc. Am. **110**, 3107–3118.

Lopez-Poveda, E. A., Plack, C. J., and Meddis, R. (**2003**). ''Cochlear nonlinearity between 500 and 8000 Hz in listeners with normal hearing,'' J. Acoust. Soc. Am. **113**, 951–960.

Meddis, R. (**1986**). ''Simulations of mechanical to neural transduction in the auditory receptor,'' J. Acoust. Soc. Am. **79**, 702–711.

Meddis, R., O'Mard, L. P., and Lopez-Poveda, E. A. (**2001**). ''A computational algorithm for computing nonlinear auditory frequency selectivity,'' J. Acoust. Soc. Am. **109**, 2852–2861.

Miller, M. I., and Sachs, M. B. (**1983**). ''Representation of stop consonants in the discharge patterns of auditory-nerve fibers,'' J. Acoust. Soc. Am. **74**, 502–517.

Miller, R. L., Schilling, J. R., Franck, K. R., and Yound, E. D. (**1997**). ''Effects of acoustic trauma on the representation of the vowel /e/ in cat auditory nerve fibers,'' J. Acoust. Soc. Am. **101**, 3602–3616.

Nelson, D. A., and Schroder, A. C. (**2004**). ''Peripheral compression as a function of stimulus level and frequency region in normal-hearing listeners,'' J. Acoust. Soc. Am. **115**, 2221–2233.

Palmer, A. R. (**1990**). ''The representation of the spectra and fundamental frequencies of steady-state single- and double-vowel sounds in the temporal discharge patterns of guinea pig cochlear-nerve fibers,'' J. Acoust. Soc. Am. **88**, 1412–1426.

Palmer, A. R., Winter, I. M., and Darwin, C. J. (**1986**). ''The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons,'' J. Acoust. Soc. Am. **79**, 100–113.

Patterson, R. D., Holdsworth, J., Nimmo-Smith, I., and Rice, P. (**1988**). ''An efficient auditory filterbank based on the gammatone function,'' (SVOS final report, Part B), MRC Applied Psychology Unit Tech. Rep. (unpublished).

Recio, A., Rhode, W. S., Kiefte, M., and Kluender, K. R. (**2002**). ''Responses to cochlear normalized speech stimuli in the auditory nerve of cat,'' J. Acoust. Soc. Am. **111**, 2213–2218.

Sachs, M. B., and Young, E. D. (**1979**). ''Encoding of steady-state vowel in the auditory nerve: Representation in terms of discharge rate,'' J. Acoust. Soc. Am. **66**, 470–479.

Sachs, M. B., Voight, H. F., and Young, E. D. (**1983**). ''AN representation of vowels in background noise,'' J. Neurophysiol. **50**, 27–45.

Sachs, M. B., Bruce, I. C., Miller, R. L., and Young, E. D. (**2002**). ''Biological Basis of Hearing Aid Design,'' Ann. Biomed. Eng. **30**, 157–168.

Secker-Walker, H. E., and Searle, C. L. (**1990**). ''Time-domain analysis of auditory nerve firing rates,'' J. Acoust. Soc. Am. **88**, 1427–1436.

Shera, C. A., Guinan, J. J., and Oxenham, A. J. (**2002**). ''Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements,'' Proc. Natl. Acad. Sci. U.S.A. **99**, 3318–3323.

Sumner, C., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2002**). ''A revised model of the inner-hair cell and auditory-nerve complex,'' J. Acoust. Soc. Am. **111**, 2178–2188.

Sumner, C., Lopez-Poveda, E. A., O'Mard, L. P., and Meddis, R. (**2003a**). ''Adaptation in a revised inner-hair cell model,'' J. Acoust. Soc. Am. **113**, 893–901.

Sumner, C., O'Mard, L. P., Lopez-Poveda, E. A., and Meddis, R. (**2003b**). ''A non-linear filter-bank model of the guinea pig cochlear nerve: Rate responses,'' J. Acoust. Soc. Am. **113**, 3264–3274.

Young, E. D., and Sachs, M. B. (**1979**). ''Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers,'' J. Acoust. Soc. Am. **66**, 1381–1403.

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.

Age effects in the human middle ear: Wideband acoustical measures[a)]

M. Patrick Feeney[b)]
Department of Speech and Hearing Science, Ohio State University, Columbus, Ohio 43210

Chris A. Sanford
Department of Speech and Hearing Sciences, University of Washington, Seattle, Washington 98195

(Received 14 June 2004; revised 23 August 2004; accepted 24 August 2004)

Studies that have examined age effects in the human middle ear using either admittance measures at 220 or 660 Hz or multifrequency tympanometry from 200 to 2000 Hz have had conflicting results. Several studies have suggested an increase in admittance with age, while several others have suggested a decrease in admittance with age. A third group of studies found no significant age effect. This study examined 226 Hz tympanometry and wideband energy reflectance and impedance at ambient pressure in a group of 40 young adults and a group of 30 adults with age $\geqslant$60 years. The groups did not differ in admittance measures of the middle ear at 226 Hz. However, significant age effects were found in wideband energy reflectance and impedance. In particular, in older adults there was a comparative decrease in reflectance from 800 to 2000 Hz but an increase near 4000 Hz. The results suggest a decrease in middle-ear stiffness with age. The findings of this study hold relevance for understanding the aging process in the auditory system, for the establishment of normative data for wideband energy reflectance, for the possibility of a conductive component to presbycusis, and for the interpretation of otoacoustic emission measurements. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1808221]

PACS numbers: 43.64.Ha [BLM] Pages: 3546–3558

I. INTRODUCTION

The middle ear serves to partially overcome the impedance mismatch between sound waves in air and the fluid-filled cochlea. An understanding of age effects in the middle ear and their affect on the input to the cochlea is critical to an understanding of the aging process in the auditory system, to an understanding of presbycusis, and to the development of clinical norms for hearing and middle-ear assessment. Age effects in the middle ear may also impact the interpretation of otoacoustic emission measurements, which depend on the forward and reverse transfer of sound energy through the middle ear.

Anatomical studies in humans have documented changes in the middle ear with age. Ruah *et al.* (1991) studied changes in the tympanic membranes harvested from 46 human cadavers ranging in age from 2 days to 91 years. They reported that with age the tympanic membrane becomes ''less vascular, less cellular, more rigid, and less elastic'' (p. 634). A histological study of the middle ear ossicles was conducted by Etholm and Belal (1974), who reported arthritic changes in the incudomalleal and incudostapedial joints in presbycusic ears. More significant arthritic changes including fusion of bone joints were seen in specimens from ears over 70 years of age. Similarly, Harty (1953) examined temporal bones of elderly patients and reported that the ossicular joints contained less elastic tissue.

Behavioral studies have suggested the presence of a conductive component to presbycusis, in which the loss is predominantly in the middle to high frequencies (Marshall *et al.*, 1983; Nixon *et al.*, 1962). Nixon *et al.* reported a 12 dB air-bone gap at 4000 Hz that they attributed to the loss of middle-ear stiffness necessary for the transmission of high frequencies. Randolph and Schow (1983) challenged this interpretation and attributed the high-frequency air-bone gap in the elderly to ear-canal collapse during audiometric headphone placement. However, Marshall *et al.* found that ear-canal collapse in their elderly subjects was not sufficient to explain the effect. They suggested that in some elderly subjects there was an apparent ''loosening in the middle-ear system'' (p. 606).

The possibility of a functional change in the middle ear with age is supported by recent studies of middle-ear function in the mouse using laser-Doppler vibrometry (LDV) that have demonstrated a decrease in mid- to high-frequency umbo velocity in aging mice (Doan *et al.*, 1996; Rosowski *et al.*, 2003). However, in a recent LDV study in humans from age 20 to 81 years, Whittemore *et al.* (2004) reported a significant age-related change in umbo velocity in only one of nine measurement frequencies, and that was a slight increase in umbo velocity with age at 6000 Hz.

Most previous studies that have examined age effects in the middle ear have used traditional 220 or 660 Hz admittance tympanometry. Overall, these studies have yielded conflicting or inconclusive results. Consider first the measurement of peak-compensated static acoustic admittance (Peak Y_{tm}), which is an estimate of the middle-ear admittance. This is obtained by subtracting the admittance measured in the ear canal under positive or negative air pressure, usually +200

[a)]Portions of this work were presented at the 145th Meeting of the Acoustical Society of America, Nashville, TN, April 2003.

[b)]Current affiliation: Otolaryngology, Head and Neck Surgery, and Virginia Merrill Bloedel Hearing Research Center, University of Washington, Seattle, WA 98195. Electronic-mail: pfeeney@u.washington.edu

or −300 decaPascals (daPa), from the admittance measured at the peak of the tympanogram (Van Camp *et al.*, 1986). Previous studies of Peak Y_{tm} include that of Beattie and Leamy (1975), who reported a significant increase in Peak Y_{tm} at 660 Hz in their older group of 20 subjects aged 60–78 years compared to a younger group of 20 subjects aged 17–29 years. A more complex effect of age on Peak Y_{tm} was reported by Jerger *et al.* (1972), who studied the effect of age and gender in 700 subjects with either normal hearing or a sensorineural hearing loss <70 dB hearing level (HL). They reported that Peak Y_{tm} measured with a 220 Hz probe tone rose to a maximum by the fourth decade and then decreased for subjects over 50 years of age. Blood and Greenberg (1977) reported a significant decrease in Peak Y_{tm} with age at 220 Hz. A decrease in Peak Y_{tm} with age was also reported by Gates *et al.* (1990) in a cohort of 1656 adults with an age range of 63–95 years. However, the reported decline in Peak Y_{tm} at 220 Hz was small, decreasing 0.0035 ± 0.0017 ml per year.

In contrast, several studies have shown little or no significant effect of age on middle ear admittance. Nerbonne *et al.* (1978) reported no significant age effect on Peak Y_{tm} using a 220 Hz probe tone in three groups of 60 men and women aged 20–29, 60–69, and 70–79 years. In a study using both 220 and 660 Hz probe tones, Thompsen *et al.* (1980) measured static acoustic conductance, susceptance, and admittance in 60 women with 10 subjects in each age decade from 20 to 79 years. No significant effect of age was found for any of the measurements with either probe frequency. Holte (1996) also found no age effect in Peak Y_{tm} at 226 Hz for men in 7 age groups with 20 subjects in each decade from 20 to 79 years and 16 subjects >79 years. Holte also examined tympanometric width (TW) in these subjects, defined as the pressure interval over which static acoustic admittance at 226 Hz was reduced by one-half. There was a moderate negative correlation ($r=-0.616$) between TW and Peak Y_{tm} for the subjects in that study, as well as a moderate positive correlation between TW and two measures of middle-ear resonant frequency derived from multifrequency tympanometry ($r=0.467$ and 0.441). Therefore, as stiffness increased, resulting in a higher resonant frequency, Peak Y_{tm} decreased and tympanometric width increased for the subjects in that study. Holte found a small but statistically significant positive correlation between TW and age, $r=0.176$. Negative findings for an age effect in Peak Y_{tm} at 226 Hz have also been reported by Uchida *et al.* (2000) in a study of 933 subjects across three decades from age 40 to 70 years and by Stenklev *et al.* (2004) in comparing a group of otologically normal subjects $\geqslant 60$ years old with a group of 20 young-adult controls.

Middle ear resonance frequency calculated by using multifrequency tympanometry (200–2000 Hz) has also been used to examine age effects in the middle ear. As part of the study reported above, Holte (1996) calculated the middle ear resonant frequency in her subjects who ranged in age from 20 to over 80 years. Since greater middle ear stiffness, as in otosclerosis, increases the resonant frequency of the middle ear, Holte reasoned that multifrequency tympanometry would be an appropriate measure of the age-related stiffness effect suggested by anatomical studies. However, she found no statistically significant age effect for this measure. Similar findings using multifrequency tympanometry were reported by Wiley *et al.* (1999), who found no age effect for middle-ear resonant frequency for 467 men and women from age 48 to 90 years. Similarly, Uchida *et al.* (2000) found no age effect for middle ear resonance frequency in their large sample of subjects from age 40 to 70 years.

These conflicting or negative findings may be due to a lack of sensitivity of clinical admittance measurements in the 200–2000 Hz bandwidth to age-related changes in middle ear function. Perhaps this is not surprising in the face of reports of normal 226 Hz tympanograms in cases of confirmed middle ear disease (Feeney, Grant *et al.*, 2003; Hunter and Margolis, 1997). Moreover, the interpretation of admittance tympanometry at frequencies between 220 and 2000 Hz has been complicated by increased errors in the calculation of ear canal volume, as well as by the appearance of notched tympanograms (Van Camp *et al.*, 1986).

A relatively new technique for middle ear measurement holds promise for the study of the aging ear by allowing the examination of the acoustic power response of the middle ear over the audiometric frequency range. Energy reflectance is the ratio of the power reflected from the middle ear to the incident power. Thus, when all of the power at a given frequency is absorbed by the middle ear, the energy reflectance is 0.0. Likewise, when all of the power is reflected by the middle ear, the energy reflectance is 1.0. Energy reflectance at ambient pressure in adults is typically near 1.0 in the low frequencies but decreases gradually with frequency to a minimum around 4000 Hz and then increases at higher frequencies (Feeney, Grant *et al.*, 2003; Keefe *et al.*, 1993; Stinson, 1990; Voss and Allen, 1994). An attractive feature of energy reflectance measurements is that, unlike traditional admittance measures, they are relatively insensitive to probe microphone location in the ear canal (Huang, *et al.* 2000b; Stinson *et al.* 1982).

Wideband energy reflectance and impedance measurements of the normal middle ear in adults and young children have been made in the frequency range of 200–10 000 Hz (Keefe *et al.*, 1993; Margolis *et al.*, 1999; Voss and Allen, 1994). Recent studies have suggested that compared to standard clinical admittance measurements, wideband energy reflectance and impedance measurements may provide a more sensitive test of middle-ear disorders and conductive hearing loss (Feeney, Grant *et al.*, 2003; Keefe and Levi, 1996; Keefe and Simmons, 2003; Piskorski *et al.*, 1999). Wideband energy reflectance may also provide a more sensitive measurement of the acoustic reflex threshold (Feeney and Keefe, 1999; 2001; Feeney, Keefe *et al.*, 2003; Feeney *et al.*, in press).

The purpose of the present study was to examine age effects in the middle ear by applying wideband acoustic measurements in a group of young adults and in a group of adults $\geqslant 60$ years of age. Measurements of wideband impedance and energy reflectance in the ear canal were obtained at ambient pressure.

II. METHOD

A. Subjects

The young-adult group consisted of 40 participants, 20 women and 20 men, ages 18–28 yr (mean=21.4 yr) who were recruited from flyers posted on campus. Young-adult subjects were included in the study if they had a negative history of middle-ear disorders and met the following inclusion criteria: (1) A normal otoscopic screening for ear canal and tympanic membrane abnormalities; (2) A normal 226 Hz admittance tympanogram defined as having a single peak within ± 10 daPa of ambient pressure and with Peak Y_{tm} between 0.3 and 1.7 mmho (Margolis and Hunter, 1999); and (3) Pure-tone air-conduction thresholds $\leqslant 15$ dB HL from 250 to 8000 Hz with air-bone gaps $\leqslant 10$ dB at octave frequencies from 250 to 4000 Hz.

The older-adult group consisted of 30 participants, 15 women and 15 men, ages 60–85 yr (mean=71.6 yr) who were recruited from flyers posted on campus and at local senior centers. These subjects also had a negative history of middle-ear disorders and met the same otoscopic and tympanometric inclusion criteria as the subjects in the young-adult group. As with the young group, audiometric air-bone gaps were $\leqslant 10$ dB at octave frequencies from 250 to 4000 Hz. However, subjects with pure-tone air conduction thresholds outside the normal range were included in the study.

B. Apparatus

With the subject seated in a sound-treated booth, pure-tone audiometry to meet the inclusion criteria was conducted using a Madsen, model 622 diagnostic audiometer calibrated for air- and bone-conduction testing to ANSI S3.6 (1996) standards. ER-3A insert earphones were used for air-conduction testing to avoid ear-canal collapse. Tympanometry was conducted using a Grason-Stadler, Inc., model 33 (GSI-33) immittance instrument calibrated to ANSI S3.39 (1987) standards.

The energy reflectance measurement system consisted of an Etymotic Research, model ER-10C microphone system; a Communication Automation & Control, model 32C data acquisition card with a 24 kHz sampling frequency; and a personal computer. The energy reflectance probe signal consisted of 40 ms electrical chirps generated by the data acquisition card, with a bandwidth extending from 200 to 10 000 Hz, which were routed to a receiver in the ER-10C probe. The microphone response was high-pass filtered at 200 Hz (48 dB per octave) to remove system and biological noise by use of a programmable filter (Tucker-Davis, model PF1) and then acquired at a sampling rate of 24 kHz with a spectral resolution of approximately 12 Hz.

C. Procedure

1. Energy reflectance calibration

The method described by Keefe *et al.* (1992) was used to calibrate the system for the measurement of energy reflectance in the ear canal. Calibration was conducted daily using a set of 6 brass tubes with inside diameters of 8 mm and lengths ranging from 486 to 918 mm with rigid terminations. During calibration a standard ER-10C foam ear tip with a diameter of 14 mm and length of 14 mm (ER10C-14A) was attached to the probe and then compressed and inserted into each calibration tube to a depth of approximately 12 mm. The calibration protocol consisted of determining the Thevenin source impedance and sound pressure of the ER-10C microphone system across frequency from 250 to 8000 Hz. This was accomplished by iterative comparison of the measured pressure response of the six tubes to a model of each including viscothermal losses.

To test the validity of the calibration, a method developed by Keefe *et al.* (1992) was employed. In this method an error function is used to determine the error in the acoustical estimate of the length of each calibration tube across frequency. An unacceptably large error term results in a failed calibration as implemented in software. When this was encountered in the present study, perhaps due to a leaky probe fit in the calibration tubes, the calibration was repeated until an acceptable error term was obtained. Keefe *et al.* (1992) reported that acceptable calibrations obtained by this method minimized the error in calculating the Thevenin equivalents such that estimates of calibration tube energy reflectance did not exceed 0.03 at any frequency across the measurement bandwidth.

Measurements of the pressure spectrum in a subject's ear canal were combined with the Thevenin parameters of the probe to derive wideband impedance at the probe tip. The characteristic impedance at the entrance of the ear canal was then estimated based on an acoustical estimate of the ear canal area. One-third octave measurements of energy reflectance from 250 to 8000 Hz were then obtained by a comparison of the impedance at the probe tip and the characteristic impedance of the ear canal using standard transformations (Keefe *et al.*, 1992).

2. Energy reflectance measurements

Following the audiometric assessments to meet inclusion criteria, energy reflectance measurements were made with the subject seated in a commercial sound-treated booth. The foam tip of the ER-10C microphone was inserted in the subject's ear canal with full insertion depth (14 mm) if possible, depending upon the ear canal shape and size. The foam tip remained in the ear canal for 2 min to allow the foam to expand and seal the ear canal before energy reflectance was measured.

The probe signal was then introduced into the subject's sealed ear canal using a receiver in the ER-10C microphone system. The overall level was set to 65 dB SPL as calibrated in a Zwislocki occluded ear simulator (Knowles Electronics, model DB-100). An artifact-reject algorithm was used to eliminate noisy responses to the chirp stimuli (Keefe and Ling, 1998). When the patient was seated quietly, the experimenter started data acquisition, which was a time-averaged response to a series of eight valid chirps. Each averaged response took approximately 1 s to acquire.

The initial averaged energy reflectance response was evaluated to determine if there was a leaky probe fit resulting in low energy reflectance at the lowest one-third octave. Based on our experience with measurements in adults with

TABLE I. Group mean air- and bone-conduction audiometric thresholds for the young-adult and elderly groups. Also displayed are the group-mean air-bone gaps (ABG). This is defined as the mean air-conduction threshold minus the mean bone-conduction threshold at each frequency. Numbers in parentheses are SD.

		Frequency (Hz)					
Group	Measure	250	500	1000	2000	4000	8000
Young	Air	7.3 (4.4)	7.8 (5.1)	5.6 (5.6)	6.3 (4.5)	3.4 (5.4)	−0.1 (6.1)
	Bone	5.8 (4.7)	2.6 (2.6)	2.1 (7.1)	0.9 (5.2)	2.0 (6.2)	
	ABG	1.5 (5.3)	5.2 (4.7)	3.5 (5.2)	5.4 (4.9)	1.4 (4.4)	
Elderly	Air	15.3 (8.1)	16.8 (9.5)	21.5 (11.5)	24.8 (13.8)	36.0 (17.5)	45.0 (23)
	Bone	10.3 (7.8)	13.2 (10.0)	17.7 (12.5)	21.2 (14.3)	29.7 (16.3)	
	ABG	5.0 (5.1)	3.6 (5.2)	3.8 (6.3)	3.6 (4.7)	6.3 (4.5)	

this system, a leaky probe fit was suspected if the energy reflectance was lower than 0.8 at 250 Hz. If such a measurement was obtained, the probe tip was removed from the ear, reinserted, and allowed to expand for an additional 2 min prior to reflectance measurement. In several cases, the ear canal could not be sealed with the standard 14 mm dia foam tip, so an 18 mm diam foam tip was used (ER10C-14C). This problem was typically experienced with oval ear canals in some elderly subjects where a leak was apparent at the superior or inferior aspect of the ear canal with the smaller foam tip. The ER10C-14C foam tip was used for 5 elderly subjects. The standard 14 mm diam foam tip was too large to be comfortably inserted in the ear canal for several subjects. In these cases, a 10 mm diam foam tip was used (ER10C-14B). This foam tip was used for 2 young subjects and 1 elderly subject.

Once an adequate fit had been confirmed, the reflectance measurement was repeated to obtain a total of three valid measurements for each ear. These were then averaged at each one-third-octave frequency to obtain the final data set for each ear. When both ears of a subject met inclusion criteria, both were measured. The data for one ear from each subject were arbitrarily selected for analysis, with the stipulation that data from equal numbers of right and left ears were used if possible. If only one of a subject's ears met the inclusion criteria, data from that ear was used for data analysis.

III. RESULTS

A. Audiometry

Responses to standard audiometric measures are shown in Table I for both groups. Data were obtained on equal numbers of right and left ears for the young-adult group, and 16 left and 14 right ears for the elderly group. For the young-adult group, the mean three-frequency pure-tone average (PTA) for air conduction for the frequencies 500, 1000, and 2000 Hz was 6.5 dB HL (SD=4.1 dB). For the elderly group, the group-mean three-frequency PTA for air conduction was 21.1 dB HL (SD=10.7 dB) with a range from 3.3 to 41.7 dB HL. To meet inclusion criteria for this study, subjects from both groups could have air-bone gaps no greater than 10 dB at the audiometric frequencies from 250 to 4000 Hz. Table I also shows the mean air-bone gaps and standard deviations for both groups. The difference in air-bone gap between groups at each frequency was tested using an independent-samples t-test. An alpha level of 0.05 was adopted for this and subsequent statistical tests, and a two-tailed test was used for all analyses. The elderly group had a significantly greater air-bone gap at 250 Hz (mean =5.0 dB) than the young group (mean=1.5 dB), $t(68)=4.61$, $p<0.01$. The air bone gap was also greater for the elderly group at 4000 Hz (mean=6.3 dB) compared to the young group (mean=1.4 dB), $t(68)=2.77$, $p<0.01$. The means were not significantly different at other frequencies.

B. 226 Hz tympanometry

Tympanometric data obtained with the GSI-33 clinical system using a 226 Hz probe tone are shown in the first three rows of Table II for young-adult and elderly groups. There were no statistically significant differences between right and left ears for these measures, so data from both ears were combined for the following statistical analyses. An independent-samples t-test was used to evaluate the differ-

TABLE II. Tympanometric and ear-canal measurements for the young and elderly groups. The means, SDs and ranges of values are shown for peak-compensated static acoustic admittance (Peak Y_{tm}), tympanometric width (TW), and equivalent ear canal volume (V_{ea}), all obtained using a 226 Hz probe tone. The ear canal diameter (EC Diameter) estimates were obtained using the energy reflectance system. Also shown are the results of the two-tailed t tests and associated p values for the mean comparisons. The p-values for significant t-test results are underlined.

	Young			Elderly			Mean Comparisons	
Measure	Mean	SD	Range	Mean	SD	Range	t	p
Peak Y_{tm} (mmhos)	0.61	0.24	0.3–1.5	0.62	0.19	0.3–1.1	0.17	0.86
TW (daPa)	76.3	19.1	50–110	78.7	17.6	50–125	0.53	0.60
V_{ea} (cm^3)	1.30	0.36	0.7–2.3	1.50	0.35	0.9–2.4	2.42	<u>0.02</u>
EC Diameter (mm)	10.67	2.87	5.5–18.0	9.25	2.78	5.4–15.3	2.07	<u>0.04</u>

TABLE III. Gender comparison for the young group for tympanometric and ear-canal measurements. The means and SDs of values are shown for peak-compensated static acoustic admittance (Peak Y_{tm}), tympanometric width (TW), and equivalent ear canal volume (V_{ea}), all obtained using a 226 Hz probe tone. Also shown are the results of the two-tailed t tests and associated p values for the mean comparisons. The p-values for significant t-test results are underlined.

	Women		Men		Mean comparisons	
Measure	Mean	SD	Mean	SD	t	p
Peak Y_{tm} (mmho)	0.55	0.17	0.67	0.28	1.55	0.13
TW (daPa)	74.0	17.1	78.7	21.0	0.77	0.48
V_{ea} (cm^3)	1.18	0.06	1.42	0.17	2.21	<u>0.03</u>

ences between groups on these measurements. There was no significant difference in Peak Y_{tm} between the young-adult group (mean=0.61 mmho) and the elderly group (mean =0.62 mmho), $t(68)=0.174$, $p=0.86$. There was also no significant difference in TW between the young-adult group (mean=76.3 daPa) and the elderly group (mean =78.7 daPa), $t(68)=0.53$, $p=0.60$. However, there was a small but significant difference between groups in equivalent ear canal volume (V_{ea}), with the young-adult group (mean $=1.3\ cm^3$) demonstrating a smaller V_{ea} than the elderly group (mean$=1.50\ cm^3$), $t(68)=2.42$, $p=0.02$.

Gender differences were also evaluated for both groups for each of the three measures derived from 226 Hz tympanometry. For the young-adult group, the V_{ea} for men (mean $=1.42\ cm^3$) was larger than for women (mean$=1.18\ cm^3$), $t(38)=2.21$, $p=0.03$; however, there were no significant gender differences for Peak Y_{tm} or TW (Table III). The V_{ea} for the elderly men (mean$=1.67\ cm^3$) was also larger than for elderly women (mean$=1.34\ cm^3$), $t(28)=2.83$, $p=<0.01$. The Peak Y_{tm} was also greater for elderly men (mean=0.70 mmho) than elderly women (mean =0.53 mmho), $t(28)=2.71$, $p=0.01$. However, there was no significant gender difference in the elderly group for TW (Table IV).

C. Wideband impedance

Wideband acoustic impedance data were accepted as satisfactory if the input resistance was positive. As was suggested by Keefe *et al.* (1993), measurements of negative resistance with this energy reflectance system are likely due to small errors in estimating the Thevenin parameters of the probe and physiological noise. For 15 young subjects and 5 elderly subjects, there were negative resistances at 8000 Hz, the highest one-third octave frequency. Due to the large amount of unsatisfactory data at this frequency, this frequency was excluded from further analysis for both groups. There were only 2 subjects in each of the young and elderly groups with negative resistances at 6350 Hz, and 1 of these subjects in the elderly group also had a negative resistance at 5040 Hz. Therefore, these negative resistance data were excluded, and the mean data for the subsequent analyses for impedance and energy reflectance were based on the remaining data at these frequencies for both groups.

TABLE IV. Gender comparison for the elderly group for tympanometric and ear-canal measurements. The means and SDs of values are shown for peak-compensated static acoustic admittance (Peak Y_{tm}), tympanometric width (TW), and equivalent ear canal volume (V_{ea}), all obtained using a 226 Hz probe tone. Also shown are the results of the two-tailed t tests and associated p values for the mean comparisons. The p-values for significant t-test results are underlined.

	Women		Men		Mean comparisons	
Measure	Mean	SD	Mean	SD	t	p
Peak Y_{tm} (mmho)	0.53	0.15	0.70	0.19	2.71	<u>0.01</u>
TW (daPa)	80.3	21.3	77.0	13.5	0.51	0.61
V_{ea} (cm^3)	1.34	0.19	1.67	0.40	2.83	<u><0.01</u>

An independent-samples t-test was used to evaluate the differences between age groups and genders on wideband measurements of impedance and energy reflectance for each of the 15 one-third-octave frequencies from 250 to 6350 Hz. This analysis assumes that the impedance and energy reflectance data for each one-third octave are independent. This approach was also used by Keefe *et al.* (1993) in their analysis of age effects and is supported by recent data showing that middle ear disorders such as ossicular disarticulation can affect a narrow frequency range while leaving other frequencies unaffected (Feeney, Grant *et al.*, 2003).

The one-third octave averaged impedance in cgs ohms measured at the probe tip is illustrated in the top panel of Fig. 1 for the young-adult and elderly groups. The error bars in Figs. 1–3, 5 and 6 represent ± 1SE, which were chosen rather than ± 1 SD to facilitate the visual comparison of mean differences. The SDs for each measurement are reported in Table V. The impedance decreased from 250 to 5040 Hz for the young-adult group consistent with a stiffness-controlled system, and increased above this frequency consistent with a mass-controlled system. The impedance for the elderly group was similar to that of the young-adult group from about 1500 to 3000 Hz, but diverged at lower and higher frequencies. There were significant differences between groups in impedance, with impedance for the elderly group averaging around 50 cgs ohms lower than that of the young group at frequencies between 315 and 1260 Hz (Table V).

The group-mean impedance phase measured at the probe tip for the young-adult and elderly groups is shown in the bottom panel of Fig. 1. Impedance phase was stiffness dominated in the low frequencies, and mass dominated above 4000 Hz for both groups. Impedance phase was significantly more negative for the elderly group at 500 Hz by around 2° but was significantly more negative for the young-adult group by around 8° over the frequency range from 1000 and 2000 Hz (Table V). Similar to the data of Keefe *et al.* (1993) there was increased variability in the phase measurements at frequencies above 2000 Hz.

The real and imaginary parts of the acoustic impedance, resistance and reactance, are displayed in Fig. 2 in cgs ohms as a function of frequency for both groups. As detailed in Table V, there was a small difference between groups in resistance with significantly lower resistance for the elderly group by an average of 19 cgs ohms between 315 and 794 Hz, and significantly higher resistance for the elderly group

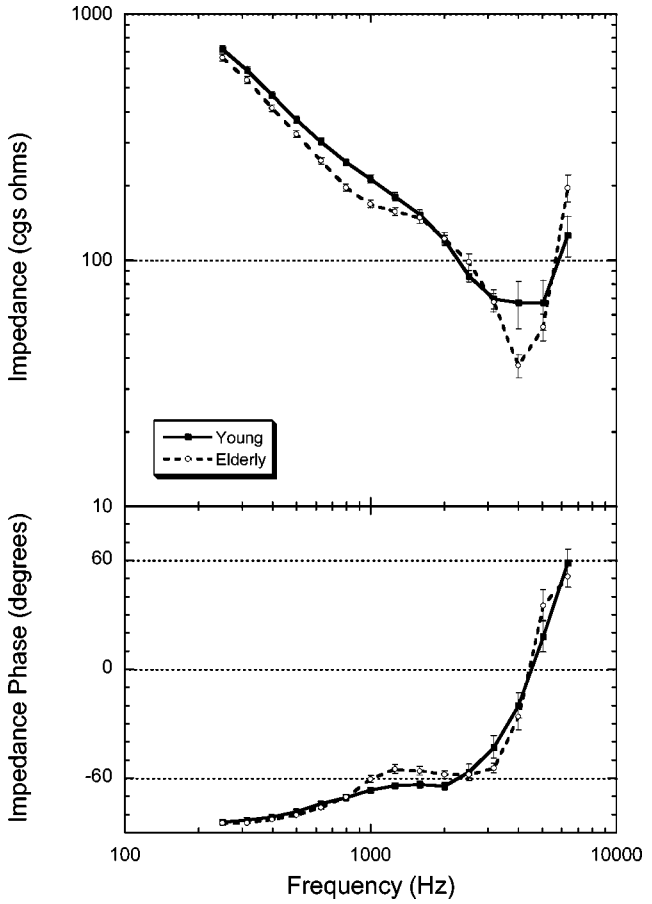


FIG. 1. The panels show the group mean one-third-octave impedance (cgs ohms, top) and impedance phase (degrees, bottom) for the young-adult group (N=40, solid line and filled squares) and the elderly group (N=30, dashed line and open circles) as a function of frequency. The error bars are ±1 SE.

of around 16 cgs ohms between 1587 and 2520 Hz, and also at 6350 Hz with a large difference of 78 cgs ohms. The reactance for the young group was significantly more stiffness-dominated from 397 to 1587 Hz by an average of 41 cgs ohms.

D. Wideband energy reflectance

An acoustic estimate of ear canal diameter for the young-adult and elderly groups was derived using a sum rule on resistance at the probe tip (Keefe *et al.*, 1992). As shown in Table II, the mean estimated ear-canal diameter for the young group (mean=10.67 mm) was larger than that of the elderly group (mean=9.25 mm) and this difference was statistically significant, $t(28)=2.07$, $p=0.04$. The characteristic impedance (Z_c) of a tube with cross-sectional area S is approximately equal to $Z_c=\rho c/S$, where ρ is the density of air and c is phase velocity of sound in air. Z_c along with the impedance measured in the ear canal, Z_{ec}, is used to calculate the pressure reflectance magnitude, $|R|=(Z_{ec}-Z_c)/(Z_{ec}+Z_c)$. The energy reflectance is equal to $|R|^2$ (Keefe *et al.*, 1992; Voss and Allen, 1994).

There were no statistically significant differences in energy reflectance between right and left ears for the young-adult group at any frequency. The only significant difference between right and left ears for the elderly group was at 6350 Hz (mean=0.72, SD=0.17) with higher energy reflectance in the right ear than the left ear (mean=0.55, SD=0.17). Therefore, the data for right and left ears were pooled for the remaining energy reflectance analyses. The group-mean one-third-octave energy reflectance for the young-adult and elderly groups is shown in Fig. 3. The energy reflectance was 0.96 at 250 Hz for the young adults and decreased with increasing frequency, dropping to 0.69 at 1000 Hz, and to a minimum of 0.24 at 4000 Hz before it increased to 0.64 at 6350 Hz. The group-mean energy reflectance for the elderly group was also 0.96 at 250 Hz, but decreased more steeply as a function of frequency, reaching a value of 0.52 at 1000 Hz, and then decreased more gradually with frequency, reaching a minimum of 0.40 at 4000 Hz before it increased to 0.62 at 6350 Hz. A *t*-test for independent samples was conducted to examine mean differences at each one-third octave frequency. The results are presented in Table V which shows the means, standard deviations and two-tailed *t*-test results. Energy reflectance was significantly lower for the elderly group by an average of 15% from 794 to 2000 Hz but significantly higher for the elderly group at 4000 Hz by 20%.

In examining individual energy reflectance plots for subjects from both groups, it was noted that the reflectance patterns for the young-adult subjects tended to have a single minimum near 4000–5000 Hz (similar to the mean data in Fig. 3), but that the reflectance patterns for the elderly subjects tended to have two minima: One between 1000 and 2000 Hz and a second between 4000 and 5000 Hz. Nineteen out of 30 elderly subjects (63%) had reflectance patterns with two energy reflectance minima lower than 0.5 with the average frequency for the low-frequency minimum at 1504 Hz (SD=463 Hz) and the average frequency for the high-frequency minimum at 4045 Hz (SD=1092 Hz). However, only 6 of the 40 young adults (15%) had secondary, low-frequency minima. This double-minima pattern for the elderly group is not evident from the mean data in Fig. 3, thus Fig. 4 shows four individual energy reflectance plots from each group representative of the majority of the patterns for each group (65% of the elderly and 85% of the young subjects).

Group-mean energy reflectance for women and men of the young-adult group is shown in Fig. 5. Women had 10% higher energy reflectance than men at 794 and 1000 Hz but 18% lower energy reflectance at 5040 Hz (Table V). The group-mean energy reflectance data were similar for women and men in the elderly group. These data are presented in Fig. 6, with a similar trend as shown for the young adults, with women having higher reflectance in the mid-to-low frequencies and lower reflectance at 4000 Hz. However, these gender differences were not statistically significant at any frequency for the elderly group.

IV. DISCUSSION

A. 226 Hz tympanometry

The subjects in this study were selected to have negative histories of middle ear disorders, normal otoscopic findings,

TABLE V. Young and elderly group means and standard deviations for impedance level and impedance phase measured at the probe tip, resistance and reactance measured at the probe tip, and energy reflectance. Mean energy reflectance and standard deviations are also shown for women and men in the young group. Also shown are the results of the two-tailed *t* tests and associated *p* values for the mean comparisons with an apha level of 0.05. The *p*-values for significant *t*-test results are underlined.

		One-third Octave Frequency (Hz)														
Measurement		250	315	397	500	630	794	1000	1260	1587	2000	2520	3175	4000	5040	6350
Impedance	Young	719	591	467	371	303	249	213	181	153	120	86	69	67	67	126
(cgs ohms)	SD	145	121	98	78	63	51	52	44	40	35	30	39	92	97	149
	Elderly	662	537	413	324	253	197	169	157	149	122	98	67	37	53	196
	SD	115	92	73	57	44	37	34	33	42	41	40	32	22	36	133
	t	1.79	2.04	2.55	2.77	3.69	4.76	4.06	2.49	0.45	0.17	1.45	0.22	1.73	0.72	1.97
	p	0.08	0.045	0.01	<0.01	<0.01	<0.01	<0.01	0.02	0.65	0.87	0.15	0.83	0.09	0.48	0.053
Impedance	Young	−84.1	−83.2	−81.3	−78.3	−74.0	−70.6	−66.6	−64.3	−63.6	−64.5	−56.8	−42.9	−20.0	18.0	58.6
Phase	SD	3.1	3.1	3.5	4.5	5.9	8.0	7.9	8.5	10.0	11.3	25.0	33.6	38.6	47.4	40.6
(degrees)	Elderly	−84.9	−84.6	−82.8	−80.5	−76.3	−70.4	−60.2	−55.1	−56.0	−58.2	−58.1	−54.5	−26.1	35.0	51.0
	SD	3.3	2.9	3.0	3.6	5.2	6.7	10.2	13.7	12.2	11.1	10.7	14.9	40.9	47.2	31.6
	t	0.96	1.95	1.82	2.12	1.68	0.15	2.93	3.44	2.84	2.33	0.28	1.77	0.63	1.47	0.81
	p	0.34	0.06	0.07	0.04	0.10	0.89	<0.01	<0.01	<0.01	0.02	0.78	0.08	0.53	0.15	0.42
Resistance	Young	72	69	69	73	81	81	85	78	67	50	40	41	51	34	50
(cgs ohms)	SD	43	34	27	27	29	33	43	29	30	22	17	34	88	45	106
	Elderly	57	49	51	53	59	65	82	89	86	66	53	37	24	25	128
	SD	34	25	22	21	23	23	33	39	47	38	30	19	13	19	134
	t	1.55	2.61	2.86	3.39	3.47	2.25	0.33	1.41	2.04	2.20	2.31	0.62	1.66	0.99	2.63
	p	0.12	0.01	<0.01	<0.01	<0.01	0.03	0.74	0.16	0.04	0.03	0.02	0.54	0.10	0.33	0.01
Reactance	Young	−715	−586	−461	−363	−290	−233	−193	−162	−135	−107	−71	−37	−9	5	89
(cgs ohms)	SD	145	121	98	79	64	52	43	41	36	34	37	46	51	104	129
	Elderly	−659	−534	−409	−319	−245	−185	−145	−125	−117	−100	−81	−55	−18	37	113
	SD	116	93	73	57	44	37	31	31	25	29	33	30	29	43	98
	t	1.75	1.96	2.45	2.56	3.32	4.39	5.20	4.11	2.33	0.98	1.14	1.81	0.84	1.56	0.81
	p	0.09	0.054	0.02	0.01	<0.01	<0.01	<0.01	<0.01	0.02	0.33	0.26	0.07	0.40	0.12	0.42
Energy	Young	0.96	0.95	0.93	0.88	0.82	0.76	0.69	0.63	0.60	0.58	0.49	0.35	0.24	0.33	0.64
Reflectance	SD	0.03	0.04	0.05	0.07	0.10	0.13	0.15	0.17	0.18	0.17	0.16	0.16	0.15	0.20	0.19
	Elderly	0.96	0.94	0.91	0.86	0.77	0.66	0.52	0.45	0.44	0.44	0.42	0.40	0.40	0.42	0.62
	SD	0.03	0.04	0.06	0.08	0.12	0.13	0.18	0.20	0.18	0.19	0.16	0.14	0.14	0.20	0.19
	t	0.92	0.74	1.30	1.20	1.66	3.12	4.30	4.04	3.53	3.36	1.76	1.50	4.76	1.99	0.31
	p	0.36	0.46	0.20	0.24	0.10	<0.01	<0.01	<0.01	<0.01	<0.01	0.08	0.14	<0.01	0.051	0.76
Energy	Women	0.97	0.96	0.94	0.90	0.85	0.80	0.74	0.67	0.64	0.62	0.51	0.35	0.19	0.24	0.61
Reflectance	SD	0.03	0.03	0.04	0.06	0.05	0.11	0.12	0.15	0.16	0.13	0.15	0.17	0.13	0.12	0.19
(Young Group)	Men	0.96	0.95	0.91	0.86	0.79	0.71	0.63	0.59	0.56	0.54	0.47	0.35	0.28	0.42	0.66
	SD	0.31	0.04	0.05	0.07	0.10	0.13	0.16	0.18	0.19	0.20	0.17	0.14	0.15	0.22	0.20
	t	0.65	1.04	1.65	1.93	1.80	2.40	2.40	1.59	1.46	1.40	0.74	0.16	2.02	3.14	0.85
	p	0.52	0.30	0.11	0.06	0.08	0.02	0.02	0.12	0.15	0.17	0.47	0.88	0.051	<0.01	0.40

and single-peaked 226 Hz tympanometry yielding Peak Y_{tm} in the normal adult range (Margolis and Hunter, 1999). The young-adult and elderly groups did not differ in respect to Peak Y_{tm} or TW, two measures that index middle-ear function, but were slightly different in V_{ea}, with the elderly group having a V_{ea} that was 0.2 cm^3 larger than that of the younger group. The difference in V_{ea} is not in agreement with Wiley *et al.* (1996) who reported that V_{ea} decreased with age. However, this trend was significant only for right ears in their subjects aged 48–90 years. In about one-quarter of the elderly subjects in the present study, the presence of an oval opening to the ear canal necessitated the use of an oval probe tip to obtain a hermetic seal for tympanometry. These tips do not fit as deeply into ear canals as circular tips; therefore, this may have contributed to the measurement of larger V_{ea} in the elderly group. In fact, group-mean estimates of ear-canal diameter obtained with the wideband reflectance system showed a significantly smaller diameter in the elderly group of 9.25 mm compared to 10.67 mm for the young-adult group (Table II). These ear-canal-diameter estimates are in good agreement with those of Keefe *et al.* (1993), obtained with the same reflectance system, of 10.4 mm for adults who ranged in age from 20 to 50 years.

A gender difference in 226 Hz tympanometry for the younger group was found only for V_{ea}. However, for the older group, men also had a significantly greater Peak Y_{tm} in addition to a larger V_{ea} than women. This is consistent with the data of Wiley *et al.* (1996) who reported that Peak Y_{tm} and V_{ea} were both greater in men than women, even when data were adjusted for age in subjects aged 48–90 years. They also reported a significantly wider TW in men of <3 daPa, which was not observed in the present study. This differs from the data of Gates *et al.* (1990) who found no gender differences in Peak Y_{tm} for their elderly subjects aged 63–95 years.

B. Wideband energy reflectance and impedance

1. Error analysis

The energy reflectance data for the young adults in the present study differed somewhat from the adult data of Keefe

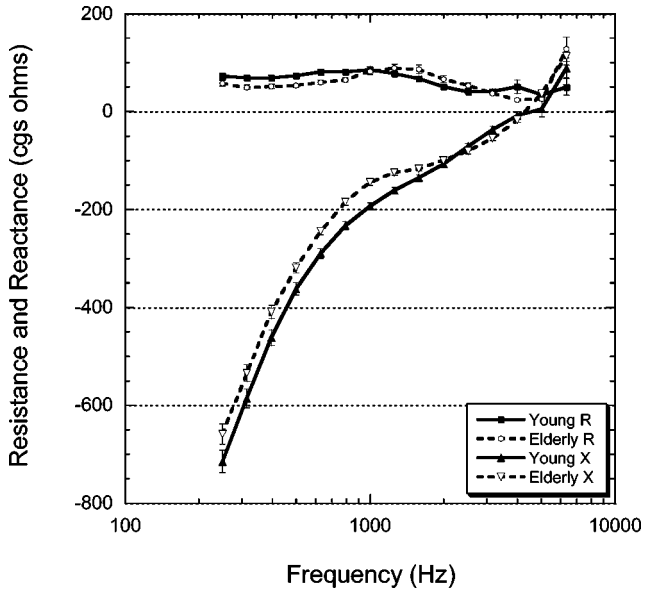


FIG. 2. The group mean one-third-octave resistance (R, cgs ohms) for the young-adult group (N=40, solid line and filled squares) and the elderly group (N=30, dashed line and open squares) as a function of frequency. Also plotted is the group mean one-third-octave reactance (X, cgs ohms) for the young-adult group (solid line and filled triangle) and elderly group (dashed line and open inverted triangle) as a function of frequency. The error bars are ±1 SE.

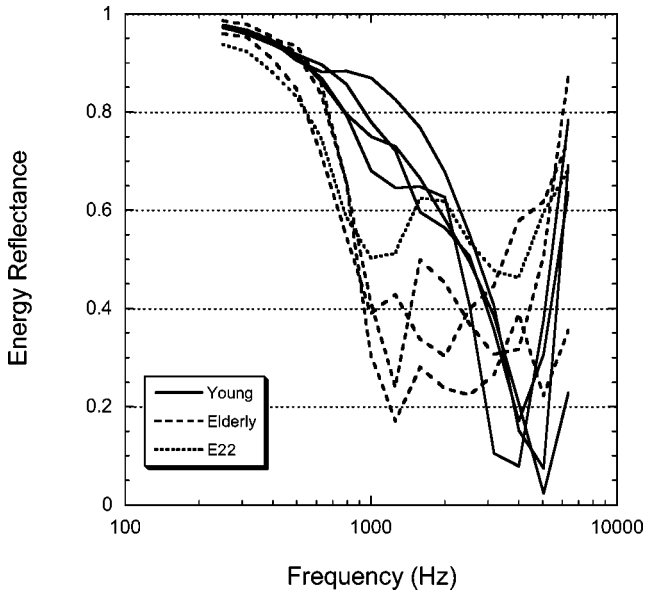


FIG. 4. Representative plots of energy reflectance for 4 young-adult subjects (solid lines) and 4 elderly subjects (dashed lines), including elderly subject E22.

et al. (1993), which were collected using the same method (Fig. 7). One difference in the results between the studies is in regard to the sharp minimum of 0.25 in energy reflectance at 4000 Hz in the present study compared to 0.37 at that frequency for the Keefe *et al.* study. However, Keefe *et al.* reported that individual subjects often had a reflectance minimum <0.20 (their Fig. 17, p. 2635) and that an energy reflectance minimum in the range of 0.1–0.2 was more typical. The minimum in energy reflectance was reported by Keefe *et al.* to occur for that study at 3600±1200 Hz, so that averaging across subjects tended to overestimate the energy reflectance minimum.

Several differences between the studies could account for these differences in energy reflectance. One difference is that the adult subjects in the Keefe *et al.* (1993) study were older than the young-adult subjects in the present study, ranging in age from 20 to 50 years old (D.H. Keefe, personal communication), so it is possible that some of the variance in the energy reflectance patterns across subjects could have been age-related. Another difference is in the type of probe microphone used. Keefe *et al.* used an Etymotic Research,

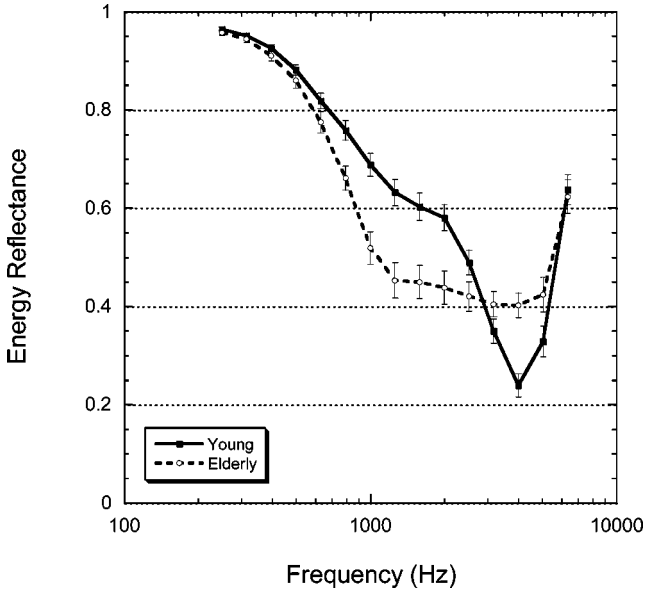


FIG. 3. The group mean one-third-octave energy reflectance for the young-adult group (N=40, solid line and filled squares) and the elderly group (N=30, dashed line and open circles) as a function of frequency. The error bars are ±1 SE.

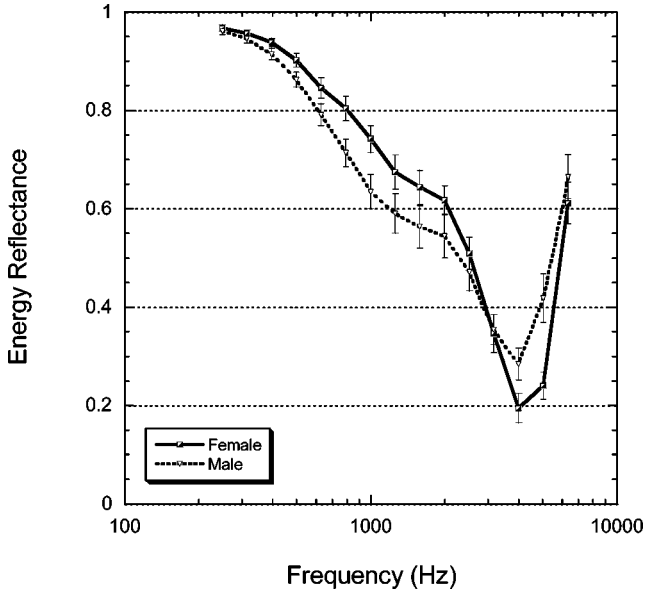


FIG. 5. The group mean one-third-octave energy reflectance for young-adult women (N=20, solid line and half-filled squares) and young-adult men (N=20, dashed line and inverted triangles) as a function of frequency. The error bars are ±1 SE.

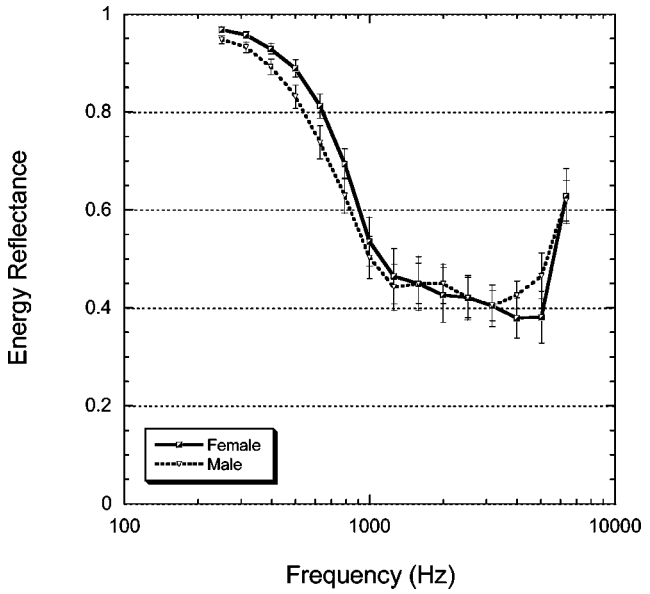


FIG. 6. The group mean one-third-octave energy reflectance for elderly women (N=15, solid line and half-filled squares) and elderly men (N=15, dashed line and inverted triangles) as a function of frequency. The error bars are ±1 SE.

ER-7C probe microphone, which had the microphone outlet extended 3 mm beyond the receiver port. This probe configuration reduces the effect of nonuniform sound waves created by the receiver-load interface, thus improving the accuracy of impedance measurements at high frequencies (Huang *et al.*, 2000a). The standard foam tip used with the ER-10C probe in the present study had no such microphone extension, and all participants in both groups were tested with the same probe. Huang *et al.* 2000a observed that in a cavity with an admittance value similar to the human ear, the ER-10C probe tip without the microphone extension introduced a greater error in measuring admittance than a probe tip with a microphone extension when compared to model predictions for frequencies above 3000 Hz [Huang *et al.* 2000a, Fig. 2(B), lower]. The measurement error resulted in higher admittance for the tip without the extension compared to the tip with the extension by about 1 dB at 3000 Hz to a maximum of 5 dB at 5000 Hz. However, since the standard ER-10C probe tip without the extension was used for subjects in both groups in the present study, it is expected that the differences between the groups would not be affected by errors in absolute measurement at the high frequencies.

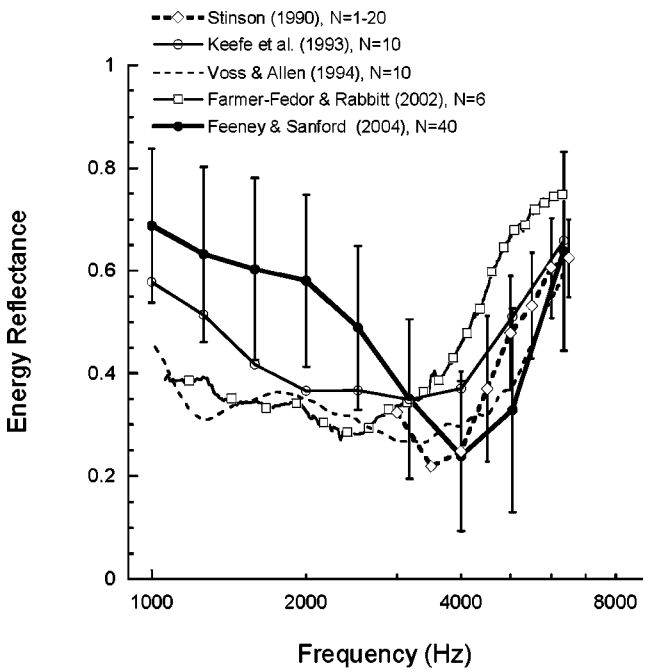


FIG. 7. Energy reflectance as a function of frequency in adult humans for 4 published studies, and young-adult data from the present study. The legend also shows the number of ears that were included in the measurements. The error bars represent ±1 SD for the present study and for the data from Stinson (1990).

The data in Fig. 7 show good agreement at frequencies of 3000 Hz and higher among the energy reflectance estimates of the present study for young adults and four published studies of adult energy reflectance. The data in Fig. 7 from Stinson (1990) were obtained from a total of 20 ear canals from 13 subjects (19–45 years, median=24 years) with data from 1 subject at 3000 and 3500 Hz, 4 subjects at 4000 Hz, and 8 or more subjects at higher frequencies. These energy reflectance data were obtained using an analysis of standing wave patterns in the open ear canal. The study by Keefe *et al.* (1993) made measurements in 10 adults (age 20–50 years) using the system of Keefe *et al.* (1992). The data from Voss and Allen (1994) were obtained for 10 subjects (18–24 years) using an energy reflectance method similar to that of Keefe *et al.* (1992) with an ER-7C microphone with a microphone extension. These measurements were obtained in sealed ear canals as in the Keefe *et al.* (1993) study, but an average value of ear canal diameter of 0.74 cm was used in the calculation of energy reflectance rather than individual estimates. Finally, the energy reflectance data from Farmer-Fedor and Rabbitt (2002) were obtained from measurements in 6 adults (age 22–34 years) using a method based on the sound pressure distribution in the open ear canal.

At frequencies of 3000 Hz and higher, the greatest difference between the mean data from the present study and the other studies shown in Fig. 7 is at 5000 Hz, which is the frequency at which the Huang *et al.* (2000a) showed the greatest deviation for the probe configuration used in this study compared to a probe with a microphone extension. However, even at this frequency, the data agree with the data from the study by Voss and Allen (1994) collected using a probe tip with a microphone extension in adults of the same age. Moreover, at around 6000 Hz the data from the present study are in agreement with most of the studies with the possible exception of the data from Farmer-Fedor and Rabbitt (2002). The greatest disagreement between the data for the present study and those of the other studies occurs for frequencies below 3000 Hz, at which the probe-microphone extension does not affect the data. For these frequencies, the present study shows higher energy reflectance than the other studies. In addition to age differences between subjects in the present study and some of the previous studies, the methodology of the present study varied slightly from that of the other two studies that measured energy reflectance in the sealed ear canal (Keefe *et al.*, 1993; Voss and Allen, 1994).

In order to avoid a leaky probe fit that might invalidate the differences between young and elderly subjects, the protocol for the present study called for a full insertion depth of the probe tip and for the tip to remain in place for 2 min to allow full expansion of the foam prior to data collection. If the energy reflectance was less that 0.8 after 2 min, the probe was reinserted and the process repeated. The result of this protocol was that the lowest energy reflectance for the 40 young and 30 elderly subjects at 250 Hz was 0.86, and the mean energy reflectance for both groups was 0.96 at 250 Hz compared to a mean of 0.92 at 250 Hz for Keefe *et al.* (1993) and 0.91 at 234 Hz for the data of Voss and Allen (1994). It is proposed that this procedural difference reduced the probability of a leaky probe fit, and as a result and yielded higher energy reflectance values at frequencies below the reflectance minimum (4000 Hz).

When the ear canal is of a different diameter than the calibration tubes, the effective Thevenin equivalents change. This can lead to errors in the calculation of wideband admittance and reflectance measurements (Huang *et al.*, 2000a, Keefe *et al.* 1992). The calibration tube set in this study had a diameter of 8 mm, which was smaller than the mean acoustic estimates of ear canal diameter for both groups. To assess the effect on group differences that extremes in ear canal diameter might have, the energy reflectance data were reanalyzed with the data from subjects with either the 5 largest or 5 smallest ear-canal diameters excluded. Gender differences were also reassessed in this manner with data from men and women with the 5 largest or 5 smallest ear-canal diameters excluded for the young-adult group. Data from men and women with the 3 largest or 3 smallest ear-canal diameters for the elderly group were excluded, with fewer data excluded for this analysis due to the smaller number of subjects of each gender in the elderly group.

In general, the results of the revised analysis did not alter the basic shape of the energy reflectance responses across frequency for either age group or gender within each age group. However, with data excluded from subjects with large-diameter canals, there were 9 frequencies at which the young and elderly groups were significantly different compared to 6 frequencies in the original analysis (Table V). All 6 of the original frequencies were still significantly different, showing the same trend, and the 3 additional frequencies at which significant differences were observed were 630 and 2520 Hz at which the young group had higher energy reflectance and 5040 Hz at which the young group had lower energy reflectance. With data excluded from subjects with small-diameter canals, there was one additional frequency that was significantly different between the groups; that being 5040 Hz for which the young group had lower energy reflectance.

The revised analysis of the effect of gender for the young group with data from subjects with large ear canals excluded revealed two additional frequencies that were significantly different between genders; those being 1260 Hz at which women had higher energy reflectance and 4000 Hz at which women had lower energy reflectance. With data excluded from subjects with small-diameter canals, there were only 2 frequencies for which significant differences existed between genders, at 500 and 1000 Hz, with women having higher energy reflectance. This is compared to 3 frequencies at which there were significant group differences in the original analysis; 794, 1000, and 5040 Hz (Table V). Finally, for the elderly group, the revised analysis of gender differences with data from subjects with large ear canals excluded revealed significantly higher energy reflectance for females at 250, 315, and 500 Hz where there had been no significant differences at any frequency in the previous analysis. With data excluded from subjects with small-diameter canals, there was significantly higher energy reflectance for females at 1260 Hz, where there had been no significant differences in the original analysis (Table V). Overall, the results of these revised analyses for age and gender suggest that the effect of including subjects with extreme values of ear canal diameter reduced the number of frequencies at which there were significant differences in energy reflectance between groups, but that the basic trends in the data were unchanged.

A problem associated with the interpretation of impedance measurements made from the ear canal at ambient pressure is the volume of air between the probe and the tympanic membrane with its associated acoustic impedance. Clinical admittance tympanometers routinely correct for the ear canal component of admittance at 226 Hz by pressurizing the ear canal and subtracting the resultant ear canal admittance from the total admittance at peak tympanometric pressure (Van Camp *et al.*, 1986). Margolis *et al.* (1999) corrected for the impedance of the ear canal in their wideband impedance data, up to 8000 Hz, using a transmission-line model of the middle ear based on the method of Lynch *et al.* (1994). This method incorporated an estimate of ear canal volume derived from a low-frequency admittance measurement at -300 daPa ear canal pressure. Margolis *et al.* (1999) obtained this admittance measurement with the same probe tip used to make the wideband impedance measurements. One limitation of the present study is that this type of ear-canal correction could not be applied to the data because the impedance measurements were obtained with a different probe tip and probe-insertion depth than the wideband impedance measurements, thus the ear canal volume estimates would be different for the two measurements.

2. Age effects

A simple model of the middle ear consisting of a series combination of resistance, stiffness, and mass can be used to predict changes in acoustic impedance and energy reflectance responses (Feeney and Keefe, 1999; Johansen, 1948; Margolis *et al.*, 1999). Figure 8 (top panel) shows the model output for the impedance level in dB ($10\times\log_{10}$ Impedance) as a function of frequency for the case where stiffness has been decreased by factors of 0.5 and 0.1. The model output can be compared with the impedance data for the young and elderly subjects in Fig. 1. The main feature of the data, which is lower impedance at frequencies below the impedance minimum for the elderly group, is accounted for in the model by a decrease in middle-ear stiffness. The model output for energy reflectance for the same stiffness conditions is also shown in Fig. 8 (bottom panel). This can be compared with the energy reflectance data for the young and elderly subjects

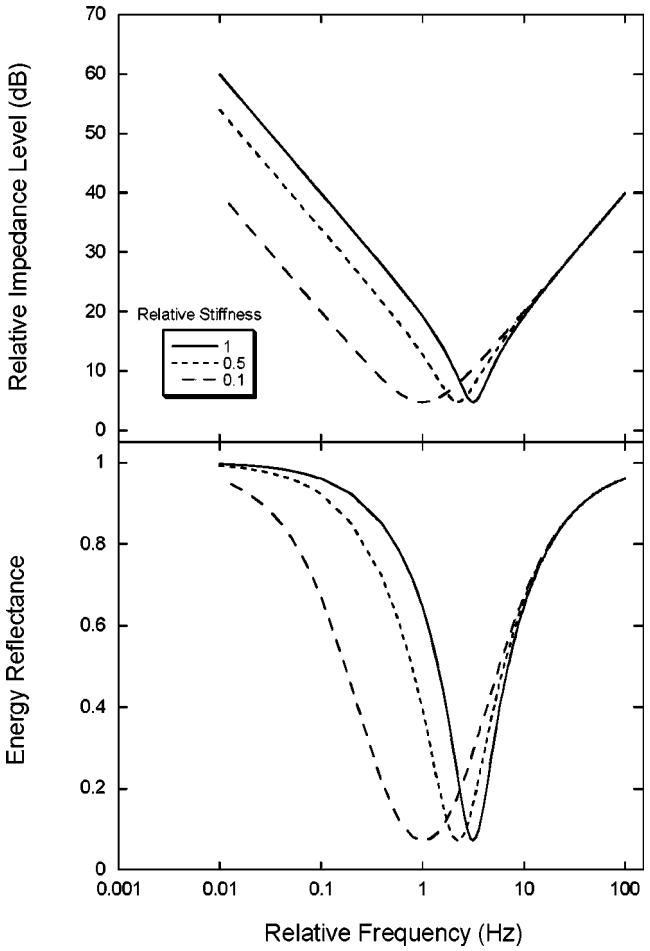


FIG. 8. Output from a simple model of the middle ear consisting of a series combination of resistance, stiffness, and mass showing the effect of changing the stiffness component by a factor of 0.5 and 0.1 on the impedance level ($10\times\log_{10}$ Impedance, upper panel) and energy reflectance (lower panel).

in Fig. 3. The main feature of the data, showing lower energy reflectance for the elderly group at frequencies below the reflectance minimum and higher reflectance at frequencies above the reflectance minimum, is observed in the model. However, the reflectance minimum for the elderly group is not as low as that observed with the model. Keefe *et al.* (1993) reported that individual subjects in their study had deep reflectance notches that occurred at various frequencies between 2400 and 4800 Hz, and that data averaging reduced the overall depth of the notch for the group data. This may have contributed to the shape of the mean energy reflectance pattern for the elderly subjects in the present study combined with the effect of multiple reflectance notches reported above.

If the difference in energy reflectance between age groups is the result of differences in middle-ear stiffness, it is somewhat surprising that a difference between groups was not observed for Peak Y_{tm} at 226 Hz; a frequency which should be affected by a change in stiffness. However, normal 226 Hz tympanometry has previously been observed in cases of ossicular disarticulation and otosclerosis (Feeney, Grant *et al.*, 2003), as well as in otitis media (Hunter and Margolis, 1997), which all affect middle ear stiffness. Therefore, the lack of sensitivity of this 226 Hz measure for age effects is not totally surprising.

Only 6 of the young-adult subjects (15%) in the present study had both low- and high-frequency minima in the pattern of energy reflectance. However, the data from 63% of elderly subjects revealed two reflectance minima, with the average frequency for the low-frequency minimum at 1504 Hz (SD=463 Hz) and the average frequency for the high-frequency minimum at 4045 Hz (SD=1092 Hz). Thus, the flatter shape of the data curve of the elderly group's mean energy reflectance for frequencies from 1000 to 4000 Hz is likely related to the presence of this second, low-frequency minimum observed in elderly ears. A marked low-frequency minimum in the energy reflectance pattern, occurring at frequencies below 1000 Hz, has been reported in ears with reduced stiffness due to a hypermobile tympanic membrane or ossicular disarticulation (Feeney, Grant *et al.*, 2003). In these cases of middle ears with decreased stiffness, the reflectance minimum indicates a frequency region where the middle ear impedance is close to the characteristic impedance of the ear canal, resulting in increased power absorption by the middle ear but with reduced power transfer to the cochlea, especially in the case of disarticulation. The secondary low-frequency reflectance minimum observed in the majority of the elderly patients in the present study may be related to a smaller reduction in middle-ear stiffness in these subjects.

The findings of this study are consistent with those of increasing admittance with age reported for Peak Y_{tm} at 660 Hz by Beattie and Leamy (1975). The findings would also appear to agree with the report of Marshall *et al.* (1983) of a ''loosening in the middle ear system'' in some of their elderly subjects. The findings of the present study are also consistent with the hypothesis put forward by Nixon *et al.* (1962) that the middle ear has a loss of stiffness with age, proposed to explain their behavioral data of a high-frequency air-bone gap at 4000 Hz in the elderly. Nixon *et al.* (1962) suggested that this occurred due to a diminished mechanical integrity of the articulations of the ossicular chain.

Of interest in this regard is the group-mean air-bone gap data in Table I for the present study. There were significantly larger air-bone gaps for the elderly group at 250 and 4000 Hz. The largest air-bone gap occurred for the elderly group at 4000 Hz, which was 6.3 dB compared to 1.4 dB for the younger group. This was not likely the result of ear-canal collapse for the subjects in the present study given the use of insert earphones for pure-tone air-conduction testing. Nixon *et al.* (1962) reported a 12 dB air-bone gap at 4000 Hz for their oldest group (50–59 years). This degree of conductive component could not have been reached with subjects in the present study because the inclusion criterion restricted the range of permitted air-bone gaps to $\leqslant$10 dB. Therefore, the results of the present study may underestimate the degree of the high-frequency conductive component in the elderly.

The difference in energy reflectance between the young and elderly groups at 4000 Hz was 2.3 dB [$10\times\log_{10}$ (Energy $\text{Reflectance}_{\text{Elderly}}$/Energy $\text{Reflectance}_{\text{Young}}$)] compared to 4.9 dB difference in air-bone gap, and there was no appreciable difference between groups in energy reflectance at 250 Hz. Thus, there does not appear to be a one-to-one relation-

ship between the change in this wideband acoustic feature and the small degree of conductive component for the subjects in this study. However, Keefe and Simmons (2003) reported that energy transmittance (1-energy reflectance) was an excellent predictor of conductive hearing loss for air-bone gaps $\geqslant$20 dB. Comparing tests at a fixed specificity of 0.90, the sensitivity of pressurized transmittance at predicting a conductive loss was 0.94 compared to only 0.28 for Peak Y_{tm} at 226 Hz. More work is needed to examine the ability of these wideband acoustic measures to predict the degree of conductive hearing loss.

The present findings, suggesting decreased stiffness in the aging middle ear, are not supported by anatomical studies showing changes in ossicular joints and in the tympanic membrane that would result in increased middle-ear stiffness with age (Etholm and Belal, 1974; Harty, 1953; Ruah *et al.*, 1991). However, Doan *et al.* (1996) proposed that the elderly middle ear may have a reduction in tonic middle-ear muscle contraction as suggested by reduced acoustic stapedius reflex amplitude (Hall, 1982; Jerger and Oliver, 1987), which would act to reduce stiffness in the ossicular chain and tympanic membrane. Doan *et al.* (1996) also suggested that changes in tonic muscle contraction, middle ear ligaments, or ossicular joints could also alter the rotational axis of the ossicles, which would alter middle-ear inertia leading to more high-frequency changes with aging.

One inclusion criterion of this study was an audiometric air-bone gap $\leqslant$10 dB to rule out significant middle-ear pathology. Therefore, the differences observed in the wideband impedance and reflectance could have no more influence on hearing than a 10 dB conductive component. Even given that restriction, the elderly group had significantly greater air bone gaps at 250 and 4000 Hz than the younger group, with the greatest conductive component at 4000 Hz in agreement with the behavioral data of Nixon *et al.* (1962). Thus, it is possible that the present study underestimated the effect of aging on the middle ear by restricting the inclusion of elderly subjects with greater high-frequency air-bone gaps, which may be the norm in this population.

One consequence of reduced high-frequency middle ear function in the elderly would be a deleterious effect on the measurement of otoacoustic emissions, the amplitudes of which rely in part on the reverse transfer function of the middle ear. In a study on the effect of aging on distortion-product otoacoustic emissions (DPOAEs), Lonsbury-Martin *et al.* (1991) reported that there were some subjects who had high-frequency hearing thresholds in the normal range, but who had a dramatic loss in DPOAE amplitudes. One possibility posed by Lonsbury-Martin *et al.* (1991) to account for this anomaly was the presence of a subtle change in the reverse transfer function of the middle ear. This might be the case for an elderly ear as shown in the data for subject E22, a 69-year-old man (Fig. 4) who had normal 226 Hz tympanometry with Peak Y_{tm} at 0.7 mmho and a PTA of 13 dB HL. His poorest air-conduction threshold was 20 dB HL at 4000 Hz, with a bone-conduction threshold at that frequency of 10 dB HL, a 10 dB air-bone gap. This subject's reflectance pattern showed the characteristic double notching discussed above for the elderly group, with energy reflectance from 1000 to 4000 Hz that is higher than the average for this age group. Although he had normal hearing (considering the 10 dB air-bone gap as within normal limits) and normal middle ear function according to traditional measures, his energy reflectance data suggest that his middle ear would be less efficient at power absorption than younger ears and, perhaps, less efficient at reverse transmission of acoustic energy.

V. CONCLUSIONS

Wideband energy reflectance and impedance measurements at ambient pressure suggested a decrease in middle-ear stiffness with age. This was observed in the absence of a middle-ear age effect with 226 Hz tympanometry. The results appear to be at variance with anatomical studies suggesting increased middle-ear stiffness with age. However, an age-related reduction in tonic middle-ear muscle contraction along with possible changes in the rotational axis of the ossicles with age could account for these findings (Doan *et al.* 1996). These results suggest the possibility of a conductive component to presbycusis, and may account for an age-related reduction in the amplitude of otoacoustic emissions.

ACKNOWLEDGMENTS

The authors thank Douglas Keefe for providing software for the measurement of wideband acoustic transfer functions and for helpful discussions throughout data collection and analysis. We thank David Mills for helpful comments on an earlier draft of this manuscript and for helpful discussion on the middle-ear model. We also thank Brenda Farmer-Fedor, Douglas Keefe, Michael Stinson, Richard Rabbitt, and Susan Voss for providing data from their reflectance studies. Data collection for this study took place at Ohio State University and Ohio University. This research was supported by a grant from NIH NIDCD, #R03 DC 04129.

American National Standards Institute (**1987**). ''Specifications for instruments to measure aural acoustic impedance and admittance (aural acoustic immittance) (ANSI S3.39-1987),'' New York: ANSI.

American National Standards Institute (**1996**). ''Specification for Audiometers (ANSI S3.6-1996),'' New York: ANSI.

Beattie, R. C., and Leamy, D. P. (**1975**). ''Otoadmittance: normative values, procedural variables, and reliability,'' J. Am. Audiol Soc. **1**, 21–27.

Blood, I. M., and Greenberg, H. (**1977**). ''Acoustic admittance of the ear in the geriatric person,'' J. Am. Audiol Soc. **3**, 185–187.

Doan, D. E., Erulkar, J. S., and Saunders, J. C. (**1996**). ''Functional changes in the aging mouse middle ear,'' Hear. Res. **97**, 174–177.

Etholm, B., and Belal, A. J. (**1974**). ''Senile changes in the middle ear joints,'' Ann. Otolaryngol. **83**, 49–54.

Farmer-Fedor, B. L., and Rabbitt, R. D. (**2002**). ''Acoustic intensity, impedance and reflection coefficient in the human ear canal,'' J. Acoust. Soc. Am. **112**, 600–620.

Feeney, M. P., Grant, I. L., and Marryott, L. P. (**2003**). ''Wideband energy reflectance measurements in adults with middle-ear disorders,'' J. Speech Lang. Hear. Res. **46**, 901–911.

Feeney, M. P., and Keefe, D. H. (**1999**). ''Acoustic reflex detection using wide-band acoustic reflectance, admittance, and power measurements,'' J. Speech Lang. Hear. Res. **42**, 1029–1041.

Feeney, M. P., and Keefe, D. H. (**2001**). ''Estimating the acoustic reflex threshold from wideband measures of reflectance, admittance, and power,'' Ear Hear. **22**, 316–332.

Feeney, M. P., Keefe, D. H., and Marryott, L. P. (**2003**). ''Contralateral acoustic reflex thresholds for tonal activators using wideband energy reflectance and admittance,'' J. Speech Lang. Hear. Res. **46**, 128–136.

Feeney, M. P., Keefe, D. H., and Sanford, C. A., (**2004**). ''Wideband reflectance measures of the ipsilateral acoustic stapedius reflex threshold,'' Ear Hear. **25**, 421–430.
Gates, G. A., Cooper, Jr., J. C., Kannel, W. B., and Miller, N. J. (**1990**). ''Hearing in the elderly: the Framingham cohort, 1983–1985. Part I. Basic audiometric test results,'' Ear Hear. **11**, 247–256.
Hall, J. W. (**1982**). ''Acoustic reflex amplitude: I. Effect of age and sex,'' Audiology **21**, 294–304.
Harty, M. (**1953**). ''Elastic tissue in the middle ear cavity,'' J. Otolaryngol. Otol. **67**, 723–729.
Holte, L. (**1996**). ''Aging effects in multifrequency tympanometry,'' Ear Hear. **17**, 12–18.
Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**2000a**). ''A noninvasive method for estimating acoustic admittance at the tympanic membrane,'' J. Acoust. Soc. Am. **108**, 1128–1146.
Huang, G. T., Rosowski, J. J., Puria, S., and Peake, W. T. (**2000b**). ''Tests of some common assumptions of ear-canal acoustics in cats,'' J. Acoust. Soc. Am. **108**, 1147–1161.
Hunter, L. L., and Margolis, R. H. (**1997**). ''Effects of tympanic membrane abnormalities on auditory function,'' J. Am. Acad. Audiol **8**, 431–446.
Jerger, J., Jerger, S., and Maudlin, L. (**1972**). ''Studies in impedance audiometry. I. Normal and sensorineural ears,'' Arch. Otolaryngol. **96**, 513–523.
Jerger, J., and Oliver, T. (**1987**). ''Interaction of age and intersignal interval on acoustic reflex amplitude,'' Ear Hear. **8**, 322–325.
Johansen, H. (**1948**). ''Relation of audiograms to the impedance formula,'' Acta Oto-Laryngol. **74**, 65–75.
Keefe, D. H., Bulen, J. C., Arehart, K. H., and Burns, E. M. (**1993**). ''Ear-canal impedance and reflection coefficient in human infants and adults,'' J. Acoust. Soc. Am. **94**, 2617–2638.
Keefe, D. H., and Levi, E. (**1996**). ''Maturation of the middle and external ears: acoustic power-based responses and reflectance tympanometry,'' Ear Hear. **17**, 361–373.
Keefe, D. H., and Ling, R. (**1998**). ''Double-evoked otoacoustic emissions. II. Intermittent noise rejection, calibration and ear-canal measurements,'' J. Acoust. Soc. Am. **103**, 3499–3508.
Keefe, D. H., Ling, R., and Bulen, J. C. (**1992**). ''Method to measure acoustic impedance and reflection coefficient,'' J. Acoust. Soc. Am. **91**, 470–485.
Keefe, D. H., and Simmons, J. L. (**2003**). ''Energy transmittance predicts conductive hearing loss in older children and adults,'' J. Acoust. Soc. Am. **114**, 3217–3238.
Lonsbury-Martin, B. L., Cutler, W. M., and Martin, G. K. (**1991**). ''Evidence for the influence of aging on distortion-product otoacoustic emissions in humans,'' J. Acoust. Soc. Am. **89**, 1749–1759.
Lynch, T. J., Peake, W. T., and Rosowski, J. J. (**1994**). ''Measurements of the acoustic input impedance of cat ears: 10 Hz to 20 kHz,'' J. Acoust. Soc. Am. **96**, 2184–2209.
Margolis, R., and Hunter, L. L. (**1999**). ''Tympanometry: Basic principles and clinical applications,'' in *Contemporary perspectives in hearing assessment*, edited by F. E. Musiek and W. F. Rintelmann (Allyn and Bacon, Boston).
Margolis, R. H., Saly, G. L., and Keefe, D. H. (**1999**). ''Wideband reflectance tympanometry in normal adults,'' J. Acoust. Soc. Am. **106**, 265–280.
Marshall, L., Martinez, S. A., and Schlaman, M. E. (**1983**). ''Reassessment of high-frequency air-bone gaps in older adults,'' Arch. Otolaryngol. **109**, 601–606.
Nerbonne, M. A., Bliss, A. T., and Schow, R. L. (**1978**). ''Acoustic impedance values in the elderly,'' J. Am. Audiol Soc. **4**, 57–59.
Nixon, J. C., Glorig, A., and High, W. S. (**1962**). ''Changes in air and bone conduction thresholds as a function of age,'' J. Laryngol. Otol. **76**, 288–298.
Piskorski, P., Keefe, D. H., Simmons, J. L., and Gorga, M. P. (**1999**). ''Prediction of conductive hearing loss based on acoustic ear-canal response using a multivariate clinical decision theory,'' J. Acoust. Soc. Am. **105**, 1749–1764.
Randolph, L. J., and Schow, R. L. (**1983**). ''Threshold inaccuracies in an elderly clinical population: ear canal collapse as a possible cause,'' J. Speech Hear. Res. **26**, 54–58.
Rosowski, J. J., Brinsko, K. M., Tempel, B. I., and Kujawa, S. G. (**2003**). ''The aging of the middle ear in 129S6/SvEvTac and CBA/CaJ mice: measurements of umbo velocity, hearing function, and the incidence of pathology,'' J. Assoc. Res. Otolaryngol. **4**, 371–383.
Ruah, C. B., Schachern, P. A., Zelterman, D., Paparella, M. M., and Yoon, T. H. (**1991**). ''Age related morphologic changes in the human tympanic membrane,'' Arch. Otolaryngol. Head Neck Surg. **117**, 627–634.
Stenklev, N. C., Vik, O., and Laukli, E. (**2004**). ''The aging ear: an otomicroscopic and tympanometric study,'' Acta Oto-Laryngol. **124**, 69–76.
Stinson, M. R. (**1990**). ''Revision of estimates of acoustic energy reflectance at the human eardrum,'' J. Acoust. Soc. Am. **88**, 1773–1778.
Stinson, M. R., Shaw, E. A., and Lawton, B. W. (**1982**). ''Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal,'' J. Acoust. Soc. Am. **72**, 766–773.
Thompsen, D. J., Sills, J. A., Recke, K. R., and Bui, D. M. (**1980**). ''Acoustic admittance and the aging ear,'' J. Speech Hear. Res. **22**, 29–36.
Uchida, Y., Nomura, H., Itoh, A., Nakashima, T., Ando, F., Niino, N., and Shimokata, H. (**2000**). ''The effects of age on hearing and middle ear function,'' J. Epidemiol. **10**, S26–32.
Van Camp, K. H., Margolis, R. H., Wilson, R. H., Creten, W. L., and Shanks, J. E. (**1986**). ''Principles of Tympanometry,'' American Speech-Language-Hearing Association, Rockville, MD.
Voss, S. E., and Allen, J. B. (**1994**). ''Measurement of acoustic impedance and reflectance in the human ear canal,'' J. Acoust. Soc. Am. **95**, 372–384.
Whittemore, Jr., K. R., Merchant, S. N., Poon, B. B., and Rosowski, J. J. (**2004**). ''A normative study of tympanic membrane motion in humans using a laser Doppler vibrometer (LDV),'' Hear. Res. **187**, 85–104.
Wiley, T. L., Cruickshanks, K. J., Nondahl, D. M., and Tweed, T. S. (**1999**). ''Aging and middle ear resonance,'' J. Am. Acad. Audiol **10**, 173–179.
Wiley, T. L., Cruickshanks, K. J., Nondahl, D. M., Tweed, T. S., Klein, R., and Klein, B. E. (**1996**). ''Tympanometric measures in older adults,'' J. Am. Acad. Audiol **7**, 260–268.

Cochlear compression: Effects of low-frequency biasing on quadratic distortion product otoacoustic emission

Lin Bian[a)]
Department of Hearing and Speech, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, Kansas 66160

(Received 28 July 2004; revised 10 September 2004; accepted 14 September 2004)

Distortion product otoacoustic emissions (DPOAEs) are generated from the nonlinear transduction in cochlear outer hair cells. The transducer function demonstrating a compressive nonlinearity can be estimated from low-frequency modulation of DPOAEs. Experimental results from the gerbils showed that the magnitude of quadratic difference tone (QDT, f_2-f_1) was either enhanced or suppressed depending on the phase of the low-frequency bias tone. Within one period of the bias tone, QDT magnitudes exhibited two similar modulation patterns, each resembling the absolute value of the second derivative of the transducer function. In the time domain, the center notches of the modulation patterns occurred around the zero crossings of the bias pressure, whereas peaks corresponded to the increase or decrease in bias pressure. Evaluated with respect to the bias pressure, modulated QDT magnitude displayed a double-modulation pattern marked by a separation of the center notches. Loading/unloading of the cochlear transducer or rise/fall in bias pressure shifted the center notch to positive or negative sound pressures, indicating a mechanical hysteresis. These results suggest that QDT arises from the compression that coexists with the active hysteresis in cochlear transduction. Modulation of QDT magnitude reflects the dynamic regulation of cochlear transducer gain and compression.
[DOI: 10.1121/1.1819501]

PACS numbers: 43.64.Jb, 43.64.Kc, 43.64.Bt [BLM] Pages: 3559–3571

I. INTRODUCTION

An extraordinary feature of the auditory system is that it maintains high sensitivity over a large range of acoustic input in nature. This wide input range over 100 dB can mostly be attributed to the transducer characteristics of the sensory hair cells in the auditory periphery (for review, see Cooper, 2004), especially the signal preprocessing of the outer hair cells (OHCs). These cells amplify low sound pressures and compress high pressures to assure a relatively constant input to the inner hair cells that control the neural transmitter release. The amplification is thought to be due to the sound-induced motile responses of the OHCs that feedback to the inner ear structures to enhance the mechanical stimulation to the inner hair cells (Yates, 1990) as evidenced by otoacoustic emissions (OAEs). However, this mechanical feedback is controlled by the electrical voltage changes produced from mechanoelectrical transduction (MET), the conversion of sound pressures to cation flows through the ion channels in the OHC stereocilia (Hudspeth, 1989). MET saturates at high sound pressures and largely dominates the nonlinearity of the inner ear. The reciprocal relation between the mechanical and electrical events in the cochlear transducer, i.e., the OHCs and their coupled basilar membrane (BM) and tectorial membrane, follows a compressive nonlinearity [Fig. 1(B)], often described by a second-order Boltzmann function (f_B, Hudspeth, 1989; Evans *et al.*, 1991)

$$y(x)=\frac{A}{1+e^{b(x+c)}\cdot[1+e^{d(x+e)}]}, \tag{1}$$

where x, y are acoustic input and mechanical output in Pa; A, b, c, d, and e are constants. Quantitative measurements of the transduction characteristics or transducer function (f_{Tr}) with OAEs provide noninvasive tools for evaluating cochlear functional and structural integrity.

One consequence of the compressive nonlinearity in cochlear transduction is the generation of intermodulation distortion products (DPs) while simultaneously presenting two pure tones ($f_1, f_2, f_1<f_2$). The DPs contain a series of frequency components of both odd- and even-order combinations or differences of the two primary frequencies. Cubic difference tone (CDT, $2f_1-f_2$) and quadratic difference tone (QDT, f_2-f_1) are the most prominent representatives of the odd- and even-order DPs, respectively. As mechanical responses from the OHCs near the f_2 place (Martin *et al.*, 1987), these DPs are directly measurable in the ear canal as distortion product otoacoustic emissions (DPOAEs). Different DPs are generated from different aspects of the sigmoid-shaped f_{Tr}, i.e., the central portion with optimal gain gives rise to CDT and the asymmetric compressions at larger inputs produce QDT (van Emst *et al.*, 1997). As a part of the response spectrum, QDT should be examined together with CDT to provide an integrated view of the cochlear processes (Cheng, 1999). Recently, a noninvasive method was introduced to quantitatively estimate the cochlear f_{Tr} from low-frequency modulation of CDT (Bian *et al.*, 2002, 2004). In this paper, the effects of low-frequency biasing on QDT are reported.

a)Electronic mail: lbian@kumc.edu

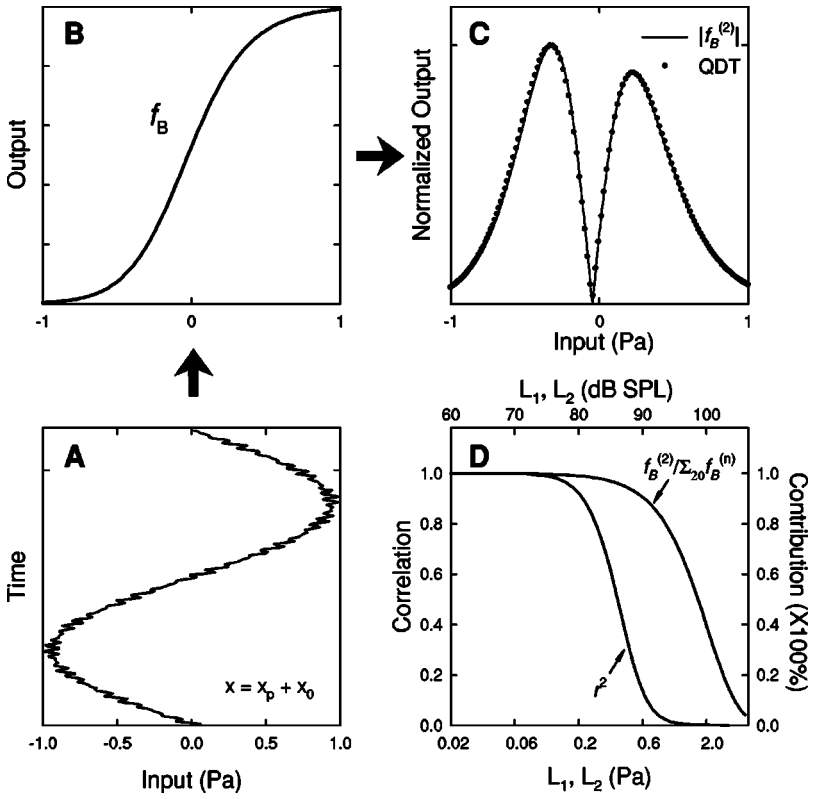


FIG. 1. Simulation: effect of low-frequency biasing on QDT magnitude. A two-tone signal (x_p) rides on a low-frequency bias tone (x_0) of large amplitude [panel (A)], then passes through a second-order Boltzmann function, f_B [panel (B)]. The normalized QDT magnitudes [dots in panel (C)] extracted from the output as a function of the instantaneous bias-tone amplitude show a shape similar to the absolute value of the second derivative of the f_B (solid line), denoted as $|f_B^{(2)}|$. Panel (D) (left): the correlation coefficient (r^2) between the QDT magnitude and the $|f_B^{(2)}|$ improves asymptotically as the primary level ($L_1=L_2$) reduces. Panel (D) (right): the contribution of the second derivative term to the QDT magnitude relative to 20 even-order terms in Eqs. (4) and (5) also increases exponentially with reduction in primary level.

II. THEORY AND EXPERIMENTS

A. Theoretical background

The principle of using low-frequency biasing to estimate cochlear f_{Tr} is that the cochlear response to sufficiently small input is modulated as the cochlear transducer structure is displaced, and the modulation pattern (MP), depending on the nature of the response, is proportional to the appropriate derivative of the f_{Tr}. In the case of QDT, the following derivations show that the MP is associated with the second derivative of the f_{Tr}. With Taylor's series expansion, the value of the nonlinear f_{Tr} in Eq. (1) at any arbitrary input x_0 can be derived as $f_{Tr}(x_0)=\Sigma_{n=0}^{\infty} a_n(x_0)\cdot(x-x_0)^n$, where $a_n(x_0)=f^{(n)}(x_0)/n!$, and $f^{(n)}(x_0)$ is the nth derivative of f_{Tr} at x_0. The point x_0, referred to as the operating point (OP), is varied by the bias tone, so that a probe signal (x_p) can be placed at different positions on the f_{Tr} [Figs. 1(A) and (B)] to estimate a complete MP. Thus, with the input $x=x_p+x_0$, the output of the f_{Tr} at an OP becomes a polynomial power series

$$f_{Tr}(x_0)=\sum_{n=0}^{\infty} a_n(x_0)\cdot x_p^n. \tag{2}$$

To obtain DPs, x_p should be a two-tone signal, $L_1\cos\theta_1+L_2\cos\theta_2$ (where $\theta_1=2\pi\cdot f_1$, $\theta_2=2\pi\cdot f_2$). Using binomial theorem and a formula for $\cos^n\theta$ (for details cf. Appendix A of Bian *et al.*, 2002), each term in the power series of Eq. (2) can be determined

$$x_p^n(\theta_1,\theta_2)=\frac{1}{2^{n-2}}\sum_{k=1}^{n}\left[L_1^h L_2^k\binom{n}{k}\cdot(\alpha_h\cos m_h\theta_1\cdot\alpha_k\cos m_k\theta_2)\right], \tag{3}$$

where $h=n-k$, $\binom{n}{k}$ is a binomial coefficient, α_h and α_k are series of binomial coefficients determined by $\binom{h}{i}$ and $\binom{k}{j}$, integer multiples $m_h=h-2i$ and $m_k=k-2j$; here, $0\leqslant i<h/2$ and $0\leqslant j<k/2$. Consequences of the products of the sinusoidal elements in each term are the creation of a series of intermodulation DPs related to that nth power term, i.e., $\cos(m_h\theta_1\pm m_k\theta_2)$, since $\cos\alpha\cdot\cos\beta=[\cos(\alpha+\beta)+\cos(\alpha-\beta)]/2$. Summation of all the terms in the power series in Eq. (2) generates an entire frequency spectrum of the DPs. Only even-order terms of the power series in Eq. (2) can produce QDT ($\theta_2-\theta_1$), i.e., n should be even. Moreover, $i=(h-1)/2$, $j=(k-1)/2$ (where h and k are odd), so that $m_h=h-2i=1$, and $m_k=k-2j=1$. Therefore, the magnitude of the QDT (M_{QDT}) given by the absolute value of the frequency component can be calculated with

$$M_{\mathrm{QDT}}(x_0)=\left|\sum_{n=2,4,\ldots}^{\infty}\left[\frac{a_n(x_0)}{2^{n-1}}\sum_{k=1}^{n}L_1^h L_2^k\binom{n}{k}\binom{h}{i}\binom{k}{j}\right]\right|, \tag{4}$$

where

$$\binom{n}{k}\binom{h}{i}\binom{k}{j}=n!\cdot\left[\left(\frac{h-1}{2}\right)!\left(\frac{h+1}{2}\right)!\left(\frac{k-1}{2}\right)!\left(\frac{k+1}{2}\right)!\right]^{-1}.$$

A more explicit form of the M_{QDT} can be derived by expanding Eq. (4) and replacing a_n with its original form

$$M_{\mathrm{QDT}}(x_0)=L_1L_2\left|\frac{f^{(2)}(x_0)}{2!}+\frac{3}{2}\frac{f^{(4)}(x_0)}{4!}(L_1^2+L_2^2)+\frac{15}{8}\frac{f^{(6)}(x_0)}{6!}(L_1^4+3L_1^2L_2^2+L_2^4)+\cdots\right|. \quad (5)$$

Therefore, QDT magnitude at an OP depends on the primary levels and the weighted second and other even-order derivatives of the f_{Tr}. When L_1 and $L_2<1$, the nth element of the series approaches zero, and the first term is the largest, since $\lim_{n\to\infty} 1/(2^{n-1}n!)=0$. If L_1 and $L_2\ll 1$, the QDT magnitude can be approximated with only the first term

$$M_{\mathrm{QDT}}(x_0)\approx\tfrac{1}{2}L_1L_2\cdot|f^{(2)}(x_0)|. \quad (6)$$

When the primary levels are held constant, M_{QDT} is essentially proportional to the absolute value of the second derivative of the f_{Tr} evaluated at different OPs, or $M_{\mathrm{QDT}}(x_0)\propto|f_{Tr}^{(2)}(x_0)|$. Therefore, it is possible to obtain the second derivative of the cochlear f_{Tr} by measuring QDT magnitudes at various OPs across the operating range of the cochlear transducer.

Numerical simulations verified the relation between QDT magnitude and the $|f_{Tr}^{(2)}(x_0)|$. As shown in Figs. 1(A)–(C), a two-tone signal x_p and a low-frequency bias tone x_0 were sent to a sigmoid-shaped f_B; QDT magnitude was computed with short-time fast Fourier transforms (FFTs) on small segments of the output waveform. Each segment had a different dc shift or an instantaneous bias pressure (BP). The QDT magnitude from each segment plotted as a function of the corresponding BPs demonstrated an MP similar to the $|f_{Tr}^{(2)}(x_0)|$ which is a dual-peak structure with a center notch [panel (C)]. Reducing the primary levels ($L_1=L_2$) below 0.2 Pa or 80 dB SPL improved the correlation coefficient (r^2) asymptotically to 1 [panel (D)]. Another simulation compared the contribution of the first term in Eqs. (4) and (5) that contains $f_{Tr}^{(2)}(x_0)$ relative to a total contribution of 20 even-order terms (including the first term). As primary level reduced ($L_1=L_2$) to less than unity (presumed as Pa), the second-derivative term accounted for almost all the contributions from the 20 terms combined [Fig. 1(D)]. Both simulations indicate that at low signal levels the QDT magnitude at an OP is proportional to the $|f_{Tr}^{(2)}(x_0)|$, i.e., the QDT is determined by the nonlinear characteristics of the f_{Tr} and the location of the OP.

B. Previous results

Effects of the low-frequency modulation on the QDT magnitude were briefly mentioned in Bian *et al.* (2002). In that study, DPOAEs were obtained from the peaks and troughs of the bias tone with various peak amplitudes. The measured QDT magnitudes evaluated as a function of the bias tone level demonstrated MPs similar to the shape of the $|f_{Tr}^{(2)}|$ consisting of a center notch and two peaks (Fig. 2). The QDT magnitude increased when the cochlear transducer is biased in either positive or negative sound-pressure directions. The shape of the MP also varies with the level of the two-tone signal, i.e., the peaks are wider and taller at higher primary levels. The center notch pinpoints the inflection point (IP) of the sigmoid-shaped f_{Tr} which is situated around the BP of -2 to -3 Pa. With little or no biasing, the QDT magnitudes present on the right side of the center notch (small notches at zero BP), implying asymmetrically situated OPs. These findings are quite typical at higher primary levels despite relatively higher intersubject variability than the CDT data.

More recently, Bian *et al.* (2004) investigated the temporal aspect of low-frequency biasing on CDT magnitude. Experimental results showed phase-dependent modulation of CDT magnitude, i.e., a typical MP similar to the $|f_{Tr}^{(3)}|$ presented during half cycle of the bias tone. Since basal BM response to low-frequency tone is linear and in phase with the stimulus (Robles and Ruggero, 2001), the BP change can be directly related to the BM displacement. Thus, the f_{Tr} derived from the low-frequency modulated DPOAEs reflects a force (sound pressure)-displacement, or a ''stress–strain'' relation of the cochlear transducer. The phase relation between the MPs and bias-tone-induced BM displacement indicated hysteresis in the mechanical responses of the cochlear transducer. Subsequently the derived force-displacement function demonstrated a hysteresis loop with counterclockwise traversal that was attributed to the timing of the feedback force related to the ''negative damping'' in OHC dynamics. Since QDT is generated from the compression portions with even symmetry on the f_{Tr}, it reflects a different perspective of the same nonlinear process that generates CDT. The QDT is more prominent at higher signal levels and could be robust under certain pathologic conditions where CDT is diminished. Moreover, QDT may be more useful in estimating the IP of the f_{Tr} (Frank and Kössl, 1996, 1997; Sirjani *et al.*, 2004) and thus may have clinical diagnostic value. Therefore, it is necessary to study the behavior of QDT under low-frequency biasing.

III. METHODS

Since this study was based on 11 Mongolian gerbils, ten of which provided the CDT data reported in Bian *et al.* (2004), the experimental procedures were the same as described previously. Briefly, after being anesthetized with Ketamine and Xylazine (100 and 2 mg/kg, i.p.), the gerbil's right pinna was removed and a speculum was sealed onto the bony ear canal. At the tip of the speculum, a subwoofer (Boston Acoustics) was connected via a silicon tube for delivery of the bias tone. A calibrated probe microphone (Etymotic Research, ER-10B) was inserted into the speculum and two earphones (ER-3A) were coupled to the ports on the ER-10B to present the two-tone signal. The two primaries and the bias tone were generated separately and the signal levels were program controlled with three attenuators (Tucker-Davis Tech., TDT PA4). Two trials of recordings were made with an interval of about 90 min. Each trial contained variations of the primary tone conditions: six levels and two pairs of frequencies.

Two-tone signals with frequencies of 3968/5120 Hz, ($f_2/f_1=1.29$) and 5888/7552 Hz, ($f_2/f_1=1.28$) and levels from 50 to 75 dB SPL ($L_1=L_2$, 5-dB step) were used in the experiment. A bias tone of 25 Hz was presented at 20 Pa peak amplitude (120 dB pSPL) via an amplifier (Optimus

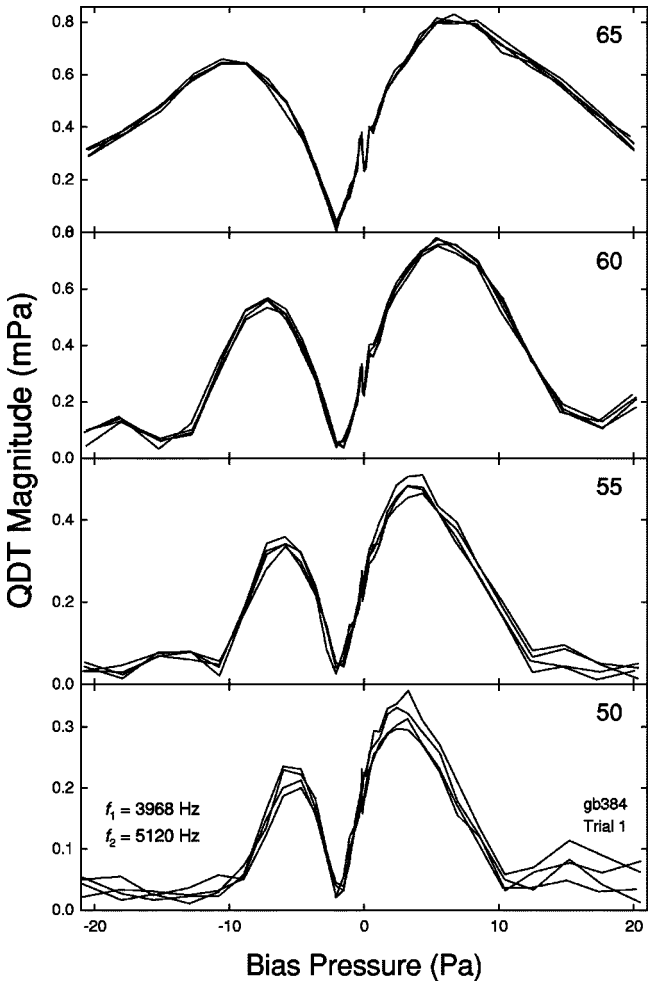


FIG. 2. Eamples of the QDT magnitude as a function of the bias tone level. Data were obtained with bias tones with various peak amplitudes. Indicated in each panel is the primary level in dB SPL ($L_1 = L_2$).

SA-155) and calibrated in the speculum with a sound-level meter (Quest 1800). The bias tone was 250 ms in duration, with a flat tail of about 40 ms. Ear-canal acoustic signal (Fig. 3) recorded from the ER-10B was amplified 29 dB (Stewart VBF 10 M), averaged 50 to 100 times, and sampled at 65 536 Hz (TDT AD2). The QDT magnitudes and corresponding BPs were derived with a moving-window method (Bian *et al.*, 2004) in MATLAB (MathWorks). Starting at time zero, a section of 512 data points (7.8 ms) was Kaiser windowed, and a QDT magnitude was extracted with FFT (Fig. 3). The data were also low-pass filtered at 100 Hz and the median was taken as the matching BP value. The next QDT and BP pair was obtained by shifting the window by 16 points (0.24 ms) until the end of the waveform was reached. Only data within the middle three cycles of the bias tone with stable peak-amplitudes were used to avoid any influence of the transient response from the recording system, the middle-, and the inner ear.

Plotting the QDT magnitudes against the corresponding BPs revealed the MPs. To characterize the MP, a series of physiologic indices was measured from the QDT and BP data using an automatic program with user interactive verifications. The indices were obtained separately for loading (increasing in BP) and unloading (decreasing in BP) processes. These indices from the three cycles of the bias tone were averaged and analyzed with respect to two trials, two pairs of primary frequencies, and six primary levels. To evaluate the effects of these signal conditions, the indices were examined separately for the two pairs of primary frequencies with a two-way (trial×primary level) or a three-way (trial×loading/unloading×level) analysis of variance (ANOVA) using STATISTICA (StatSoft).

IV. RESULTS

A. Modulation of QDT magnitude

To ensure accuracy and reliability, the time-domain QDT data obtained from the moving-window FFT were also verified with those from an alternative method: bandpass filtering. The ear-canal acoustic signal was bandpassed with a zero-phase Butterworth filter centered at the QDT frequency with a 300-Hz bandwidth. The bandpassed signals for different primary frequencies are sinusoids of the QDT with modulated envelopes [Figs. 4(A)–(B)]. Compared with QDT magnitude obtained from the moving-window method, the envelope of the bandpassed QDT waveform matched precisely with the QDT magnitude from FFT. Such comparisons were made in all the animals for various experimental conditions, and results from the two methods were consistent. This indicates that the moving-window FFT did not introduce distortions in the QDT magnitude, nor phase- or time delays in the data.

Closely examining the temporal characteristics of the QDT waveforms in panels (A) and (B) of Fig. 4 with respect to the bias tone in panel (C), two observations can be made: (1) QDT magnitude is modulated by the bias tone in a phase-dependent manner; (2) in the tail portion of the signal where

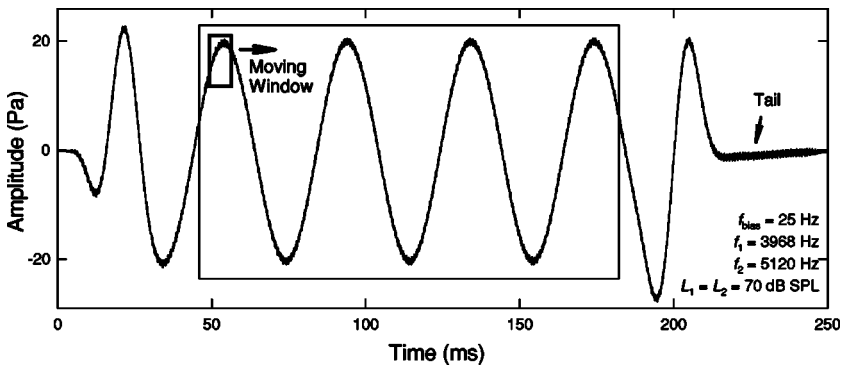


FIG. 3. Ear canal acoustical signal and the moving-window method. The peak amplitude of the bias tone is 120 dB pSPL (20 Pa). The tail portion (about 40 ms) of the signal is flat where bias level is zero. The QDT data are obtained from middle three complete cycles (large rectangle) of bias tone where the peak amplitude of the bias tone is stable. A window of 512 points (small rectangle) shifting along the time waveform in 16-point steps is used to extract the QDT with FFT.

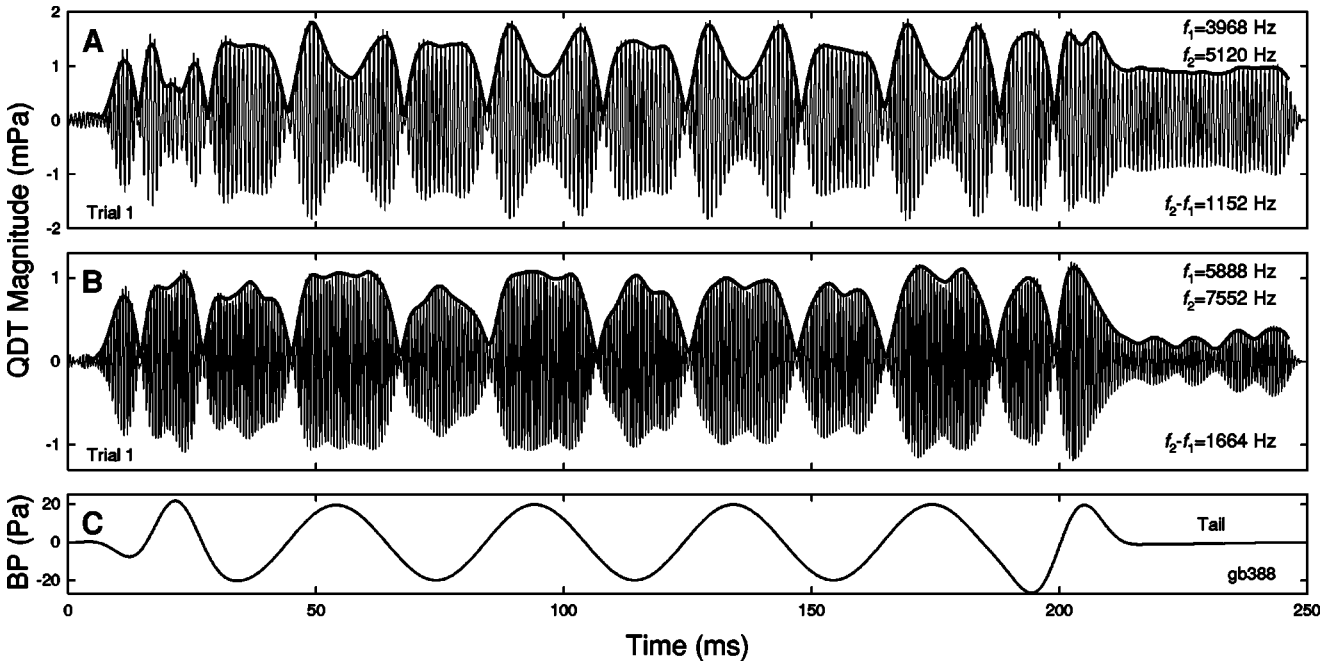


FIG. 4. Low-frequency modulation of QDT in the time domain. Panels (A) and (B): comparison between two methods, the moving-window FFT (thick line) and a narrow-band filtering at the QDT frequency (thin line). The primary levels in both panels are 70 dB SPL ($L_1=L_2$). Panel (C) The corresponding low-frequency bias tone serves as a reference. Note: notches in QDT waveforms occur roughly at the zero crossings of the bias tone, and enhancements of the QDT (compared to the tails) relate to the peaks and troughs of the bias tone. The enhancement is greater at positive peaks of the bias tone.

the BP is zero, the QDT magnitude is moderate, stable, and generally greater for lower primary frequencies. The effects of the bias tone were either enhancing or suppressing of the QDT magnitude or envelope. The enhancement occurred at the peaks and troughs of the bias tone, while the suppression occurred near zero crossings of the BP. In the time domain, the modulation of QDT magnitude formed a series of lobes of different sizes presented alternately. In addition, the lobes with greater magnitudes and widths corresponded to the peaks of the bias tone, whereas the lobes of smaller sizes related to the troughs. Between the lobes deep notches occurred periodically. Despite considerable variations, these observations were typical among all the animals across different signal conditions, especially at primary levels above 50 dB SPL.

Focused on one period of the bias tone where BP rises and falls monotonically, the QDT magnitude showed a double MP containing two shapes of the $|f_B^{(2)}(x)|$ [Fig. 5(A)], i.e., the typical ''V-shaped'' MP appeared for every half cycle of the bias tone with the notches located around the zero crossings of the bias tone. The MP of the QDT in relation to the bias tone can be better observed by plotting the QDT magnitude as a function of the BP [panel (B)]. The double MP of the QDT is marked by the separation of the center notches of the two typical MPs. The center notch or the minimal QDT magnitudes occur at a more positive sound pressure during BP rising or loading of the cochlear transducer, whereas the MP centered in the negative pressures are related to BP falling or unloading. Since the two typical MPs of the QDT connected at the maximal BPs [Fig. 5(B)], the corresponding f_{Tr} related to the double MP should contain two sigmoid-shaped curves forming a hysteresis loop.

B. Characterization of QDT double MP and hysteresis

To directly find inference of the f_{Tr} from the experimental data, and for the ease of comparisons among experimental conditions, the double MP of QDT magnitude was characterized with a set of physiological indices. As indicated in Fig. 6(A), the typical MP directly relates to the shape of the cochlear f_{Tr}. Three points on the typical MP were identified to determine the shape of the MP: the center notch and two peaks on each side. The peaks of QDT magnitude correspond to the points where the cochlear transducer is driven into compression in either negative or positive sound pressures. The absolute amplitudes of the peaks in dB SPL at negative and positive pressures were measured as P_1 and P_2, respectively. They represent the enhanced QDT magnitudes due to the low-frequency biasing. Values of P_1 and P_2 indicate the maximal reduction rate in cochlear gain or maximal compression, since $f_{Tr}^{(2)}$ is the slope of the $f_{Tr}^{(1)}$ which reflects the cochlear transducer gain. Locations of P_1 and P_2 also correspond to the two notches of the ''W-shaped'' MP of CDT (Bian *et al.*, 2002). Thus, the distance between the two peaks on the MP of QDT is equivalent to the modulation width (MW) measured from the CDT (Bian *et al.*, 2004). The MW is a measure of the central portion of the transducer operating range where the transduction is close to optimal.

The location of the center notch, evaluated with the BP at the notch, corresponds to the IP of the sigmoid-shaped f_{Tr}, where the $|f_{Tr}^{(2)}|$ equals zero and the $|f_{Tr}^{(1)}|$ is maximal. Index BP at the IP, or BP_{IP}, indicates the optimal OP of the cochlear transducer with maximal gain. Nonzero BP_{IP} represents the sound pressure required for the transducer to operate optimally. The QDT magnitude in dB SPL evaluated at the notch may reflect the general compression from the feed-

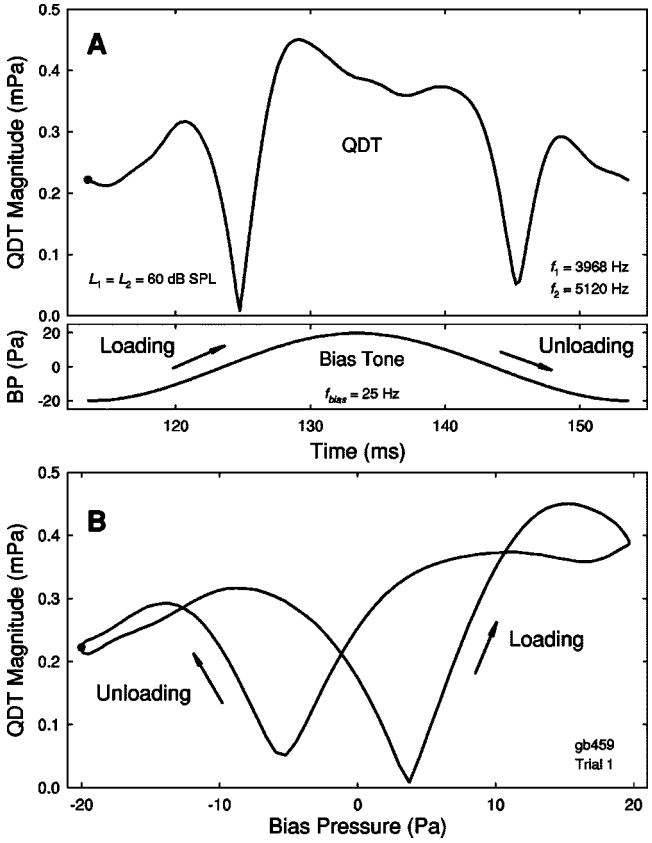


FIG. 5. Modulation pattern (MP) of the QDT magnitude. (A) Temporal MP. For a complete cycle of the bias tone (lower trace), there are two ''V-shaped'' MPs (upper trace) with notches corresponding to the zero crossings of the bias tone. Larger peaks of the QDT magnitude present when the BP is positive, while smaller peaks occur at the negative BPs. (B) Double MP of the QDT and hysteresis. The QDT magnitudes within one bias tone period plotted as a function of the bias tone pressure show a ''butterfly-like'' double MP with separated center notches. During loading (increasing in the BP), the center notch shifts to the right (positive BP), whereas the center notch shifts to the left (negative BP) during unloading. Dots represent the starting points of the traces.

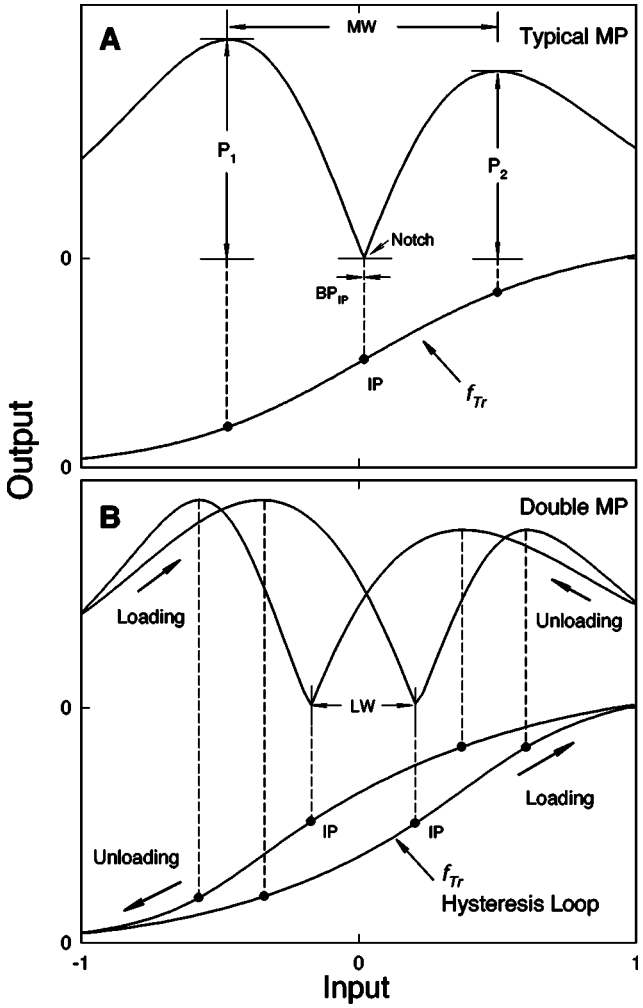


FIG. 6. Characterization of the MP of QDT magnitude. (A) Typical MP and f_{Tr}. The typical MP contains a center notch and two peaks on each side. Notch, P_1, and P_2 are the amplitudes of the center notch and the two peaks, respectively. MW: the distance between the peaks. BP_{IP}: location of the inflection point (IP) measured with the BP at the center notch. Corresponding locations of the indices on the f_{Tr} are shown on the lower curve. (B) Double MP and f_{Tr} hysteresis. The separation of the two center notches on the double MP is used to determine the loop width (LW) of the hysteresis. Corresponding locations of the center notches and peaks of the MP on the f_{Tr} hysteresis loop are indicated.

back mechanism governing the transduction at a given signal level. The heights of the two peaks, P_1 and P_2, relative to the notch were, measured as PN_1 and PN_2 in dB, respectively. These relative measures indicate the compression due to cochlear f_{Tr} in addition to the general compression. The asymmetry of the f_{Tr} was accessed by the ratio of the magnitude of the two lobes P_1 and P_2 (in Pa), with values greater or less than 1 indicating asymmetric transducer.

When evaluated with respect to BP, the MP of QDT showed a separation of the center notches [Fig. 5(B)], i.e., the location of the center notch depended on the direction of the change in BP. The center notch followed the same direction of the BP change. The corresponding f_{Tr} is a hysteresis loop with separated IPs [Fig. 6(B)], where response for loading was smaller and unloading response greater. During loading of cochlear transducer, the IP of f_{Tr} is located in the positive BPs, whereas the IP shifts in the negative sound-pressure direction during unloading, thus giving rise to a *counterclockwise* hysteresis loop. Since the center notch of the MP of QDT corresponds to the IP of the sigmoid-shaped f_{Tr}, the distance between the two notches on the double MP measures the loop width (LW) of the cochlear hysteresis.

C. Effects of primary level on double MP

Even with considerable variability at lower primary levels, the typical MP of QDT generally presented in all animals for two trials of both primary frequencies. Clear presence of the MP was often observed from 60 to 75 dB SPL primary levels. As the QDT data from an animal shown in Fig. 7, the double MP becomes better defined as the QDT magnitude rises with primary level and progressive changes in the MP were noticeable across primary levels.

1. Separation of center notches and hysteresis

Separation of the center notches on the double MP was observed in all animals at various primary levels. Different from other indices, a three-way (frequency×trial×level) ANOVA was used to analyze the LW data for the lack of loading effect. Due to no significant difference between the repeated measures over 90 min, LWs from the two trials

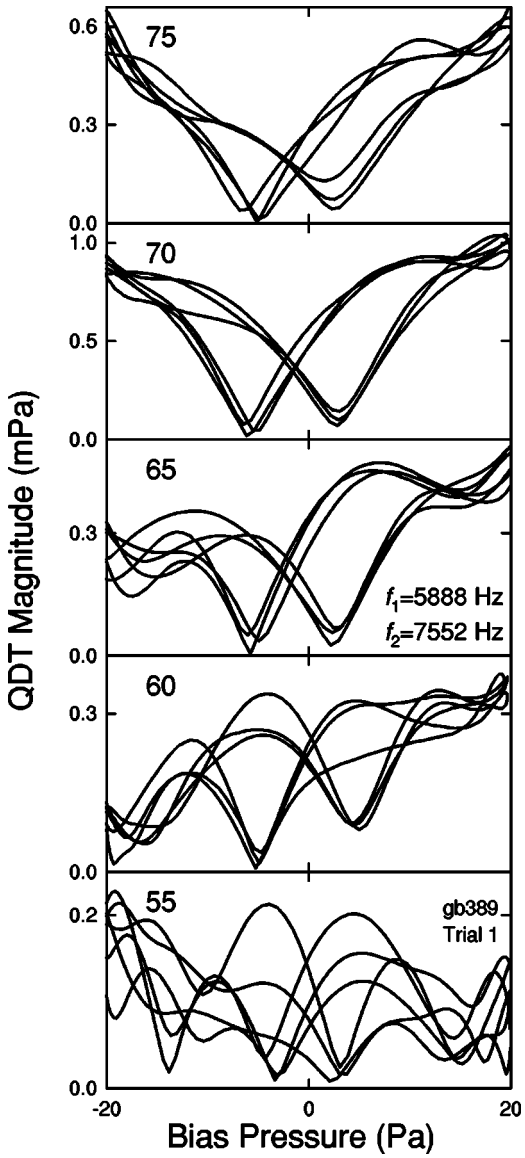


FIG. 7. Examples of the QDT double MP. The primary level in dB SPL ($L_1=L_2$) is indicated in each panel. The QDT magnitudes are extracted from three complete cycle of the bias tone. Note the progressive changes in the MPs with the primary levels.

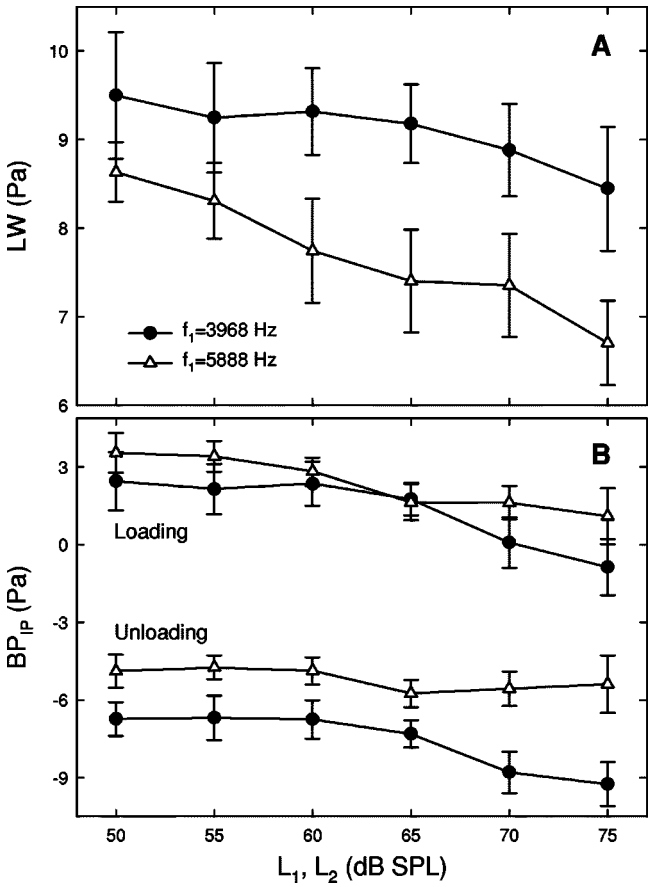


FIG. 8. Separation of the center notches on the QDT double MP. Mean ±standard error (SE) of the LW, the distance between the center notches, and BP_{IP}, the location of the IP, are displayed in panels (A) and (B), respectively. The LW reduces with primary level while the IPs of the transducer hysteresis shift to the negative sound pressures. Data reflect the average across two trials ($n=22$).

were combined to compute the mean and standard error (SE). The averaged LWs were about 7 to 9 Pa for six primary levels and two pairs of primary frequencies [Fig. 8(A)]. The LW of higher primary frequencies ($f_1=5888$ Hz) was significantly smaller than that of the lower frequencies ($f_1=3968$ Hz). The separation of the center notches was also evidenced with the locations of the BP_{IP} [Fig. 8(B)]. For increasing the BP or loading of the cochlear transducer, the BP_{IP} averaged across trials (ANOVA, $p>0.05$) was about 3–0 Pa, whereas the BP_{IP} for unloading is located at about −6–−9 Pa. At both primary frequencies, the BP_{IP} decreased with increase in primary levels. The separation of BP_{IP} between loading and unloading was consistent with the measure of LW.

2. Changes in shape of the MP

It can be observed from Fig. 7 that the double MP of QDT magnitude changes with the primary level. Since the ANOVA showed significant effects of the primary level on the physiological indices of the MP but no differences between the two trials, the data were pooled to evaluate the level and loading effects. Indices notch, P_1, and P_2 as functions of primary levels showed variations of a general growth pattern (Fig. 9). The QDT magnitudes at the notch (left panel) were very small and grew slowly up to about 60–65 dB SPL. At higher primary levels (>65 dB SPL), however, the notch magnitude increased at a greater 1-dB/dB, rate. The transition of the two growth rates was indicated with a ''knee point.'' As shown in Fig. 9, the magnitudes of P_1 and P_2 were greater and generally increased at a linear rate (1 dB/dB) within the input range of 55–70 dB SPL with the knee points located at about 55–60 dB SPL. Compared to the notch, the primary levels at the knee points were lower, and the change of growth rates across the knee points were smaller. In addition, P_1 and P_2 growth functions saturated above 70 dB SPL. Moreover, the QDT magnitudes were smaller for the higher primary frequencies ($f_1=5888$ Hz), especially within the linear growth range. For this frequency, the primary levels at the knee points were higher, or the ranges of low growth rate were wider. Except for P_2, the QDT magnitudes are greater during loading process than unloading.

The relative magnitude of the two peaks on the MP, PN_1

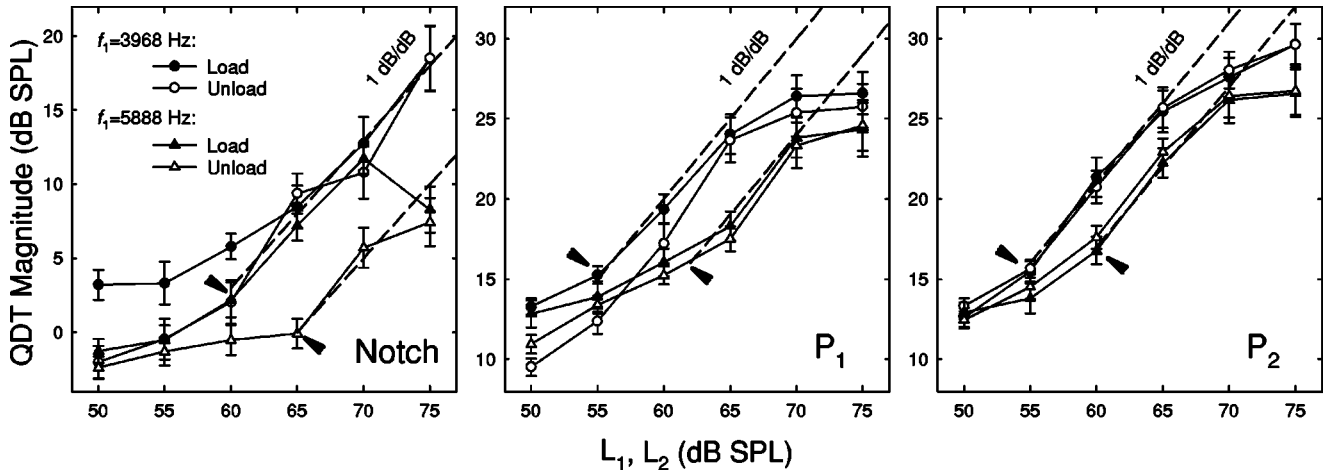


FIG. 9. Effects of primary level on physiologic indices: growth patterns of notch, P_1, and P_2. Two stages of growth at low to moderate levels are indicated by the "knee points" (short arrows). Dashed lines represent the linear growth at moderate levels. The knee points for higher frequency shift to higher primary levels. The range of lower growth rate is wider for the notch than P_1, and P_2. Growth functions of P_1, and P_2 saturate at high primary levels. Data represent mean±SE of the indices averaged across the two trials ($n=22$).

and PN_2, grew gradually up to 65–70-dB SPL primary levels, and followed by a decline at 75 dB SPL [Figs. 10(A) and (B)], indicating that effects of the bias tone reached the maximal. The data were more variable than the absolute measure of the peak amplitudes, P_1 and P_2. It can be observed from Figs. 10(A) and (B), also from Fig. 9, that the magnitude of P_2 is greater than P_1. This unevenness in the sizes of the two peaks on the MP was described by the asymmetry index, ratio of P_1/P_2. As shown in Fig. 10(C), the asymmetry generally is less than 1, except at the two lowest primary levels during the loading process. A common trend can be found that the P_1/P_2 value reduces with primary level. In other words, the size of QDT peak in the positive sound pressures (P_2) is greater than the peak in the negative side (P_1) at higher levels. Index MW, the distance between the two peaks on the MP, expanded from about 15–30 Pa as primary level rose from 50–70 dB SPL [Fig. 10(D)], suggesting that the two peaks are further apart.

V. DISCUSSION

A. Temporal MP of QDT magnitude

Phase-dependent modulation of QDT magnitude is the major effect of the low-frequency biasing. In the time domain, the magnitude of QDT demonstrates enhancements when the BP is increased or decreased, and suppressions at around zero crossings of the BP [Fig. 4 and Fig. 5(A)]. These findings are in agreement with previous measurements in

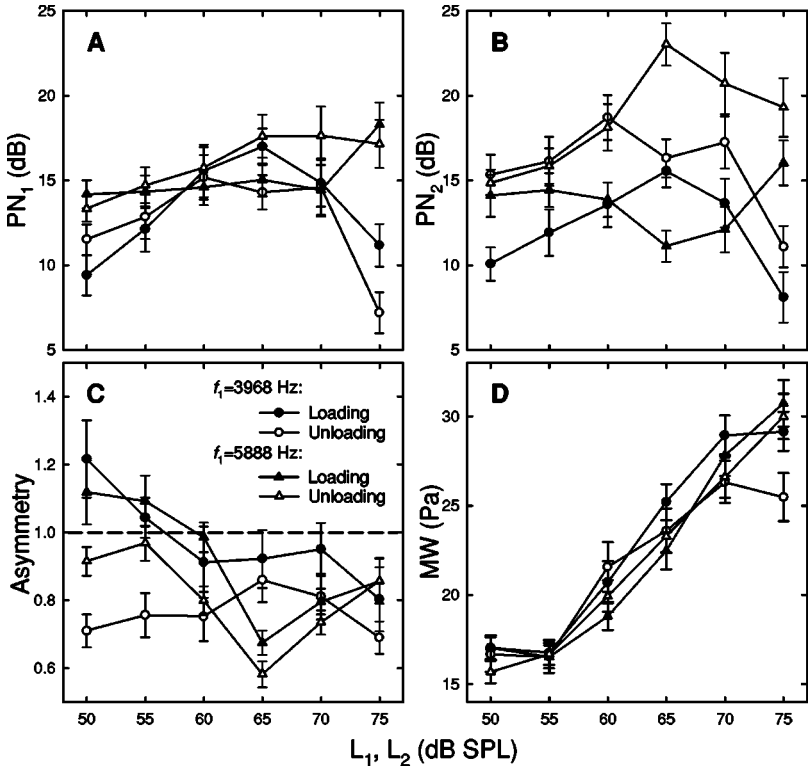


FIG. 10. The effect of primary level on physiologic indices. (A), (B) Growth functions of the magnitudes of P_1, and P_2 relative to the notch, PN_1, and PN_2. (C) Asymmetry index (P_1/P_2) reduces with level. Dashed line indicates perfect symmetry in the MP. (D) MW expands as primary level rises ($L_1=L_2$). Data represent mean±SE of the indices averaged across the two trials ($n=22$).

gerbils on acoustical low-frequency biasing of DPOAEs by Frank and Kössl (1996, 1997), who found a modulation of QDT magnitude up to 16 dB with a 5-Hz bias tone above 105 dB SPL. Furthermore, they noticed that the QDT magnitude was enhanced when the BM was displaced towards either scala tympani (ST) or scala vestibuli (SV) and minimized at zero crossings of the bias tone. However, the present study provides a more detailed characterization of the low-frequency modulation of QDT (Figs. 4, 5, and 7). For example, positive BPs give rise to larger QDT magnitudes, indicating that biasing of the BM towards ST produces greater enhancement of the QDT.

In humans, the low-frequency modulation of CDT magnitude has been measured (Scholz *et al.*, 1999), but the effect on QDT has never been reported. However, phase-dependent modulation of tone-burst-evoked OAEs was observed using a 45-Hz tone (Zwicker and Manley, 1981). When tone bursts are placed at different phases within one period of the bias tone, the variation of the OAE magnitude demonstrates an MP that is a mirror image of the masking period pattern (MPP), a temporal modulation of threshold by a low-frequency masker (Zwicker and Sherer, 1987). The typical MPP with sinusoidal masker shows two ''V-shaped'' patterns with notches at zero crossings of the masker (Zwicker, 1977). Psychoacoustical MPP and MP of OAEs are directly comparable to the MP of QDT [Fig. 5(A)], since the notches and the peaks all occur near zero BP, where the cochlear gain is highest. As proposed by Patuzzi *et al.* (1984), the modulations of auditory sensitivity originate from the nonlinearity in the MET of OHCs, where compression of transducer gain elevates hearing threshold. It is necessary to point out that the MPP of threshold and MP of the tone-burst-evoked OAEs are related to the $f_{Tr}^{(1)}$, while the MP of QDT are related to the $f_{Tr}^{(2)}$. Therefore, the same double MP of DPOAEs observed in gerbils could be found in humans. More importantly, the MPs of OAEs may alter under pathologic conditions as indicated by the MMP derived from amplitude-modulated maskers (Wojtczak *et al.*, 2001).

B. QDT double MP: Comparison with CDT

The QDT and CDT largely originate from different portions on the sigmoid-shaped f_{Tr} (van Emst *et al.*, 1997), representing compression and gain of the transducer, respectively. Consequently, QDT and CDT demonstrate somewhat opposite behaviors during low-frequency biasing. Therefore, the characteristics of QDT double MP and their variations with signal level are discussed in comparison with the results of CDT (Bian *et al.*, 2002, 2004).

1. Separation of center notches

An important finding of the low-frequency biasing of QDT is the double MP marked by the separation of the center notches (Fig. 8). This effect indicates that the cochlear f_{Tr} has hysteresis, and the loop is wider at low levels. The hysteresis is possibly a consequence of the active force generation from the OHCs which is more prominent at low levels (Yates, 1990). Direct measurements of the cochlear fluid pressure during acoustic stimulation (Olson, 2004) show that harmonic distortions are generated from a force production which coincides with maximal velocity of the BM motion. Such an active force, termed ''negative damping'' due to the proportionality to velocity, is reduced at the extremes of the BM displacement, thus producing greatest effect at zero crossings of the BM displacement. Therefore, the deviation of the output from that of a single f_{Tr} is most prominent at zero displacement and forms a hysteresis loop. As analyzed previously (Bian *et al.*, 2004), the counterclockwise traversal of the hysteresis indicates energy gain during the transduction.

Slight phase- or time delay in the force production could generate a deviation of the maximal forcing away from zero displacement which could also contribute to the separation of the center notches and hysteresis. Consequently, the LW can be evaluated as time delay, i.e., the amount of *backward* time shift to overlap the center notches (Bian *et al.*, 2004). Averaged across primary levels, the QDT delays for the two pairs of primary frequencies were about 1.4 and 1.1 ms, respectively. These QDT delays are generally consistent with similar measures for CDT (Bian *et al.*, 2004) and group delays of CDT in gerbils (Faulstich and Kössl, 2000), suggesting the same source of generation for the two different DPs. Moreover, the QDT delay at the higher primary frequencies is slightly shorter than CDT delay (1.2 ms). Since the frequency of QDT is always lower than the CDT ($f_2/f_1 < 1.5$), the shorter delay could be explained by a dispersive backward-traveling-wave where low frequencies travel faster. However, there is no difference between the delays of QDT and CDT at the lower primary frequencies. A recent finding (Ren, 2004) indicates that the DPOAEs propagate in acoustic pressure wave with virtually no delay. Therefore, the tuning characteristics of the cochlear transducer that generates DPOAEs could account for the difference between the delays of QDT and CDT. Since the transducer with higher characteristic frequency (CF) is more sharply tuned, the response phase around the CF is steeper (Geisler and Rhode, 1982) so that the phase difference between CDT and QDT is greater. These filter effects can be attributed to the time-dependent properties of the MET channels (Housley and Ashmore, 1992) and the membrane motors of the OHCs (Santos-Sacchi, 1992; Ospeck *et al.*, 2003) that determine the activation time and frequency limit.

2. Shapes of the MPs

The shape of the QDT magnitude MP is fundamentally different from that of the CDT. Despite the dependence on the bias tone phase, these MPs can be better compared with the averaged QDT and CDT magnitudes in dB SPL obtained from the peaks and troughs of the three cycles of bias tones with various peaks amplitudes (Fig. 11; see also Fig. 2). The major difference is located near zero BP, where CDT magnitude is almost maximal and QDT is relatively small. A small shift of the OP in the negative sound-pressure direction can produce a drastic reduction in QDT magnitude marked by a deep notch exceeding 20 dB, whereas the CDT magnitude remains relatively constant. This can explain the dissociation of QDT magnitude from CDT after prolonged stimulation of low-level primaries (Brown, 1988; Whitehead *et al.*, 1991), or increase in QDT magnitude following a brief

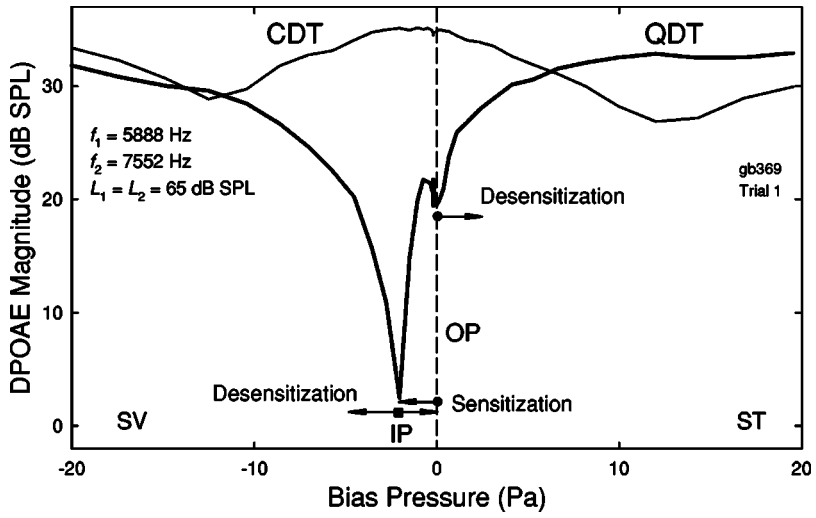


FIG. 11. Different MPs of QDT and CDT magnitudes. The MP of QDT magnitude resembles the shape of the $|f_{Tr}^{(2)}|$ marked by a deep notch near -2-Pa bias level, whereas the MP of CDT is similar to the $|f_{Tr}^{(3)}|$ with two notches located around ± 10 Pa. Dashed line indicates the location of the OP when there is no or little biasing (BP=0). Note: the small notch on the MP of QDT near zero BP is caused by the variability of QDT measured at several small bias levels. The deep notch on the magnitude of QDT corresponds to the IP of the cochlear f_{Tr}, where the transducer gain is optimal. A shift of the OP towards the IP results in large reduction in QDT magnitude with little change in CDT. For acoustic input without biasing, this is equivalent to the adjustment of the IP to zero BP, or "sensitization." An OP shift in the opposite direction, corresponding to the movement IP farther away from zero BP, is associated with "desensitization." Dot arrows indicate shifts of the OP from zero BP; arrows with a square denote adjustments of the IP. Positive BP indicates BM displacement towards scala tympani (ST); negative BP: scala vestibuli (SV).

exposure to a low-frequency tone (Kirk and Patuzzi, 1997; Sirjani *et al.*, 2004). The mechanism of the OP shift is likely to be the voltage-dependent length change of the OHCs, since the odd- and even-order DPs in the electromotile response of the isolated OHCs with varying holding potentials demonstrate different MPs (Takahashi and Santos-Sacchi, 1999) similar to those in Fig. 11. Increasing the driving voltage of the cochlear transducer by injecting positive current into scala media produces enhanced MP of QDT (Frank and Kössl, 1997) and BM vibration at the CF (Parthasarathi *et al.*, 2003). Therefore, the compressive nonlinearity in the OHC transduction underlies the different behaviors of the odd- and even-order DPs. Dependence of QDT on cochlear transducer driving voltage implies that compression and amplification are two inseparable aspects of the "active" process.

If the inputs to the f_{Tr} are acoustic stimuli without a bias tone, i.e., they are independent of OP shift or BP is always zero, then an alternative explanation of the OP shift would be that the cochlear transducer adjusts the location of its IP. Given that the initial OP for the primaries alone (BP=0) is situated on the rising slope of P_2 (Fig. 11), an OP shift in the negative sound-pressure direction or towards SV is equivalent to an adjustment of the transducer IP towards zero BP. This would result in "sensitization" of cochlear transducer, since the cochlear gain would be optimal at the zero BP. In this case, which may occur with prolonged stimulation of low-level primaries, the QDT magnitude is reduced and hearing threshold improved. Shift of the OP in positive direction or ST is associated with "desensitization," since the transducer IP is moved farther away from zero BP (Fig. 11). If this happens, e.g., after brief noise exposure, the transducer gain is reduced and signal is in asymmetrical compression (Bian and Chertoff, 1998), resulting in greater QDT magnitude and elevated threshold. Regulations of the cochlear f_{Tr} can also be observed when the signal level is varied and are discussed in the following sections.

3. Effects of primary level on indices of cochlear transduction

Physiologic indices that characterize the cochlear f_{Tr} show significant effects of the primary level. Notch, P_1, and P_2 exhibit at least two stages of growth, a shallower (<0.5 dB/dB) and a steeper (1 dB/dB) stage separated by a knee point at about 55 to 65 SPL (Fig. 9). The slow versus linear growths observed at lower and higher levels are in agreement with Brown (1993), who showed gradual increase of QDT magnitude below 65 dB SPL and a more rapid growth above 70 dB SPL. Compared with the CDT growth functions obtained from low-frequency biasing (Bian *et al.*, 2004) and standard methods (Dorn *et al.*, 2001; Mills, 2002), the slow QDT growth at lower levels corresponds to the linear growth of CDT that is attributable to higher transducer gain. This is consistent with a comparison that CDT magnitude starts to saturate and QDT begins to rise linearly around 60 dB SPL (Kim *et al.*, 1980). Compressive growth of CDT and corresponding linear growth of QDT have also been found in perceived DPs (Zwicker, 1979). Moreover, linear growth of QDT magnitude coincides with peaks of PN_1 and PN_2. Similar to the notch on CDT growth function, the knee point indicates the transition from amplification at low input levels to compression at higher levels. Combined measures of CDT and QDT could provide a complete estimate of cochlear transduction.

Index asymmetry smaller than 1 indicates that biasing towards positive sound pressures produces a greater QDT magnitude (P_1). This is consistent with the result that the dc

component of the round-window electrical response or summating potential is greater during low-frequency biasing in the positive pressure direction (Choi *et al.*, 2004), indicating greater compression while biasing the BM towards ST. Under loading condition at low primary levels ($\leqslant$55 dB SPL), a reversal of the symmetry is noticeable [Fig. 10(C)], i.e., biasing towards negative sound pressures produces greater QDT magnitude. This may be comparable to the observation that electrical biasing produced opposite change in the magnitude of the electrically evoked OAEs at low versus high levels (Roddy *et al.*, 1994). Indicator of the IP on the f_{Tr}, BP_{IP}, also shifts in the negative sound-pressure direction with primary level [Fig. 2 and Fig. 8(B)]. Shift of the IP away form zero BP contributes to the asymmetry of the f_{Tr} at higher levels. Therefore, the shape of the cochlear f_{Tr} alters dynamically depending on the input level. There is a small but sometimes significant effect of loading and unloading on the MP of QDT, i.e., loading produces greater QDT magnitude than unloading, especially the center notch at the higher primary frequencies [Fig. 9(A)]. Brown (1993) noticed that descending signal levels resulted in smaller QDT magnitude than ascending. If sequence of signal presentation is equivalent to loading process in the biasing, then its effect on QDT can be attributed to the hysteresis in cochlear f_{Tr}.

4. Regulation of cochlear transduction

Regulation of the f_{Tr} characteristics and shift of the IP with signal level may underlie the mechanisms of the ''cochlear amplifier.'' A local feedback from the voltage-dependent motile response of the OHCs in the cochlear transducer may actively control the gain of the f_{Tr} for effective energy conversions. In the feedback loop, OHC receptor potentials from MET act as feedback signal that enhances the mechanical response of the cochlear transducer (Yates, 1990), thus further influencing the voltage changes that control the force production. In fact, MET functions derived from the cochlear microphonic are level dependent (Patuzzi and Moleirinho, 1998; Bian and Chertoff, 1998, 2001). The dc component in the electric response could induce a tonic motile response of the OHCs (Evans *et al.*, 1991), and subsequently a static position shift of the BM depending on the input level (LePage, 1987; Cooper and Dong, 2003). This, in turn, could alter the IP of the cochlear f_{Tr}, resulting in sensitization at threshold or low input levels and desensitization at higher levels. The reciprocal relation between electrical and mechanical responses of the OHCs could produce the dynamical change in cochlear f_{Tr} and hysteresis (Bian *et al.*, 2004). Adaptation of the OHC transduction channels may also play a role in the active adjustment of the f_{Tr}. It is known that the adaptation of MET is Ca^{2+}-dependent (Fettiplace and Ricci, 2003), and blocking Ca^{2+} channels can prevent the reduction of QDT magnitude after prolonged stimulation of the primaries (Kujawa, 1996). Adaptation of myosin motors in the stereocilia could adjust the angle of the bundle that controls the overall open probability of the MET channels, thus resulting in alterations in the IP location and symmetry of the f_{Tr}. In mammals, adaptation of OHC transduction channel could influence the positive feedback from the somatic motility for optimal sensitivity at low levels and compression for high input.

Evidences from QDT measurement suggest that the efferent system may also involve in the regulation of cochlear f_{Tr} (Brown, 1988; Kirk and Johnstone, 1993; Chang and Norton, 1997). As indicated by summating potential, stimulating the efferent system can produce tonic activities of the OHCs (Konishi and Slepian, 1971) that could affect the IP of the f_{Tr}. Decrease in DPOAEs by activating efferent feedback is possibly due to the reduction in gain and adjustment of the IP of f_{Tr} in adaptation to the presence of the stimuli (Liberman *et al.*, 1996; Kim *et al.*, 2003). However, these efferent effects on DPOAEs are usually small ($<$6 dB) and slow (time constant $>$10 ms). In contrast, the changes in QDT magnitude in the present study are over 20 dB and on a cycle-by-cycle basis. Moreover, the data for analysis were taken from the middle portion of the stimuli to avoid the onset and offset of the bias tone (Fig. 3), where efferent reflex is more likely to occur. Therefore, the low-frequency modulation of DPOAEs is not an efferent effect. Although the efferent effect on the modulation of QDT is limited (Kujawa, 1995), the exact role of efferent system on the regulation of cochlear f_{Tr} remains to be determined.

C. Relation to the model of hysteresis

The adequacy of the ''negative damping'' model that fit to the MP of CDT magnitude (Bian *et al.*, 2004) was checked with the QDT data. The parameters derived from the CDT data were used to generate an output from which QDT magnitudes were extracted with the moving-window method. The model predicted the QDT double MP and hysteresis (Fig. 12), indicating that the same mechanism generates the two DPs. However, without adjusting model parameters, the predicted QDT magnitudes were greater than the actual data. When fitting the model to QDT data, reducing the parameter A [Eq. (1)] significantly improved the fit. From six animals ($r^2=0.5-0.9$), A was reduced 2–6 times depending on the primary levels, and adjustments to other parameters were within the variability from the CDT data. As evidenced from the electromotile response of the isolated OHCs (Takahashi and Santos-Sacchi, 1999), compressive nonlinearity of Boltzmann type usually produces greater QDT than CDT, especially when the OP is away from the IP (Lukashkin *et al.*, 2002). Although adding a negative damping can enhance CDT magnitude more than the increase in QDT, the model's overestimate of the QDT indicate that an additional component maybe involved in shaping the frequency spectra of the cochlear response. There could be a bandpass filtering during the DP-generating processes so that only frequency components closer to the CF, or f_2, can be amplified (Brown *et al.*, 1992). In mammals, tectorial membrane could be a structural basis for the filtering of DPOAEs (Lukashkin *et al.*, 2004), so that the low-passed OHC motility could be more efficient near the CF as shown in a feedback model (Cooper, 2004). Therefore, parameter A may reflect the filter shape of the feedback component.

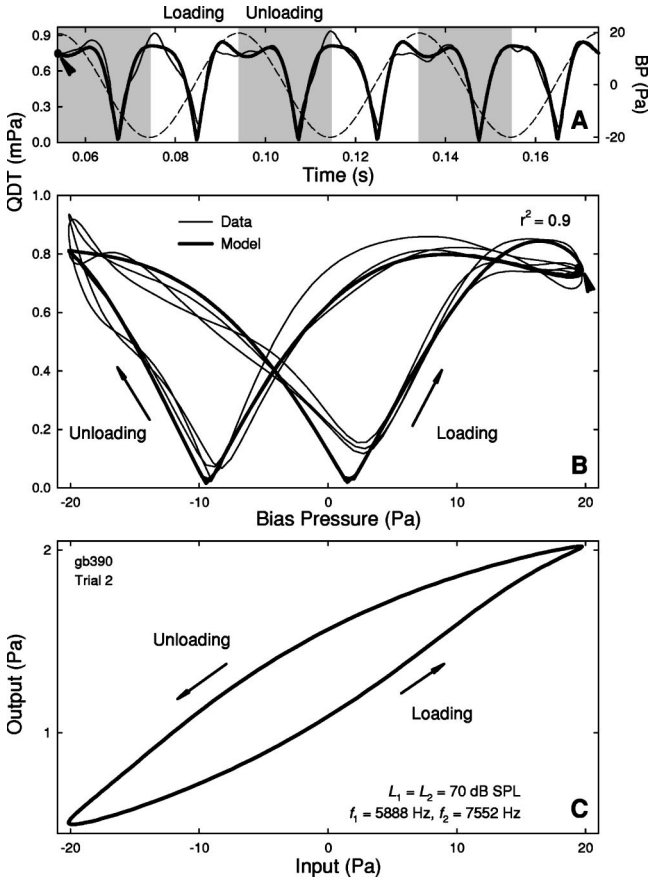


FIG. 12. Model results. (A) Temporal MP of QDT magnitude produced by the model (thick line) versus experimental data (thin line). Dashed line indicates the bias tone. (B) QDT double MP. The QDT magnitudes in panel (A) plotted as a function of BP. The short arrows indicate the starting points of the traces. (C) The hysteresis loop of the cochlear f_{Tr} derived from the model fit to the data in panels (A) and (B).

VI. SUMMARY AND CONCLUSION

Low-frequency modulation of QDT in gerbils showed enhancement and suppression depending on the phase of the bias tone. Suppression of QDT occurred around the zero crossings of the bias tone, and enhancement corresponded to the peaks and troughs of the BP. For each half cycle of bias tone with monotonic increase or decrease in BP, the MP of QDT magnitude demonstrated a typical ''V-shaped'' $|f_{Tr}^{(2)}|$. When the QDT magnitude was plotted as a function of the BP, a double MP consisting of two typical shapes of the $|f_{Tr}^{(2)}|$ presented with a separation of the center-notches. These experimental results are consistent with the observation from the CDT (Bian *et al.*, 2004) that the nonlinear cochlear f_{Tr} has hysteresis where the traversal is counterclockwise. These results also confirm that both odd and even-order DPs are generated from the same nonlinearity: the OHC transduction (Fahey *et al.*, 2000; Lukashkin *et al.*, 2002). Therefore, the QDT magnitude MP could be used with CDT in differential diagnosis of hearing problems related to cochlear transduction. Moreover, the enhanced QDT at higher levels may be less vulnerable to cochlear damage than the CDT. The QDT, a consequence of compression in cochlear transduction and an indicator of the IP on the f_{Tr}, would be useful for the early detection of Ménière's disease (Sirjani *et al.*, 2004), where position of the BM is displaced due to the enlargement of endolymphatic volume. Furthermore, since the frequency of the bias tone is closer to the DP frequencies (QDT and CDT), it could suppress the secondary generation of DPs at their CF regions (e.g., Konrad-Martin *et al.*, 2001). This could reduce the complexity of DPOAE measures caused by multiple sources of generation that limit the clinical application and interpretation (Shaffer *et al.*, 2003; Shera, 2004). Therefore, low-frequency modulated DPOAEs can be used to accurately estimate the nonlinear f_{Tr} at the primary source.

ACKNOWLEDGMENTS

The author is indebted to Mark Chertoff for help in the experiments and valuable comments on the manuscript. The author would like to thank Alec Salt for useful discussions and Jeff Lichtenhan for suggestions. This work was supported by a grant (R03 DC006165) from the National Institute on Deafness and Other Communication Disorders of the NIH.

Bian, L., and Chertoff, M. E. (**1998**). ''Differentiation of cochlear pathophysiology in ears damaged by salicylate or a pure tone using a nonlinear systems identification technique,'' J. Acoust. Soc. Am. **104**, 2261–2271.

Bian, L., and Chertoff, M. E. (**2001**). ''Distinguishing cochlear pathophysiology in 4-aminopyridine and furosemide treated ears using a nonlinear systems identification technique,'' J. Acoust. Soc. Am. **109**, 671–685.

Bian, L., Chertoff, M. E., and Miller, E. (**2002**). ''Deriving a cochlear transducer function from low-frequency modulation of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **112**, 198–210.

Bian, L., Linhardt, E. E., and Chertoff, M. E. (**2004**). ''Cochlear hysteresis: Observation with low-frequency modulated distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **115**, 2159–2172.

Brown, A. M. (**1988**). ''Continuous low level sound alters cochlear mechanics: An efferent effect?'' Hear. Res. **34**, 27–38.

Brown, A. M. (**1993**). ''Distortion in the cochlea: Acoustic f_2-f_1 at low stimulus levels,'' Hear. Res. **70**, 160–166.

Brown, A. M., Gaskill, S. A., and Williams, D. M. (**1992**). ''Mechanical filtering of sound in the inner ear,'' Proc. R. Soc. London, Ser. B **250**, 29–34.

Chang, K. W., and Norton, S. J. (**1997**). ''Efferently mediated changes in the quadratic distortion product (f_2-f_1),'' J. Acoust. Soc. Am. **102**, 1719–1733.

Cheng, J. (**1999**). ''Estimation of active cochlear nonlinearity by multicomponent distortion-product otoacoustic emissions,'' Acustica **85**, 721–727.

Choi, C.-H., Chertoff, M. E., Bian, L., and Lerner, D. (**2004**). ''Constructing a cochlear transducer function from the summating potential using a low-frequency bias tone,'' J. Acoust. Soc. Am. **116**(5).

Cooper, N. P. (**2004**). ''Compression in the peripheral auditory system,'' in *Compression: From Cochlea to Cochlear Implants*, edited by S. P. Bacon, R. R. Fay, and A. N. Popper (Springer, New York), pp. 18–61.

Cooper, N. P., and Dong, W. (**2003**). ''Baseline position shifts and mechanical compression in the apical turns of the cochlea,'' in *Biophysics of the Cochlea: from Molecules to Models*, edited by A. W. Gummer, E. Dalhoff, M. Nowotny, and M. P. Scherer (World Scientific, Singapore), pp. 261–267.

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., Cyr, E., and Gorga, M. P. (**2001**). ''Distortion product otoacoustic emission input/output functions in normal-hearing and hearing-impaired human ears,'' J. Acoust. Soc. Am. **110**, 3119–3131.

Evans, B. N., Hallworth, R., and Dallos, P. (**1991**). ''Outer hair cell electromotility: The sensitivity and vulnerability of the dc component,'' Hear. Res. **52**, 288–304.

Fahey, P. F., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**2000**). ''Nonlinear interaction that could explain distortion product interference response areas,'' J. Acoust. Soc. Am. **108**, 1786–1802.

Faulstich, M., and Kössl, M. (**2000**). ''Evidence for multiple DPOAE components based upon group delay of the $2f_1-f_2$ distortion in the gerbil,'' Hear. Res. **140**, 99–110.
Fettiplace, R., and Ricci, A. J. (**2003**). ''Adaptation in auditory hair cells,'' Curr. Opin. Neurobiol. **13**, 446–451.
Frank, G., and Kössl, M. (**1996**). ''The acoustic two-tone distortions $2f_1-f_2$ and f_2-f_1 and their possible relation to changes in the operating point of the cochlear amplifier,'' Hear. Res. **98**, 104–115.
Frank, G., and Kössl, M. (**1997**). ''Acoustic and electrical biasing of the cochlear partition. Effects on the acoustic two tone distortions f_2-f_1 and $2f_1-f_2$,'' Hear. Res. **113**, 57–68.
Geisler, C. D., and Rhode, W. S. (**1982**). ''The phases of basilar-membrane vibrations,'' J. Acoust. Soc. Am. **71**, 1201–1203.
Housley, G. D., and Ashmore, J. F. (**1992**). ''Ionic currents of outer hair cells isolated from the guinea-pig cochlea,'' J. Physiol. (London) **64**, 1386–1391.
Hudspeth, A. J. (**1989**). ''How the ear's works work,'' Nature (London) **341**, 397–404.
Kim, D. O., Molnar, C. E., and Matthews, J. W. (**1980**). ''Cochlear mechanics: Nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal pressure,'' J. Acoust. Soc. Am. **67**, 1704–1721.
Kim, D. O., Yang, X. M., and Neely, S. T. (**2003**). ''Effects of medial olivocochlear reflex on cochlear mechanics: Experimental and modeling studies of DPOAE,'' in *Biophysics of the Cochlea: From Molecules to Models*, edited by A. W. Gummer, E. Dalhoff, M. Nowotny, and M. P. Scherer (World Scientific, Singapore), pp. 506–514.
Kirk, D. L., and Johnstone, B. M. (**1993**). ''Modulation of f_2-f_1: Evidence for a GABA-ergic efferent system in apical cochlea of the guinea pig,'' Hear. Res. **67**, 20–34.
Kirk, D. L., and Patuzzi, R. B. (**1997**). ''Transient changes in cochlear potentials and DPOAEs after low-frequency tones: The 'two-minutes bounce' revisited,'' Hear. Res. **112**, 49–68.
Konishi, T., and Slepian, J. (**1971**). ''Summating potential with electrical stimulation of crossed olivocochlear bundles,'' Science **172**, 483–484.
Konrad-Martin, D., Neely, S. T., Keefe, D. H., Dorn, P. A., and Gorga, M. P. (**2001**). ''Sources of distortion product otoacoustic emissions revealed by suppression experiments and inverse fast Fourier transforms in normal ears,'' J. Acoust. Soc. Am. **109**, 2862–2879.
Kujawa, S. G., Fallon, M., and Bobbin, R. P. (**1995**). ''Time-varying alterations in the f_2-f_1 DPOAE response to continuous primary stimulation. I. Response characterization and contribution of the olivocochlear efferents,'' Hear. Res. **85**, 142–154.
Kujawa, S. G., Fallon, M., Skellett, R. A., and Bobbin, R. P. (**1996**). ''Time-varying alterations in the f_2-f_1 DPOAE response to continuous primary stimulation. II. Influence of local calcium-dependent mechanisms,'' Hear. Res. **97**, 153–164.
LePage, E. L. (**1987**). ''Frequency-dependent self-induced bias of the basilar membrane and its potential for controlling sensitivity and tuning in the mammalian cochlea,'' J. Acoust. Soc. Am. **82**, 139–154.
Liberman, M. C., Puria, S., and Guinan, Jr., J. J. (**1996**). ''The ipsilaterally evoked olivocochlear reflex causes rapid adaptation of the $2f_1-f_2$ distortion product otoacoustic emission,'' J. Acoust. Soc. Am. **99**, 3572–3584.
Lukashkin, A. N., Lukashkina, V. A., and Russell, I. J. (**2002**). ''One source for distortion product otoacoustic emissions generated by low- and high-level primaries,'' J. Acoust. Soc. Am. **111**, 2740–2748.
Lukashkin, A. N., Lukashkina, V. A., Legan, P. K., Richardson, G. P., and Russell, I. J. (**2004**). ''Role of the tectorial membrane revealed by otoacoustic emissions recorded from wild-type and transgenic Tectra$^{\Delta ENT/\Delta ENT}$ mice,'' J. Neurophysiol. **91**, 163–171.
Martin, G. K., Lonsbury-Martin, B. L., Probst, R., Scheinin, S. A., and Coats, A. C. (**1987**). ''Acoustic distortion products in rabbit ear canal. II. Sites of origin revealed by suppression contour and pure-tone exposures,'' Hear. Res. **28**, 191–208.
Mills, D. M. (**2002**). ''Interpretation of standard distortion product otoacoustic emission measurements in light of the complete parametric response,'' J. Acoust. Soc. Am. **112**, 1545–1560.
Olson, E. S. (**2004**). ''Harmonic distortion in intracochlear pressure and its analysis to explore the cochlear amplifier,'' J. Acoust. Soc. Am. **113**, 442–452.
Ospeck, M., Dong, X., and Iwasa, K. H. (**2003**). ''Limiting frequency of the cochlear amplifier based on electromotility of outer hair cells,'' Biophys. J. **84**, 739–749.
Parthasarathi, A. A., Groth, K., Zheng, J., and Nuttall, A. L. (**2003**). ''Effect of current stimulus on *in vivo* cochlear mechanics,'' J. Acoust. Soc. Am. **113**, 442–452.
Patuzzi, R., and Moleirinho, A. (**1998**). ''Automatic monitoring of mechanoelectrical transduction in the guinea pig cochlea,'' Hear. Res. **125**, 1–16.
Patuzzi, R., Sellick, P. M., and Johnstone, B. M. (**1984**). ''The modulation of the sensitivity of the mammalian cochlea by low frequency tones. III. Basilar membrane motion,'' Hear. Res. **13**, 19–27.
Ren, T. (**2004**). ''Reverse propagation of sound in the gerbil cochlea,'' Nat. Neurosci. **7**, 333–334.
Robles, L., and Ruggero, M. A. (**2001**). ''Mechanics of the mammalian cochlea,'' Physiol. Rev. **81**, 1305–1352.
Roddy, J., Hubbard, A. E., Mountain, D. C., and Xue, S. (**1994**). ''Effects of electrical biasing on electrically-evoked otoacoustic emissions,'' Hear. Res. **73**, 148–154.
Santos-Sacchi, J. (**1992**). ''On the frequency limit and phase of outer hair cell motility: Effects of the membrane filter,'' J. Neurosci. **12**, 1906–1916.
Scholz, G., Hirschfelder, A., Marquardt, T., Hensel, J., and Mrowinski, D. (**1999**). ''Low-frequency modulation of the $2f_1-f_2$ distortion product otoacoustic emissions in the human ears,'' Hear. Res. **130**, 189–196.
Shaffer, L. A., Withnell, R. H., Dhar, S., Lilly, D. J., Goodman, S. S., and Harmon, K. M. (**2003**). ''Sources and mechanisms of DPOAE generation: Implications for the prediction of auditory sensitivity,'' Ear Hear. **24**, 367–379.
Shera, C. A. (**2004**). ''Mechanisms of mammalian otoacoustic emission and their implications for the clinical utility of otoacoustic emissions,'' Hear. Res. **25**, 86–97.
Sirjani, D. B., Salt, A. N., Gill, R. M., and Hale, S. A. (**2004**). ''The influence of transducer operating point on distortion generation in the cochlea,'' J. Acoust. Soc. Am. **115**, 1219–1229.
Takahashi, S., and Santos-Sacshi, J. (**1999**). ''Distortion component analysis of outer hair cell motility-related gating charge,'' J. Membr. Biol. **169**, 199–207.
van Emst, M. G., Klis, S. F. L., and Smoorenburg, G. F. (**1997**). ''Identification of the nonlinearity governing even-order distortion products in cochlear potential,'' Hear. Res. **114**, 93–101.
Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). ''Slow variation of the amplitude of acoustic distortion at f_2-f_1 in awake rabbits,'' Hear. Res. **51**, 293–300.
Wojtczak, M., Schroder, A. C., Kong, Y.-Y., and Nelson, D. A. (**2001**). ''The effect of basilar-membrane nonlinearity on the shapes of masking period patterns in normal and impaired hearing,'' J. Acoust. Soc. Am. **109**, 1571–1586.
Yates, G. K. (**1990**). ''Basilar membrane nonlinearity and its influence on auditory nerve rate-intensity functions,'' Hear. Res. **50**, 145–162.
Zwicker, E. (**1977**). ''Masking-period patterns produced by very-low-frequency maskers and their possible relation to basilar-membrane displacement,'' J. Acoust. Soc. Am. **61**, 1031–1040.
Zwicker, E. (**1979**). ''Different behaviour of quadratic and cubic difference tones,'' Hear. Res. **1**, 283–292.
Zwicker, E., and Manley, G. (**1981**). ''Acoustical responses and suppression-period patterns in guinea pigs,'' Hear. Res. **4**, 43–52.
Zwicker, E., and Scherer, G. (**1987**). ''Correlation between time function of sound pressure, masking, and OAE suppression,'' J. Acoust. Soc. Am. **81**, 1043–1049.

Distortion product otoacoustic emission ($2f1-f2$) suppression in 3-month-old infants: Evidence for postnatal maturation of human cochlear function?

Carolina Abdala[a)]
Children's Auditory Research and Evaluation Center, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 26 May 2004; revised 31 August 2004; accepted 10 September 2004)

The complete timeline for maturation of human cochlear function has not been defined. Distortion product otoacoustic emission (DPOAE)-based measures of cochlear function show non-adult-like responses from premature and term-born neonates at high $f2$ frequencies; however, older infants were not included in these studies. In the present experiment, previously collected DPOAE ipsilateral suppression data from premature neonates were combined with new data collected from adults, term-born neonates, and 3-month-old infants to further examine the time course for maturation of cochlear function. DPOAE suppression tuning curves (STC) and suppression growth patterns were measured in the three age groups at $f2=6000$ Hz, L1=65, L2=55 dB SPL, with an $f2/f1$ of 1.2. Results indicate that term-born neonates and 3-month-old infants have non-adult-like STC width, slope on the low-frequency flank, and tip features. However, the two infant groups are not significantly different from one another. Suppression growth patterns for low-frequency suppressor tones show a clear developmental progression. In general, the younger the infant, the more shallow and compressive the suppression growth for the lowest suppressor frequencies. These findings suggest a high-frequency postnatal immaturity in cochlear function as measured by DPOAE suppression. Results may have been influenced by noncochlear factors, such as middle-ear immaturity. These factors are reviewed and considered.
[DOI: 10.1121/1.1811472]

PACS numbers: 43.64.Jb, 43.64.Kc [BLM] Pages: 3572–3580

I. INTRODUCTION

Several characteristics of $2f1-f2$ DPOAE suppression are non-adult-like in premature neonates. Research using both cross-sectional and longitudinal experimental designs have shown that premature neonates (ranging from 31 to 39-weeks postconceptional age) have narrower suppression tuning curves than adults at $f2=1500$ and 6000 Hz, and atypically shallow growth of suppression for suppressor tones lower in frequency than $f2$ (Abdala, 1998; 2001; Abdala and Chatterjee, 2003). These age effects are most robust and reliably present at $f2=6000$ Hz and have been interpreted to indicate an immaturity in cochlear function just prior to term birth.

In contrast to the consistent DPOAE-based immaturities observed in premature neonates, term-born neonates have produced somewhat mixed results. They appear to have adult-like DPOAE suppression tuning at $f2=1500$ Hz, but narrower, sharper tuning than adults at $f2=6000$ Hz (Abdala *et al.*, 1996; Abdala, 1998). They show excessively shallow suppression growth for low-frequency suppressor tones at both 1500 and 6000 Hz. Therefore, although term-born neonates show some adult-like features of cochlear function as measured by DPOAE suppression, they also show systematic immaturities. These immaturities suggest that continued development of the cochlear response may persist beyond term birth in humans.

Because immaturities in DPOAE suppression are present at term birth, ''older'' infants need to be studied to define the time course for complete maturation of human cochlear function. Much of the developmental work examining cochlear function in humans thus far has focused on the period prior to full gestation (<37 weeks). However, recent findings have indicated that DPOAE-based assays of cochlear function change little between 33-weeks postconceptional age (PCA) and term status (37–40 weeks) (Abdala, 2003); therefore, it will be necessary to examine the months following term birth to describe the timeline for maturation of cochlear function as measured by DPOAE suppression. Preliminary work from our laboratory has shown that 1-month-old infants have immature DPOAE suppression tuning and suppression growth at 6000 Hz, although this finding was based on a very small number of observations (Abdala, 2003). Additionally, normal-hearing children ranging in age from 8 to 12 years have completely adult-like DPOAE suppression at both 1500 and 6000 Hz (Abdala, 2001). Therefore, we have initiated the study of postnatal cochlear maturation by measuring DPOAE suppression in 3-month-old infants and comparing their results to newborns and to normal-hearing adults. In the present study, we: (1) compiled existing DPOAE suppression data from premature neonates ranging from 31-weeks PCA to 1 month corrected age and (2) collected new DPOAE suppression data from healthy term newborns, 3-month-old infants, and adults. The inclusion of 3-month-old infants allowed for the study of

a)Electronic mail: cabdala@hei.org

cochlear development during this early postnatal period. The research question addressed in this study is whether cochlear function, as measured by DPOAE ipsilateral suppression, is adult-like by the first months of postnatal life.

II. METHODS

A. Subjects

Subjects tested for this study included: a group of ten healthy term-born neonates, a group of ten healthy 3-month-old infants (term-born), and a group of five normal-hearing adults. The data from these five adults were combined with data from five adults previously tested with the same paradigm to reach an $n=10$. Each subject was tested during a one-time test session. In addition, data from subjects participating in completed experiments were included in many of the analyses and figures for comparison. DPOAE suppression data previously collected included: nine prematurely born newborns tested at weekly intervals for 6–8 test sessions (three of these nine subjects were tested again at 1 month corrected age) and five adult subjects tested at weekly intervals for 5–6 test sessions (see Abdala, 2003 for detail *re*: adult and premature subjects).

The term neonates (eight males; two females) were tested on average within 4 days (97 h) of birth; their mean 1- and 5-min APGAR scores were 8.2 and 9.1, respectively. Four left and six right ears were tested. Average birth weight was 3506 grams (i.e., 7.7 lbs.). The 3-month-old infants (six males; four females) were tested on average 104 days after term birth (i.e., 14.8 weeks or 3.5 months). Seven right ears and three left ears were tested. These infants had an unremarkable birth history and normal postnatal development as per parent interview and postnatal medical examinations. All infant subjects passed a click-evoked ABR hearing screening (35 dB HL) at birth to be included in this study. In addition, a DP-gram was performed as a screening evaluation prior to the test session to ensure adequate signal-to-noise ratio (+5 dB) at the test frequency.

B. Instrumentation and signal processing

An Ariel DSP16+ signal processing and acquisition board housed within a Compaq Prolinea 590 personal computer with Pentium processor was used to generate stimuli and acquire data. The Ariel board was connected to an Etymotic Research ER-10C probe system and to an analog high-pass filter (12-dB/oct; 710-Hz high-pass cutoff). The ER-10C probe contains two output transducers and a low-noise microphone. The two primary tones and the suppressor tone were generated by the DSP processor. The primary tone at $f1$ was generated by one D/A converter and delivered via one transducer. The primary tone at $f2$ and the suppressor tone (f_s) were produced by the second D/A converter and output through the second transducer. The signal at the probe microphone was high-pass filtered and sampled at a rate of 50 kHz with a sweep length of 4096 samples, giving a frequency resolution of 12.2 Hz. Twenty-five sweeps of the microphone signal were added and comprised one block.

C. Data acceptance criteria

Acceptance criteria were as follows: (1) noise measurements for three frequency bins (12.2 Hz wide) on either side of the $2f1-f2$ frequency had to be <0 dB SPL to assure appropriate subject state; and (2) the measured DPOAE level had to be at least 5 dB above the average noise measured in the same six bins around the distortion product frequency to be accepted into the grand average. In addition, sweeps were accepted into a block of data only when the estimated rms level in that sweep did not exceed a user-controlled artifact rejection threshold. This level was set for each subject based on observations of baseline activity level determined early in the test session, and modified if necessary during the experiment.

D. Calibration

Intermodulation distortion produced by the recording system at $2f1-f2$ was measured with the probe in a Zwislocki coupler for all test conditions. The mean level of distortion was -21 dB SPL. In no case did the level exceed -17 dB SPL. The recording system noise floor was determined using a similar method with no tones present. The level of system noise floor ranged between -22 and -27 dB SPL depending on frequency.

An *in situ* calibration procedure was conducted on both output transducers before each subject was tested. Tones of fixed voltage were presented to the transducers at 250 Hz intervals from 500 to 10 000 Hz and the resulting SPL of these tones recorded in each ear canal. Based on this information, an equalization of output levels was performed for each subject to achieve target stimulus levels across test frequencies.

E. Procedure

Following the acquisition of informed parental consent, term-born neonates and 3-month-old infants were tested at Women and Children's Hospital Los Angeles County+USC Medical Center in a quiet room. Newborns were swaddled and fed if necessary, then placed in a hospital isolette for testing. Three-month-old infants were tested either in the isolette, in the mother's arms, or sleeping in an infant seat. Adult subjects were tested within an IAC sound-attenuated booth at the House Ear Institute in the Children's Auditory Research Laboratory while sitting comfortably in a lounge chair, reading, or napping.

DPOAE suppression data were collected at $f2=6000$ Hz using 15 different suppressor tones. The most marked age effects for DPOAE suppression have been previously observed at 6000 Hz. For this reason, 6000 Hz was chosen as the test frequency to enhance the probability of detecting age effects. DPOAEs were recorded with an $f2/f1$ ratio of 1.2 and primary tone levels were presented at 65 (L1)–55 (L2) dB SPL. Custom-designed software for the collection of DPOAE suppression tuning curves was developed at the Children's Auditory Research and Evaluation Center, House Ear Institute.

Once the newborn was asleep, he/she was positioned either on his/her back or slightly angled onto their side with

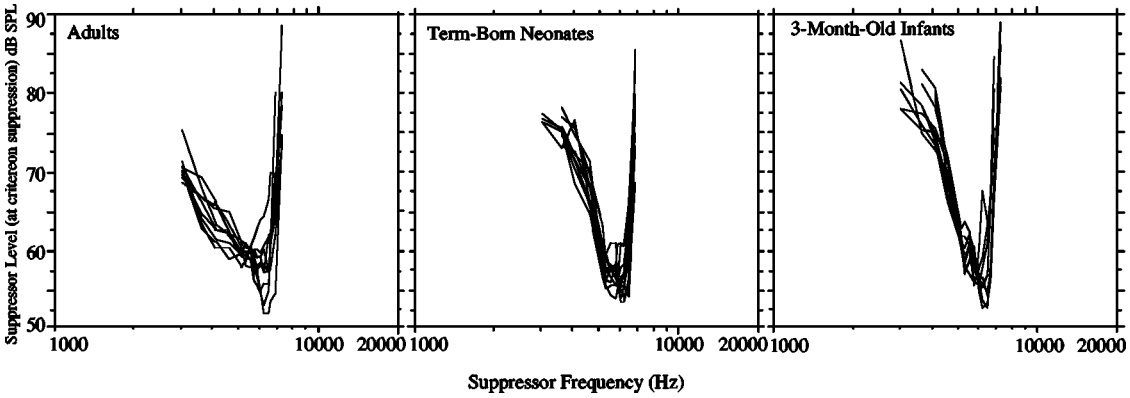


FIG. 1. Ten superimposed DPOAE suppression tuning curves (STC) recorded at $f2=6000$ Hz for three age groups. A 6-dB suppression criterion was used to plot each STC.

blankets for support at their back. The positioning was conducted to ensure optimal accessibility to the external auditory meatus for placement of the probe. When the probe was placed at the meatus snugly and the cable fixed to the isolette or the blankets with surgical tape to minimize movement, the test was initiated. If the probe tip slipped out minimally during testing (as noted by a 2–3-dB drop in stimulus and DPOAE levels), it was nudged back into place until appropriate levels were achieved and the test continued. If the probe tip came out of the ear completely during the recording of a DPOAE STC, it was refit and levels were recalibrated before continuing.

At the beginning of the test, an unsuppressed DPOAE was initially measured and then a suppressor tone (F_s) was presented simultaneously with the primary tones and its level increased in 5-dB steps over a range of intensities (40 to 85 dB SPL). Fifteen suppressor tones with frequency ranging from one octave below to 1/4 octave above $f2$ were presented at intervals between 25 and 150 cents (one octave $=1200$ cents). The list of suppressor frequencies used in this experiment is shown in the Appendix (left-most column). To generate iso-suppression tuning curves, the suppressor level reducing DPOAE amplitude by 6 dB was determined for each suppressor tone using linear interpolation (between data points just prior to and following 6 dB of suppression) and then plotted as a function of its frequency.

F. Data analysis

DPOAE suppression data were analyzed in the following manner: (1) DPOAE suppression tuning curve width was quantified 10 dB above the tip by calculating a Q_{10} value; (2) slope of the low- and high-frequency flank of the STC was determined by fitting a regression line from the tip of the curve to end points on the low- and high-frequency boundaries; (3) STC tip frequency and suppressor level corresponding to the tip were measured; (4) the growth pattern of DPOAE suppression was calculated for each suppressor tone by measuring the slope of the DPOAE amplitude $\times$ suppressor level function using a linear regression equation. Any initial amplitude plateau was eliminated and only the linear portion of this function was included. The measured segment began 1–2 dB down from the unsuppressed value and included all points with sufficient SNR. If the last points in the function showed an increase in DPOAE amplitude they were eliminated.

III. RESULTS

A. Suppression tuning curves

Figure 1 shows DPOAE STCs collected with $f2=6000$ Hz from ten adults, ten term-born neonates, and ten 3-month-old infants. The suppression tuning curves collected from term neonates and adults are typical in morphology and their quantitative features (width, slope, and tip values) fall well within the normal range of suppression values as reported in previous literature (Abdala *et al.*, 1996; Abdala, 1998). It is clear from Fig. 1 that the low-frequency slope and the width of the adult STC were different than the STCs collected from either infant group. The adult tuning curve appears to be broader with a more shallow low-frequency flank.

Figures 2 and 3 display STC Q_{10} and low-frequency flank data from all subject groups and includes longitudinal and single-session data. The longitudinally collected data from premature neonates and adults were collected, analyzed, and described in a previous publication (Abdala, 2003). In those reports, repeated measures ANOVAs (restricted to the age range from 33–37 weeks) indicated no effect of test session or postconceptional age on any STC feature, but a significant effect of age group; adults had significantly broader tuning (smaller Q_{10}) and a less steep low-frequency flank than premature neonates (Q_{10}: $f=24.51$; $p=0.0008$; Low-frequency slope: $f=21$; $p=0.0001$). As is evident from Figs. 2 and 3, STC data from term, 1-month-old, and 3-month-old infants, show the same trend; the DPOAE STC is narrower in infants than adults and has a steeper low-frequency flank. Mean data are included in the inset graphs and suggest that there may be a progression toward adult-like values of STC width and steepness as the infants approach 3 months of age.

The single-session data collected for the three primary age groups (term newborns, 3-month-old infants, and adults) were statistically analyzed for age effects [Figs. 4(a)–(c)]. One-way ANOVAs were conducted for STC Q_{10}, slope, tip level, and tip frequency. Age effects were observed for Q_{10},

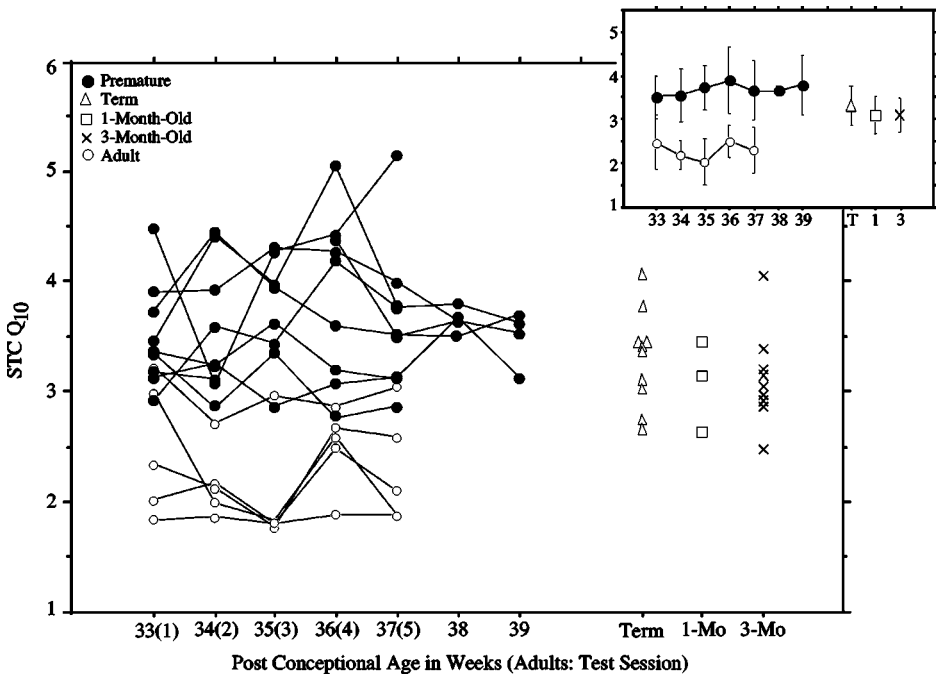


FIG. 2. DPOAE STC width (Q_{10}) plotted as a function of repeated test session for premature neonates and adults. Data from one single test session are plotted for term-born neonates and 3-month-old infants. The open squares reflect data from three infant subjects born prematurely and tested at 1 month corrected age. Inset graph displays mean Q_{10} from each age group.

low-frequency slope, and tip level ($f=30.2$; $p=0.0001$; $f=58.42$; $p=0.0001$; $f=3.79$; $p=0.03$, respectively). *Post hoc* t-tests further revealed that both the term group and the 3-month-old infant group had significantly larger Q_{10} values and significantly steeper low-frequency flank than the adult group. In contrast, neither tuning curve width nor steepness on the low-frequency flank was significantly different between term newborns and 3-month-old infants. Although non-adult-like, the mean Q_{10} and low-frequency values from the 3-month-old group appear to be moving in an adult-like direction.

Tip level also showed an age effect, indicating that the level of the suppressor tone at the tip of the STC was generally elevated for adults relative to either term newborns or 3-month-old infants [see Fig. 4(c)].

B. Suppression growth

Figure 5 shows mean growth of suppression (slope in dB/dB) plotted as a function of suppressor frequency for adults, 3-month-old infants, term newborns, and data previously collected from prematurely born infants at four different ages (33-, 35-, 38-weeks PCA and 1 month corrected age). Standard deviations of the mean are provided in the Appendix. Data points to the left of the dashed vertical line represent suppressor tones lower in frequency than $f2$ ($<f2$); those to the right, suppressor tones higher in frequency than $f2$ ($>f2$). Data points below the gray horizontal line reflect shallow/compressive growth of suppression (<1.0); points above this line reflect steep/linear growth (>1.0). Adults show the expected, previously reported pat-

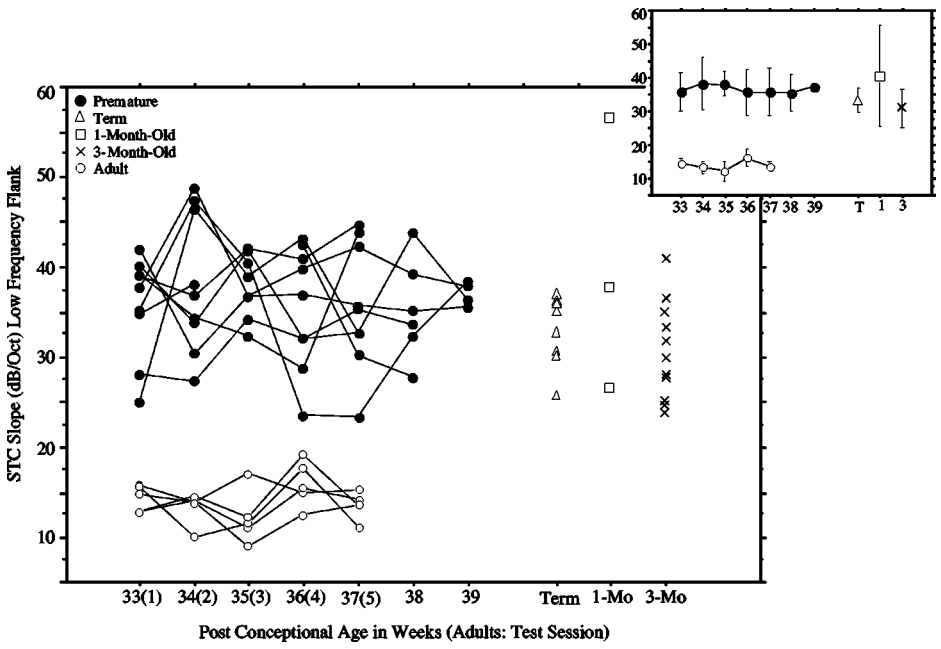


FIG. 3. Slope on the low-frequency flank of the DPOAE STC as a function of repeated test session for premature neonates and adults. Data from one single test session are plotted for term-born and 3-month-old infants. The open squares reflect data from three infant subjects born prematurely and tested at 1 month corrected age. Inset graph displays mean slope from each age group.

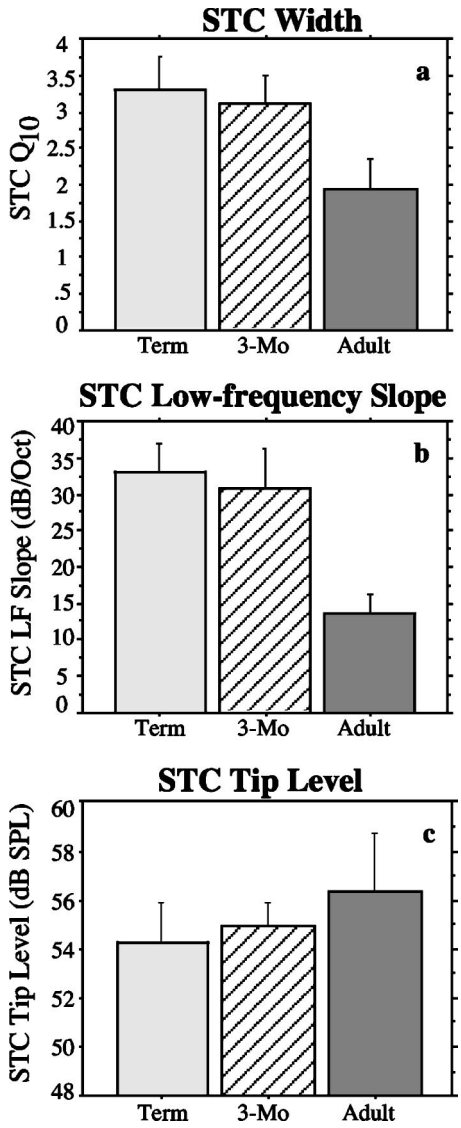


FIG. 4. (a)–(c) Mean DPOAE STC (a) width, (b) low-frequency slope, and (c) tip level for three age groups. Each mean value is comprised of ten observations.

tern of suppression growth. The steepest growth is observed for suppressor tones$<f2$ and shallow growth of suppression for tones$>f2$. This pattern accurately reflects basilar-membrane nonlinearity since a suppressor lower in frequency than the probe (i.e., $f2$ tone) should produce linear or near-linear response/suppression growth. This typical pattern of DPOAE suppression growth in adults is consistent with physiologic and psychoacoustic studies of response growth in laboratory animals and humans.

One-way ANOVAs were performed only on the four lowest suppressor frequencies to test for age effects. These four frequencies were targeted because they are the only data points indicating a potential age effect as seen in Fig. 5. There was a significant effect of age group on suppression growth for 3044, 3620, and 4063 Hz, indicating that infants show more shallow growth of suppression than adults at these frequencies ($f=8.95$; $p=0.0001$; $f=2.74$; $p=0.046$; $f=3.28$; $p=0.02$, respectively). Premature neonates at 33-, 35-, and 38-weeks postconceptional age show a peak of steep suppression growth for frequencies around $f2$, but increasingly shallow growth for tones$<f2$. At the very lowest frequencies, these premature groups show very compressive growth of suppression, indicating that the DPOAE does not suppress easily as suppressor level is raised. In contrast, 3-month-olds (X) show a pattern of suppression growth similar to adults for all but the lowest frequency suppressor tone (3044 Hz). At this frequency, their growth of suppression remains excessively shallow. Term neonates and 1-month-old infants show intermediate slope values (falling between neonate and adult values), suggesting a developmental progression in the pattern of suppression growth for low-frequency suppressor tones.

C. Summary of results

Premature neonates that have reached 39–40-weeks PCA, term-born neonates, and 3-month-old infants show various features of DPOAE suppression tuning and suppression growth that are non-adult-like. DPOAE STCs are narrower and steeper on the low-frequency flank, show a slightly reduced tip level, and the growth of suppression for the lowest frequency suppressor tones is more shallow in infants than adults.

IV. DISCUSSION

A. Immaturity in cochlear function

The results suggest that cochlear function, as measured by DPOAE ipsilateral suppression, is still not completely mature in the early postnatal period. These results, combined with previously published findings, suggest that cochlear immaturities are subtle in nature, and data appear to become systematically more adult-like with increasing postnatal age. This developmental pattern is particularly striking when examining growth of suppression patterns (see Fig. 5). Shallow, compressive growth of suppression for low-frequency suppressor tones is not observed in the healthy, mature cochlea. The adult cochlea shows the expected pattern of very effective, at least linear suppression of the primary DPOAE source (around $f2$) when a suppressor tone is presented at a more apical region than the probe ($f2$). The portion of the traveling wave vibration pattern overlapping with $f2$ behaves linearly because it is far from the compressive peak region. Suppressor tones$>f2$ cannot effectively suppress the DPOAE, even as their level is raised, because of compressive motion in the peak region of this traveling wave (Abdala and Chatterjee, 2003). This pattern represents normal response/suppression growth on the basilar membrane and is consistent with reported nonlinearities in physiological and psychoacoustic paradigms (Oxenham and Plack, 1997; Rhode and Cooper, 1993; Cooper, 1996).

Infants, on the other hand, do not show linear growth of suppression when suppressors are $<f2$. They show compressive growth of suppression (<1.0). Paradoxically, this result suggests that vibratory motion on the infant basilar membrane is more focal around CF than adult motion and more *nonlinear* as level is increased. Past work has, in fact, led us to hypothesize that the newborn cochlear amplifier is excessively nonlinear and may go through a developmental ''overshoot'' period. This type of overshoot of nonlinearity has

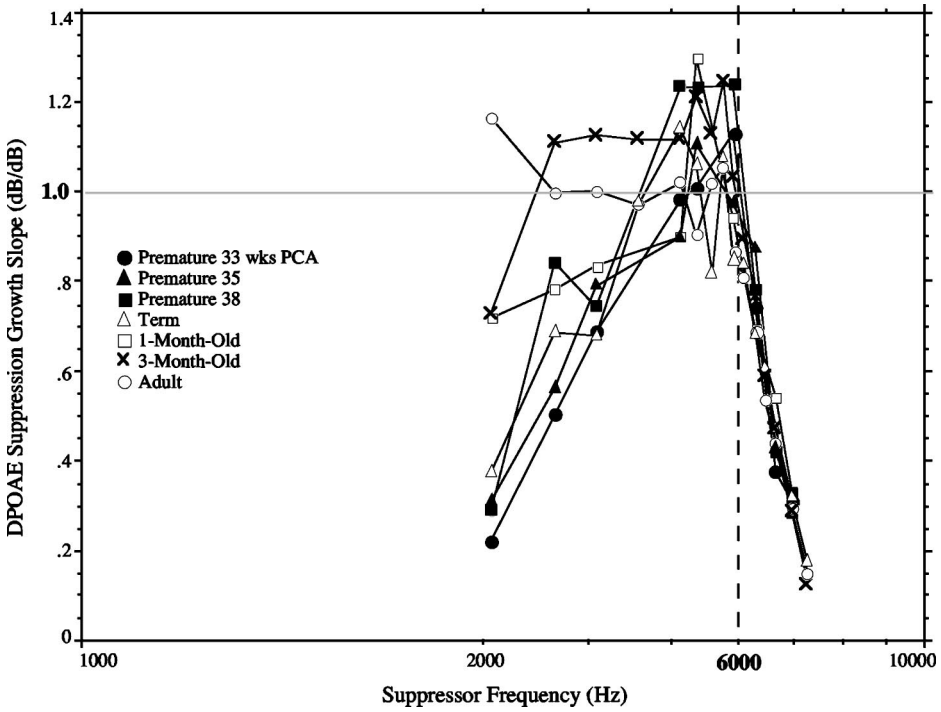


FIG. 5. DPOAE suppression growth slope for three age groups and premature neonates at four different ages (33-, 35-, and 38-weeks postconceptional age and 1 month corrected age). Points to the left of the vertical, dashed line represent slope values that are lower in frequency than $f2$ and points to the right, values that are higher in frequency. Data at or above the gray horizontal line show linear, steep growth of suppression, whereas data below the gray line indicate compressive, shallow growth.

been reported in gerbils during development and has been attributed to variations in the endocochlear potential (Mills and Rubel, 1998). At present, the exact mechanism or source of non-adult-like DPOAE suppression growth in human infants is not clearly understood; however, these consistently present and highly replicable age effects may be indicating a subtle immaturity in cochlear function, even into the early postnatal months.

This immaturity in cochlear function is most likely related to immaturities in the *cochlear amplifier*, since the last maturational stages of morphological development in the human fetal cochlea support cochlear micromechanics (Lavigne-Rebillard and Pujol, 1987). However, an alternative possibility must be considered. It is possible that passive vibratory characteristics of the infant cochlea related to basilar-membrane mass and stiffness remain immature into the postnatal period. We have typically considered these to be early maturing aspects of cochlear anatomy and function. However, animal research has shown concurrent development of active and passive cochlear function in gerbils during the early weeks of postnatal life.

Mills and Rubel (1996) reported that the passive cutoff frequency (i.e., frequency above which passive waves do not propagate on the basilar membrane) to be approximately 8000 Hz at 14 days after birth (dab) in gerbil. Only a few days later (18–23 dab), the cutoff frequency was almost adult-like at 32 kHz. Mills and Rubel stated that this 2-octave shift in base cutoff frequency is consistent with a reported 1.5-octave shift of characteristic frequencies in the basal turn of the gerbil during this same time period (Arjmand *et al.*, 1988). The work of Mills and Rubel suggests that passive constraints may require higher stimulus levels at high frequencies to activate areas of the cochlea that have mature cochlear amplifier function. That is, what appears to be immature cochlear amplifier function may actually be directly related to elevated passive thresholds of basilar-membrane motion. If this is the case in human newborns, stimulus levels reaching the infant cochlea may not provide an optimum trigger for active processes, even if the cochlear amplifier is fully mature. In contrast, the same stimulus levels probably evoke active cochlear function effectively in adults.

Available human morphological data do not shed light on this question of passive vibratory motion of the basilar membrane. Bredberg (1968) reported on light microscope studies of the postmortem fetal cochlea, and found that at 5 fetal months, fluid spaces were still immature. He also described one cochlea in which the mesenchymal layer under the basilar membrane was thickened (relative to adults). Lindeman *et al.* (1971) found that the cat cochlea looks basically mature by 2–3 weeks after birth, but that some changes in the detachment of the tectorial membrane can be seen up to 4 postnatal weeks. Likewise, decades earlier, Wada (1923) found that the rat cochlea showed slight changes well after most structures had achieved adult size. Any subtle, late immaturities in mass of the basilar membrane could potentially affect stiffness and, consequently, vibratory patterns. Although the DPOAE suppression paradigm we used for the current experiment was effective in detecting age effects related to cochlear function, it does not allow for a distinction between immaturities based in passive versus active basilar-membrane motion.

Putting aside the issue of whether functional cochlear immaturities are due to active or passive aspects of basilar-membrane vibration, it is not known whether these subtle age differences produce any perceptual consequences for infants. Are these cochlear immaturities represented and preserved at higher levels of the auditory system? Both behavioral and auditory brainstem response (ABR) measures of frequency resolution have shown immaturities in frequency

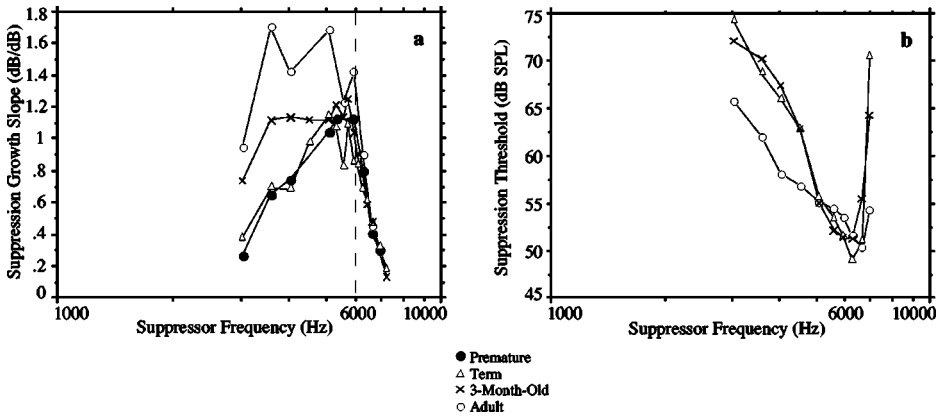


FIG. 6. (a) Suppression growth for three infant groups and adults. Infant data have been collected with primary tones presented 10 dB higher (L1=65; L2=55 dB SPL) than adult primary tones (L1=55; L2=45 dB SPL). (b) Suppression thresholds are defined as the level of suppressor required to just initiate suppression; that is, to produce 2 dB of suppression. Suppression thresholds are shown here for three age groups.

coding for 3-month-old infants. ABR-derived tuning curves and bandpass functions are non-adult-like in 3-month-olds at 8000 Hz, but appear to be mature at lower frequencies. Both low and high frequencies are adult-like by 6 months of age (Folsom and Wynne, 1987; Abdala and Folsom, 1995). Likewise, psychoacoustic tuning curves recorded with high-frequency probe tones (8000 Hz) are immature in 3-month-old infants, although they also become adult-like by 6 months of age (Spetner and Olsho, 1990). Whether the source of ABR and psychoacoustic tuning immaturities in 3-month-old infants is the same as the source of DPOAE suppression tuning immaturities (i.e., cochlea) is not known. It is possible that the ABR and behavioral results simply reflect and preserve these peripheral limitations. However, it is also possible that they are reflecting limitations and immaturities that are only neural in nature and unrelated to cochlear immaturity observed in the current study.

B. Noncochlear factors

It is also possible that noncochlear factors contribute to some extent to the non-adult-like DPOAE suppression results observed in infants. Immature infant middle-ear function is one of these potential factors. Power-based measures of middle-ear function, such as acoustic reflectance, change with increasing age in infants (Keefe *et al.*, 2000). Infants tend to have lower middle-ear compliance and higher resistance, leading to less effective power transfer into the infant middle ear and possible attenuation of sound (Keefe *et al.*, 1993). Attenuated primary tone and suppressor tone levels in infant ears could produce age effects by altering the operating point of cochlear amplifier function in infants relative to adults. The cochlear amplifier functions in a level-dependent fashion; therefore, differing levels reaching the infant versus adult cochlea could selectively impact the shape of the infant DPOAE STC and suppression growth pattern.

The possibility of middle-ear involvement can be considered in various ways. Because the middle-ear system behaves linearly, and immaturity of function should produce a simple, relatively flat level attenuation, we have simulated its effect by manipulating primary tone level.

(1) If the immature middle ear produces a 10-dB attenuation of primary tones reaching the cochlea of infants, then presenting these tones 10 dB higher in infants should effectively equalize stimulus levels for these two groups. To this end, suppression growth at $f2=6000$ Hz was measured in neonates and adults using level compensation. Figure 6(a) shows that, even when primary tones are presented at a +10-dB level in infants, their slope of suppression growth for tones$<f2$ remains shallow and non-adultlike.

(2) It has been shown that infants generally require a higher suppressor level to initiate DPOAE suppression with low-frequency suppressor tones (Abdala and Chatterjee, 2003). Figure 6(b) illustrates this trend. Suppression thresholds (i.e., the suppressor level required to produce 2 dB of suppression) are plotted as a function of suppressor frequency for adults and two infant groups. Infant suppression thresholds are elevated for low-frequency suppressor tones; that is, it takes more suppressor level to just initiate DPOAE amplitude reduction in infants than in adults. This threshold elevation may be due to attenuation of suppressor levels produced by an immature middle ear, and could result in a reduced range of suppression growth available for measurement.

If the immature middle ear differentially influences which segment of the suppression growth function is measured to estimate slope, it might produce age effects unrelated to cochlear function. Does this abbreviated suppression growth function in infants produce an inaccurate estimate and bias data toward shallow slope values? In order to address this question, statistical correlations were conducted between suppression thresholds (at the four lowest suppressor frequencies) and slope of suppression growth for subjects in this study. If an elevated suppression threshold (and resulting abbreviated growth function) always produces shallow suppression growth, the correlations should elucidate this relationship. Of the four suppressor frequencies tested, only the lowest suppressor tone (3044 Hz) showed a modest correlation between suppression threshold and slope of suppression growth ($r=0.43$; $f=4.7855$; $p=0.04$). The more elevated the suppression threshold, the more shallow the slope of suppression growth. It could be argued that an immature middle ear produces elevated suppression thresholds, thus leading to the suppression growth values that are biased toward shallow slope.

Even with this modest correlation, data in Fig. 5 indicate that an elevated suppression threshold is not *sufficient* to reliably produce shallow measures of suppression

growth slope. Three-month-old infants show adult-like growth of suppression for most suppressor tones below $f2$ (as shown in Fig. 5), yet they have elevated thresholds at these same suppressor frequencies as shown by the data presented in Fig. 6(b). It is too simplistic to assume that suppression growth functions that include fewer data points will always produce shallow slope values. This was not shown to be the case. Nevertheless, it is likely that this factor, which may at its root be of middle-ear origin, influences the suppression growth slope to some extent for low suppressor frequencies.

(3) Finally, it is also possible that the segment of the suppression growth function available for measurement in infants is abbreviated because the infant noise floor is higher than the adult noise floor. This noise-floor elevation could also lead to artificially shallow suppression growth measurements in much the same way as middle-ear-based level attenuation. To test this possibility, statistical correlations were conducted between unsuppressed DPOAE amplitude and suppression growth slope. It is assumed that a robust, high-amplitude DPOAE would provide adequate noise immunity (SNR) so as to record a complete suppression growth function, comprised of several data points. The more complete and detailed the function, the less likely the contamination of slope estimates by noncochlear factors like the middle ear and elevated noise. No significant correlations were found between unsuppressed DPOAE amplitude and slope of suppression growth.

V. CONCLUSION

The non-adult-like DPOAE suppression observed in infants as old as 3 months of age is hypothesized to stem from subtle, postnatal immaturities in cochlear function. In the near future, we will expand this research to examine DPOAE suppression through 6 months of age. An intriguing question remains about the origin of these non-adult-like DPOAE suppression findings in 3-month-olds: are they due to immaturities in active cochlear mechanics or in the passive vibratory motion of the basilar membrane? To answer this question, it will be important to develop experimental paradigms that can distinguish between these two modalities of cochlear function and assess the relative contribution of each factor to the experimental results. Additionally, postmortem anatomical studies of the human fetal basilar membrane during various stages of development might shed light on this issue.

In addition to investigating the cochlear source of these DPOAE-based immaturities, it will be important to better define how noncochlear influences might contribute to DPOAE suppression results. Our preliminary analyses do not suggest that these factors greatly impacted the findings or impacted them in a clear and systematic fashion. The most effective way to address this question, however, is to concurrently measure DPOAE suppression and middle-ear function during development and examine the relationship between the two measures. This study is currently ongoing in our laboratory.

ACKNOWLEDGMENTS

The author would like to thank Leslie Visser-Dumont and Ellen Ma for their outstanding contribution to data collection and management. This experiment was funded by a grant from the National Institutes of Health, NIDCD R01 DC003552-6, and by the House Ear Institute.

APPENDIX: STANDARD DEVIATIONS OF MEAN SUPPRESSION GROWTH SLOPE

f_s (Hz)	P33	P35	P38	Term	1 Month	3 Month	Adult
3044	0.060		0.025	0.238		0.397	0.520
3620	0.284	0.248	0.386	0.362	0.107	0.448	0.321
4063	0.430	0.455	0.312	0.258	0.291	0.423	0.304
4560				0.365		0.266	0.230
5119	0.300	0.346	0.301	0.384	0.031	0.377	0.339
5350	0.223	0.413	0.367	0.308	0.374	0.437	0.208
5582				0.234		0.380	0.189
5746				0.239		0.328	0.248
5914	0.466	0.519	0.366	0.293	0.371	0.284	0.191
6087				0.194		0.199	0.181
6266	0.099	0.303	0.191	0.217	0.067	0.291	0.285
6449				0.148		0.274	0.090
6638	0.120	0.129	0.101	0.196	0.219	0.103	0.055
6932	0.129	0.125	0.227	0.080	0.057	0.034	0.037
7239				0.067		0.018	0.046

Abdala, C. (**1998**). ''A developmental study of DPOAE ($2f1-f2$) suppression in premature human neonates,'' Hear. Res. **121**, 125–138.

Abdala, C. (**2001**). ''DPOAE suppression tuning: Cochlear immaturity in premature neonates or auditory aging in normal-hearing adults?,'' J. Acoust. Soc. Am. **110**(6), 3155–3162.

Abdala, C. (**2001**). ''Maturation of the human cochlear amplifier: Distortion product otoacoustic emission suppression tuning curves recorded at low and high primary tone levels,'' J. Acoust. Soc. Am. **110**, 1465–1476.

Abdala, C. (**2003**). ''A longitudinal study of distortion product otoacoustic emission ipsilateral suppression and input/output characteristics in human neonates,'' J. Acoust. Soc. Am. **114**, 3239–3250.

Abdala, C., and Chatterjee, M. (**2003**). ''Maturation of cochlear nonlinearity as measured by distortion product otoacoustic emission suppression growth in humans,'' J. Acoust. Soc. Am. **114**, 932–943.

Abdala, C., and Folsom, R. (**1995**). '' The development of frequency resolution in human adults and infants as revealed by the ABR recorded with notched-noise masking.'' J. Acoust. Soc. Am. **98**(2), Pt. 1:921–930.

Abdala, C., Sininger, Y., Ekelid, M., and Zeng, F.-G. (**1996**). ''Distortion product otoacoustic emission suppression tuning curves in human adults and neonates,'' Hear. Res. **98**, 38–53.

Arjmand, E., Harris, D., and Dallos, P. (**1988**). ''Developmental changes in frequency mapping of the gerbil cochlea: comparison of two cochlear locations,'' Hear. Res. **32**, 93–96.

Bredberg, G. (**1968**). ''Cellular pattern and nerve supply of the human organ of Corti,'' Acta Oto-Laryngol., Suppl. **236**, 1–135.

Cooper, N. (**1996**). ''Two-tone suppression in cochlear mechanics,'' J. Acoust. Soc. Am. **99**, 3087–3098.

Folsom, R., and Wynne, M. (**1987**). ''Auditory brainstem responses from human adults and infants: Wave-V tuning curves,'' J. Acoust. Soc. Am. **81**, 412–417.

Keefe, D., Bulen, J., Hoberg Arehart, K., and Burns, E. (**1993**). ''Ear-canal impedance and reflection coefficient in human infants and adults,'' J. Acoust. Soc. Am. **94**, 2617–2638.

Keefe, D., Folsom, R., Gorga, M., Vohr, B., Bulen, J., and Norton, S. (**2000**). ''Identification of neonatal hearing impairment: Ear-canal measurements of acoustic admittance and reflectance in neonates,'' Ear Hear. **21**, 443–461.

Lavigne-Rebillard, M., and Pujol, R. (**1987**). ''Surface aspects of the developing human organ of Corti,'' Acta Oto-Laryngol., Suppl. **436**, 43–50.

Lindeman, H., Ades, H., Bredberg, G., and Engstro, H. (**1971**). ''The sensory hairs and the tectorial membrane in the development of the cat s organ of Corti. A scanning electron microscopic study,'' Acta Oto-Laryngol. **72**, 229–242.
Mills, D., and Rubel, E. (**1996**). ''Development of the cochlear amplifier,'' J. Acoust. Soc. Am. **100**, 428–441.
Mills, D., and Rubel, E. (**1998**). ''Development of the base of the cochlea: Place code shift in the gerbil,'' Hear. Res. **122**, 82–96.
Oxenham, A., and Plack, C. (**1997**). ''A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing,'' J. Acoust. Soc. Am. **101**, 3666–3675.
Rhode, W., and Cooper, N. (**1993**). ''Two-tone suppression and distortion production on the basilar membrane in the hook region of cat and guinea pig cochleae,'' Hear. Res. **66**, 31–45.
Spetner, N., and Olsho, L. (**1990**). ''Auditory frequency resolution in human infancy,'' Child Dev. **61**, 632–652.
Wada, T. (**1923**). ''Anatomical and physiological studies on the growth of the inner ear of the albino rat,'' Wistlar Inst. Anat. & Biol. Memoirs, 10.

Human temporal auditory acuity as assessed by envelope following responses[a)]

David W. Purcell[b)] and Sasha M. John
The Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Ontario, M6A 2E1 Canada

Bruce A. Schneider
Department of Psychology, University of Toronto at Mississauga, Mississauga, Ontario, L5L 1C6 Canada

Terence W. Picton
The Rotman Research Institute, Baycrest Centre for Geriatric Care, Toronto, Ontario M6A 2E1 Canada

(Received 29 March 2004; revised 31 July 2004; accepted 4 August 2004)

Temporal auditory acuity, the ability to discriminate rapid changes in the envelope of a sound, is essential for speech comprehension. Human envelope following responses (EFRs) recorded from scalp electrodes were evaluated as an objective measurement of temporal processing in the auditory nervous system. The temporal auditory acuity of older and younger participants was measured behaviorally using both gap and modulation detection tasks. These findings were then related to EFRs evoked by white noise that was amplitude modulated (25% modulation depth) with a sweep of modulation frequencies from 20 to 600 Hz. The frequency at which the EFR was no longer detectable was significantly correlated with behavioral measurements of gap detection ($r=-0.43$), and with the maximum perceptible modulation frequency ($r=0.72$). The EFR techniques investigated here might be developed into a clinically useful objective estimate of temporal auditory acuity for subjects who cannot provide reliable behavioral responses. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1798354]

PACS numbers: 43.64.Ri, 43.66.Mk [BLM] Pages: 3581–3593

I. INTRODUCTION

Prosody (rhythmic and stress variations in the amplitude envelope of a speech signal) is crucial to speech intelligibility. The envelopes of speech signals contain enough information, even without their spectral content, to permit a listener to identify tokens (Van Tasell *et al.*, 1987; Rosen, 1992). Shannon *et al.* (1995) showed that a listener can recognize words, phrases, and sentences using amplitude envelopes of the speech signal in each of four frequency regions to modulate corresponding bandlimited noises. They found that speech intelligibility remained high when they recombined these four bands of amplitude-modulated noise. The envelopes of the speech signal in different spectral regions therefore play an important role in speech understanding.

Anything that reduces a listener's sensitivity to these amplitude modulations (auditory temporal acuity) is likely to lead to problems in speech understanding. Temporal acuity deteriorates in old age (Schneider *et al.*, 1998; Snell and Frisina, 2000; Gordon-Salant and Fitzgibbons, 1999), and this deficit may explain many age-related hearing problems (Schneider, 1997; Schneider *et al.*, 2002; Snell *et al.*, 2002). Certain disease processes can also result in decreased temporal acuity. For example, patients with auditory neuropathy (Zeng *et al.*, 1999, 2001) or multiple sclerosis (Rappaport *et al.*, 1994) have difficulty perceiving rapid changes in auditory input. Temporal auditory acuity can be abnormal in sensorineural hearing loss (e.g., Fitzgibbons and Wightman, 1982; Formby, 1987), although if one adjusts the intensity and bandwidth of the sounds to compensate for a subject's hearing loss, there is no impairment (Moore, 1995; Grose *et al.*, 2001). Decreased sensitivity to amplitude fluctuations may also contribute to developmental language disorders (Benasich and Tallal, 2002), but the nature of this relationship is not clear (Bishop *et al.*, 1999; Amitay *et al.*, 2002). Given the importance of envelope cues to speech intelligibility, tests to evaluate temporal acuity are essential.

The two most prevalent subjective ways to measure a listener's sensitivity to envelope fluctuations are to determine how sensitive they are to sinusoidal modulations in the amplitude of a sound (the temporal modulation transfer function, TMTF), or how sensitive they are to a gap in an otherwise continuous sound or between two sounds (gap detection). In the first procedure, a sound (often a band of noise) is sinusoidally modulated at a given frequency, with the depth of modulation (the ratio of the difference between the peak and trough of the modulated wave and their sum) systematically adjusted until the listener can no longer detect the modulation. Intensity adjustments are typically made to the modulated stimulus to control for intensity differences between modulated and unmodulated stimuli (Viemeister, 1979). The frequency of modulation is then changed and the procedure repeated (Viemeister and Plack, 1993).

In young, normal-hearing human listeners, the behavioral TMTF for white noise can be modeled as a single-pole low-pass filter with a 3-dB cutoff near 55 Hz (Viemeister,

[a)]Part of these results was presented at the International Evoked Response Audiometry Study Group meeting in Tenerife, Spain, June 2003, and at the XI International Symposium on Audiological Medicine, Padua, Italy, October 2002.

[b)]Author to whom correspondence should be addressed. Electronic mail: d.purcell@utoronto.ca Currently affiliated with the Department of Psychology, Queen's University, Kingston, Ontario, K7L 3N6 Canada.

1979). This means that sensitivity to sinusoidal variations in the amplitude envelope remains roughly constant so long as the frequency of modulation is less than approximately 55 Hz, and then begins to decline steadily. However, the rate of decline is slow enough that listeners can normally detect a 25% amplitude modulation up to frequencies of about 500 Hz. The highest frequency at which a subject can discriminate 100% amplitude modulation is approximately 2.5 kHz, if the stimulus is designed to minimize high-frequency intensity effects (Viemeister and Plack, 1993).

Gap detection tests measure the shortest perceptible duration of a gap in an ongoing sound, or between two sounds. In young, normal-hearing human listeners, gap detection thresholds are about 1 ms for white noise. For pure tones the gap threshold increases with decreasing frequency from 2.3 ms at 8000 Hz to 22.5 ms at 200 Hz (Shailer and Moore, 1983). Since two cycles of a 100% modulated sound roughly resemble a gap between two short sounds (sound off, sound on, gap, sound on, sound off), it might be expected that the two tasks engage somewhat similar mechanisms. Note, however, that 100% modulation of a sound is similar to a continuous string of gaps between adjacent, identical, short-duration sounds. Hence, the modulation detection task allows for integration of information about periodic level changes over a time period that is much longer than that involved in detecting a gap between two short sounds. Accordingly, a reduced correspondence would be expected between modulation detection and detection of a gap in an otherwise continuous stimulus, or detection of a gap between two spectrally different sounds. Indeed, in the latter case, gap detection thresholds can be as much as ten times longer than when the markers are the same on both sides of the gap (Phillips *et al.*, 1997). In the former case, auditory processes (such as neural adaptation to the first marker) are likely to affect gap duration thresholds.

These two measures of temporal acuity are also likely to be related to different aspects of speech recognition. Phonemic contrasts can be cued by differences in the size of gaps between bursts that signal the presence or absence of a stop consonant (e.g., *slit* versus *split*). Hence, gap detection acuity is likely to be most relevant at the segmental or phonemic level. The TMTF, on the other hand, limits the processing of amplitude modulations at the syllabic or suprasegmental level, and is therefore likely to be more directly relevant to speech prosody. However, to the extent that the detection of a gap between two identical short-duration sounds can be conceived of as two cycles of an amplitude modulated envelope, knowledge of the TMTF could also be informative about the ability of a listener to detect a gap.

Most of the information that is known about temporal acuity has been collected using behavioral paradigms. An objective electrophysiological measure of human temporal auditory acuity would be very helpful in testing subjects who are unable to give accurate behavioral responses (e.g., young children, individuals with language impairment, or patients with cognitive difficulties). The present study addresses this need. In the research reported here, auditory steady-state response methods (Picton *et al.*, 2003a) were adapted to objectively estimate the TMTF. The results from this electrophysiological method were compared to behavioral measures of the TMTF.

The spectrum of the electrophysiological response to an amplitude-modulated signal contains energy at the modulation frequency. Thus, if a 1-kHz tone is modulated at a rate of 80 Hz, the spectrum of the response will reveal energy at 80 Hz and at overtones of 80 Hz (due to nonlinearities in the auditory system), even though there is no energy in the stimulus at the modulating frequency. The strength of the 80-Hz component can be taken as an index of the brain's sensitivity to an 80-Hz modulation of the stimulus, and the modulation depth at which the 80-Hz component is first detected can be considered as a measure of the modulation threshold at that frequency.

Auditory steady-state responses must be recorded across a wide band of modulation frequencies in order to obtain an electrophysiological representation of the TMTF. While multiple separate responses can be sequentially recorded at selected modulation frequencies (Stapells *et al.*, 1984; Rees *et al.*, 1986; Sapsford *et al.*, 1996), a more efficient approach might be to use a continuous sweep of modulation frequency through the band of interest. Sweep techniques were originally employed for the measurement of visual evoked potentials (Regan, 1966, 1989; Norcia *et al.*, 1989). They have also been used for auditory stimuli, where both intensity (Picton *et al.*, 1984; Linden *et al.*, 1985; Rodriguez *et al.*, 1986), and modulation rate (Linden *et al.*, 1985) have been varied. In the sweep technique, some aspect of the stimulus is continuously changed or ''swept'' across a specified range. Since the stimulus changes over time, the instantaneous amplitude and phase of the response is continuously changing. Referring to these as ''auditory steady-state responses'' is therefore not appropriate. However, since the modulation frequency of the stimuli used here was fast relative to the rate of change of modulation frequency, the responses examined in this paper are much more akin to auditory steady-state responses than to transient evoked responses. For simplicity, they will henceforth be referred to with the more general term of envelope following response (EFR).

This paper describes the results of a set of experiments examining how well the human EFR can be recorded by sweeping through a range of modulation frequencies. Several preliminary experiments evaluated the effects of different stimuli and subject states on the EFRs recorded using the sweep technique. The results of these experiments helped shape the experimental parameters for the main experiment, which compared objective EFR measurements of the TMTF and subjective measurements of temporal acuity in younger and older subject groups.

II. METHODS

A. Subjects

Two groups of adult subjects participated in this study. The younger group ($n=25$, 20 females) had normal thresholds of hearing ($\leqslant$20 dB HL) for octave frequencies from 500 to 4000 Hz, and varied in age from 18 to 43 years. Fourteen of these subjects (11 females) participated in the main experiment comparing younger and older subjects. The

older group ($n=13$, 8 females) ranged in age from 60 to 78 years. They had mildly elevated thresholds relative to those of young normal-hearing adults, but did not use hearing aids. On average, their thresholds were 20.8 dB HL (over the range from 500 to 4000 Hz) with standard deviation (SD) 12.6 dB. Only three individuals had any thresholds >35 dB HL. For the preliminary experiments, individuals from the younger group were employed. The main experiment, involving both EFRs and behavioral measurements, used both subject groups.

During the main experiment, it was found that the background noise in the older group's electroencephalogram (EEG) was approximately 1.5 times that in the younger group (probably due to increased muscle activity associated with wakefulness; older subjects tended to have more difficulty sleeping for the duration of the experiment). In order to approximately equalize the background EEG-noise levels to which the averaged EFRs are compared, the younger group attended one measurement session, whereas the older group attended two. Therefore, the older group had twice as many repetitions in the averaged EEG data (and the noise levels were reduced by $\sqrt{2}$).

B. Auditory stimuli

Acoustic stimuli were generated using a MASTER research system (John and Picton, 2000; see www.mastersystem.ca) running the SWEEP_V1 software module which permits using external sound files. Amplitude modulation was applied to either a uniform white-noise carrier or a 1-kHz tone. The depth of modulation was 100%, 50%, or 25% in the preliminary experiments, and 25% in the main experiment. Each sound file consisted of a single stimulus ''sweep'' containing 30 1.024-s epochs. For the first 15 epochs, the modulation rate increased linearly from the minimum to the maximum instantaneous frequency. In the second half of the sweep, the modulation rate decreased from maximum to minimum in symmetry with the first half. In some test conditions, the modulation rate was fixed in frequency throughout the 30.72-s sweep. The sweeps were repeated without any pause between them, since the sounds were designed so that there was no discontinuity at the transition between the end of one sweep and the beginning of the next.

Digital-to-analog conversion was performed using a National Instruments 6052E input/output board at 32 kHz with 16-bit precision. The electrical stimulus amplitude was adjusted using a Grason-Stadler model 16 audiometer prior to transduction by a pair of Etymotic ER-2 transducers with a flat frequency response at the eardrum up to 10 kHz. Generally, a single channel was used to present the stimulus monaurally using a foam ear insert. A Knowles DB-100 Zwislocki coupler and Brüel & Kjær sound-level meter were used to calibrate the stimulus.

C. Recordings

Measurements were performed in an Industrial Acoustics Company (IAC) sound-insulated room. During the experiments, the participants sat in a comfortable chair and watched a silent subtitled movie. The chair could be reclined for protocols wherein the subjects slept. The experiments generally took between 1 and 2 h. Gold-plated Grass electrodes were used to record the EEG from the vertex (Cz), and just below the hairline at the posterior midline of the neck (reference), with a ground on the collarbone. All electrode impedances were below 8 kOhm at 10 Hz. Responses were amplified with gain 10 000 and bandpass filtered 1–300 Hz or 1–1000 Hz using a Grass P50 battery-powered amplifier. The acquisition board applied a further gain of 5, and data were digitized at 16-bit resolution, and stored on disk at a rate of 1 (preliminary studies) or 2 (main experiment) kHz.

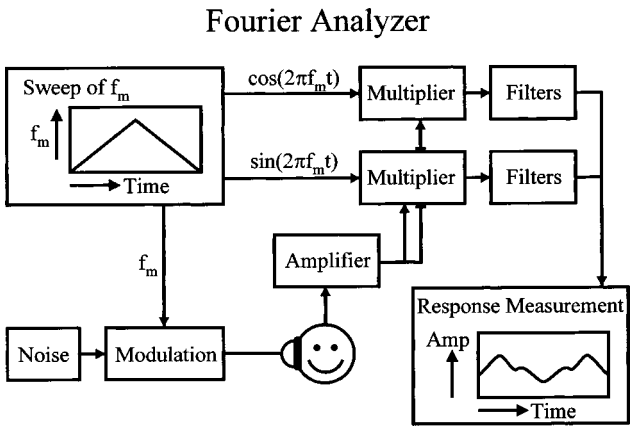


FIG. 1. Fourier analyzer schematic. Diagram showing how modulation rate, f_m, is swept over time, and how the instantaneous frequency is used to modulate the stimulus and extract the response in a Fourier analyzer.

D. EFR analysis

Data analysis was performed offline after the experimental recordings were completed. Data from each sweep were synchronously averaged in the time domain. The number of sweeps averaged was between 50 and 100. A noise rejection algorithm was employed to exclude 1.024-s epochs of data if a threshold noise level was exceeded in a broad frequency band containing the response frequency. Prior to averaging, the mean and SD of the noise in all epochs were estimated. The noise rejection threshold for any given epoch was then set as 1.5 SDs above the mean noise. Given that a sweep was divided into 30 1.024-s epochs, each ''epoch slot'' could contain a different number of epochs in the average. Typically, fewer than five epochs would be rejected from each slot when 50 sweeps were averaged.

The EFR was extracted from the average EEG sweep using a Fourier analyzer (FA) with orthogonal reference sinusoids that matched the instantaneous frequency of the stimulus (Regan, 1989). The complex outputs of the analyzer were filtered using simple 1.024-s boxcar filters (applied twice per multiplier) as shown in Fig. 1. The second half of the analyzed sweep was subsequently vector averaged with the first half of the sweep. This ''fold-and-average'' operation was justified since the modulation frequency was reversed between the two halves of the sweep. Given the 30.72-s sweep length, and the filter windows of 1.024 s, anticipated physiological delays (which could have been up to 60 ms) were neglected during the averaging operation.

The instantaneous response amplitude at each frequency of interest was compared to a noise estimate, derived from the discrete Fourier transform (DFT) of the average sweep folded in the time domain, in order to determine whether a given response signal was statistically different from the background EEG-noise level estimate. This noise estimate was calculated using ± 60 DFT frequency bins (± 3.9 Hz) surrounding the instantaneous modulation rate of the swept stimulus. Since the FA passes more noise than the DFT (due to its wider bandwidth), a scaling factor was determined using simulated noise. The EEG noise in FA estimates effectively had an amplitude 3.24 times greater than in the DFT estimates. An F ratio was then employed to test whether the FA response amplitude was significantly different from the scaled DFT EEG-noise estimate (John and Picton, 2000).

In the main experiment, the highest frequency at which the EFR could be reliably measured was estimated for comparison with the behavioral thresholds. The measured EFR amplitude never reaches zero due to noise passing through the FA. The EFR amplitude was therefore phase weighted using an expected phase. This phase-weighted amplitude can reach zero, and a threshold can therefore be determined.

The phase-weighted EFR amplitude was calculated from the projection of the complex FA output onto an expected phase (Picton *et al.*, 2001)

$$A_w = A \cos(\theta - \theta_e), \tag{1}$$

where A_w is the phase-weighted amplitude, A is the FA output amplitude, θ_e is the expected phase, and θ is the FA output phase. The analysis assumes that a single EFR source with a modulation-frequency-independent delay is dominating the measured EFR, and that the phase therefore changes linearly with the instantaneous modulation frequency. The expected phase was determined by linear regression of the phase versus frequency data for frequencies between 100 Hz and the highest frequency at which a response was judged significantly different from background EEG noise using an F ratio. For one subject the regression was based in the 40-Hz range because this subject had no significant responses at frequencies over 100 Hz. The maximum frequency at which an EFR was recognizable was then taken as the highest frequency at which the phase-weighted amplitude was significantly different from zero using a t test.

Apparent latencies were calculated using linear regression as the slope of the phase by frequency plot over a range of frequencies (Regan, 1966, 1989). No correction was made for the small acoustic delay of approximately 0.9 ms from the Etymotic ER-2 transducer to the ear canal through a short sound tube (292 mm, including foam insert).

E. Behavioral measurements

Two behavioral measurements of temporal auditory acuity were employed. A two-alternative, forced-choice paradigm was used, and thresholds were estimated using a procedure derived from PEST (Taylor and Creelman, 1967). The initial increase in difficulty was one step per correct answer, until the first error. Subsequently, the difficulty was increased one step every three correct answers, and decreased one step for every error. A complete test was composed of 105 trials, and took approximately 10 to 20 min to finish depending on the task and the individual. The behavioral threshold for a given test was determined as the average of all available reversals, excluding the first.

The threshold for detecting 25% amplitude modulation of a white-noise carrier was found by asking the participant to choose which of two randomized 1-s sounds was modulated. The nonmodulated sound was of equal duration and power (Viemeister, 1979). For most individuals, the modulation frequency started at 50 Hz and increased in equal steps of 50 Hz. Some of the participants had a relatively low threshold, and were tested beginning at 10 Hz, and increasing in equal steps of 10 Hz. Sounds were presented at 60 dB SPL, but could be increased for the comfort of listeners in the older group. This should have no effect on the detection threshold at these sensation levels (Bacon and Viemeister, 1985).

The threshold for detecting a gap between brief white-noise markers was found by asking the participant to choose which of two randomized sounds contained a gap. The gap target was created with two 5-ms markers with Gaussian-shaped rise and fall times (SD 0.167 ms) separated by a short interval. The nontarget sound was white noise of equal duration and total energy, and had similar rise and fall envelopes (Schneider and Hamstra, 1999). The gap duration started at 20 ms, and decreased in equal steps of 1 ms to a minimum of 1 ms. For training purposes, larger gap durations were available for subjects who initially had trouble with 20 ms. Sounds were presented at 75 pSPL, but could be increased if requested by listeners in the older group. Changes in sensation level should have no effect on the detection threshold at these levels (Schneider *et al.*, 1998; Schneider and Hamstra, 1999).

All participants performed training trials until it was clear that they understood how to respond. For each task, training was typically about 10 trials. The younger subject group was generally familiar with behavioral testing, and performed the modulation and gap detection tasks once (105 trials each). The older group performed both tasks twice, and thresholds were determined from the results of the second tests.

F. Statistical analyses

Repeated-measures analyses of variance (ANOVAs) were performed using Greenhouse–Geisser corrections. In the main experiment, a correlation analysis was performed between the behavioral thresholds, and values derived from the EFR measurements. This was to determine if physiological measurements could serve as useful indices of the behavioral responses. The behavioral thresholds were (a) the maximum frequency (Hz) at which 25% amplitude modulation of noise could be detected; (b) the minimum gap (ms) that could be detected between brief noise markers; (c) the reciprocal of this gap detection threshold (Hz); (d) the mean pure-tone hearing threshold (dB HL) from 500 to 4000 Hz. The EFR measurements were (a) the highest frequency (Hz) at which the phase-weighted amplitude of the EFR was statistically different from noise; (b) the apparent latency (ms) of the EFR in the frequency band used to calculate the expected

phase; (c) the frequency (Hz) at which amplitude peaked between 30 and 50 Hz; (d) the amplitude (nV) of the peak frequency between 30 and 50 Hz; (e) the apparent latency (ms) between 30 and 50 Hz; (f) the mean amplitudes (nV) from 30–50 Hz, 80–100 Hz, 100–200 Hz, 200–300 Hz, 300–400 Hz, and 400–500 Hz; and (g) the optimal cutoff frequency of a first-order low-pass filter model fit to the amplitude data above 80 Hz, using a fixed gain of mean amplitude 80–100 Hz and a least-squares cost function. The correlations between all of these measures and age (in years) were also investigated. Two-tailed separate variance *t* tests were employed to assess the null hypotheses that the group means were equal for each variable above.

Similar *t* tests were also used to evaluate at what frequency the phase-weighted EFR amplitude was no longer significantly different from zero. In this implementation of the FA, the output was determined from data averaged in the time domain. Therefore, multiple signal estimates were not available at a given response frequency, and the variance of the amplitude estimate had to be calculated indirectly for use in the *t* tests. The variance across ± 60 DFT EEG-noise bins near the response frequency was employed instead, after appropriate scaling. Simulations showed that the variance of multiple estimates from the FA is similar to the DFT variance across multiple noise bins, after scaling by the effective bandwidths of the two analyzers.

III. RESULTS

A. Preliminary experiments

1. Effects of carrier and level

The purpose of the first experiment was to evaluate the influence of stimulus carrier and level on the response. Three different amplitude-modulated stimuli were presented in a balanced design: white noise presented at 60 and 50 dB SPL, and a 1-kHz tone presented at 60 dB SPL. Figure 2(A) plots the average amplitude of the frequency component in the evoked response corresponding to the modulation frequency in the stimulus. These average amplitudes represent the grand vector average of 58 sweeps from each of five waking individuals in response to 100% amplitude modulation at frequencies from 35 to 100 Hz. In the band 35 to 60 Hz, the modulated noise elicited a larger response than the modulated pure tone. Since these averages were determined during a sweep over the modulation frequencies employed, also shown are the responses to noise at 60 dB SPL that was modulated at fixed rates of 40 and 80 Hz (squares), and analyzed using the DFT. The amplitudes measured during the sweep are the same as those obtained using a fixed modulation frequency.

Also shown in Fig. 2(B) is the average phase of the frequency component in the evoked response corresponding to the modulation frequency. All responses in the grand average were statistically different from the background EEG noise, with the exception of the region between 65 and 75 Hz. This loss of significance corresponds to the reduced response amplitude and rapid phase changes observed near 70 Hz. Apparent latencies were calculated for the changing modulation rate stimuli. In the band 38 to 60 Hz, the latencies were 22.4, 22.7, and 29.9 ms for the 60- and 50-dB SPL noise, and the 60-dB SPL 1-kHz tone, respectively. Peak amplitudes between 35 and 50 Hz were significantly larger at 60 dB SPL for the noise than for the 1-kHz tone ($F=5.96$; $df=2, 8$; $p<0.03$; *post hoc t* test).

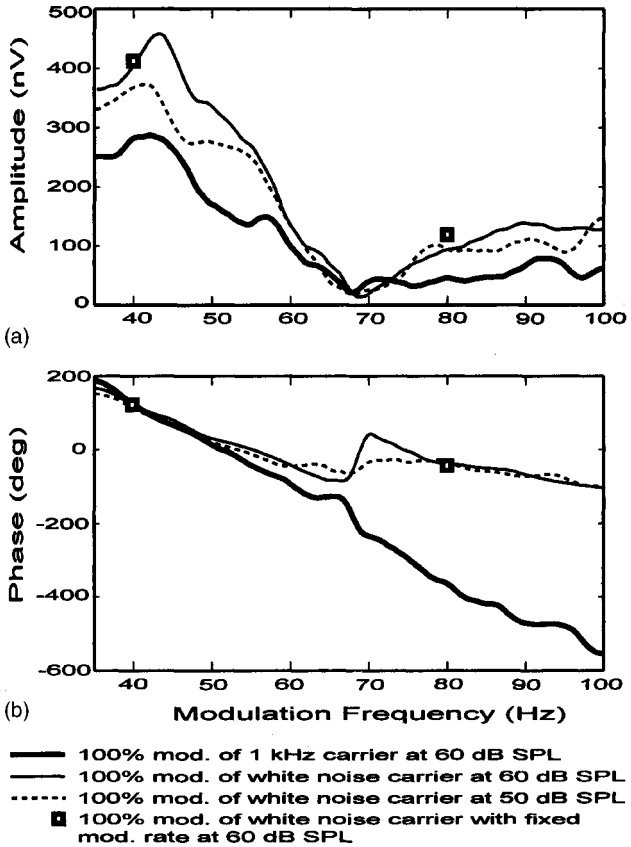


FIG. 2. Grand average of responses from five waking individuals. Panel (A) shows the response amplitude, and panel (B) the response phase as the rate was swept from 35 to 100 Hz for 100% amplitude modulation. The continuous thick line is the response to a 1-kHz pure-tone carrier presented at 60 dB SPL. The continuous thin and dashed lines are the responses to white-noise carriers presented at 60 and 50 dB SPL, respectively. The open squares show the responses to noise modulated at fixed rates of 40 and 80 Hz, and presented at 60 dB SPL.

2. Comparing swept and fixed modulation frequencies

The purpose of the second experiment was to confirm that the sweep stimulus elicited the same response as fixed modulation rate stimuli, including at higher modulation rates. The grand vector average of 56 sweeps from each of 10 waking individuals is shown in Fig. 3. The stimulus was 60 dB SPL 100% amplitude modulated noise presented in a balanced design for the two ranges 20 to 100 Hz, and 70 to 200 Hz. Also shown are the DFT results for fixed modulation rates of 40, 80, and 160 Hz. Responses were highly significant with the exception of about a 5-Hz band centered on 27 Hz, where there was a null in the amplitude response. The apparent latency was 20.9 ms for the band 35 to 55 Hz, and 8.6 ms from 80 to 190 Hz.

A repeated-measures ANOVA of the EFR amplitude at the frequencies 40, 80, and 160 Hz with factors of frequency and modulation type (fixed or swept) showed a significant effect of modulation frequency ($F=115.385$; $df=2, 18$; p

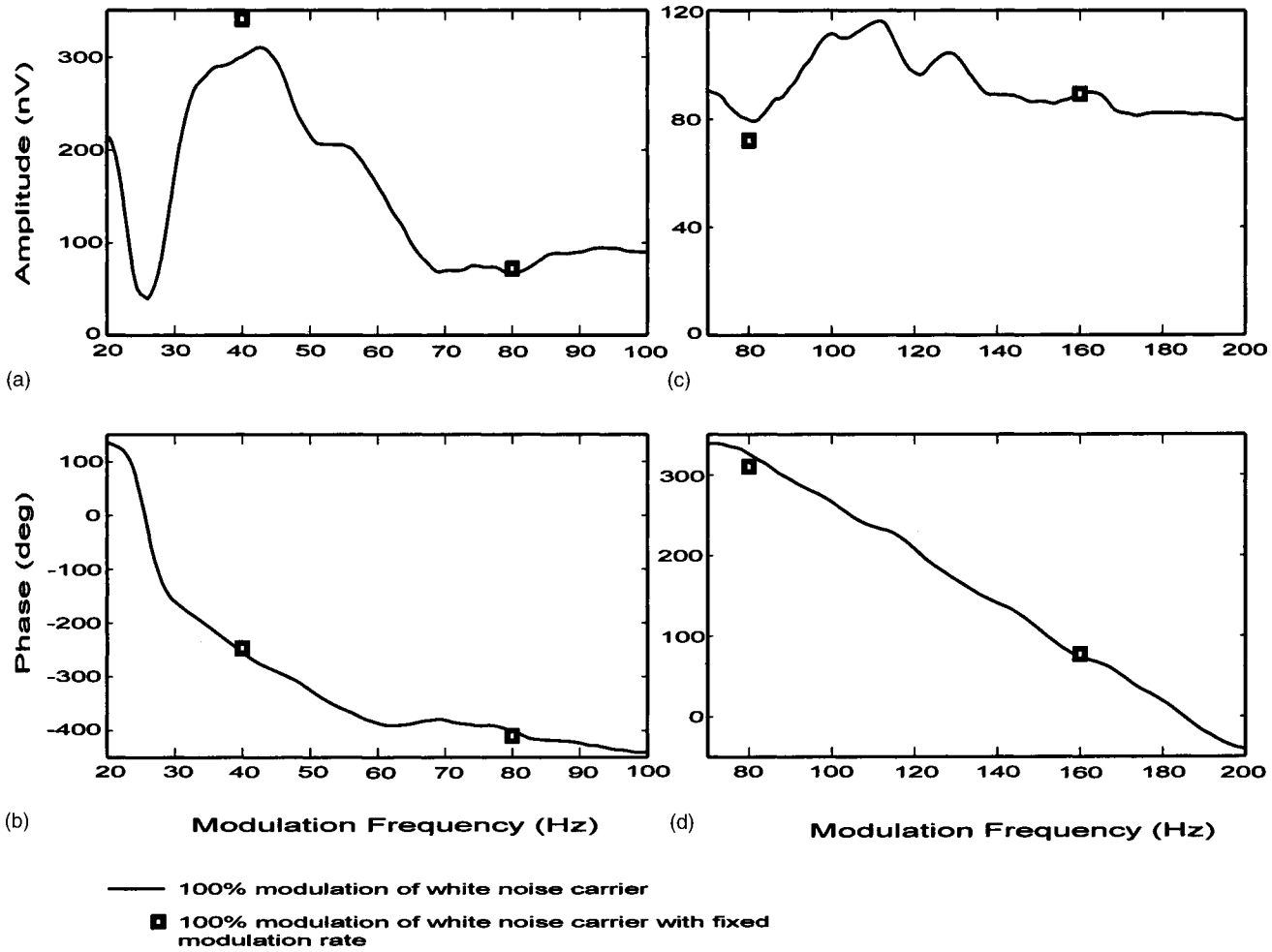


FIG. 3. Grand average of responses from 10 waking individuals. Panels (A) and (B) show the amplitude and phase responses for the range 20 to 100 Hz for 100% amplitude modulation of a white-noise carrier presented at 60 dB SPL. Similarly, panels (C) and (D) show the responses for the range 70 to 200 Hz. The open squares show the responses to noise modulated at fixed rates of 40, 80, and 160 Hz and presented at 60 dB SPL.

<0.001), but no effect of modulation type and no interaction. The fixed-rate response amplitude at 40 Hz was a little larger than the swept-rate measurement, but the difference was not significant using a *t* test.

In two subjects, responses were examined near 70 Hz, where the amplitude was low, using swept modulation frequencies. Five measurements were made with fixed modulation rates distributed within a few hertz of 70 Hz, and responses were analyzed using the DFT. There was no clear difference between swept and fixed modulation responses, indicating an amplitude minimum near 70 Hz that was independent of the recording technique.

3. Electrical artifacts

The purpose of the third experiment was to verify that there were no electrical artifacts influencing the EFR measurements. The EFR was obtained in a balanced design with amplitude-modulated noise either delivered to the ear, or delivered to a Zwislocki coupler on the subject's shoulder. Responses were recorded to a variety of sweeps (20–100 and 70–200 Hz modulated 100% at 60 dB SPL, and 100–700 Hz modulated 50% at 65 dB SPL) in four subjects. All subjects had highly significant responses when the stimulus was in the ear. When the stimulus was delivered to the Zwislocki coupler, the EFR was significant with an average incidence during the sweep across the different subjects and conditions of 5.4% (range 0%–11%). This result was expected since a statistical criterion of $p<0.05$ was used to evaluate whether a given EFR amplitude was significantly different from an estimate of the background EEG noise. When no stimulus was delivered to the ear, a false positive detection would be expected at the EFR frequency 5/100 times, which is close to the 5.4% found here.

4. Effects of sleep

The purpose of the fourth experiment was to investigate the effects of sleep on the EFR. Three individuals were evaluated both asleep and awake using 25% amplitude modulated noise presented at 60 dB SPL for the ranges 20 to 100 Hz, and 100 to 600 Hz. Data from a single individual are shown in Fig. 4. For the 20- to 100-Hz range while awake, the responses were highly significant except near the amplitude minimum of 70 Hz. In sleep, the responses remained highly significant except below 33 Hz, and near the broader amplitude minimum at 66 Hz. For the 100- to 600-Hz range, both waking and sleeping responses became insignificant after 500 Hz. There was a deep valley in the amplitude response near 27 Hz during both sleep and wakefulness.

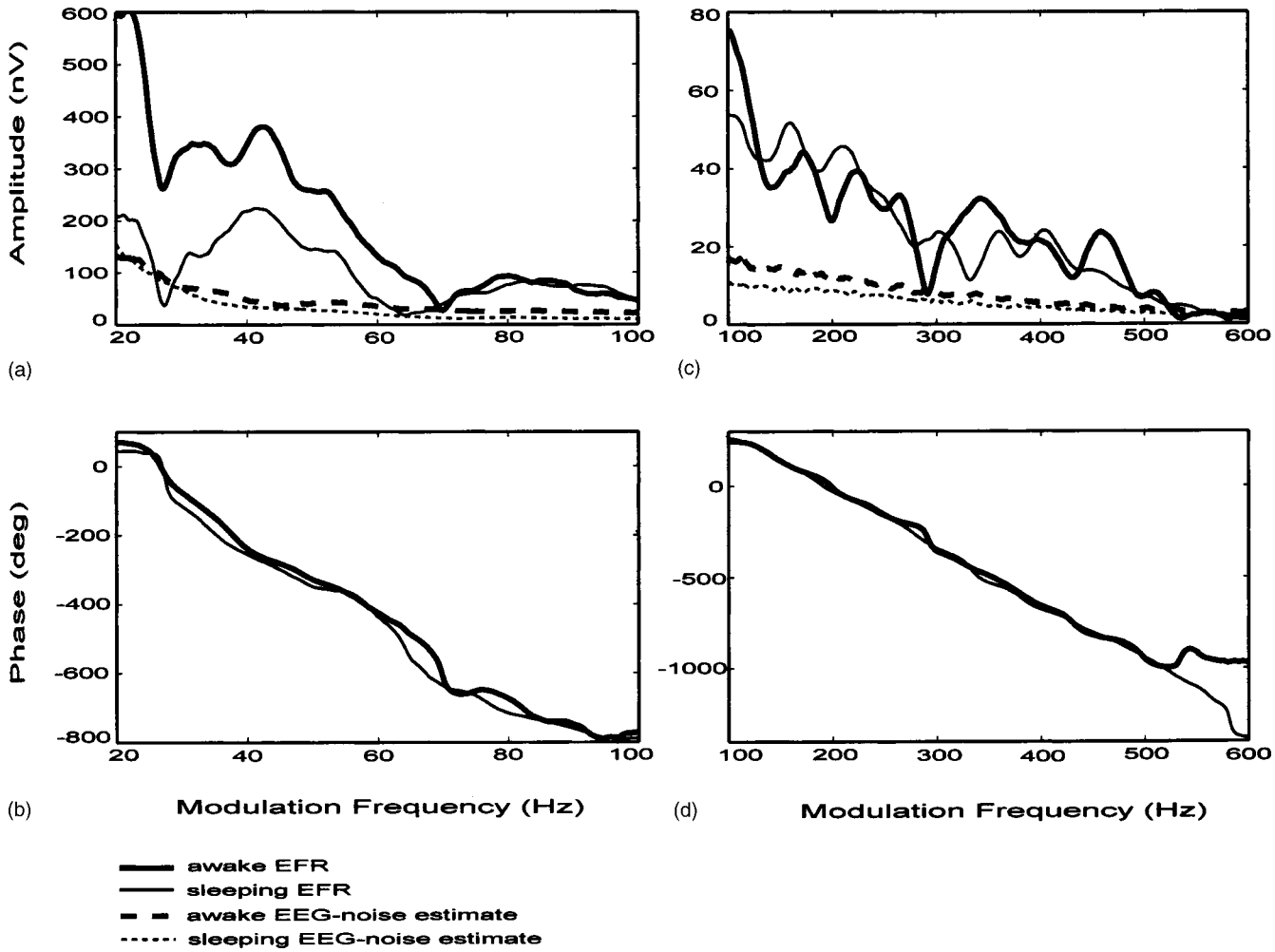


FIG. 4. Individual response during waking and sleeping. Panels (A) and (B) show the amplitude and phase for the range 20 to 100 Hz from a single individual in response to 25% amplitude modulated white noise presented at 60 dB SPL. Similarly, panels (C) and (D) are for a 60-dB SPL stimulus 25% modulated from 100 to 600 Hz. The continuous thick line indicates the responses when the individual was awake, and the continuous thin line shows the responses during sleep on a different date. The dashed thick and thin lines are the EEG-noise estimates at the response frequency during wakefulness and sleep, respectively.

The peak amplitude of the EFR was significantly larger during wakefulness than during sleep for all three subjects (sign test between 35 to 50 Hz, $p<0.000\,01$). At higher frequencies, there was no significant difference in response amplitude between sleep and wakefulness for two of three subjects (including the subject in Fig. 4). For the third subject, the amplitude during wakefulness was on average 4 nV larger than during sleep. The mean EEG-noise estimate was smaller during sleep for all three subjects.

For the subject shown in Fig. 4, the apparent latencies in the band 35 to 50 Hz were 27.1 and 31.2 ms during sleep and wakefulness. The other two subjects had apparent latencies of 12.0 and 25.4 ms while asleep, and 23.6 and 28.7 ms while awake. In the frequency band 100 to 400 Hz, the subject shown in Fig. 4 had apparent latencies of 9.0 and 8.8 ms while asleep and awake. For the other two subjects, the apparent latencies were correspondingly 9.3 and 9.4 ms during sleep, and 9.5 and 9.1 ms during wakefulness.

B. Main experiment

Based on the preliminary data, it was decided to use white noise at 60 dB SPL as the stimulus, since this gave the highest signal-to-noise ratio (SNR) of the three stimuli tested in the first experiment. The sweep approach was justified since there were no significant differences between swept and fixed rates of modulation. Finally, it was decided to have the subjects sleep through the recording since the residual EEG noise was lower during sleep, and since the response at the higher frequencies was not affected by the subject state. The stimulus was 25% amplitude-modulated white noise in two sweep ranges: from 20 to 100 Hz, and from 100 to 600 Hz. The reason for using 25% modulation depth was that this depth evoked responses that were typically statistically significant only up to about 500 Hz. A larger modulation depth would have increased this frequency, and thereby increased the maximum modulation rate required in the stimulus and the test time needed. The stimulus for the behavioral modulation detection task also requires less adjustment for intensity effects at lower frequencies (Viemeister, 1979). By employing 25% modulation, the detection threshold was at a frequency low enough to minimize intensity artifacts.

Figure 5 shows two grand vector averages: one of 14 individuals from the younger group, and the other of 13 individuals from the older group. During each measurement session, 60 and 90 sweeps were averaged for the two ranges 20 to 100, and 100 to 600 Hz. As noted above, the younger

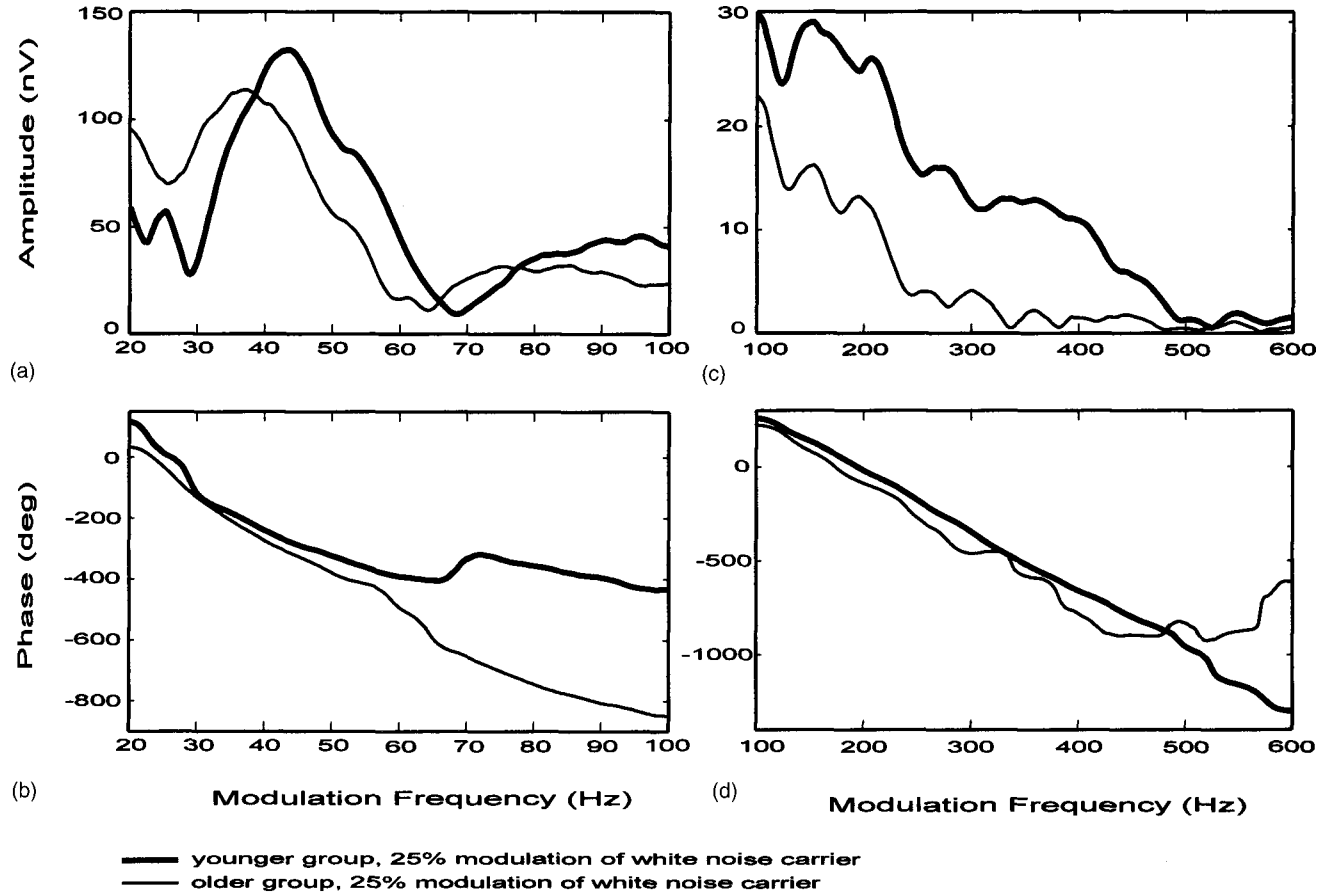


FIG. 5. Grand average of responses from sleeping individuals. Panels (A) and (B) show the responses to 25% amplitude modulated white noise presented at 60 dB SPL in the range 20 to 100 Hz. Similarly, panels (C) and (D) show the responses for the range 100 to 600 Hz. The thick lines indicate the average response from 14 individuals in the younger group. The thin lines show the average response from 13 participants in the older group.

group attended one session, whereas the older group attended two. For the younger group in the 20- to 100-Hz range, responses were highly significant except below 31 Hz, and near the response amplitude minimum at 69 Hz. A reversal in phase slope also occurred near 69 Hz. Between 35 and 55 Hz, the apparent latency was 24.3 ms, whereas it was 11.2 ms between 75 and 90 Hz. For the higher modulation range from 100 to 600 Hz, the responses were highly significant until 485 Hz. The apparent latency between 110 and 450 Hz was 8.8 ms.

Responses were highly significant for the older group in the 20- to 100-Hz range, with the exception of responses near the minimum at 64 Hz where the phase slope changed, but did not reverse. The apparent latency was 29.3 ms between 35 and 55 Hz, and 19.0 ms between 75 and 90 Hz. In the higher frequency range, the response was highly significant until 235 Hz. Between 120 and 235 Hz, the apparent latency was 9.1 ms.

Both groups showed a valley in the magnitude response below 30 Hz, and a peak near 40 Hz. Regarding the amplitude minima near 70 Hz, all but one subject had at least one region where the EFR amplitudes were not significantly different from the noise estimates, between 50 and 80 Hz. In the younger group, nine participants had a single amplitude minimum near 70 Hz, whereas in the older group only four had simple distinct minima. Many older participants had more complex amplitude functions with multiple local minima between 50 and 90 Hz.

Table I summarizes the results for those variables where the group means were significantly different. While the frequency of the amplitude peak in the band 30 to 50 Hz was significantly lower for the older group [see Table I and Fig. 5(A)], the peak amplitude itself was not significantly different between the two groups. In the same band, the older group had a significantly longer apparent latency than the younger group. This is evident in Fig. 5(B), where the older group's phase versus modulation frequency slope is steeper than for the younger group. The mean amplitudes were also significantly larger for the younger group in the bands between 100 and 500 Hz, as visible in Fig. 5(C).

The behavioral responses for the two subject groups are plotted in Fig. 6. Both the threshold for detecting 25% modulation of a white-noise carrier, and the threshold for detecting a gap between brief white-noise markers are plotted against the objectively determined frequency at which the phase-weighted EFR amplitude was no longer reliably different from noise. The number of reversals available for calculating the behavioral thresholds varied with the individual. On average, 9.8 and 10.3 reversals (SD 2.2 and 1.6 reversals) were available for the older and younger groups, respectively. As shown in Fig. 6, the two groups were largely separated with only a couple of individuals overlapping. Two participants

TABLE I. Significant results from two-tail separate variance *t* tests that evaluated whether the means for the younger and older groups were equal for different variables. These variables included the behavioral thresholds, age, and values derived from the EFR measurements. A low probability indicates that the means were unequal for the younger and older groups. The one nonsignificant result is included because this variable derived from the EFR was significantly correlated with a behavioral measurement in Table II.

	Modulation detection (Hz)	Gap detection (Hz)	Gap detection (ms)	Pure-tone hearing threshold (dB HL)	Maximum EFR frequency (Hz)	30–50–Hz frequency of amp. peak (Hz)	30–50-Hz apparent latency (ms)	100–600-Hz apparent latency (ms)	100–200-Hz mean amp. (nV)	200–300-Hz mean amp. (nV)	300–400-Hz mean amp. (nV)	400–500-Hz mean amp. (nV)	Low-pass model cutoff frequency (Hz)	Age (yrs)
Younger group mean (SD)	567 (95)	460 (197)	2.7 (1.5)	2.1(3.9)	494(114)	41(5)	25.6(6.2)	8.4(0.5)	31(10)	21(8)	15(6)	8(4)	260(108)	28(6)
Older group mean (SD)	264 (131)	274 (77)	4.0 (1.6)	20.8(12.6)	294(131)	37(4)	35.5(4.8)	8.6(1.5)	18 (8)	10(4)	6(4)	3(2)	165 (39)	69(6)
Significance	***	**	*	***	***	*	***		**	***	***	***	**	***

*$p<0.05$.
**$p<0.01$.
***$p<0.001$.

from the younger group and one from the older group had maximum EFR frequencies at the highest frequency present in the amplitude-modulated stimulus.

The younger group performed significantly better in the behavioral measurements (see Table I). The maximum EFR frequency was also significantly higher for the younger group. Table II summarizes the significant correlations between the objective EFR and behavioral variables. The best correlations between objective and behavioral measurements were obtained between the maximum EFR frequency and behavioral modulation detection ($r=0.72, p<0.001$), or mean pure-tone hearing threshold ($r=-0.76, p<0.001$). When correlations were calculated for the older and younger groups separately, some variables were significant within a given group, but none was significantly correlated in both groups.

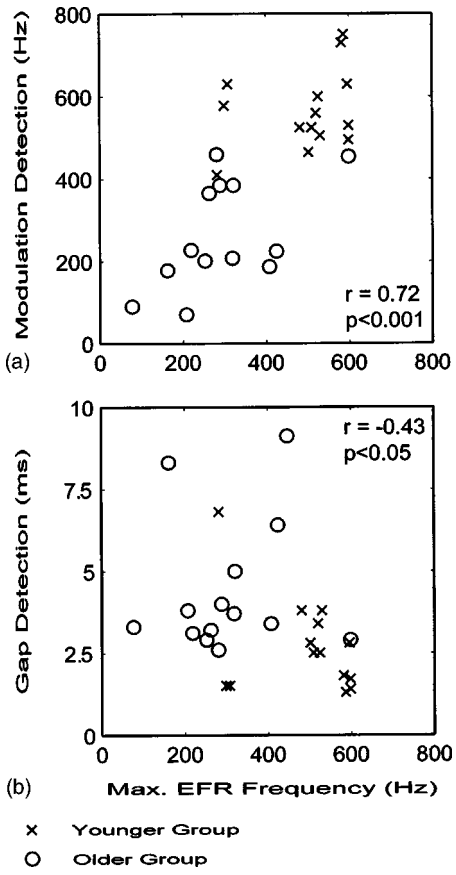


FIG. 6. Behavioral responses versus maximum EFR frequency. Panel (A) shows the behavioral threshold for detecting 25% amplitude modulation of white noise plotted against the highest frequency at which the phase-weighted EFR amplitude was reliably different from zero. Similarly, panel (B) shows the threshold for detection of a gap between brief Gaussian white-noise markers versus the same EFR threshold. The open circle ○ demarks each participant from the older group, and crossed symbol X shows the individuals from the younger group.

IV. DISCUSSION

A. Preliminary experiments

1. Different carriers

The results shown in Fig. 2 demonstrate the effects of carrier type and level. The largest responses were obtained for amplitude-modulated noise, presented at 60 dB. The 50-dB SPL modulated noise also generated a larger response than the 1-kHz carrier tone presented at 60 dB SPL. This is presumably because the noise will stimulate a larger region of the cochlea, and hence activate more afferent fibers (Picton *et al.*, 2003a; John *et al.*, 2003). While these may not all add in phase, the net result is a larger response than for the modulated tone.

The apparent latency of the modulated tone was larger than either of the noise stimuli. This may be because the dominant latency in the noise response was from a higher frequency band, and therefore more basal region of the cochlea, than for the 1-kHz carrier. The traveling wave delays are longer for the lower frequency pure-tone carrier.

2. Swept and fixed modulation rates

Similar responses were obtained using swept or fixed rate modulation of noise carriers. As shown in Figs. 2 and 3, the correspondence of the phase measurements between the FA (swept rate) and DFT (fixed rate) analyses was excellent. The amplitude results were also similar, and no significant differences were found. This suggests that the rate of change of the sweep stimulus was sufficiently low that response attenuation did not occur. An altered response might be ex-

TABLE II. Relationship (Pearson r) between behavioral thresholds, age, and values derived from the EFR measurements. Variables that are shown were significantly correlated with at least one behavioral threshold.

Behavioral threshold	Modulation detection	Gap detection in Hz	Gap detection in ms	Pure-tone hearing threshold	Maximum EFR frequency	30–50-Hz frequency of amp. peak	30–50-Hz apparent latency	100–600-Hz apparent latency	100–200-Hz mean amp.	200–300-Hz mean amp.	300–400-Hz mean amp.	400–500-Hz mean amp.	Low-pass model cutoff frequency	Age
Modulation detection	1	0.64***	−0.54**	−0.74***	0.72***	0.42*	−0.59***	−0.34	0.65***	0.67***	0.63***	0.64***	0.42*	−0.81***
Gap detection in Hz		1	−0.84***	−0.42*	0.42*	0.12	−0.19	−0.07	0.59**	0.54**	0.49**	0.44*	0.16	−0.51**
Gap detection in ms			1	0.36	−0.43*	−0.24	0.11	−0.02	−0.58***	−0.54***	−0.51***	−0.44*	−0.14	0.42*
Pure-tone hearing threshold				1	−0.76***	−0.26	0.53**	0.63***	−0.42*	−0.46*	−0.51**	−0.45*	−0.35	0.70***

*$p<0.05$.
**$p<0.01$.
***$p<0.001$.

pected at very high sweep rates, since the sources of the EFR in the auditory nervous system could have difficulty tracking very rapid changes in modulation rate. However, Artieda *et al.* (2004) have recently described a technique that can follow the brain's response to modulated sounds by averaging responses to a 1.6-s ''chirp'' that increased in modulation frequency from 1–120 Hz. They were able to demonstrate clear responses with maximum amplitude near 40 Hz.

The EFR amplitude null observed near 27 Hz in Figs. 3(A) through 5(A) may be due to the superposition of middle latency responses (MLRs) and auditory brainstem responses (ABRs, Galambos *et al.*, 1981), since the MLR and ABR waves occur with interpeak intervals of approximately 25 ms. Therefore, the responses will be in phase for successive occurrences elicited at 40 Hz, and out of phase when elicited every 37.5 ms, or at the rate of 26.7 Hz. Similar interference might occur for successive responses elicited every 12.5 ms, or at the rate of 80 Hz, but at these rates the MLR is likely too attenuated to have any effect, leaving only a periodic ABR.

3. Subject state

The response amplitude near 40 Hz was larger during wakefulness than sleep. This is consistent with other studies that show a reduction in the 40-Hz response with sleep (Linden *et al.*, 1985; Cohen *et al.*, 1991; Picton *et al.*, 2003b), or sedation (Plourde and Picton, 1990; Dobie and Wilson, 1998). At frequencies above 100 Hz, the three subjects tested here had the same EFR amplitudes during sleep and wakefulness, but noise estimates were consistently lower during sleep. The goal of the main experiment was to compare behavioral thresholds with relatively high-frequency EFRs. It was therefore decided to measure the EFR during sleep for the main experiment, in order to optimize SNR for a given measurement duration.

B. Main experiment

1. Age effects on psychophysical measurements of auditory temporal acuity

The behavioral tasks were performed with relatively few trials for a psychoacoustic measurement, in order to maintain a reasonable test time for one or two sessions. Fewer reversals were therefore available for calculating the behavioral thresholds, and some increased variability is expected. Nevertheless, a clear decrease in temporal auditory acuity was found with age, replicating the reports of others. Although most studies have not used noise stimuli, similar decreases in gap detection threshold with increasing age have been found. Schneider and Hamstra (1999), for example, determined that the gap detection threshold increased from 1.7 to 3.4 ms from young adults to old adults using 2-kHz tones of 5-ms duration.

2. Age effects on the EFR

There was no age-related difference in the amplitude of the EFR measured in the frequency range of 30–50 Hz. Similar findings have been obtained in other studies (Muchnik *et al.*, 1993; Boettcher *et al.*, 2001). Under some conditions the 40-Hz response to frequency modulation may be larger in elderly subjects (Boettcher *et al.*, 2002). One problem in assessing age-related changes is the fact that the response is very susceptible to drowsiness and sleep, and elderly subjects may be less able to sleep through the recordings than younger subjects (Dimitrijevic *et al.*, 2004). The age-related decrease in the peak frequency in this range and the increase in the apparent latency has not been reported before. These changes suggest some change in the timing and responsiveness of the cortical generators. Both effects are significantly related to the decrease in the maximum modulation frequency that can be perceived (Table II).

The EFR at frequencies greater than 100 Hz was significantly smaller in the elderly subjects, but this was not the case in the band 80 to 100 Hz. The apparent latency for these responses was, however, not affected by age. Boettcher *et al.*

(2001) and Dimitrijevic *et al.* (2004) also found no change in the amplitude of the 80-Hz response in normal-hearing elderly subjects. The age-related decrease in amplitude found here above 100 Hz may have been related to the mild hearing loss in the elderly subjects, but the correlations between amplitude and pure-tone thresholds were low (Table II). One might therefore postulate that the decreased size of the response reflects the decreased temporal acuity of the aging nervous system—an inability of the auditory brainstem to follow frequencies near 100 to 200 Hz as well as in the younger subjects.

3. Relations between the EFR measurements and the psychophysical findings

A good correlation between the maximum EFR frequency and modulation detection thresholds is shown in the data of Fig. 6. There was a clear separation of groups and a wide spread of individual thresholds across the test range for the modulation detection task. The gap detection thresholds were more closely clustered at values less than 5 ms. Interindividual discrimination may have improved if the step changes between gap stimuli had been smaller than 1 ms. In Fig. 6, three participants had maximum EFR frequencies that were at the ceiling of 600 Hz. In these individuals, it would probably have been possible to measure the EFR to higher frequencies, but the highest modulation rate in the stimulus was 600 Hz.

The lack of significant psychophysiological correlations within each group separately was likely due to the small numbers of subjects, and the lack of variation of the psychophysical findings within the groups. Unfortunately, the results therefore do not have sufficient power to disentangle the effects of temporal acuity from other effects of aging.

The EFRs generated in the brainstem or cortex cannot distinguish between peripheral (cochlear and auditory nerve) and brainstem causes for decreased auditory acuity. It may, however, be possible to distinguish between brainstem and cortical problems if different frequency bands of the response are evaluated (e.g., 30–50 Hz, and 100–200 Hz). Amplitude and latency values are different for the brainstem and cortical sources, and this may help evaluate changes in each region.

C. Model of EFR sources

For the correlation analysis, a simple first-order low-pass filter model was fit to the EFR amplitude data above 80 Hz. The optimal cutoff frequencies were significantly different between subject groups (Table I), but the correlation with behavioral measures was relatively poor (Table II). The model gain was calculated as the mean amplitude from 80 to 100 Hz, and was neither significantly different between groups nor correlated. The physiology is more complex than can be easily fit with such a model. There are at least two subsystems (brainstem and cortex) with different response characteristics and latencies.

The grand vector average amplitude functions had nulls near 70 Hz [Fig. 5(A)], and the presence of two distinct apparent latencies. Similar decreases in amplitude have been reported previously (see Fig. 10, Picton *et al.*, 2003a). This morphology suggested a two-component model of the EFR (Fig. 7). The model postulates two separate sinusoidal generators, each responding at the envelope frequency, one in the brainstem and one in the cortex, with latencies of 7.3 and 29 ms, respectively. The cortical source has a constant amplitude of 85 nV up to 50 Hz, and then decreases linearly to 0 nV at 95 Hz. The brainstem source had constant amplitude of 35 nV across the plotted frequency band, but would be expected to decrease to zero from 100 to 500 Hz [see Fig. 5(C)]. The total net response measured at a hypothetical electrode on the scalp was determined at each simulation frequency by summing the steady-state sinusoidal responses of the two sources in the time domain, and then obtaining the amplitude and phase of the resultant net sinusoid. The amplitude of each source in the sum was determined from the amplitude versus frequency plot in the top half of Fig. 7, and the phase of each source across frequency was determined by the source's latency. Summation was justifiable since the scalp fields for the two sources are similar and superimpose without interaction (Herdman *et al.*, 2002). Since the phase of each generator will change when the latency is constant and the envelope frequency varies, the sources sum constructively in some frequency regions, and destructively in others.

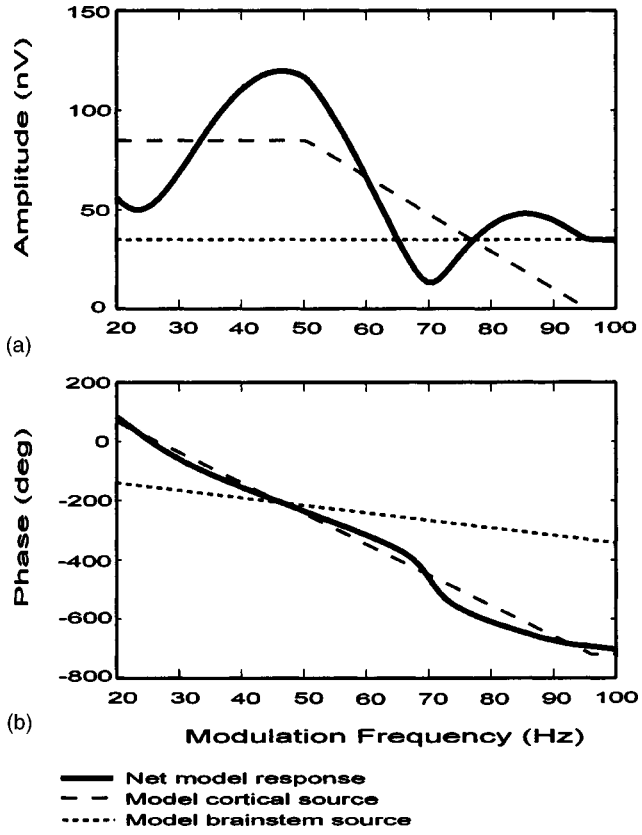


FIG. 7. Model EFR resulting from the sum of two sources. Hypothetical cortical and brainstem sources are shown with long dashed and short dashed lines, respectively. Amplitude is shown in panel (A), and phase in panel (B). The cortical source had a true delay of 29 ms, and constant amplitude from 20 to 50 Hz of 85 nV. The amplitude decreased linearly to zero from 50 to 95 Hz. The brainstem source had a true delay of 7.3 ms and constant amplitude of 35 nV. The continuous thick line indicates the sum of the two sources as might be recorded at scalp electrodes.

No correction was attempted for transmission from the two sources to the hypothetical electrode.

For the total net response, an amplitude minimum occurred near 70 Hz due to destructive interference between the two sources, and the apparent latency was 22.8 ms for the band 35 to 55 Hz. It can be seen in Fig. 7(B) that the net phase slope was determined by the relative amplitudes of the cortical source with its relatively long delay (steeper slope), and the brainstem source with its shorter delay (shallower slope). In other words, when the cortical source amplitude is large relative to the brainstem source (e.g., near 45 Hz) then the total net phase slope is closer to that of the cortical source, and this is reflected in the apparent latency estimate. This simple model reproduces the peak near 45 Hz and null near 70 Hz that occur in the grand vector average amplitude response of the younger group shown in Fig. 5(A).

If the model source amplitudes and delays are adjusted, the net output can be made to resemble the average response of the older group. The frequencies of the amplitude peak and null are controlled by the relative delays of the two sources. As shown in Table I, the older group had a peak response in the 40-Hz region that was significantly lower in frequency than for the younger group. Their apparent latency in that region was also significantly longer. While the amplitude in the 40-Hz region was not significantly different between groups, the mean amplitude between 100 and 200 Hz was significantly larger for the younger group. For the older group, the model may be adjusted such that both sources have lower cutoff frequencies, and steeper decay. The relative amplitudes of the sources could also be manipulated to emphasize the cortical source.

The average experimental apparent latency results are congruent with the simple model. For the younger group, the measured EFR amplitude between 100 and 200 Hz suggests that a brainstem source may make a larger relative contribution to the net response in the 40-Hz region than for the older group. The apparent latency of the presumed brainstem source measured at high frequencies ($>$100 Hz) was low for both groups ($<$10 ms). For the younger group, the relatively large amplitude of this source may in part be responsible for the lower apparent latency estimated in the 40-Hz region.

In the fourth preliminary experiment, the apparent latency in the 40-Hz region was shorter during sleep for three subjects. This could be due in part to the relatively larger contribution of the brainstem source when the cortical source amplitude was attenuated by sleep.

V. CONCLUDING REMARKS

As discussed in the Introduction to this paper, it would be very helpful to have a simple objective test of temporal auditory acuity that could be used to assess patients who cannot give reliable data from psychophysical testing. Werner *et al.* (2001) have shown that auditory brainstem responses can be recorded to gaps in noise, and that the thresholds for recognizing these responses were related to psychophysical gap detection thresholds in adults. They found similar ABR thresholds in infants, although their psychophysical thresholds were much longer. The results presented here show the maximum frequency for recognizing the EFR correlates well with the psychophysical measures of the maximum perceptible threshold. This could then serve as an objective test of temporal auditory acuity.

The techniques investigated here for objectively assessing temporal acuity are neither simple nor rapid. The test could be made faster by only eliciting the response from a narrower range of frequencies (e.g., 100 to 600 Hz). Even then, the measurement could take 45 to 90 min (i.e., two sessions for the older subjects) depending on the background EEG noise of the recording. The limiting factor is that the system must determine when a near-threshold response is not present, and time is required to reduce the background EEG noise sufficiently to make this judgment. Accordingly, a better approach might be to derive a simple measurement at suprathreshold levels that relates well to the psychophysical measurement. Shorter testing durations can be achieved when the evoked responses are larger, since the background EEG noise need not be reduced as much to achieve a given SNR. The peak frequency of the 40-Hz response is significantly related to the maximum perceptible modulation frequency, but the correlation is not high.

This study relied upon 25% modulation, which elicits a relatively small EFR. Future studies could examine the utility of recording responses with stimulus parameters that are optimized for short recording times. For example, these could employ a brief sweep of frequencies (e.g., 100–300 Hz) using 50% or 100% amplitude modulation, to produce larger responses. The amplitudes over this limited range of modulation rates might be used to predict the threshold. Response amplitudes can also be increased by using bandlimited noise (e.g., 1000–2000 Hz) rather than broadband noise (John *et al.*, 2003). Further, since the behavioral gap threshold increases with decreasing carrier frequency for pure-tone markers, using pure tones or bandpass noise may be anticipated to cause a more pronounced decrease in EFR amplitude with increasing modulation frequency. Additionally, when testing both ears, the testing time can be reduced to that of one single ear, since both ears can likely be tested simultaneously using different modulation rates in each ear. Ultimately, one would like a measurement that would reliably indicate good or bad temporal auditory acuity within about 10 min.

Despite these reservations about test time, the sweep EFR techniques investigated here can provide an objective measurement of human auditory temporal acuity that is reasonably well correlated to behavioral measurements.

ACKNOWLEDGMENTS

The authors would like to thank Mercedes Sarudiansky and Patricia Van Roon for their assistance in data collection and analysis. Research was supported by the Canadian Institutes of Health Research.

Amitay, S., Ahissar, M., and Nelken, I. (**2002**). ''Auditory processing deficits in reading disabled adults,'' J. Assoc. Res. Otolaryngol. **3**, 302–320.

Artieda, J., Valencia, M., Alegre, M., Olaziregi, O., Urrestarazu, E., and Iriarte, J. (**2004**). ''Potentials evoked by chirp-modulated tones: A new technique to evaluate oscillatory activity in the auditory pathway,'' Clin. Neurophysiol. **115**, 699–709.

Bacon, S. P., and Viemeister, N. F. (**1985**). ''Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners,'' Audiology **24**, 117–134.
Benasich, A. A., and Tallal, P. (**2002**). ''Infant discrimination of rapid auditory cues predicts later language impairment,'' Behav. Brain Res. **136**, 31–49.
Bishop, D. V., Carlyon, R. P., Deeks, J. M., and Bishop, S. J. (**1999**). ''Auditory temporal processing impairment: Neither necessary nor sufficient for causing language impairment in children,'' J. Speech Lang. Hear. Res. **42**, 1295–1310.
Boettcher, F. A., Madhotra, D., Poth, E. A., and Mills, J. H. (**2002**). ''The frequency-modulation following response in young and aged human subjects,'' Hear. Res. **165**, 10–18.
Boettcher, F. A., Poth, E. A., Mills, J. H., and Dubno, J. R. (**2001**). ''The amplitude-modulation following response in young and aged human subjects,'' Hear. Res. **153**, 32–42.
Cohen, L. T., Rickards, F. W., and Clark, G. M. (**1991**). ''A comparison of steady-state evoked potentials to modulated tones in awake and sleeping humans,'' J. Acoust. Soc. Am. **90**, 2467–2479.
Dimitrijevic, A., John, M. S., and Picton, T. W. (**2004**). ''Auditory steady-state responses and word recognition scores in normal-hearing and hearing-impaired adults,'' Ear Hear. **25**, 68–84.
Dobie, R. A., and Wilson, M. J. (**1998**). ''Low-level steady-state auditory evoked potentials: Effects of rate and sedation on detectability,'' J. Acoust. Soc. Am. **104**, 3482–3488.
Fitzgibbons, P. J., and Wightman, F. L. (**1982**). ''Gap detection in normal and hearing-impaired listeners,'' J. Acoust. Soc. Am. **72**, 761–765.
Formby, C. (**1987**). ''Modulation threshold functions for chronically impaired Meniere patients,'' Audiology **26**, 89–102.
Galambos, R., Makeig, S., and Talmachoff, P. J. (**1981**). ''A 40-Hz auditory potential recorded from the human scalp,'' Proc. Natl. Acad. Sci. U.S.A. **78**, 2643–2647.
Gordon-Salant, S., and Fitzgibbons, P. J. (**1999**). ''Profile of auditory temporal processing in older listeners,'' J. Speech Lang. Hear. Res. **42**, 300–311.
Grose, J. H., Hall III, J. W., and Buss, E. (**2001**). ''Gap duration discrimination in listeners with cochlear hearing loss: effects of gap and marker duration, frequency separation, and mode of presentation,'' J. Assoc. Res. Otolaryngol. **2**, 388–398.
Herdman, A. T., Lins, O., Van Roon, P., Stapells, D. R., Scherg, M., and Picton, T. W. (**2002**). ''Intracerebral sources of human auditory steady-state responses,'' Brain Topogr **15**, 69–86.
John, M. S., and Picton, T. W. (**2000**). ''MASTER: A Windows program for recording multiple auditory steady-state responses,'' Comput. Methods Programs Biomed. **61**, 125–150.
John, M. S., Dimitrijevic, A., and Picton, T. W. (**2003**). ''Efficient stimuli for evoking auditory steady-state responses,'' Ear Hear. **24**, 406–423.
Linden, R. D., Campbell, K. B., Hamel, G., and Picton, T. W. (**1985**). ''Human auditory steady state evoked potentials during sleep,'' Ear Hear. **6**, 167–174.
Moore, B. C. (**1995**). ''Effects of cochlear damage on temporal resolution and temporal integration,'' in *Perceptual Consequences of Cochlear Damage*, edited by B. C. Moore (Oxford University Press, Oxford), pp. 88–108.
Muchnik, C., Katz-Putter, H., Rubinstein, M., and Hildesheimer, M. (**1993**). ''Normative data for 40-Hz event-related potentials to 500-Hz tonal stimuli in young and elderly subjects,'' Audiology **32**, 27–35.
Norcia, A. M., Tyler, C. W., Hamer, R. D., and Wesemann, W. (**1989**). ''Measurement of spatial contrast sensitivity with the swept contrast VEP,'' Vision Res. **29**, 627–637.
Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., and Mossop, J. E. (**1997**). ''Detection of silent intervals between noises activating different perceptual channels: Some properties of central auditory gap detection,'' J. Acoust. Soc. Am. **101**, 3694–3705.
Picton, T. W., Dimitrijevic, A., John, M. S., and Van Roon, P. (**2001**). ''The use of phase in the detection of auditory steady-state responses,'' Clin. Neurophysiol. **112**, 1698–1711.
Picton, T. W., John, M. S., Dimitrijevic, A., and Purcell, D. (**2003a**). ''Human auditory steady-state responses,'' Int. J. Audiol. **42**, 177–219.
Picton, T. W., John, M. S., Purcell, D. W., and Plourde, G. (**2003b**). ''Human auditory steady-state responses: The effects of recording technique and state of arousal,'' Anesth. Analg. (Baltimore) **97**, 1396–1402.
Picton, T. W., Hink, R. F., Perez-Abalo, M., Linden, R. D., and Wiens, A. S. (**1984**). ''Evoked potentials: How now?,'' Electrophysiol. Technol. **10**, 177–221.
Plourde, G., and Picton, T. W. (**1990**). ''Human auditory steady-state response during general anesthesia,'' Anesth. Analg. (Baltimore) **71**, 460–468.
Rappaport, J. M., Gulliver, J. M., Phillips, D. P., Van Dorpe, R. A., Maxner, C. E., and Bhan, V. (**1994**). ''Auditory temporal resolution in multiple sclerosis,'' J. Otolaryngol. **23**, 307–324.
Rees, A., Green, G. G., and Kay, R. H. (**1986**). ''Steady-state evoked responses to sinusoidally amplitude-modulated sounds recorded in man,'' Hear. Res. **23**, 123–133.
Regan, D. (**1966**). ''Some characteristics of average steady-state and transient responses evoked by modulated light,'' Electroencephalogr. Clin. Neurophysiol. **20**, 238–248.
Regan, D. (**1989**). *Human Brain Electrophysiology: Evoked Potentials and Evoked Magnetic Fields in Science and Medicine* (Elsevier Science, New York), pp. 70–98, 112–123, 273–275.
Rodriguez, R., Picton, T., Linden, D., Hamel, G., and Laframboise, G. (**1986**). ''Human auditory steady state responses: Effects of intensity and frequency,'' Ear Hear. **7**, 300–313.
Rosen, S. (**1992**). ''Temporal information in speech: acoustic, auditory and linguistic aspects,'' Philos. Trans. R. Soc. London, Ser. B **336**, 367–373.
Sapsford, D. J., Pickworth, A. J., and Jones, J. G. (**1996**). ''A method for producing the coherent frequency: A steady-state auditory evoked response in the electroencephalogram,'' Anesth. Analg. (Baltimore) **83**, 1273–1278.
Schneider, B. (**1997**). ''Psychoacoustics and aging: Implications for everyday listening,'' J. Speech-Lang. Pathol. Audiol. **21**, 111–124.
Schneider, B. A., and Hamstra, S. J. (**1999**). ''Gap detection thresholds as a function of tonal duration for younger and older listeners,'' J. Acoust. Soc. Am. **106**, 371–380.
Schneider, B. A., Daneman, M., and Pichora-Fuller, M. K. (**2002**). ''Listening in aging adults: From discourse comprehension to psychoacoustics,'' Can. J. Exp. Psychol. **56**, 139–152.
Schneider, B., Speranza, F., and Pichora-Fuller, M. K. (**1998**). ''Age-related changes in temporal resolution: Envelope and intensity effects,'' Can. J. Exp. Psychol. **52**, 184–191.
Shailer, M. J., and Moore, B. C. (**1983**). ''Gap detection as a function of frequency, bandwidth, and level,'' J. Acoust. Soc. Am. **74**, 467–473.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Snell, K. B., and Frisina, D. R. (**2000**). ''Relationships among age-related differences in gap detection and word recognition,'' J. Acoust. Soc. Am. **107**, 1615–1626.
Snell, K. B., Mapes, F. M., Hickman, E. D., and Frisina, D. R. (**2002**). ''Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity,'' J. Acoust. Soc. Am. **112**, 720–727.
Stapells, D. R., Linden, D., Suffield, J. B., Hamel, G., and Picton, T. W. (**1984**). ''Human auditory steady-state potentials,'' Ear Hear. **5**, 105–113.
Taylor, M. M., and Creelman, C. D. (**1967**). ''PEST: Efficient estimates on probability functions,'' J. Acoust. Soc. Am. **41**, 782–787.
Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). ''Speech waveform envelope cues for consonant recognition,'' J. Acoust. Soc. Am. **82**, 1152–1161.
Viemeister, N. F. (**1979**). ''Temporal modulation transfer functions based upon modulation thresholds,'' J. Acoust. Soc. Am. **66**, 1364–1380.
Viemeister, N. F., and Plack, C. J. (**1993**). ''Time analysis,'' in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York), pp. 116–154.
Werner, L. A., Folsom, R. C., Mancl, L. R., and Syapin, C. L. (**2001**). ''Human auditory brainstem response to temporal gaps in noise,'' J. Speech Lang. Hear. Res. **44**, 737–750.
Zeng, F. G., Oba, S., Garde, S., Sininger, Y., and Starr, A. (**1999**). ''Temporal and speech processing deficits in auditory neuropathy,'' NeuroReport **10**, 3429–3435.
Zeng, F. G., Oba, S., Garde, S., Sininger, Y., and Starr, A. (**2001**). ''Psychoacoustics and speech perception in auditory neuropathy,'' in *Auditory Neuropathy*, edited by Y. Sininger and A. Starr (Singular Thomson Learning, San Diego), pp. 141–164.

The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation

Murat Aytekin, Elena Grassi, Manjit Sahota, and Cynthia F. Moss
Department of Psychology and Institute for Systems Research, Neuroscience and Cognitive Science Program, University of Maryland, College Park, Maryland 20742

(Received 21 March 2004; revised 26 August 2004; accepted 12 September 2004)

Directional properties of the sound transformation at the ear of four intact echolocating bats, *Eptesicus fuscus*, were investigated via measurements of the head-related transfer function (HRTF). Contributions of external ear structures to directional features of the transfer functions were examined by remeasuring the HRTF in the absence of the pinna and tragus. The investigation mainly focused on the interactions between the spatial and the spectral features in the bat HRTF. The pinna provides gain and shapes these features over a large frequency band (20–90 kHz), and the tragus contributes gain and directionality at the high frequencies (60 to 90 kHz). Analysis of the spatial and spectral characteristics of the bat HRTF reveals that both interaural level differences (ILD) and monaural spectral features are subject to changes in sound source azimuth and elevation. Consequently, localization cues for horizontal and vertical components of the sound source location interact. Availability of multiple cues about sound source azimuth and elevation should enhance information to support reliable sound localization. These findings stress the importance of the acoustic information received at the two ears for sound localization of sonar target position in both azimuth and elevation. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1811412]

PACS numbers: 43.66.Qp, 43.66.Pn, 43.80.Lb, 43.64.Bt, 43.64.Ev [JAS] Pages: 3594–3605

I. INTRODUCTION

Echolocating bats, as auditory specialists, present a valuable model for the study of spatial hearing. These flying mammals produce ultrasonic vocalizations and listen to the echoes reflected from objects around them.[1] Their vital activities, such as spatial orientation, prey detection, localization, and capture can be accomplished entirely with the use of a biological sonar system. An important question is whether specialized sound localization capacity is required for successful orientation by echolocation in bats. If so, what specializations are evident that can help to deepen our understanding of spatial hearing?

Azimuthal discrimination experiments by Simmons *et al.*[2] and horizontal tracking experiments by Masters *et al.*[3] report about 1.5° accuracy in the echolocating big brown bat, *Eptesicus fuscus*. This species uses broadband frequency-modulated (FM) echolocation signals, which are well suited for spatial localization.[4] Lawrence and Simmons[5] measured the minimum audible angle (MAA) along the vertical axis and report discrimination by *E. fuscus* of about 3°. In a later study, Wotton and Simmons[6] closely studied the bat's MAA in the midsagittal plane at 5 different elevations ($+/-40°$, $+/-20°$, and 0°). They reported that the bat's MAA falls between 2.9° and 4.1° at around 0° and $+/-20°$ elevation and increases to 6.7° at around $-40°$ and 8.3° at around 40°. In the same study, when the bat's tragus was deflected, the MAA increased significantly, 18.3°–20.1°, at around $-20°$ elevation; however, the MAA in the upper hemisphere did not show a measurable change with the tragus manipulation. In these studies, the bats used biosonar to perform the localization tasks.

Studying bats in a passive listening localization task, Koay *et al.*[7] used a conditioned avoidance procedure to measure left–right azimuthal sound source discrimination by *E. fuscus*. They reported that the bat's threshold for discrimination is 14°, comparable to a number of other mammalian species they tested. Further, they reported that this bat species could not discriminate narrow-band sounds that were positioned at 30° off the midline at low frequencies (5.6 to 8 kHz). They concluded that *E. fuscus* is not able to use interaural time difference (ITD) cues at sound frequencies where interaural phase differences would be available. Koay *et al.* estimated the maximum ITD experienced by the bats during the experiment to be approximately 55 μs. Although the results of this study suggest that *E. fuscus* cannot use interaural phase differences for sound localization, the result does not eliminate the possibility of ITD localization using the envelope of wideband sounds.

To explore sound localization one step further, attention has been given to the directional properties of the external ears. Studies of the directionality of sound-pressure transformation at the pinna of several different bat species have been conducted.[8–10] In these studies, sound pressure at the tympanic membrane was measured at different frequencies by moving a narrow-band sound source in space and recording signals at the base of the ear canals. Disturbance of the isopressure contours and a decrease in the gain of the sound pressure for all frequencies tested was reported after removing the pinna and the tragus together.[8,10]

Using a similar approach, Fuzzessery[11] measured acoustic pressure transformation by the intact ears of the pallid bat, *Antrazous pallidus*. He reported that monaural spectral features of the HRTF of this species show parallel patterns to what has been reported in other mammals, such as humans and cats. He also showed that binaural difference cues changed with both azimuth and elevation. Grinnell and

Grinnell[12] reported a similar observation in the directional dependence of evoked potential recordings from the inferior colliculus (IC) of FM bats, *Plecotus townsendii* and *Myotis lucifugus*. They found that directional sensitivity of the neural responses changed with small changes in sound source angle in both the horizontal and vertical planes. Moreover, they reported that directional sensitivity was different at different frequencies. Based on these observations and assuming that directional responses of the left and right ICs are similar, they hypothesized that bats can use interaural level differences (ILD) obtained from several different frequencies to localize both sound source azimuth and elevation. Fuzessery and Pollak[13] conducted cochlear microphonic recordings in mustached bats, *Pteronotus parnellii*, a species that produces signals with a combination of constant frequency and FM components. These researchers also proposed a similar role for ILD in horizontal and vertical sound localization. These important findings stand apart from the common view that interaural differences are used for azimuthal localization. Thus, the role of binaural cues for localization of sound source elevation warrants further study.

Sounds reaching the ear are transformed in a directionally dependent manner as a result of their interaction with the body, the head, and the pinnae. The direction-dependent transformations can be modeled as a linear time-invariant systems transfer function, commonly known as the head-related transfer function (HRTF). Contrary to what the name suggests, the torso, in particular neck and shoulders, also introduce strong effects on the properties of the HRTF.[14]

To our knowledge, the most detailed published measurement of the HRTF in bats was reported by Wotton *et al.*[15] The study focused on elevation cues: in particular, spectral notches and peaks in the HRTF of the big brown bat, *Eptesicus fuscus*. They measured the HRTF of full and half-head preparations and reported a linear relationship between center frequency (CF) of spectral notches (30 to 50 kHz) and sound source elevations ($-30°$ to $0°$). They also reported that the relation between notch CF and elevation was disrupted when the bat's tragus was deflected. Based on their findings, Wotton *et al.* suggested that position of the spectral notches along the frequency axis could contribute to vertical localization in bats, as is the case in other animals that have been studied, such as cats[16] and humans.[17]

Wotton *et al.*[6,18] provided behavioral evidence for the role of the position of a spectral notch in sound source localization along the vertical plane. They trained the bats to discriminate a linear FM sound with a spectral notch at a particular frequency, and they demonstrated that spectral notches in the HRTF are perceived by the bat.[18] Bats showed a significant decrease in their performance in this behavioral task when the target sound source was presented at a particular elevation at which the filtering characteristics of the external ear produced a spectral notch at the same frequency. In another experiment, they showed that deflection of the bat's tragus resulted in an increase in vertical MAA.[6] Based on these and other MAA measurements[2,5,6] they suggested that there are distinct sound localization cues for azimuth and elevation, consistent with the classic dichotomy of separate cues for horizontal and vertical sound localization.

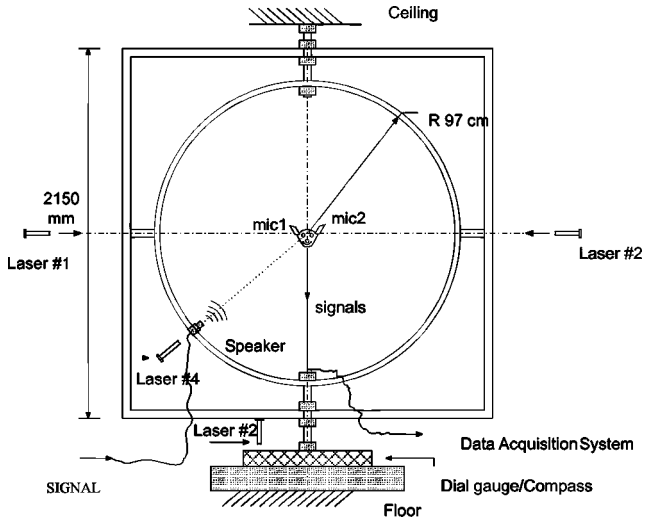


FIG. 1. Schematic of the frame used to suspend the bat, with microphones implanted in the ear canals, and to deliver acoustic signals from changing azimuths and elevations.

In our paper, we present the HRTF of an echolocating FM bat, *Eptesicus fuscus*. Our first goal was to confirm earlier reports by Wotton *et al.*[15] and extend observations to intact preparations that preserve the head, neck, body, and wings. The measurements were taken from the entire frontal hemisphere at a high spatial and spectral resolution to capture a complete and detailed HRTF. We carried out analyses of HRTF data that enabled us to investigate the interactions between spatial and spectral features of the HRTF that were not previously reported. By comparing measured transfer functions in the absence of pinna and tragus, we examined the contribution of the external ear structures to the directional properties of the HRTF. In our analysis presented here, we emphasize the acoustic features that are believed to play a role in mammalian sound localization. This approach has led us also to consider a role for binaural cues for vertical localization by animals that hear in the ultrasonic range.

II. METHODS

We constructed an apparatus and developed the methods to permit measurements of the head related transfer function (HRTF) of the intact bat, in the frontal hemisphere. A 1-m-radius hoop was used to carry out the acoustic measurements (Fig. 1). The hoop was mounted upright and fixed to a rectangular frame, which was attached to the ceiling and floor of a large room (7×6 m) lined with acoustic foam. This device could rotate around its vertical axis and could be locked on a reference dial at any selected azimuth angle with $1°$ resolution. A custom 1.6-cm-diameter ultrasound loudspeaker (design by Lee Miller, Odense, Denmark, frequency response flat $+/-3$ dB between 20 and 100 kHz) was attached to a sliding fixture that could be locked at any position on the hoop. The hoop was marked in $5°$ steps, indicating elevation coordinates.

A. Animal preparation

Measurements were taken from four intact *Eptesicus fuscus* specimens (S1, S2, S3, S4). Each bat was perfused with 4% paraformaldehyde, followed by a saline flush. A

microphone with a 1.5-mm-diameter metal probe (FG 3453-Knowles Electronics) was inserted into each of the bat's ear canals after rupturing the tympanic membrane. Lithium batteries (1.5 V) were used to power the microphones. The microphones were glued in place (Loctite 411 with Zipkicker accelerator) and sealed tightly around the ear canal to avoid any leak of sound pressure. Cross-talk between the microphones was measured before each experiment, and it was no greater than −60 dB. The bat was suspended horizontally with its wings spread, as in flight, and the pinnae were preserved in an upright position.

The bat's head was positioned at the center of the hoop pointing at the 0° azimuth and elevation. Photographs of the bat were taken to make necessary coordinate transformations to align the bat's eye–nostril axis to 0° azimuth and 0° elevation. The coordinate transformations that were applied to S1, S2, S3, and S4 were 40°, 30°, 0°, and 22° downward tilt, respectively. Two laser pointers were installed at positions −90° and 90° azimuth, each aimed at the tip of the corresponding tragus. These laser beams were permanently illuminated to monitor the position of the bat and to ensure the correct position of the frame in the calibration stage. Another laser pointer was used to adjust the directionality of the speaker with respect to the bat's head any time the elevation of the speaker was changed.

To investigate the contribution of the pinna and the tragus to the directional properties of the HRTF, a separate set of measurements for tragus-absent and pinna-absent ears was conducted. Tragus-absent measurements were taken from S3 and S4 by keeping the right ear intact (control ear) and deflecting the tragus of the left ear. Pinna-absent measurements were taken from S2, S3, and S4 by cutting away the pinna and returning the tragus of the left ear to its natural position. Measurements of the HRTF from the intact right ear allowed comparisons across different sets of measurements to control for any time-related changes that may have occurred in the condition of the bat. During data collection from S3, an unintended change in the orientation of the bat after the first set of measurements prevented a detailed comparison of the HRTF features between the intact and manipulated ears.

B. Data acquisition

The data were recorded for sounds broadcast from a loudspeaker at a distance of 84 cm from the center of the bat's head. Microphone recordings from 685 different speaker positions on the frontal hemisphere were taken. Spatial positions were sampled from a double-polar coordinate system with 5° resolution in both dimensions. Speaker positions were specified by their azimuth and elevation values. Azimuth values changed from −90° to 90° from left to right of the midline, while elevation values changed from −90° to 90° from below to above the horizon.

Computer-generated signals, consisting of 2-ms-duration downward linear frequency-modulated (FM) sweeps, were broadcast at a rate of 33 Hz. The duration and interpulse interval of the signals were selected to prevent overlap of the broadcast signals with echoes from the hoop and surrounding walls. The signals recorded with the implanted microphones were amplified and bandpass filtered (10–100 kHz) using a commercial filter unit (Stewart). The signals were sampled through a data acquisition system (IOTech Wavebook/512) connected to a Pentium III PC computer. The outputs from the microphones were monitored at all times on a two-channel oscilloscope (LeCroy).

C. Signal processing

Fifty signals per channel (left and right ears) were recorded, with a sampling rate of 500 kHz for each speaker position. To improve SNR, these signals were time aligned and averaged, making sure that the relative time delay between channels was unchanged. Then, HRTFs for the left and the right ears at each source direction were calculated as the ratio of the Fourier transform (2048-point FFT) of the averaged signals to the Fourier transform of the averaged signals obtained from free-field sound recordings. Free-field sound recordings were taken from the speaker position 0° azimuth and 0° elevation with both microphones placed at the center of the coordinate frame before they were implanted in the bat's ears.

The magnitude of the HRTF was smoothed in logarithmic frequency domain using a rectangular sliding window. The width of the sliding window was chosen to obtain a quality factor of 20. In this case, the quality factor is defined as the proportion of the window's center frequency to its width. This quality factor was chosen based on measured tuning curves of cells in the inferior colliculus of *E. fuscus*[19,20] to achieve biologically realistic frequency resolution.

For some analyses, the HRTF was divided into direction-dependent and direction-independent linear time-invariant systems. The magnitude of the direction-independent component (here referred to as the common transfer function, CTF) was calculated as the spatial average of the magnitude function across the entire set of HRTF measured. The phase of this component was computed, based on a minimum-phase assumption.[21] The direction-dependent component, also called direction-dependent transfer function (DTF), was calculated by dividing the HRTF by the CTF.

Spatial interpolation was applied to the measured HRTF at each frequency using spherical harmonic expansion.[22] This process allowed us to obtain HRTF for the directions that were not sampled during the measurements. After interpolation, iso-level contours were obtained using built-in MATLAB functions.

III. RESULTS

A. Acoustic gain

Acoustic gain can be defined as the ratio of sound intensity at the ear canal to sound intensity in free field at a particular frequency. Thus, acoustic gain is the magnitude of the HRTF at that frequency. The direction of the maximum gain of HRTF (or DTF) across azimuth and elevation, at a particular frequency, is commonly known as the acoustic axis.

Figure 2(a) depicts the maximum gain (MG) provided by the HRTFs of the left (solid) and the right (dashed) ear of S1 as a function of frequency, across all the directions. Each curve shows distinct peaks and troughs, the acoustical modes

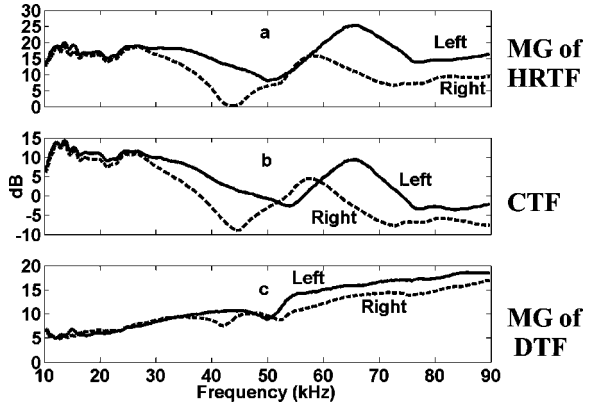


FIG. 2. (a) Maximum gain (MG) of head related transfer functions (HRTF). (b) Common transfer functions (CTF). (c) MG of directional transfer functions (DTF). Functions for the left and the right ears of S1 given as solid and dashed lines, respectively.

of the external ears. A plateau at the lower frequencies spans a frequency region between 10 and 30 kHz. A peak can be seen at the higher end of the spectral range. The peak point of this mode is around 57 and 67 kHz for the right and the left ears, respectively. A trough in between the plateau and the peak occurs at 43 kHz for the right ear and 50 kHz for the left. Across most of the frequency range, except the trough, the gain is positive and reaches a maximum of 25 dB for the left ear.

1. Asymmetry between the left and the right ears

Another interesting feature of the data is that the left ear's and the right ear's MG of HRTF and DTF are not identical, suggesting asymmetry between the ears. This asymmetry may be a result of structural and/or orientation differences between ears. Orientation differences between the two pinnae could be natural or might have been created during preparation of the animals for acoustic measurements. However, the asymmetry might also be an artifact caused by the placement of the microphones in the ear canals. Many HRTF studies have investigated effects of microphone position in the ear canal on the directional changes in HRTF.[23–25] These studies have reported that transformation in the ear canal is independent of sound direction. Thus, microphone placement in the ear canal likely does not contribute to directional changes in the HRTF. To evaluate unintended effects of microphone placement in both ears that may potentially produce asymmetry, we examine the common transfer function (CTF) and the MG of the direction-dependent component of the HRTF, i.e. the direction-dependent transfer function (DTF). Effects of the ear canal and its interactions with the microphones on the HRTF are accounted for in CTF [Fig. 2(b)]. Any structural differences between the ears, before the ear canal opening, would still be present in DTF (for a technical discussion, see Refs. 17 and 26). However, some asymmetry in the MG of DTF still persists between the two ears [see Fig. 2(c)]. In particular, above 35 kHz the left ear shows a higher gain than the right ear.

The local troughs in both DTFs [Fig. 2(c)] are the result of differences in center frequencies of the troughs in MG functions and CTFs, shown in Figs. 2(a) and (b), respectively. The center frequencies of the troughs for both ears shift to a lower frequency in the CTF. We hypothesize that the asymmetry between the left and the right ear DTF is a result of structural and/or orientation differences between the two ears.

2. Pinna and tragus manipulations

Figure 3 illustrates the left and right MG of HRTF, CTF and MG of DTF of S4. The left ear of S4 was manipulated after the intact ear measurement, while the right remained untouched to serve as a control. Figures 3(a) and (d) show MG of HRTFs of the left and right ears, respectively.

As expected, the right ear characteristic functions overlap across different measurement sessions, ensuring that no significant decay of the preparation occurred [Figs. 3(d), (e), and (f)]. Shaded areas represent the range of change of characteristic functions along the frequency axis across three sets of measurements. Notice also that CTFs of the left ear show only small variations across the different manipulations. In contrast, the MG of the left ear HRTF for the intact and the tragus-absent cases differ above 57 kHz, where tragus removal causes a drop in the gain [Figs. 3(a), (b), dashed line]. Notice that this effect persists in the maximum gain of the DTF of the left ear [Fig. 3(c)]. This observation suggests that the presence of the tragus affects the acoustical mode, caus-

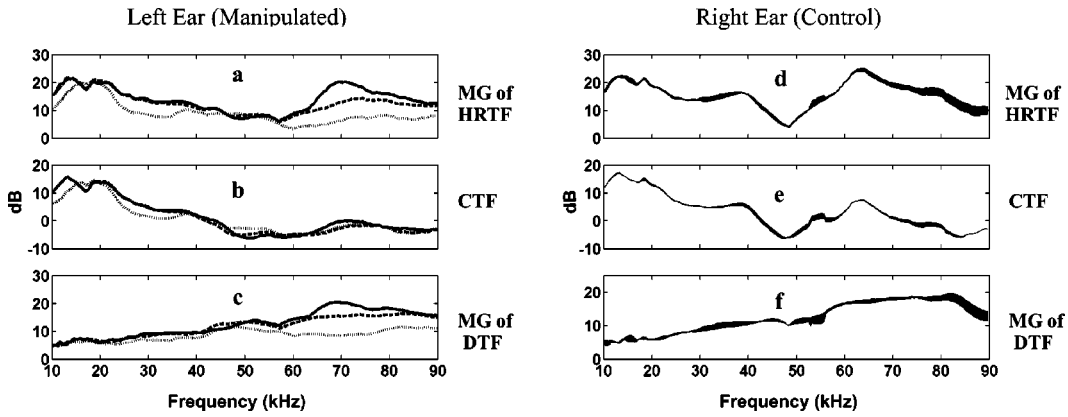


FIG. 3. Transfer functions for S4: Intact (solid), tragus-removed (dashed), and pinna-removed (dotted). The right ear remained intact as a control. Characteristic functions for the right ear depicted as lines with varying thickness, representing the range of change of the functions across different measurements. The characteristic functions, from top to bottom, are maximum gain (MG) of HRTF, CTF, and MG of DTF.

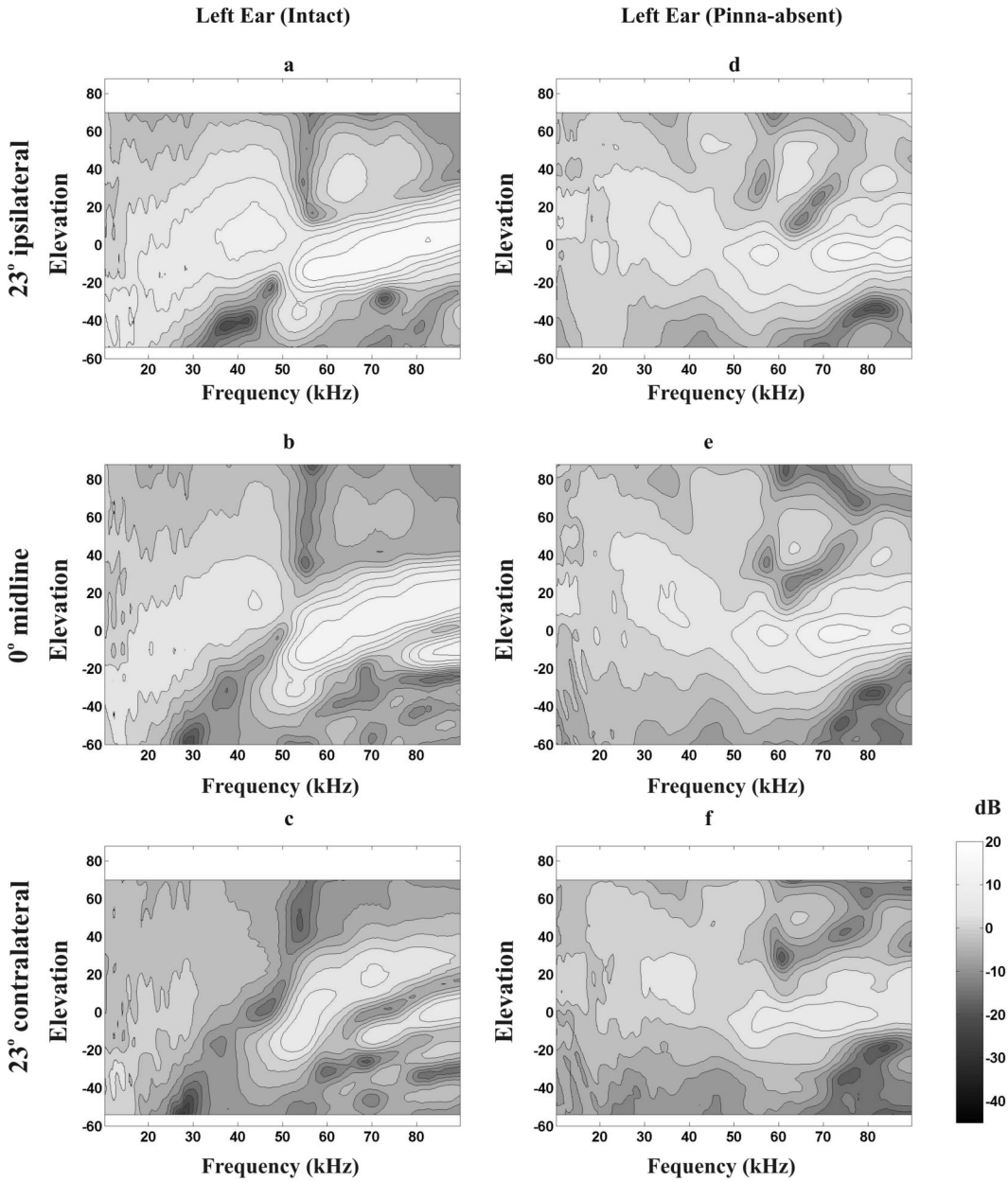


FIG. 4. Left ear DTF contour plot for sagittal planes at 23° ipsilateral (top row), 0° (middle row), and 23° contralateral (bottom row). Intact (a), (b), (c) and pinna-absent (d), (e), (f) cases for S2. Contour lines are plotted with 3-dB separation.

ing an improvement in gain and directionality and a small change in the frequency of the resonant mode.

Following removal of the pinna on the left ear, the maximum gain drops up to 6 dB between 20 and 35 kHz and up to 14 dB at frequencies above 57 kHz. The maximum gain of the DTF, after the pinna was removed, shows a very shallow increase with frequency [Fig. 3(c), dotted line].

MG of HRTFs obtained from S4 [Figs. 3(a) and (d)] and S1 [Fig. 2(a)] were slightly different from each other, not only across subjects but also across ears in the same subject. These differences may be a result of natural variations in ear shape and/or differences in pinna orientation and microphone placement in the subjects' ear canals. However, MG of DTFs show a similar pattern across subjects.

B. Monaural spectral properties

The HRTF includes direction-dependent spectral features that can be used for the computation of sound localization by the auditory system. Spectral notches are the most systematic of these features, and are believed to play an important role in vertical sound localization.[15] Figure 4 shows the contour plots of the DTF magnitudes measured from the left ear of S2, as a function of frequency, on three sagittal planes, 23° ipsilateral (top), midsagittal 0° (middle), and 23° contralateral (bottom).

1. Position of spectral notches along the frequency axis changes with elevation

DTFs at different sagittal planes share a similar profile. Relatively higher intensities are observed for the planes that are more ipsilateral (negative azimuths in this case) as a result of head shadow [Fig. 4(a)]. A peak covering almost the entire frontal hemisphere below 55 kHz is interrupted by a primary spectral notch that shifts from 30 to 55 kHz with increasing elevation [Figs. 4(a), (b) and (c)] and remains

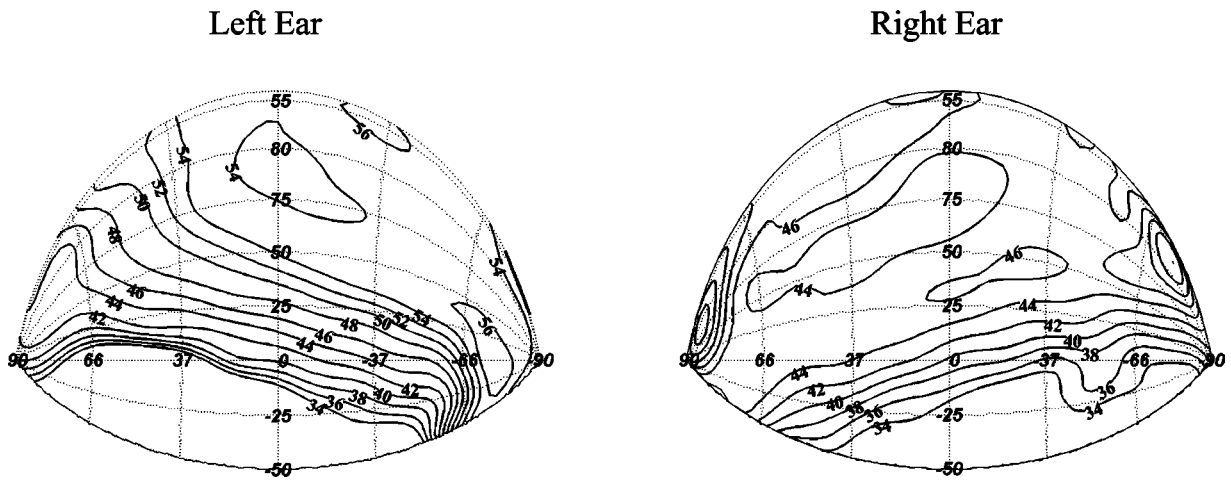


FIG. 5. Primary spectral notch center frequency as a function of elevation and azimuth, for the intact left and right ears of S1. Iso-frequency contour lines were obtained using spherical harmonic interpolation.

fixed at approximately 55 kHz at higher elevations. This notch changes roughly linearly for elevations −60° to 20° [Figs. 4(b) and (c)]. A secondary notch appears between 55 and 90 kHz over the elevation range of −60° to 0° [Fig. 4(b)] and and −60° to 20° [Fig. 4(c)], respectively. A third notch trend is observed between −50° and −20° elevations along the lower border of a second peak for frequencies above 55 kHz [Figs. 4(a) and (b)].

2. The pinna is essential for the elevation-dependent nature of the spectral notches

Figures 4(d), (e), and (f), illustrate the changes in DTF as a result of pinna removal in the same subject. As can be seen, primary and secondary notch trends are no longer present for the pinna-absent ear. Starting at 70 kHz, a linear notch trend, less visible in the intact ear measurements [Figs. 4(a), (b), (c)], shows similar elevation-dependent behavior as the primary and secondary notch trends in the lower hemisphere. Moreover, at high elevation and frequency, a different group of spectral notches surrounds a peak in this region. With the removal of the pinna, less directional dependence of spectral profile is observed.

3. Spectral notches are azimuth- and elevation dependent

The frequency where a notch reaches its lowest point is defined as the notch center frequency. In Fig. 5 the spatial changes in the frequency of the primary notch, obtained from the HRTF of S1, for both ears are represented as a function of azimuth and elevation. The contour lines in the figure represent directions that share the same primary notch center frequency in the HRTF. To distinguish the primary notch, from other notches, a clustering algorithm in the three-dimensional domain of azimuth, elevation, and frequency was applied. This algorithm groups points that are close together in this 3D space. Interpolation between the obtained points was achieved using spherical harmonic decomposition. Iso-frequency contour lines of the center frequencies for both ears extend from ipsilateral low elevations to contralateral high elevations. The center frequencies monotonically increase with elevation in the medial-frontal space in both ears.

C. Monaural spatial features

1. Spatial patterns of HRTF consist of directional lobes separated by deep spatial nulls that are co-localized with the spectral notches

In Fig. 6, intact and pinna-absent DTF magnitude contour plots at selected frequencies are compared. The figures are based on data collected from the left ear of S2. Contour maps of the intact ear differ from the ones corresponding to the pinna-absent manipulation. In the intact case, a large lobe covers almost the entire ipsilateral and part of the contralateral frontal hemisphere for frequencies below 40 kHz. Above 40 kHz, two sidelobes, in addition to the main lobe, appear. Deep spatial nulls separate the main lobes and the sidelobes.

In the pinna-absent case, the main lobe for all frequencies covers a much larger spatial area in the frontal hemisphere. At lower frequencies (32 kHz), the main lobe occupies a large portion of the contralateral side. The acoustic axis falls between 0° and −34° of azimuth, around the 0° elevation plane, for almost the entire frequency region. In general, for the pinna-absent ear not only is the spatial selectivity low (larger lobes) but also the gain across frequency. This observation is in agreement with the effect of the pinna removal on MG of HRTF and DTF in Fig. 3.

It is interesting to note the relation between the spatial nulls in the HRTF magnitude at a particular frequency, and the presence of spectral notches at that frequency. Figure 6 (left column) illustrates this relationship via overlaid spatial contour maps and spatial position of the spectral notches, white dots, at given frequencies. Note that spectral notches are co-localized with spatial nulls. This is a result of a rapid change in spatial location of the deep nulls in a small frequency range.

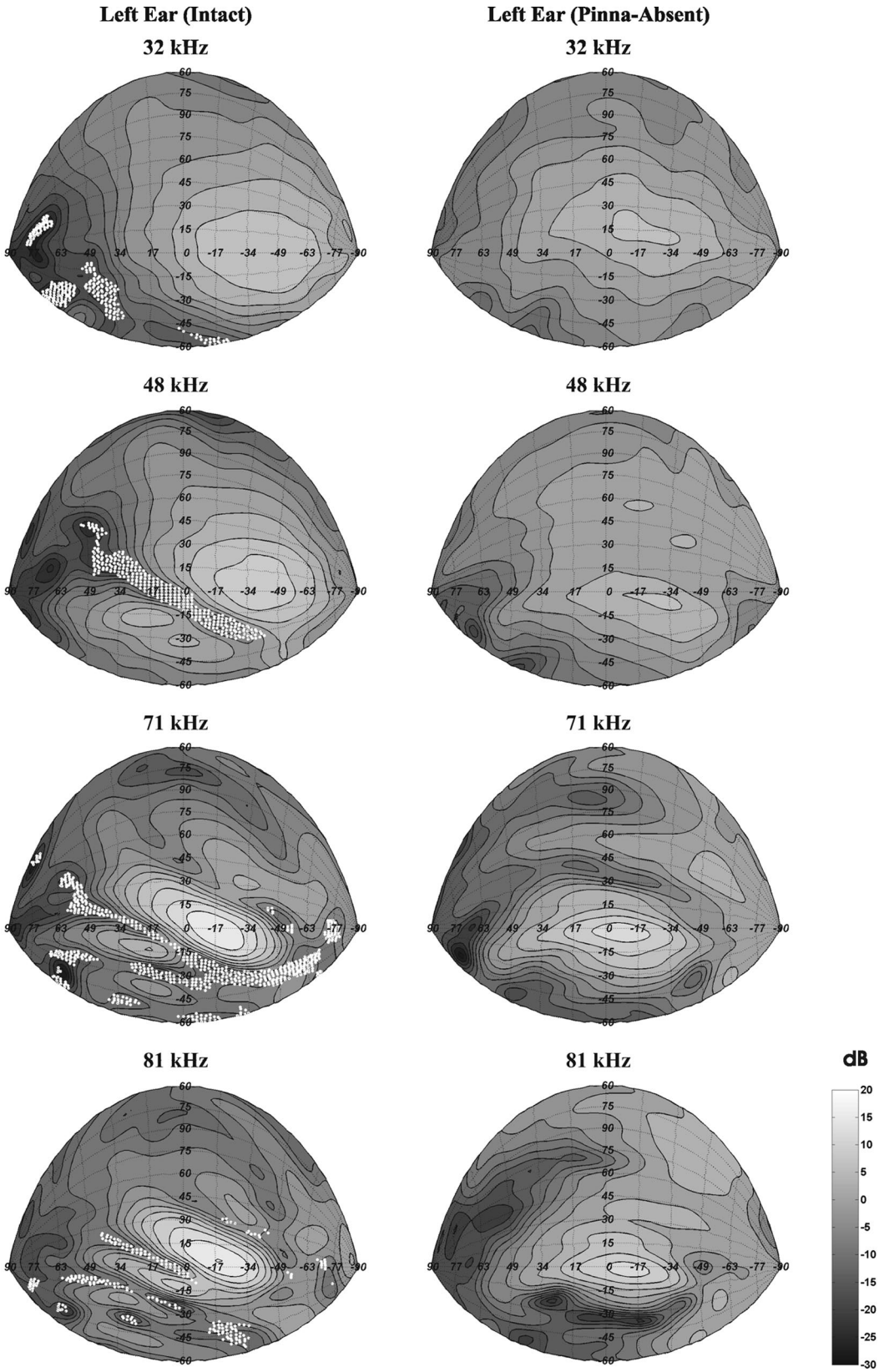


FIG. 6. Spatial contour map of HRTF magnitude at selected constant frequencies (32, 48, 71, 81 kHz) for intact and pinna-absent cases of S2. Contour lines are plotted with 3-dB separation. Overlaid on the left ear plots are white dots representing the existence of spectral notches in the corresponding HRTF, at that particular frequency.

2. Direction of the acoustic axis changes with frequency

Another important observation is that, in the intact case, the acoustic axis (maximum point of the lobe) moves to higher elevations and lower azimuths as frequency increases (Fig. 7). Figures 7(a) and (b) show the elevation and the azimuth of the acoustic axis as a function of frequency obtained from the HRTF of S2. Acoustic axis elevation shows a

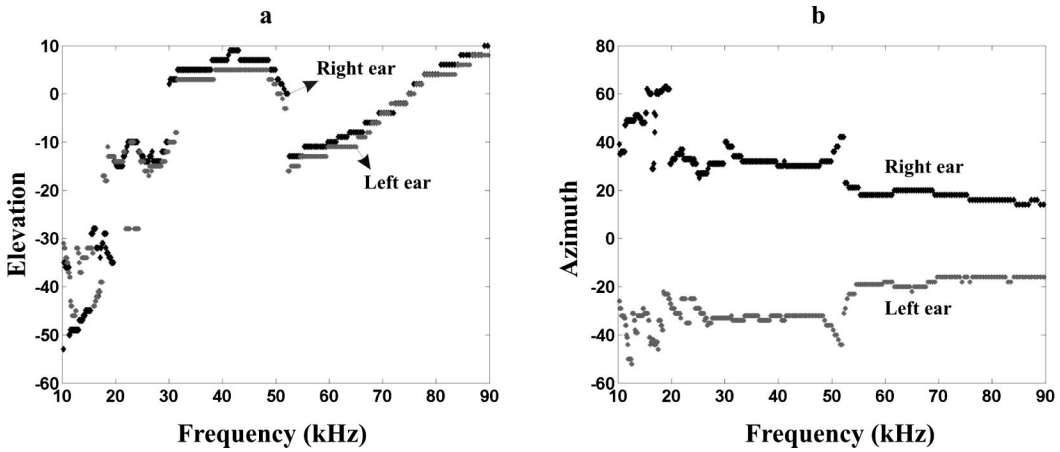


FIG. 7. Frequency-dependent changes in acoustical axis elevation (a), and azimuth (b) for the left and the right ear of S2.

linear increase from −50° at 10 kHz to 10° at 42 kHz. A second linear trend between 50 and 90 kHz also shows an increase in acoustic axis elevation from −15° to 10°. At the frequency region (40 to 50 kHz) where the first linear trend is interrupted a second spatial lobe becomes prominent, as can be seen in Fig. 6 (48 kHz). Note that behavior of the acoustic axis is very similar in both ears.

The azimuth of the acoustic axis shifts towards 0° azimuth (from +/−40° to +/−10°) as frequency increases from 20 to 90 kHz. From Fig. 7 one can conclude that external ears provide more sensitivity to the sounds arriving from directly in front of the bat at frequencies above 50 kHz.

3. The pinna and the tragus contribute to the directionality of the HRTF

One can observe in Fig. 6 that the main lobes and the sidelobes become more directional with sound frequency. We quantified the frequency-dependent increase in directionality using a directivity index.[27] The directivity index is defined as the log ratio of the acoustic power delivered by a directional receiver on the acoustic axis to that of an omnidirectional with the same sensitivity in a diffused sound field. The directionality of the spatial maps of the left ears of S1, S2, and S4, as a function of frequency, are given in Fig. 8 for manipulated (left) and control (right) ears. For all subjects spatial directionality of intact ears increases monotonically with frequency (Fig. 8, solid lines), indicating that acoustic energy is collected from a restricted directional range as the frequency increases. Thus, this finding is consistent with the observation that the HRTF becomes more spatially selective with sound frequency. Using linear regression, an increase in the directionality was computed for the intact left and right ears, respectively: 0.148 and 0.121 dB/kHz for S1, 0.13 and 0.135 dB/kHz for S2, and 0.150 and 0.139 dB/kHz for S4. Note that the slopes were slightly different between the left and the right ears of S1 and S4, consistent with the asymmetry between the left and right MGF described earlier (Fig. 2).

The effects of the manipulation of the left ear of S2 (pinna-absent) and S4 (pinna-absent and tragus-absent) on directionality of the HRTF as a function of frequency can also be seen in Fig. 8 (dashed lines for tragus-absent and dotted lines for pinna-absent cases). For both S2 and S4 [Figs. 8(b) and (c)] directionality still increases monotonically, but removal of the pinna decreases directionality. The decrease in directionality is small below 30 kHz and more apparent above 57 kHz for S4; however, for S2 it occurs over

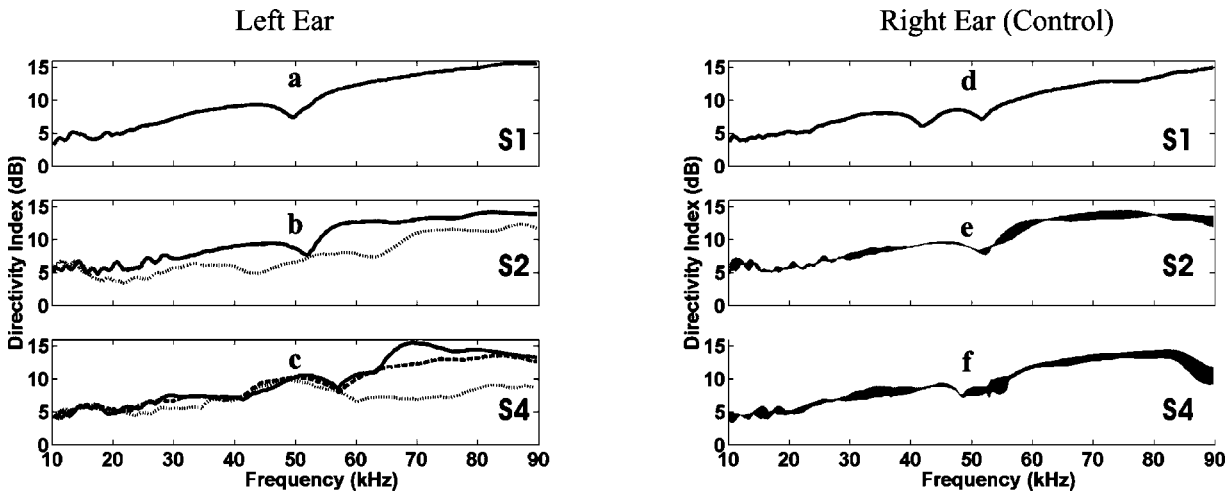


FIG. 8. Directivity index of S1, S2, and S4 for left (a), (b), (c) and right (d), (e), (f) ear. Pinna and tragus manipulations were only applied to the left ears of S2 and S4. Data are shown from the right ear as a control. Intact (solid), pinna-absent (dotted), and tragus-absent (dashed) cases are illustrated. Directivity index for the right ear depicted as lines with varying thickness representing the range of change of the functions across (different) measurements.

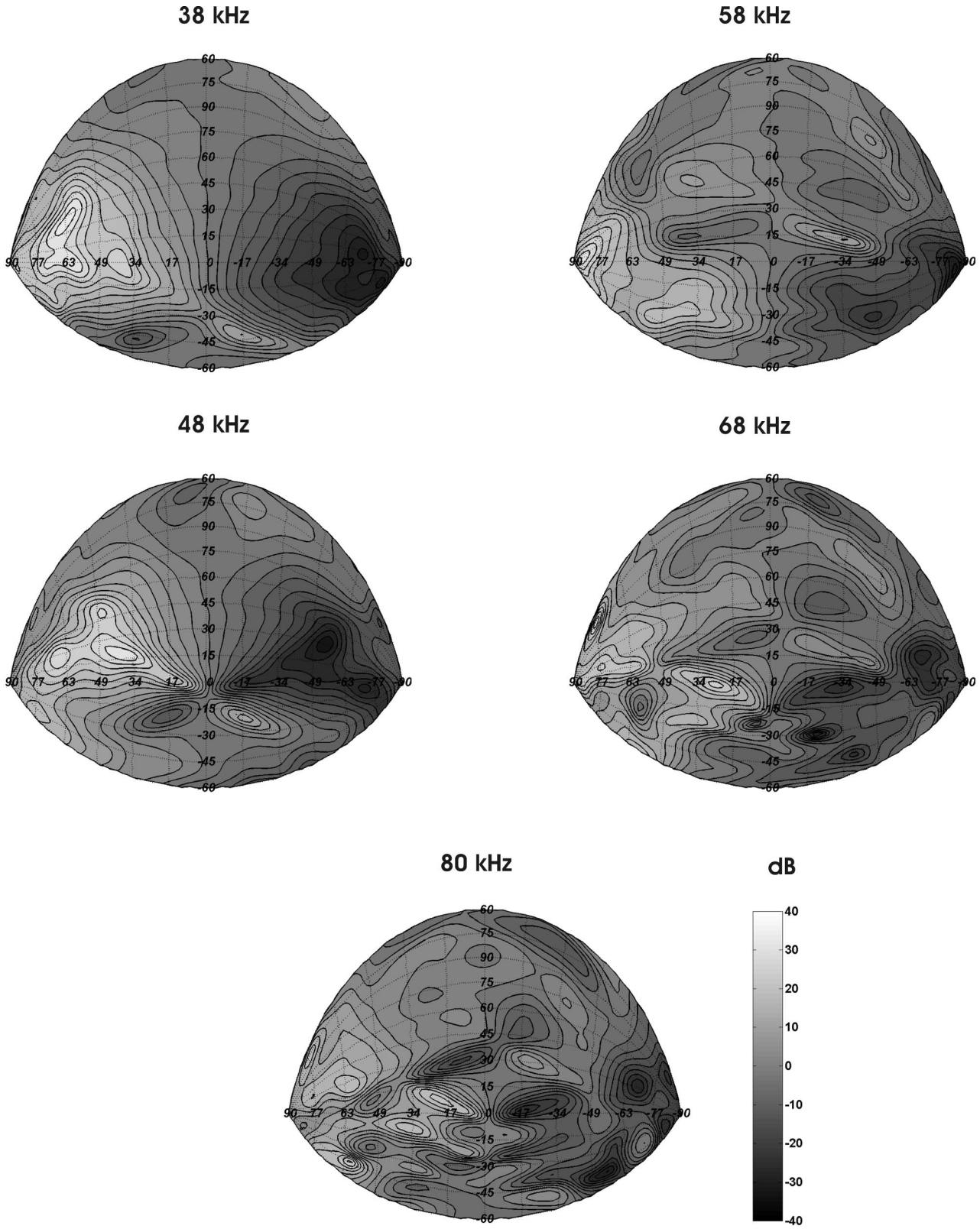


FIG. 9. Interaural level difference (ILD) spatial contour maps for 38, 48, 58, 68, and 80 kHz of S2. Contour lines are plotted with 3-dB separation.

a much larger frequency range. Removal of the tragus also causes a decrease in the directionality over 57 kHz for S4.

D. Interaural level differences

1. Spatial characteristics of the ILD becomes complex with sound frequency

Spatial contour maps of interaural level differences (ILD) at different frequencies are displayed in Fig. 9 for S2. For frequencies below 40 kHz, ILD changes monotonically from the right to the left, as a function of azimuth (Fig. 9, 38 kHz). In contrast, at higher frequencies this monotonic change no longer holds for the frontal hemisphere. In fact, at high frequencies negative ILDs (Ref. 25) were observed: negative ILD refers to measurements for which sound pressure at the contralateral ear is larger than the ipsilateral ear.

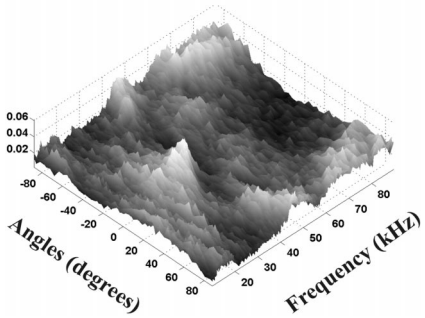


FIG. 10. Angle of spatial gradient vectors of ILD to the vector perpendicular to cone of confusion plane at each frequency for S4.

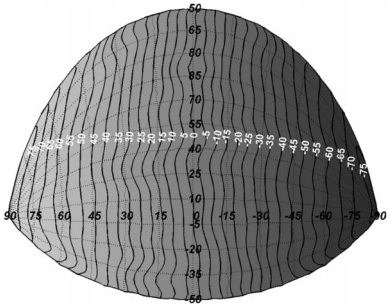


FIG. 11. Interaural time difference (ITD) spatial contour map for S1.

The contour plot for 48 kHz (Fig. 9) shows that the monotonic ILD trend along the azimuth was disrupted with a negative ILD pattern around the center (0° azimuth and elevation.). Above this frequency, negative and positive ILD patterns alternate along the vertical axis around the midsagittal plane. The occurrence of this alternating ILD pattern increases with sound frequency. Note that large ILD values also appear at directions around the mid-segittal axis.

2. ILD changes on the cones of confusion

Figure 10 shows the distribution of the angle of the spatial gradient vector of ILD with respect to the vector perpendicular to the cone of confusion surface, for each HRTF measurement location, as a function of frequency, for S4. A gradient vector's norm and orientation indicate the amount and the relative direction of composite local changes of the ILD at a given sound source direction and frequency. Having all the local gradient vectors orienting perpendicular to the surface of the cone of confusion would indicate that ILD change is independent from sound source elevation. Distributions are obtained as normalized histograms of the angles at each frequency. Normalization was undertaken by weighting each angle by a factor. The weight for each position was obtained as the ratio of the local gradient vector norm to the spatial integral of the norm of all the gradient vectors. This way the local ILD change at a particular position contributed to the distribution in proportion to the overall change in ILD on the frontal hemisphere. The histogram should show a peak around 90° if ILD changes on the cone of confusion, and it should peak around 0° if it does not. The distributions show a peak around 0°, below 40 kHz, extending to 20° on each side, suggesting a constant ILD along the cone of confusion. Above 40 kHz, histograms show peaks around 90°, suggesting that local change of ILD is dominant on the cones of confusion.

E. Interaural time differences

Figure 11 illustrates the interaural time differences (ITD) obtained from S1. ITD was calculated based on the spectrogram correlation of the impulse responses.[21] For this process a 100-channel, second-order Butterworth bandpass filter bank, consisting of filters with 1/3-octave bandwidth and envelope detectors, was used. The cross correlation obtained from outputs of the right-and the left ear frequency channel pairs was summed across frequency. The time lag of the maximum of this summed cross correlation was taken as the ITD for that sound source location. ITD changed approximately between -75 and 75 μs as the sound source moved from right to left for all the subjects.

IV. DISCUSSION

Accurate sound localization is essential for bats when they negotiate complex acoustic environments. Extraordinary sensitivity to echo delay, used by the bat for target range discrimination, has been reported (for a review, see Ref. 28). However, less is known about the acoustic cues used by bats to localize sound source direction. Several behavioral studies reported accurate azimuthal and elevational sound localization performance in echolocating bats.[2,3,6] The bats' accuracy in sound localization depends on robust localization cues created by the directional properties of the ears and/or specializations of acoustic information processing that take place in the auditory system. Studying the HRTF is an important step in our understanding of sound localization, as localization cues available to all mammals are captured entirely by the HRTF. Investigating bat sound localization from the HRTF is limited to a situation where the pinnae are immobile during passive listening. *E. fuscus* has mobile pinnae, and thus, HRTF-based analysis cannot capture the role of the pinna mobility in sound localization. Yet, our results can still apply to the localization of brief sounds when the pinnae may be immobile during the sound reception. Bats have been shown to localize brief sounds with high accuracy.[29] Thus, our results are likely to be relevant to the understanding of sound localization in echolocating bats.

Our analysis of the echolocating bat HRTF suggests that the pinna plays a role in shaping its directional features. Comparisons between the intact-and the pinna-absent ears showed that the pinna improves the directionality and the gain of the HRTF. The pinna also affects the directional properties of the spectral notches.

Spectral notches systematically changed with sound direction in the bat HRTF. There were three prominent notch trends, which were between 30 and 50 kHz, 55 and 90 kHz for sound source positions on the lower hemisphere, and 55 and 90 kHz for sound source positions on the upper hemisphere. A closer look at the spatial distribution of these notches showed that iso-frequency lines of notch center frequencies extend diagonally from contralateral at low elevations to ipsilateral at high elevations, exhibiting both elevation and azimuth dependence. Iso-frequency lines of notch

center frequencies from both ears result in a grid-like organization of these lines, as observed earlier in cats by Rice *et al.*[30]

Another salient feature of the HRTF is the presence of spatial nulls. These nulls were found in between the main lobe and the sidelobes of the HRTF. A comparison between the spatial positions of the spectral notches and the spatial nulls indicates that the two are overlapping (Fig. 6, left column). In addition to our observations, many earlier studies on the directional properties of the bat external ear also mention that the acoustical axis moves in the frontal hemisphere and increases in directionality with sound frequency.[8–10] In relation to these characteristics of the HRTF, the spatial nulls also move with the acoustical axis. These changes in spatial nulls within a small frequency region are related to a deep and sharp notch in the HRTF. Ear manipulations in this study revealed that the pinna contributes to this phenomenon.

We observed that the ILD also shows both azimuth and elevation dependence. An azimuth-dominant change in ILD appears at low frequencies, between 10 and 40 kHz. For frequencies above 40 kHz, the ILD spatial distribution was more complex and sensitive to the elevation of the sound source.

Wotton *et al.*[15] reported that the absence of the tragus causes disturbances in the primary notch trend of the bat HRTF. In a related behavioral study these authors reported tragus deflection produced a dramatic increase in the bat's vertical MAA in the frontal-lower hemisphere, where they observed the primary notch trend of the HRTF from the intact ear. They hypothesized that bats uses the primary notch trend as a cue for localization of the sound source elevation.

We found that the tragus contributes to the gain and the directionality of the HRTF at frequencies between 70 and 90 kHz. Aside from these effects, we did not observe disturbances in the vertical dependence of spectral notches in the frequency range of 30–50 kHz following tragus removal, as reported in the Wotton *et al.* study.[15]

ITD based on the measured HRTF changed between -75 and 75 μs. Although this range of change is relatively small, we believe that ITD could be a reliable cue for sound localization during echolocation. The range difference discrimination threshold of *E. fuscus* is 30–60 μs;[28] thus, it may be possible for a bat to detect echo arrival differences from the echo delay difference between the two ears using FM sweeps.

1. A model for sound localization

Earlier theoretical and experimental studies of sound localization have asserted that mammals localize sound using binaural difference cues (ILD and ITD) for azimuth and monaural spectral cues, such as spectral notches, for elevation[31] in the frontal (lower) hemisphere. However, neither the monaural cues that are investigated here nor the ILD in our data behaves in a way that is consistent with this localization scheme. By contrast, both the ILD and the spectral notches were dependent on sound source azimuth and elevation.

Our observations are in agreement with some earlier reports. Grinnell and Grinnell[12] and Fuzzessery and Pollak[13] suggested a role for ILD in vertical localization in other bat species. In addition to bats, earlier studies report elevation-dependent ILD changes in humans[32] and other mammals such as ferret.[25] However, these reports did not lead to further studies of the role of ILD on vertical sound localization in bats and other animals.

Wotton *et al.*[15] briefly mentioned a weak relation between the primary notch trend and sound source azimuth. They mentioned a possible role for the spectral notch in binaural sound localization, as was suggested in cats;[30] however, they focused on the role of the spectral notch in vertical sound localization. They reported that elevation-dependent notch trends are limited to the lower hemisphere, and suggested that this observation can explain the vertical MAA difference between high and low elevations along the vertical plane. Moreover, disturbance of the primary notch trend and the increase in the MAA at lower elevations in bats with deflected tragus were considered to provide further support for the role of the spectral notch in vertical sound localization. Combining their behavioral and HRTF data, Wotton *et al.* suggested separate sound localization cues for vertical and horizontal components of the sound source position.

We observed that the pinna contributes more to the direction-dependent spectral patterns than does the tragus. In addition, we showed that spectral notches are not limited to the lower hemisphere. Spectral notches mainly surround the acoustic energy peaks in the HRTF spatial patterns, and they can appear for sound sources at high elevations. We found that there is a coupling between the spatial nulls and the spectral notches and most important, the spatial nulls give rise to complex ILD patterns. Consequently, elevation dependence of the ILD is related to the directional properties of the spectral notches. Thus, ILD and spectral notches should not be viewed as entirely separate cues that could be used exclusively for horizontal and vertical sound localization, respectively.

We propose that both the ILD and the spectral notches contribute to localization of sound source azimuth and elevation, probably in cooperation with other spectral features of the HRTF. The fact that both cues can provide horizontal and vertical information about sound source position improves the robustness of computation for sound localization. This new scheme can replace the simple dichotomy of acoustic cues for sound localization in azimuth and elevation.

Several recent psychoacoustical findings showed that human subjects' vertical localization performance was not in accordance with the single-notch theory.[33] The single-notch theory predicts that subjects should perceive a sound source at a vertical position that depends on the frequency of the spectral notch embedded in the sound source. This prediction is based on the relationship between the sound source elevation and the notch frequency in the HRTF. Ladgendjik and Bronkhorst,[34] using virtual acoustic techniques, observed that human subjects' localization acuity did not change when the spectral notch cue was removed from HRTFs.

Our data suggest a model of sound localization in which the absence of the primary notch cue should not cause loss of acuity in vertical localization outside the midsagittal plane, since the ILD cue is still available. The fact that in our model

the spectral features in the HRTF from both ears contribute to the computation of sound source localization is also in agreement with the study by Hotman *et al.*[35] In this study it was shown that spectral cues from left and right ears are weighted to construct an elevation percept around the midsagittal plane.

We suggest that computation of sound location in the bat auditory system should be based on information received from both ears. Localization cues that were investigated here are not distinctly separate for sound source azimuth or elevation, but rather can contribute together to the estimation of sound source location.

V. SUMMARY AND CONCLUSIONS

A spatial-spectral analysis of the HRTF of *E. fuscus* is reported. HRTFs obtained from intact, pinna-absent, and tragus-absent ears are compared to understand contributions of external ear structures to the directional transformation of the sounds at the ear canals.

We made key observations that may have important consequences to our understanding of bat sound localization. We found that both ILD and spectral notches, believed to be important localization cues, are dependent on sound source azimuth and elevation

These findings lead us to a localization model for *E. fuscus*, for which both of these cues contribute to localization of sound source direction. This model provides robustness in sound localization by employing multiple cues for sound source azimuth and elevation.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Science Foundation (IBN-0111973) and by a P-30 Center Grant from the National Institute of Deafness and Other Communication Disorders (P30 DCO4664). We thank Kari Bohn, Kaushik Ghose, Timothy Horiuchi, Amy Kryjak, Aaron Schurger, Shiva Sinha, Jonathan Simon, and Rose Young for their help with data collection. We also thank Dr. Bradford May for his valuable comments on an earlier version of this manuscript.

[1] D. R. Griffin, *Listening in the Dark* (Yale University Press, New Haven, 1958).
[2] J. A. Simmons, S. A. Kick, C. Hale, C. Bard, and B. Escudie, ''Acuity of horizontal angle discrimination by the echolocating bats,'' J. Comp. Physiol. [A] **153**, 321–330 (1983).
[3] W. M. Masters, A. J. Moffat, and J. A. Simmons, ''Sonar tracking of horizontally moving targets by the big brown bat, *Eptesicus fuscus*,'' Science **228**, 1331–1333 (1985).
[4] J. A. Simmons and R. A. Stein, ''Acoustic imaging in bat sonar: Echolocation signals and the evolution of echolocation,'' J. Comp. Physiol. [A] **172**, 533–547 (1980).
[5] B. D. Lawrence and J. A. Simmons, ''Echolocation in bats: The external ear and perception of the vertical positions of targets,'' Science **218**, 481–483 (1982).
[6] J. M. Wotton and J. A. Simmons, ''Spectral cues and perception of the vertical position of targets by the big brown bat, *Eptesicus fuscus*,'' J. Acoust. Soc. Am. **107**(2), 1034–1041 (2000).
[7] G. Koay, D. Kearns, H. E. Heffner, and R. S. Heffner, ''Passive sound-localization ability of the big brown bat (*Eptesicus fuscus*),'' Hear. Res. **119**(1–2), 37–48 (1998).
[8] P. H. Jen and D. M. Chen, ''Directionality of sound pressure transformation at the pinna of echolocating bats,'' Hear. Res. **34**(2), 101–117 (1988).
[9] M. K. Obrist, M. B. Fenton, J. L. Eger, and P. A. Schlegel, ''What ears do for bats: A comparative study of pinna sound pressure transformation in chiroptera,'' J. Exp. Biol. **180**, 119–152 (1993).
[10] A. Guppy and R. B. Coles, ''Acoustical and neural aspects of hearing in the Australian gleaning bats, *Macroderma gigas* and *Nyctophilus gouldi*,'' J. Comp. Physiol., A **162**, 653–668 (1988).
[11] Z. M. Fuzessery, ''Monaural and binaural spectral cues created by the external ears of the pallid bat,'' Hear. Res. **95**(1–2), 1–17 (1996).
[12] A. D. Grinnell and V. S. Grinnell, ''Neural correlates of vertical localization by echo-locating bats'' J. Physiol. (London) **181**(4), 830–851 (1965).
[13] Z. M. Fuzessery and G. D. Pollack, ''Neural mechanisms of sound localization in an echolocating bat,'' Science **225**, 725–728 (1984).
[14] J. Blauert, *Spatial Hearing* (MIT Press, Cambridge, MA, 1997).
[15] J. M. Wotton, T. Haresign, and J. A. Simmons, ''Spatially dependent acoustic cues generated by the external ear of the big brown bat, *Eptesicus fuscus*,'' J. Acoust. Soc. Am. **98**(3), 1423–1445 (1995).
[16] A. Y. Huang and B. J. May, ''Sound orientation behavior in cats. II. Mid-frequency spectral cues for sound localization,'' J. Acoust. Soc. Am. **100**(2), 1070–1080 (1996).
[17] P. J Bloom, ''Creating source elevation illusions by spectral manipulations,'' J. Audio Eng. Soc. **25**, 560–565 (1977).
[18] J. M. Wotton, T. Haresign, M. J. Ferragamo, and J. A. Simmons, ''Sound source elevation and external ear cues influence the discrimination of spectral notches by the big brown bat, *Eptesicus fuscus*,'' J. Acoust. Soc. Am. **100**(3), 1764–1776 (1996).
[19] M. J. Ferragamo, T. Haresign, and J. A. Simmons, ''Frequency tuning, latencies, and responses to frequency-modulated sweeps in the inferior colliculus of the echolocating bat, *Eptesicus fuscus*,'' J. Comp. Physiol., A **182**(1), 65–79 (1998).
[20] E. Covey and J. H. Casseday ''The lower brainstem auditory pathways,'' in *Hearing by Bats*, edited by A. N. Popper and R. R. Fay (Springer, New York, 1995), pp. 235–295.
[21] J. C. Middlebrooks, J. C. Makous, and D. M. Green, ''Directional dependence of interaural envelope delays,'' J. Acoust. Soc. Am. **87**, 2149–2162 (1990).
[22] G. Wahba, ''Spline interpolation and smoothing on the sphere,'' SIAM (Soc. Ind. Appl. Math.) J. Sci. Stat. Comput. **2**(1), 5–16 (1981); **3**(3), 385–386(E) (1982).
[23] J. C. Middlebrooks, J. C. Makous, and D. M. Green, ''Directional sensitivity of sound-pressure levels in the human ear canal,'' J. Acoust. Soc. Am. **86**, 89–108 (1989).
[24] C. H. Keller, K. Hartung, and T. T. Takahashi ''Head-related transfer functions of the barn owl: Measurement and neural responses,'' Hear. Res. **118**(2), 13–34 (1998).
[25] S. Carlile, ''The auditory periphery of the ferret. I. Directional response properties and the pattern of interaural level differences,'' J. Acoust. Soc. Am. **88**(5), 2180–2195 (1990).
[26] D. Hammershoi and H. Moller, ''Sound transmission to and within the human ear canal,'' J. Acoust. Soc. Am. **100**(1), 408–427 (1996).
[27] L. L. Beranek, *Acoustics* (Published for the Acoustical Society of America by the American Institude of Physics, Woodbury, NY, 1988).
[28] C. F. Moss and H.-U. Schnitzler, ''Behavioral studies of auditory information processing,'' in *Hearing by Bats*, edited by A. N. Popper and R. R. Fay (Springer, New York, 1995), pp. 87–145.
[29] Z. M. Fuzessery, P. Buttenhoff, B. Andrews, and J. M. Kennedy, ''Passive sound localization of prey by the pallid bat (*Antrozous p. pallidus*),'' J. Comp. Physiol., A **171**(16), 767–777 (1993).
[30] J. J. Rice, B. J. May, G. A. Spirou, and E. D. Young, ''Pinna-based spectral cues for sound localization in cat,'' Hear. Res. **58**, 132–152 (1992).
[31] R. R. Fay and A. N. Popper, ''Evolution of hearing in vertebrates: The inner ears and processing,'' Hear. Res. **149**(1–2), 1–10 (2000).
[32] J. C. Middlebrooks, J. C. Makous, and D. G. Green, ''Directional sensitivity of sound pressure levels in the human ear canal,'' J. Acoust. Soc. Am. **86**(1), 89–108 (1989).
[33] E. A. Macpherson, ''Spectral cue processing in the auditory localization of sounds with wideband non-flat spectra,'' Doctoral dissertation, Department of Psychology, University of Wisconsin-Madison (May 1998).
[34] E. H. A. Ladgendjik and A. W. Bronkhorst, ''Contribution of spectral cues to human sound localization,'' J. Acoust. Soc. Am. **112**(4), 1583–1596 (2002).
[35] P. M. Hofman and A. J. Van Opstal, ''Binaural weighting of pinna cues in human sound localization,'' Exp. Brain Res. **148**, 458–470 (2003).

Relative contributions of temporal and place pitch cues to fundamental frequency discrimination in cochlear implantees[a)]

Johan Laneau[b)] and Jan Wouters
Lab. Exp. ORL, K.U.Leuven, Kapucijnenvoer 33, B 3000 Leuven, Belgium

Marc Moonen
ESAT-SCD, K.U.Leuven, Kasteelpark Arenberg 10, B 3001 Leuven, Belgium

(Received 9 February 2004; revised 24 September 2004; accepted 1 October 2004)

The effect of the filter bank on fundamental frequency ($F0$) discrimination was examined in four Nucleus CI24 cochlear implant subjects for synthetic stylized vowel-like stimuli. The four tested filter banks differed in cutoff frequencies, amount of overlap between filters, and shape of the filters. To assess the effects of temporal pitch cues on $F0$ discrimination, temporal fluctuations were removed above 10 Hz in one condition and above 200 Hz in another. Results indicate that $F0$ discrimination based upon place pitch cues is possible, but just-noticeable differences exceed 1 octave or more depending on the filter bank used. Increasing the frequency resolution in the $F0$ range improves the $F0$ discrimination based upon place pitch cues. The results of $F0$ discrimination based upon place pitch agree with a model that compares the centroids of the electrical excitation pattern. The addition of temporal fluctuations up to 200 Hz significantly improves $F0$ discrimination. Just-noticeable differences using both place and temporal pitch cues range from 6% to 60%. Filter banks that do not resolve the higher harmonics provided the best temporal pitch cues, because temporal pitch cues are clearest when the fluctuation on all channels is at $F0$ and preferably in phase. [DOI: 10.1121/1.1823311]

PACS numbers: 43.66.Ts, 43.66.Fe, 43.66.Hg [GDK] Pages: 3606–3619

I. INTRODUCTION

Although cochlear implants (CI) have helped many deaf people in their daily lives by providing a means for oral communication, the sound that CI patients perceive is very different from the sounds normal-hearing subjects perceive. Cochlear implant patients have, for instance, problems discriminating pitch (Gfeller and Lansing, 1992; Pijl, 1997; Geurts and Wouters, 2001). Geurts and Wouters found just-noticeable differences (jnd) of fundamental frequency ($F0$) from 4% up to 13% in four Laura cochlear implant subjects (Geurts and Wouters, 2001). These jnd's are very poor compared to normal-hearing subjects who typically perceive differences of less than 0.5% (Moore and Glasberg, 1990). However, pitch perception is very important for intonation, speaker identification, music perception, and speech perception in tonal languages. In an attempt to improve the $F0$ discrimination for cochlear implant recipients, this study assessed the effects of the filter bank design on the discrimination of the fundamental frequency of speech-like harmonic sounds. The filter banks tested in the present experiment differ in cutoff frequency distribution in the low frequencies, amount of overlap between filters, and filter shape.

Pitch cues can be elicited using two independent mechanisms in cochlear implants: place pitch cues and temporal pitch cues (Tong *et al.*, 1982; Shannon, 1983; Townshend *et al.*, 1987; McKay *et al.*, 2000). Temporal pitch cues arise when the repetition rate of stimulation on one channel changes. The temporal pitch sensation rises with increasing rate up to 300 Hz but saturates at higher rates (Shannon, 1983). Place pitch cues arise when the site of stimulation is changed while keeping the stimulation rate constant, with more basal stimulation eliciting higher pitches.

It has been shown that both temporal and place pitch cues enable CI recipients to some degree to perceive $F0$ differences of synthetic harmonic sounds with currently used sound processors (Geurts and Wouters, 2001), and in normal-hearing subjects with acoustic models of CI processors (Green *et al.*, 2002). Geurts and Wouters (2001) measured the discrimination of $F0$ of synthetic steady-state vowels in subjects implanted with the 8-channel Laura implant. Their results indicated that even without temporal pitch cues, only preserving place pitch cues, some discrimination of $F0$ was possible although the jnd's were very large. Without temporal pitch cues the jnd's were on average around 50% but varied over conditions and subjects from 5% to more than 100%. With the inclusion of temporal information, the jnd's decreased significantly, as would be expected. Green *et al.* (2002) examined the discrimination of CI simulations of frequency glides in sawtooth waves and synthetic diphthongs in normal-hearing subjects. They studied the influence of both place and temporal pitch in modulated noise-band vocoder CI simulations, and concluded that both pitch mechanisms contribute to the discrimination of $F0$. Subjects were able to integrate both cues as the inclusion of both pitch cues produced the highest discriminative performance. However, when the original sounds were diphthongs, the inherent variation of formant structures perturbed the place pitch cue.

The current sound processing in cochlear implants has a

[a)]Portions of this work were presented in "Pitch perception in cochlear implants with different filter bank designs," Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Cancun, Mexico, 2003.

[b)]Electronic mail: johan.laneau@med.kuleuven.ac.be

TABLE I. Relevant information about each of the subjects who participated in the experiments. The last column shows the results for a phoneme recognition task with monosyllabic consonant–vowel–consonant (CVC) words of the NVA list (Wouters *et al.*, 1994).

Subject	Duration of profound deafness (y)	Etiology	Implant type	Speech processor	Clinical speech processing strategy	Age (y)	Implant experience (y;m)	Phoneme recognition (%)
S1	11	Hereditary	Nucleus CI24R(ST)	ESPrit	Speak	19	1;11	74%
S2	3	Unknown	Nucleus CI24R(ST)	ESPrit	Speak	21	3;2	86%
S3	30	Progressive	Nucleus CI24R(CS)	SPrint	ACE	54	0;8	84%
S4	32	Progressive	Nucleus CI24R(CS)	ESPrit 3G	ACE	49	0;6	52%

negative effect on $F0$ discrimination, since CI subjects appear to be able to estimate musical intervals more accurately for pulse trains than for real musical sounds processed by their speech processors (Pijl, 1997). Part of this limited $F0$ discrimination ability in CI recipients using their speech processors may be due to the limited transmission of the fine temporal information in current speech processors, since most current speech processors only extract the slowly varying envelope and use this to modulate a constant rate pulse train (Smith *et al.*, 2002).

Because of this limited use of fine temporal pitch cues in current CI sound processors, previous studies have tried to improve the transmission of temporal pitch cues in order to improve the $F0$ discrimination in cochlear implants (Jones *et al.*, 1995; Geurts and Wouters, 2001; Barry *et al.*, 2002). However, no study resulted in a significant improvement of $F0$ discrimination. Explicit presentation of the fundamental frequency by stimulating the most apical channel with a pulse train at the rate of $F0$ did not improve the perception of suprasegmental cues such as intonation (Jones *et al.*, 1995). Barry *et al.* (2002) assessed the effect of pulse rate on the perception of tones in Cantonese-speaking children by comparing a group of children using the SPEAK strategy to a group of children using the higher rate ACE strategy. The hypothesis was that higher rates better preserve the temporal code (Wilson, 1997; Rubinstein *et al.*, 1999), but the authors found no significant differences between the strategies. Indeed, the group of children with the ACE strategy performed slightly (but nonsignificantly) worse than the group of SPEAK users. Geurts and Wouters (2001) developed a variation to the CIS sound-processing scheme that enhanced the modulation depth of the beating of the output of the filters, and also did not find a significant difference when compared to the standard CIS in the discrimination of $F0$ of synthetic vowels.

In a recent study, Geurts and Wouters (2004) showed that a new filter bank can improve the place pitch cues for $F0$ discrimination of synthetic vowels. Their newly proposed filter bank had more resolution in the lower frequencies and more filter overlap between different channels compared to the reference filter bank in that study. An important improvement in $F0$ discrimination based only upon place pitch cues was obtained but it remains, however, unclear whether this is due to the increased frequency resolution, the smooth transition between filters, or the increased filter overlap. With the inclusion of temporal cues, the improvement was small or even absent.

The present study investigates which aspect of the filter bank (cutoff frequency distribution, filter shape, or filter overlap) has the strongest effect on $F0$ discrimination for both place pitch cues and temporal pitch cues. It also investigates which physical properties of the stimuli are related to the place pitch cues that facilitate $F0$ discrimination, and which lead to clearer temporal pitch cues.

The effect of four filter banks on the voice-range pitch discrimination is compared in cochlear implant users. In order to separate the effects of place pitch cues and temporal pitch cues on complex pitch perception, two experimental conditions were created. In the first condition, temporal pitch cues were excluded by low-pass filtering the stimulation envelopes at 10 Hz, and consequently only place pitch cues were present. In the second condition both place and temporal pitch cues (up to approximately 200 Hz) are present. Performance on $F0$-discrimination task was measured as a function of relative difference of fundamental frequency, using a constant-stimuli 2I2AFC pitch-ranking experiment. Four Nucleus CI24 cochlear implant subjects took part. The harmonic signals were synthetic stylized vowel-like single-formant stimuli.

II. METHODS

A. Subjects

Four experienced postlingually deafened users of the Nucleus CI24 cochlear implant participated in this study. All subjects performed all tests and were paid for their collaboration. Some relevant details about the subjects can be found in Table I. Subjects S1 and S2 are implanted with the straight version of the Nucleus electrode array CI24R(ST) Nucleus24k and use a SPEAK strategy in daily life. Subjects S3 and S4 are implanted with the perimodiolar version of the Nucleus electrode array CI24R(CS) Nucleus24 Contour and

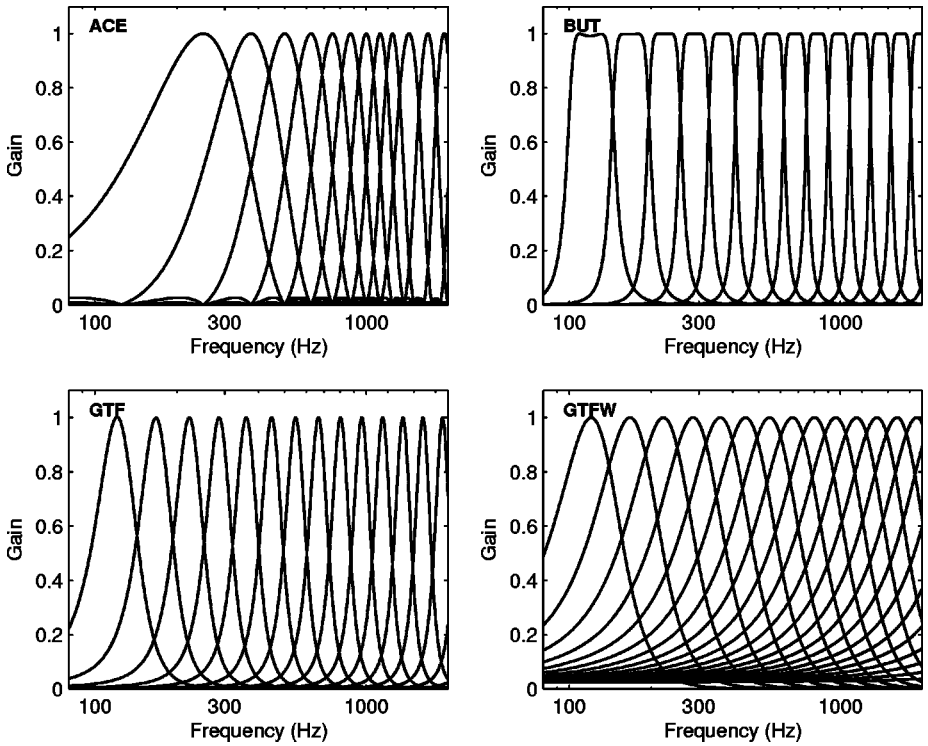


FIG. 1. The frequency response of the different filter banks examined in the present study. The ACE filter bank is the filter bank implemented in the commercial Sprint speech processor. The GTF filter bank is based upon the physiology of the normal ear and has a higer resolution in the lower frequencies. The BUT filter bank is based upon the GTF filter bank but with more rectangular filter shapes. The GTFW is also based upon the GTF filter bank but the bandwidth of the filters of the GTFW filter bank were set not to resolve any overtone. Because the stimuli in the present study contained virtually no energy in higher frequencies, the frequency response is represented up to 2000 Hz for clarity.

use an ACE strategy (900 pps per channel) in daily life. All subjects are relatively good performers with their implants in daily life.

B. Sound processing

Four different filter banks were used to process the stimuli. The frequency response of the four filter banks is depicted in Fig. 1. The filter banks differ in spectral resolution for the low frequencies and in the shape of the individual filters. All filter banks have 22 bands corresponding to the number of channels in the Cochlear Nucleus CI24 using monopolar stimulation mode. One filter bank (ACE) is at present most commonly used clinically for Nucleus CI24, and is included in the present experiment as a reference for current performance of cochlear implant subjects. The three other filter banks are specifically designed for this experiment and resemble more closely the frequency analysis of the normal human ear. All three filter banks have more resolution in the low frequencies compared to the ACE filter bank. The GTF filter bank is based upon a model for the filtering of the normal ear (Patterson *et al.*, 1995). The BUT and GTFW are derived from this filter bank by changing the filters' shape and bandwidth, respectively.

The first filter bank (ACE) is the filter bank of the standard commercial speech-processing strategy (Cochlear Corp: ACE) of the Cochlear SPrint speech processor implemented in MATLAB (Cochlear Ltd, 2002). It consists of a 128-point FFT with a Hamming window, leading to 64 frequency bins. The first two bins, containing the dc and the lowest frequencies, are discarded. The following eight bins are assigned to the first eight channels of the filter bank, respectively. The higher bins are summed in magnitude in groups of two or more to form the rest of the channels, leading to an approximately logarithmic frequency distribution in the higher channels.

The GTF, BUT, and GTFW filter bank all consist of 22 fourth-order IIR bandpass filters. The filters in the GTF filter bank were fourth-order gammatone filters using Slaney's implementation (Slaney, 1993). The center frequencies of the gammatone filters were set corresponding to the middle of 22 equally long sections along the cochlea ranging from 100 to 8000 Hz in total. The transformation from frequency to distance along the cochlea was calculated by Greenwood's formula for a cochlear length of 35 mm and $\alpha = 0.06$ (Greenwood, 1990). The bandwidth of the filter was set to the equivalent rectangular bandwidth (ERB) of the auditory filter corresponding to its center frequency (Glasberg and Moore, 1990).

The BUT filter bank is similar to the GTF filter bank; however, the shape of the filters is more rectangular (a flat frequency response in the passband and 24-dB/octave slopes). The BUT filter bank consists of 22 fourth-order Butterworth filters, of which the cutoff frequencies were equidistantly spaced along the length of the cochlea with cutoff frequencies ranging from 100 to 7800 Hz in total. The highest cutoff frequency was set slightly below the Nyquist frequency to avoid numerical problems. Both the GTF and the BUT filter bank resolve one or two of the lowest harmonics for most of the stimuli in the present study. The GTFW filter bank is based upon the GTF filter bank but the bandwidth of the filters is increased so that all overtones are unresolved. This caused the output of every filter for a harmonic sound to beat at the fundamental frequency. The GTFW was implemented identically to the GTF filter bank except that the bandwidth was set at half the filter's center frequency.

In order to separate the effects of temporal and place pitch cues, the envelopes of the filter outputs were obtained

in two different ways. In one condition, including both temporal and place pitch cues, the envelope contained modulations up to 200 Hz. In the other condition, including place pitch cues but without temporal pitch cues, the envelope only contained modulations up to 10 Hz. The deletion of modulations at frequencies above 10 Hz removes all fluctuations that generate temporal pitch percepts, although listeners can still understand speech, at least when presented in a quiet background (Drullman *et al.*, 1994; Shannon *et al.*, 1995).

In the ACE filter bank, the envelope is extracted by taking the magnitude of the complex-valued output. The bin width of the FFT filter bank limits the maximal modulation frequency after this envelope extractor to approximately 200 Hz. In the condition without temporal pitch cues these envelopes were, in addition to the standard ACE processing, low-pass filtered with a fourth-order Butterworth filter at 10 Hz.

The envelopes for the GTF, BUT, and GTFW were extracted using a half-wave rectifier and a fourth-order Butterworth low-pass filter. The cutoff frequency of the low-pass filter was set at 200 Hz when temporal pitch cues were transmitted and at 10 Hz when temporal pitch cues were removed.

The subsequent processing was identical for all filter banks. First, the signals were resampled from the audio sampling rate (16 kHz) to the channel stimulation rate (900 Hz). Then, only the eight most intense channels per time slot were selected for stimulation. Finally, the amplitude of the signals is compressed using the standard compression function of the Nucleus MATLAB Toolbox (Cochlear Ltd., 2002), in order to map the acoustic input level onto the subjects' electrical dynamic range. The compression function is defined by

$$y(x)=\begin{cases} 0 & \text{for } 0\leqslant x<0.0156 \\ \log\left(1+416.2\dfrac{x-0.0156}{0.5703}\right)\Big/\,6.033 & \text{for } 0.0156\leqslant x<0.5859. \\ 1 & \text{for } 0.5859\leqslant x\leqslant 1 \end{cases} \tag{1}$$

In formula (1) x is the amplitude of the extracted envelopes for each channel and y is the compressed amplitude expressed in proportion of the dynamic range expressed in clinical current units of the subject for that channel. The obtained envelopes were used to modulate pulse trains of 900 pulses per second with biphasic pulses of 25-μs phase width and 8-μs interphase gap. The pulses were delivered in monopolar mode with both external return electrodes active (MP1+2).

C. Stimuli

Original stimuli before sound processing were synthetic harmonic complexes produced by the excitation of a single-formant resonator by a periodic signal, and resemble highly stylized vowels. The generation of the single-formant stimuli (SFS) was based upon the Klatt speech synthesizer (Klatt, 1980). The stimuli were generated by filtering a 500-ms pulse train sampled at 16 kHz and ramped on and off with a raised cosine over 10 ms with a cascade of two filters.

The first filter models the voice source and its output is similar to the typical glottal volume velocity waveform. It is implemented as a second-order low-pass filter with a cutoff frequency of 50 Hz (Klatt, 1980). The second filter is a resonator that shapes the single formant of the stimuli. It is implemented as a second-order bandpass filter, of which the transfer function is defined by

$$H(z)=\frac{A}{1-Bz^{-1}-Cz^{-2}}$$

$$\text{with}\quad \begin{aligned} C&=-\exp(-2.\pi B_w T) \\ B&=2\exp(-\pi B_w T)\cos(2\pi F_r T), \\ A&=1-C-B \end{aligned} \tag{2}$$

where T is the sampling period (62.5 μs) and F_r and B_w are the filter's center frequency and bandwidth, respectively (Klatt, 1980).

The fundamental frequency of the stimuli used in the present study ranged from 133 Hz up to 558 Hz. The center frequencies of the bandpass filter, or formant frequencies ($F1$), were 300, 350, 400, 450, or 500 Hz, and the bandwidth of the formant resonator was fixed at 100 Hz for all stimuli. The stimuli were first equalized in rms and then scaled with a common factor so that most of the signals nearly optimally used the input dynamic range of the compression function [Eq. (1)] but none of the signals was clipped. Figure 2 shows the spectrum of two representative stimuli.

D. Psychophysical procedure

Fundamental frequency discrimination was measured in a two-interval, two-alternative forced choice (2I2AFC) procedure using the method of constant stimuli. The subjects were instructed to indicate which of the two intervals contained the stimulus with higher pitch, while ignoring the loudness cues. The intervals were separated by a 500-ms silent gap and no feedback was provided, in order to force the subjects to base their judgments on a perceptual dimension that corresponded to their intuitive sense of pitch. This further reduced the likelihood of any residual loudness cues affecting their judgments.

Both stimuli in a trial had the same formant frequency. The $F0$ of one stimulus in each 2I2AFC trial was equal to the reference frequency, and that of the other was set at a higher frequency. This reference frequency was 133 or 165 Hz. These frequencies lie in the range of normal adult voice,

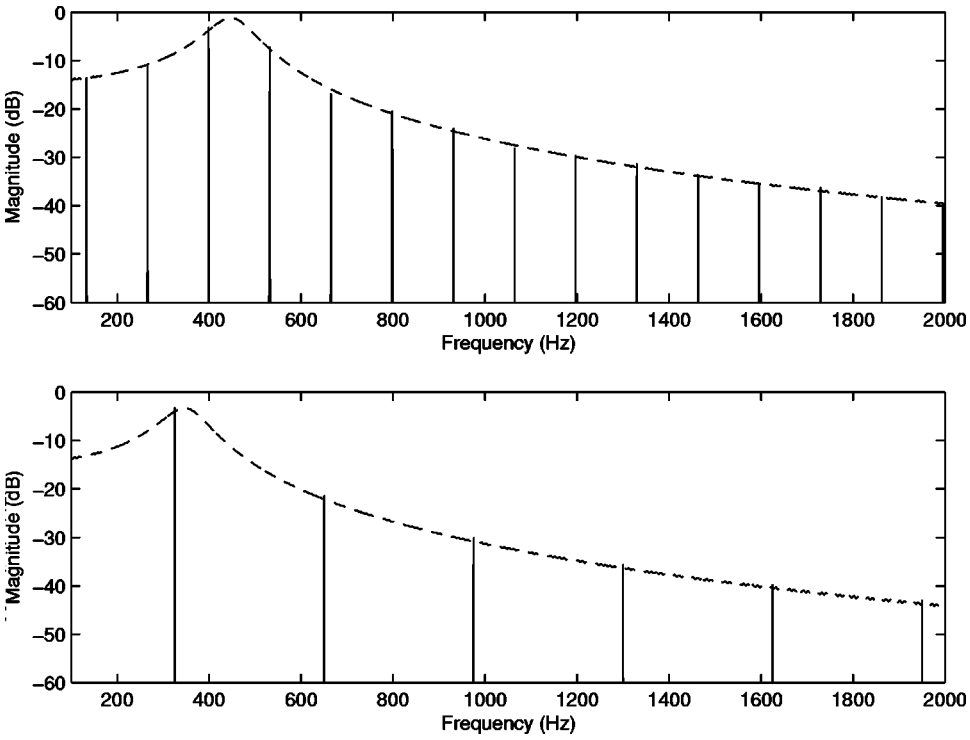


FIG. 2. The spectra of two representative stimuli used in the present study. The upper panel shows the spectrum of a single-formant stimulus with formant frequency 450 Hz and fundamental frequency 133 Hz. The lower panel shows the spectrum of a single-formant stimulus with formant frequency 350 Hz and fundamental frequency 325 Hz. The dotted line in each panel shows the combined frequency response of the two filters that shape the spectrum (see the text). The plot of the frequency response is limited up to 2000 Hz for clarity.

and 133 Hz is set near the first crossover between filters for the IIR filter banks, and 165 Hz is set near the center of the second filter of the IIR filter banks. The $F0$ frequencies of the eight comparison stimuli were 140, 150, 165, 180, 210, 250, 325, and 450 Hz for the reference stimuli with an $F0$ of 133 Hz. For the reference stimuli with an $F0$ of 165 Hz, the eight comparison stimuli had $F0$ frequencies of 174, 186, 205, 223, 261, 310, 403, and 558 Hz. The relative differences between $F0$ are the same for both reference $F0$ conditions and range from 5.2% up to 238%. The formant frequency was kept constant within trials because stimuli with the same $F0$ but with their main energy in different frequency regions can lead to differences in pitch (Ohgushi, 1978).

The trials were presented in blocks. All 16 conditions resulting from the combinations of type of envelope extraction (with or without temporal pitch cues), filter bank (ACE, GTF, BUT, or GTFW), and reference $F0$ (133 or 165Hz), were presented in separate blocks. Each block contained trials with all eight comparison frequencies for all five formant frequencies. These 40 different trials were repeated twice, leading to 80 trials per block. Different formant frequencies were included per block in order to prevent subjects from identifying and learning the reference stimulus based upon other cues. Every block was presented twice to every subject. Care was taken that blocks processed with the ACE, GTF, and BUT filter bank were presented in random order. The blocks processed with the GTFW filter bank were presented at the end of the experiment.

To reduce the influence of loudness cues, amplitude was roved by multiplying the amplitude of all stimuli with a random factor between 0.85 and 1.1, with amplitude expressed in clinical current units and relative to the dynamic range. An exception was made for subject S4 in the condition of the ACE filter bank without temporal pitch cues and the reference $F0$ of 165 Hz, where the subject judged the loudness of the stimuli as uncomfortably loud using the default amplitude roving. In this one case, the amplitude was roved with a factor between 0.77 and 1.

III. RESULTS

The proportion of times the comparison stimulus was judged higher than the reference stimulus, pooled over all formant frequencies, is shown in the absence and presence of temporal pitch cues by Figs. 3 and 4, respectively. The top four rows of panels in each figure show the results for a single subject, and the mean data are shown in the bottom, fifth row of panels. The left column in every figure shows the results for the condition with a reference $F0$ of 133 Hz and the right column shows the results for the condition with a reference $F0$ of 165 Hz. The proportions depicted in Figs. 3 and 4 are based upon 20 presentations per data point.

A repeated measures analysis of variance (ANOVA) was performed on the proportions of correctly ranked trials using five factors. The independent variables were presence of temporal pitch cues (absent or present), reference $F0$ (133 or 165 Hz), filter bank (ACE, GTF, BUT, or GTFW), relative $F0$ difference (eight levels), and formant frequencies (five different $F1$). The main effects of presence of temporal pitch cues ($p=0.002$) and relative $F0$ difference ($p=0.007$) were significant. The other main factors were not significant: reference $F0$ ($p=0.678$), filter bank ($p=0.251$), and formant frequency ($p=0.122$). The latter result allows averaging over the different formant frequencies in the presentation of the data in Figs. 3 and 4. Only the interaction term between filter bank and presence of temporal pitch cues was significant ($p=0.015$). No other interaction terms were significant, but there were significant intersubject differences ($p<0.001$).

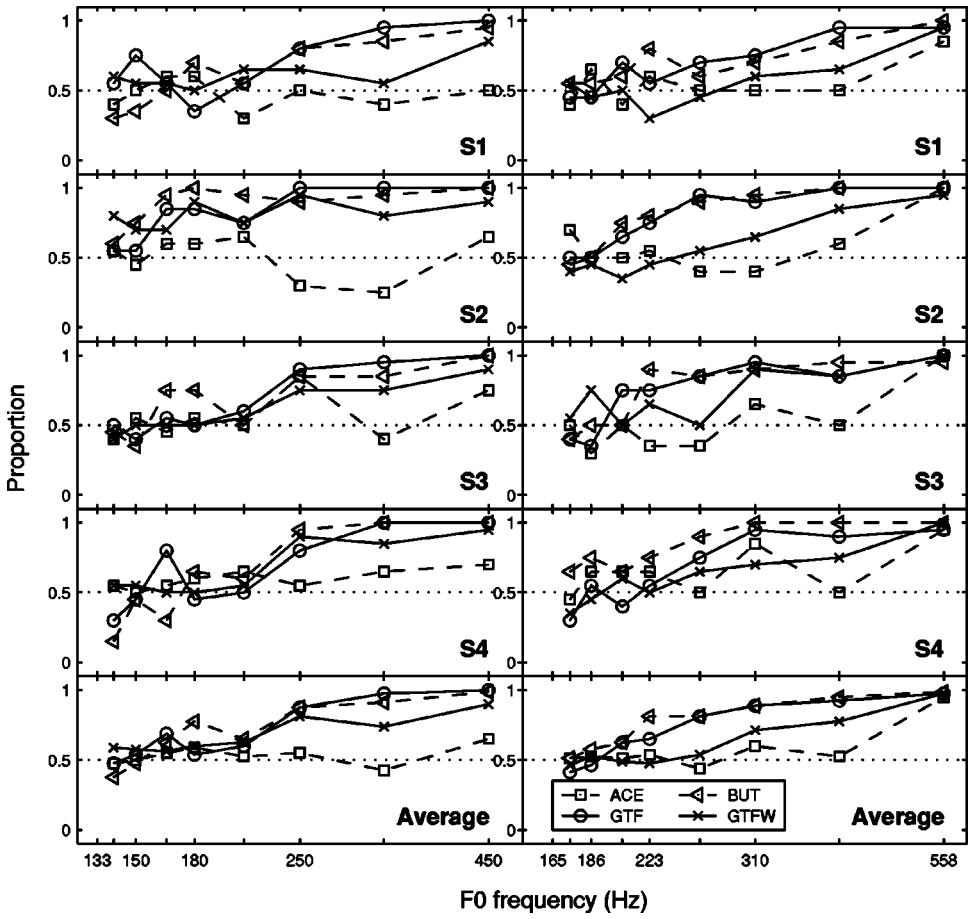


FIG. 3. The average proportion of correctly pitch-ranked stimuli pooled over the various formant frequencies. The different lines in the figures represent the results for the various filter banks examined in the present study, with all temporal fluctuations above 10 Hz filtered out. The fundamental frequencies of the signals are plotted on the horizontal axis on a log scale. The left and right columns of panels depict the results for the condition where the reference $F0$ was 133 and 165 Hz, respectively. Rows 1 to 4 represent a different subject. The bottom panels represent the average of the four subjects. The dotted line indicates the chance level.

Figures 3 and 4 indicate that subjects were able to rank the stimuli based upon their fundamental frequency. They also show that as the relative $F0$ difference increases the difference is perceived more clearly. Also, comparing Figs. 3 and 4, the results indicate that providing temporal fluctuations in the envelope of the channels up to 200 Hz improves the ability of the subjects to discriminate $F0$ differences. This improvement is, however, dependent on the filter bank used, and the largest improvement is found for the ACE filter bank.

Because of the significant interaction between presence/absence of temporal pitch cues and the filter banks, the results are presented in two parts: with or without temporal pitch cues.

A. Without temporal pitch cues

A four-factorial repeated-measures analysis of variance was performed on the proportions of correctly pitch-ranked trials that were collected with stimuli without temporal pitch cues. The main effects of filter banks ($p=0.030$) and relative $F0$ difference ($p=0.010$) were significant. The effects of formant frequency and reference $F0$ were not significant, as well as all interaction terms. The between-subject differences were significant ($p<0.001$). Pairwise comparisons were performed using the least-significant difference (LSD) on the filter bank design. The GTF and BUT filter bank produced significantly better performance than the ACE filter bank ($p=0.035$ and $p=0.018$, respectively for GTF and BUT) and the GTFW filter bank ($p=0.049$ and $p=0.015$, respectively for GTF and BUT), but there was no significant difference between BUT and GTF. The GTFW filter bank allowed for significantly better performance than the ACE filter bank ($p=0.046$).

The results indicate that even without temporal pitch cues it was still possible for the subjects to rank the stimuli based on $F0$. However, it must be emphasized that this was only possible for relatively large differences in $F0$ of about one octave for the BUT and GTF filter bank and more than one octave for the GTFW filter bank. For the ACE filter bank, the subjects found it almost impossible to use place pitch cues to rank $F0$, even when stimuli differed more than 20 semitones.

B. Including temporal pitch cues

A four-factorial repeated measures ANOVA was performed on the proportions of correctly ranked trials where the stimuli now contained temporal pitch cues. The main factors reference $F0$ ($p=0.006$) and relative $F0$ difference ($p=0.013$) had a significant effect. No other main effects were significant. The effect of the filter bank was not significant ($p=0.126$).

The discrimination was significantly poorer at the higher reference $F0$ tested. This result agrees with previous studies showing that the effectiveness of temporal pitch cues decrease with increasing overall rate (Shannon, 1983).

Three of the four subjects (S1, S3, and S4) did not reach perfect performance for the ACE filter bank and reference

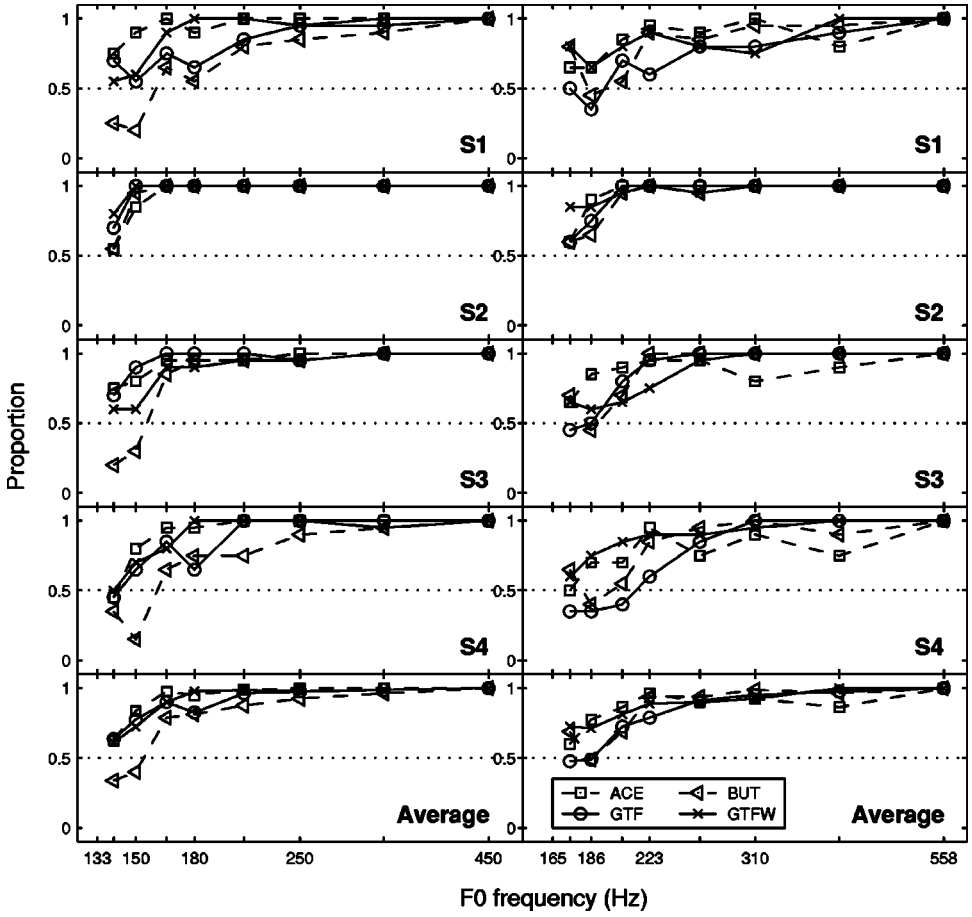


FIG. 4. The average proportion of correctly pitch-ranked stimuli pooled over the various formant frequencies. The different lines in the figures represent the results for the various filter banks examined in the present study, with all temporal fluctuations included up to 200 Hz. The fundamental frequencies of the signals are plotted on the horizontal axis on a log scale. The left and right columns of panels depict the results for the condition where the reference $F0$ was 133 Hz and 165 Hz, respectively. Rows 1 to 4 represent a different subject. The bottom panels represent the average of the four subjects. The dotted line indicates the chance level.

$F0$ of 165 Hz until the relative $F0$ difference was more than one octave. Nevertheless, the performance was above chance level for relatively small $F0$ differences. This indicated that further increasing the $F0$ difference did not make the two processed stimuli perceptually more dissimilar. This was probably due to the poor resolution of the ACE filter bank in the $F0$ range causing $F0$ to be encoded solely by temporal pitch cues, whose effectiveness was limited at high $F0$'s. This limitation might have been caused by the limited modulation depth of the envelopes at higher frequencies or by the reduced perceptual differences for temporal pitch cues at higher rates. The three IIR filter banks (GTF, BUT, and GTFW) did not suffer from this shortcoming because, for larger $F0$ differences, the pitch discrimination was most probably also mediated by place pitch. The difference in results between the condition without temporal pitch cues and the condition with both place and temporal pitch cues quantifies the influence of temporal pitch cues present in the presented stimuli. To assess this, the proportions of correctly ranked trials were averaged over the four subjects and converted into d' values for all conditions and all stimuli. To prevent infinite d' values for perfect responses, the perfect responses are adjusted by introducing an error of half a trial as suggested by MacMillan and Creelman (1991). The amount of perceptual difference, introduced by the temporal pitch cues, is the difference of the d' values between the condition with temporal pitch cues ($d'_{\text{place+temp}}$) and the condition without temporal pitch cues (d'_{place}). Assuming that both temporal and place pitch cues are optimally used and independent (McKay *et al.*, 2000), and taking into account the possibility of pitch reversals, the amount of temporal pitch cues is calculated by

$$d'_{\text{temp}} = \text{sign}(X) * \sqrt{|X|} \quad \text{with} \quad X = \text{sign}(d'_{\text{place+temp}}) \times (d'_{\text{place+temp}})^2 - \text{sign}(d'_{\text{place}})(d'_{\text{place}})^2. \tag{3}$$

This formula simplifies to $d'_{\text{temp}} = \sqrt{d'^2_{\text{place+temp}} - d'^2_{\text{place}}}$ for discrimination experiments as stated in McKay *et al.* (2000) because in discrimination experiments only positive d' values are possible. However, in a pitch-ranking experiment pitch reversals can occur and these lead to negative d' values. Moreover, in a pitch-ranking test the subjects are forced to project their percepts onto a single axis (''the combined pitch axis''). When temporal and place pitch cues of two compared stimuli are conflicting the two stimuli are discriminable (McKay *et al.*, 2000), but ranking them in pitch is at least very hard and might possibly lead to a d' value close to zero.

All data points where both conditions, with and without temporal pitch cues, resulted in perfect discrimination were discarded because these points did not provide any information on the effect of adding temporal pitch cues. The additional perceptual effect based upon the temporal pitch cues is shown in Fig. 5. Data points represent averages over formant frequency and reference $F0$. The temporal pitch cues become clearer as the $F0$ difference increases for all filter banks, but the clarity of the temporal pitch cues differs between filter banks. The ACE and GTFW filter bank, having broad filters,

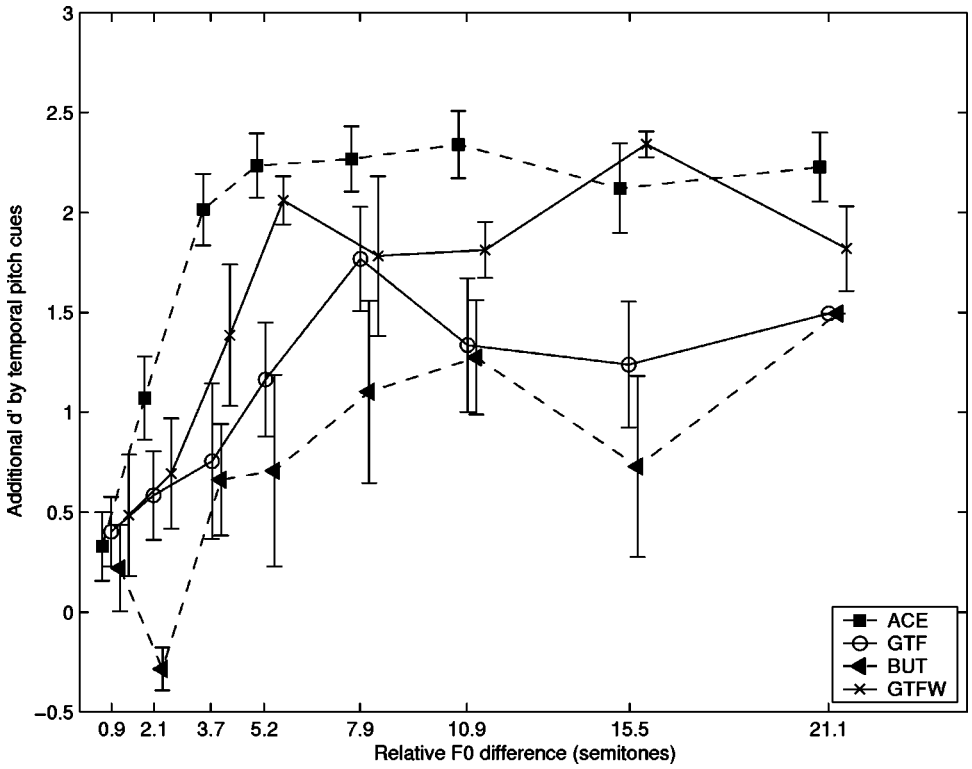


FIG. 5. The additional d' caused by the inclusion of temporal pitch cues in the processed stimuli, for the four filter banks tested in this study. The additional d' values are calculated based upon the average response of the four CI subjects, and are derived from the difference in d' between the corresponding data points with and without temporal pitch cues (see the text for details). The additional d' is averaged over the five different formant frequencies and over both reference $F0$ conditions. The error bars indicate the standard error of the mean. The data points are slightly shifted along the x axis for clarity.

allow for the largest increase in performance based upon temporal pitch cues. The GTF and BUT filter bank, which have narrower filters, exhibit less increase of the discrimination performance possibly because the baseline performance was lower.

IV. DISCUSSION

A. Without temporal pitch cues

Previous research has shown that $F0$ discrimination can take place in the absence of temporal pitch cues in CI subjects (Geurts and Wouters, 2001; Geurts and Wouters, 2004). This is confirmed by our results. Additionally, it is shown that this ability to discriminate $F0$ based on place pitch cues is highly dependent on the filter bank used. The ACE filter bank did not allow the $F0$ information to be coded by place pitch cues. However, the BUT, GTF, and GTFW filter banks, which have more resolution in the lower frequencies, did enable the subjects to perceive $F0$ differences based upon place pitch. The GTFW has slightly less resolution compared to GTF and BUT due to the large filter overlap, and consequently the $F0$ discrimination using this filter bank is somewhat lower compared to the latter two filter banks. The actual shape of the filter bank, being more rectangular with steep slopes and less overlap (BUT), or more triangular with shallower slopes and more overlap (GTF), did not significantly change the performance. This contrasts with the results of Geurts and Wouters (2004), who found an improved performance for a triangular-like filter bank. However, given the present results, the improved performance of the triangular filter bank in that study was probably facilitated by the different cutoff frequencies used in the triangular filter bank, leading to more resolution in the lower frequency register. For subjects who have poor temporal pitch sensitivity but good place pitch sensitivity, the filter banks with high resolution in the $F0$ register may enable them to better discriminate $F0$.

Some of the psychometric curves of Fig. 3 are nonmonotonic and contain pitch reversals, some of which are consistent over the various subjects. In addition, when pooling the data for the different subjects and plotting the separate psychometric curves for the different $F1$ values, some significant pitch reversals can be observed, e.g., the psychometric curve for $F1=500$ Hz in Fig. 6(A). An analysis of all conditions that resulted in significant pitch reversals after pooling the data over the different subjects reveals a consistent pattern of the average channel amplitude before compression. This pattern is most easily interpreted by means of an example. Figure 6(B) shows the average amplitude of the 22 channels for a stimulus with $F1=500$ Hz, processed with the GTF filter bank. The thick line represents the normalized average amplitude for the reference stimulus having an $F0$ of 133 Hz. The two thin lines represent the normalized average amplitude for two comparison stimuli, $F0=150$ Hz (solid line) and $F0=250$ Hz (dashed line). The former stimulus resulted in a pitch reversal and the latter stimulus resulted in a correct pitch order [the arrows in Fig. 6(A)]. When comparing $F0$'s 133- and 150 Hz, the position of the fundamental frequency moves from the first to the second channel (towards the base), but the peak caused by the harmonics under the formant frequency shifts downward from channel 7 to channel 6, towards the apex. So, the pitch percept was not determined by the place of the fundamental frequency, but was largely determined by the place of the largest excitation along the electrode array. In the case where the channels with the largest amplitude coincide, as in Fig. 6(B) for $F0$ 133 Hz (thick solid line) and $F0$ 250 Hz (thin

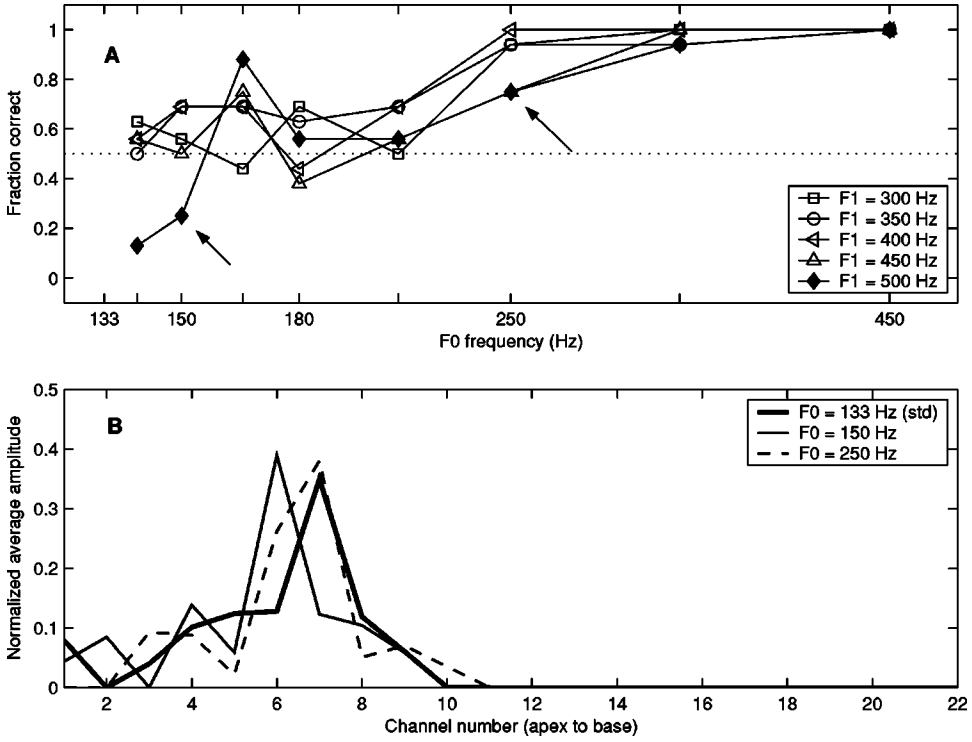


FIG. 6. Top panel: The proportion of correctly pitch-ranked stimuli pooled over all subjects for the condition with reference $F0$ 133 Hz and stimuli processed with the GTF filter bank and all temporal modulations above 10 Hz removed. The different lines in the figure depict the results for the various formant frequencies used. The fundamental frequencies of the signals are plotted on the horizontal axis on a log scale. A significant pitch reversal is noticed for the case where the formant frequency is 500 Hz (filled symbols) for the two lowest tested $F0$ frequencies (140 and 150 Hz). Bottom panel: The average channel amplitudes for the single-formant stimuli with formant frequency 500 Hz and fundamental frequencies 133, 150, and 250 Hz when processed with the GTF filter bank and with all temporal modulations above 10 Hz removed. The thick line represents the average amplitude of the reference stimulus with $F0$ 133Hz, $F1$ 500 Hz, and processed with the GTF filter bank (i.e., the reference stimulus of the line with the filled symbols in the upper panel). The thin solid line represents the average channel amplitude of the comparison stimulus mentioned in the upper panel and causing a pitch reversal ($F0=150$ Hz). The peak of the average amplitude has shifted towards the apex with respect to the average amplitude of the reference stimulus (thick line), causing the comparison stimulus to be perceived at a lower pitch. The thin dotted line represents the average amplitude of a comparison stimulus correctly ranked in pitch ($F0=250$ Hz). Although the peak of the average amplitudes of the reference and comparison stimulus are at the same location, most probably the correct ranking is due to the shift of the fainter channels towards the base.

dashed line), the pitch difference was most probably mediated by the excitation shift along the electrode array of the smaller peaks. Summarizing the two previous effects, we hypothesize that the pitch of the different sounds was determined by a weighted position of the peak of the excitation pattern. The simplest form of a weighted peak position was the center of gravity or the centroid, and can be expressed as

$$C=\frac{\Sigma_{e=1}^{22} e.m(e)}{\Sigma_{e=1}^{22} m(e)} \tag{4}$$

In Eq. (4) C is the centroid, e indicates the channel number (ranked from apex to base), and $m(e)$ is the average amplitude on channel e. The centroid is computed based upon the average amplitude of the envelopes (the compression accommodating for loudness growth is not applied). According to the proposed model, the discriminability of two sounds (based upon place pitch cues) was determined by the distance between the centroids of the excitation patterns of the two sounds. The centroid has the additional benefit of being independent of overall intensity. Figure 7 shows a scatter plot of the model prediction of all stimuli and the respective average proportion of correctly ranked trials, for $F0$'s of 133 Hz (top panel) and 165 Hz (bottom panel). The solid line in Fig. 7 represents the normal cumulative distribution fit to the pooled data over all conditions and subjects. This function was fitted to the average proportion using a Gauss–Newton nonlinear least-squares fitting method.

Similarly, a normal cumulative distribution function was fitted to the scatter plots for every subject separately. The slope of this fitted normal cumulative distribution function was taken as a performance measure of place pitch sensitivity. This fitting procedure is similar to a linear regression with the data points converted into d' values. Just-noticeable differences (jnd) were derived from the fitted curves as the minimal distance expressed in number of electrodes to obtain 75%-correct pitch rankings.

Figure 8 depicts the jnd estimates that were obtained from these fitted curves for every subject. The place pitch sensitivity of the subjects was also measured using an electrode ranking experiment in a separate study (Laneau and Wouters, 2004). These jnd estimates are also depicted in Fig. 8. In the latter study the subjects were presented two stimuli that differed in site of stimulation along the electrode array, and had to indicate the highest one. The jnd's shown in Fig. 8 are the mean of the jnd's for conditions with different numbers of active channels. No significant effect was found

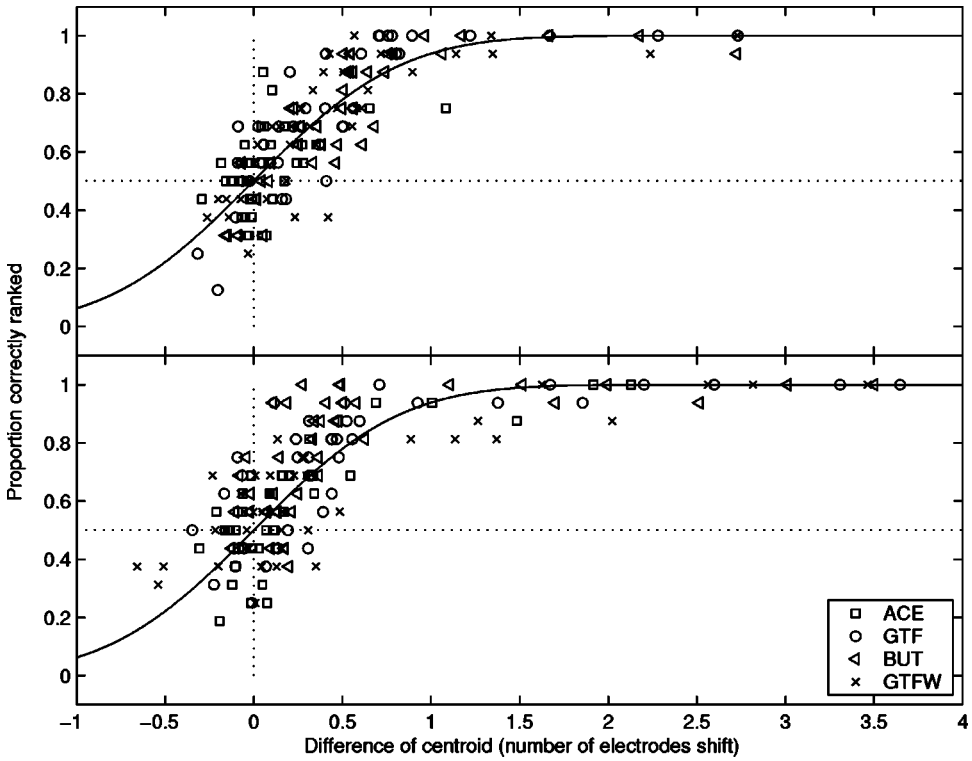


FIG. 7. The proportion of correctly ranked stimuli pooled over all subjects plotted as a function of the difference in centroid. Only data of stimuli that allowed no temporal pitch cues were included in this picture. For the sake of clarity the data for the two reference $F0$'s are depicted in separate panels ($F0$ 133 Hz in the upper panel and $F0$ 165 Hz in the lower panel). A normal cumulative distribution function was fitted to the pooled data and is represented by the solid line in both panels.

of the number of active channels of the stimuli on place pitch sensitivity in the latter study.

The jnd's estimated from the $F0$ discrimination data and the jnd's estimated from the place pitch measurements are nearly equal for all subjects (within error bars). However, for subject S3, the place pitch sensitivity is somewhat worse when compared to the centroid sensitivity. This can be due to the fact that we arbitrarily choose two electrode locations to estimate the place pitch sensitivity. At other locations the place pitch sensitivity for subject S3 might be better.

Previous support for the centroid model has been presented by McDermott and McKay (1994) for the case of two stimulation channels in cochlear implants. They investigated the pitch perceived when two bipolar channels were stimulated simultaneously and interleaved using constant current levels. The pitch of the two-channel stimuli varied monotonically between the pitches of the stimuli on the two single electrodes, as the relative current amplitudes were changed in orderly fashion. They interpreted their results by stating that the centroid of the excitation pattern shifted from one electrode location to the other as relative currents varied, and that this centroid shift caused the varying pitch. Their results also indicated that the pitch of two interleaved stimulated channels was perceived as one-dimensional, at least up to separations of 3 mm between stimulating electrodes. McKay *et al.* (1996) later confirmed and extended this result using dissimilarity matrices and multidimensional scaling analysis. They concluded that there exist two basic perceptual dimen-

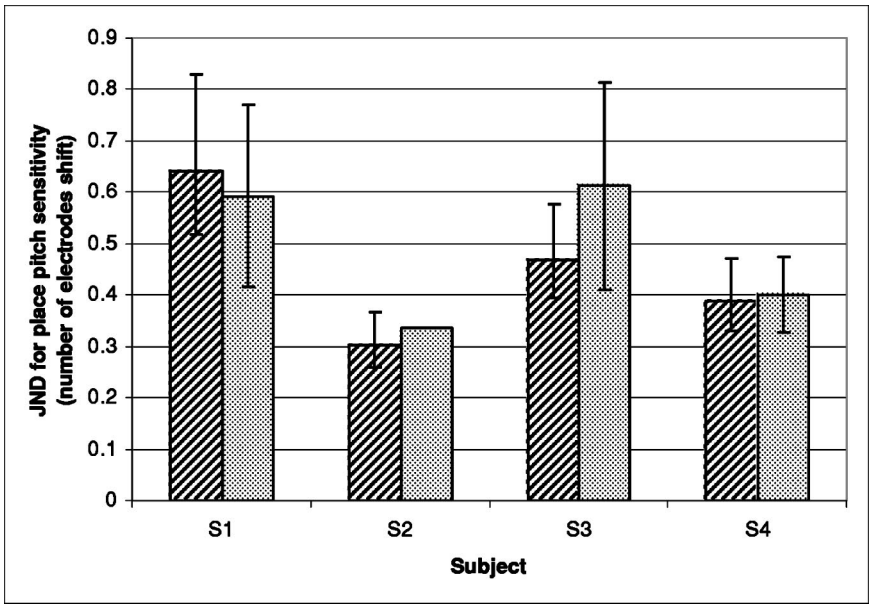


FIG. 8. Estimated just-noticeable differences (jnd's) for the four subjects. The dashed bars present the jnd's from the centroid differences calculated from the $F0$-discrimination data. The jnd's estimated from site of stimulation discrimination of the multichannel stimuli in the same patients (Laneau and Wouters, 2004) are depicted by means of the dotted bars. The error bars indicate the 95% confidence interval on the jnd estimated given by the fitting routine for the centroid jnd's. The error bars on the place pitch sensitivity of the multielectrode stimuli indicate ± 1 standard deviation.

sions with two-channel stimulation. The first percept was associated with the middle of the two active channels, corresponding to the centroid of the excitation pattern. The second percept was associated with the separation of the two active channels. However, McKay *et al.* also argued that their experiment did not rule out that the dimensions could also be interpreted as simply being associated with the location of the two channels. The results of McKay *et al.* (1996) were consistent for separations up to 12 mm, the largest separation tested, between active channels. Due to the nature of the acoustic stimuli used in the present study and the maxima selection in the sound-processing stage, the pulse sequences only stimulated channels that were at most ten electrodes apart, resulting in a maximal separation of 7.5 mm between active channels.

Cohen *et al.* (1996) examined the direct relationship between the forward-masking pattern and the numerical pitch estimates that single-channel stimuli exert. They found a clear qualitative correspondence between the centroid of the forward-masking pattern and the numerical pitch estimates, for all stimulation modes. They were also able to predict the reversed (compared to normal tonotopy) pitch ranking of some pairs of electrodes based upon the centroid of the forward-masking pattern for several subjects.

In normal-hearing subjects, models based upon the centroid of the spectrum (Stover and Feth, 1983; Anantharaman *et al.*, 1993) have been used to explain the results of pitch ranking and discrimination experiments with noise bands (Dai *et al.*, 1996; Versfeld, 1997). These models accurately described the discriminability of narrow-band sounds, but the models had to be extended for the discrimination of sounds with larger bandwidths.

It is, however, questionable whether the place pitch is truly a pitch percept or an aspect of the sounds' timbre. According to the ANSI definition of pitch (American National Standards Institute, 1994), place pitch is pitch as subjects are able to rank stimuli from high to low. Nevertheless, there is a striking resemblance between the proposed model for place pitch and the model for sharpness in normal-hearing subjects (Zwicker and Fastl, 1999). Both models calculate the centroid of the excitation pattern.

The results of the present study, and the studies of Geurts and Wouters (2001; 2004) have shown that $F0$ discrimination in cochlear implant recipients is possible solely based upon place pitch cues. These studies used stationary synthetic signals while running speech is nonstationary and formant frequencies change over time. Green *et al.* (2002) measured the discrimination of direction of sweeps of fundamental frequency using noise-band vocoders in normal-hearing subjects. They showed that for stationary signals the discrimination was possible based solely on place pitch cues, but the discrimination became almost impossible when the signals had nonstationary formant frequencies. Moreover, the model presented above to account for the $F0$-discrimination results based upon place pitch cues indicates a strong dependence of the place pitch cues on the overall spectral shape of the stimuli. This might limit the usefulness of place pitch cues to encode $F0$ in cochlear implants during running speech.

On the other hand, although formant frequencies and $F0$ frequencies originate from physically independent processes, formant frequencies and $F0$ frequencies are often correlated. For example, both the formant frequencies and the $F0$ frequencies of a female speaker are on average higher than those of a male speaker. The combined effect of higher formant frequencies and higher fundamental frequencies will shift the centroid of the excitation pattern toward the base. This shift might be detectable for cochlear implant subjects even during running speech. This might help the subject listening to running speech to discriminate pitch differences.

Moreover, the possible problem for $F0$ discrimination based upon place pitch cues coming from the interaction between overall spectral shape and $F0$ cues is less likely to exist for some musical sounds, such as woodwinds, because these musical notes have stationary spectral shapes throughout their duration (Gray, 1977). Second, the dimensions and form of the instrument, defining the resonance frequencies, are fixed so the spectral shape is fixed across different notes. Consequently, the differences in place pitch between two different notes are related to $F0$ differences.

B. Including temporal pitch cues

Comparing Figs. 3 and 4, the improvement in performance from adding the temporal pitch cues was largest for the ACE and GTFW filter banks. However, the range for improvement was much larger for the ACE filter bank than for the other filter banks because the latter already enabled some $F0$ discrimination without temporal pitch cues.

The limited improvement obtained by adding temporal pitch for the GTF and BUT filter bank might have been due to the fact that higher harmonics of the test stimuli is also resolved in those filter banks. Resolved harmonics prevented the output of filters from beating at the fundamental frequency, and the only fluctuation seen at the output was at the frequency of the respective harmonic. The filter outputs with the largest amplitude were filters having center frequencies close to the formant frequencies (between 300 and 500 Hz). In this range the BUT and GTF filter banks resolved the individual harmonics, so the largest output of the filters fluctuated at multiples of the fundamental frequency. As an example, the modulations of channels 1 to 7 of one of the sounds processed with the BUT filter bank and after envelope detection are depicted in Fig. 9. The most intense channels in this example do not beat at $F0$ frequency but at some higher harmonic frequency. Although this higher frequency fluctuation was partly attenuated by the envelope detection low-pass filter, it is still clearly present in the channel output. Only the output of the lower channel, containing the fundamental frequency, and possibly higher channels, containing unresolved harmonics, provide the temporal pitch cues related to $F0$. The channels containing resolved higher harmonics are modulated at multiples of $F0$ and consequently elicit higher pitches because higher modulation frequencies lead to higher pitches (McKay *et al.*, 1994).

The ACE filter bank did not suffer from this shortcoming because all the bins in the FFT filter bank are approximately 250 Hz wide, so all harmonics was unresolved (at least for the reference stimuli), causing all filter outputs to

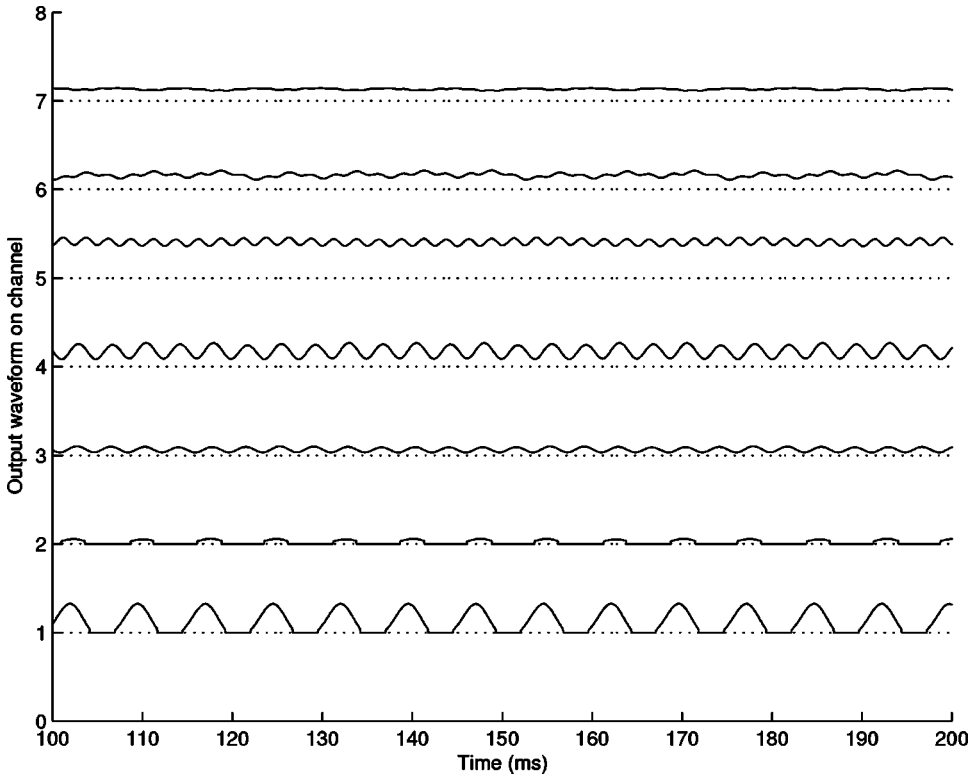


FIG. 9. Time fragment of the amplitudes of channels 1 to 7 across time for the SFS stimulus with $F0=133$ Hz and $F1=400$ Hz, processed with the BUT filter bank. Channels 1 and 2 only contain the first harmonic, and consequently the outputs beat at the fundamental frequency. Channels 3 and 4 only contain the second harmonic, leading to a modulation at the second harmonic frequency. Channel 5 contains only the third harmonic, and consequently the output beats at the third harmonic frequency. The beating at higher harmonics makes the temporal pitch cues less clear.

beat at the fundamental frequency. Similarly, the GTFW filter bank was designed to only resolve the fundamental frequency and with all overtones unresolved. This also made the output of all channels beat at $F0$ for the GTFW filter bank.

In order to assess the effect of the modulation depth of the beatings at $F0$ in the different channels, the average modulation depths of all stimuli were computed. We also assess the effect of the phase differences of the beating across channels. These interchannel phase differences of the beating at $F0$ are caused by the phase differences between the different harmonics of the original sound and by the phase response across frequency of the filters in the filter bank. The average modulation depth over all channels is computed in two different ways to assess the effect of the phase differences. In the first way, all phase information of the channels was discarded (''without phase''). In the second way the phase information was included and contributed to the average as out-of-phase channels could interfere destructively (''with phase'').

As a first step, the average spectrum of the modulations in all channels was calculated. For the ''without phase'' approach, the spectrum of all channels was calculated separately and the average of the magnitude of the spectrum over the channels was taken, discarding the phase components of the modulations. In the ''with phase'' approach, the time signal was averaged over all channels and then the magnitude of the spectrum of this average time signal was calculated. In the second step, the modulation depth was calculated for both approaches from the respective average spectrum, using Eq. (5)

$$\text{modulation depth} = \frac{2*\|S(F0)\|}{\|S(F0)\|+\|S(0)\|}, \tag{5}$$

where $\|S(F0)\|$ was the magnitude of the spectrum at the fundamental frequency and $\|S(0)\|$ was the magnitude of the dc component of the spectrum.

As a third step, the obtained modulation depths were correlated with the amount of temporal pitch cues (d'_{temp}) for all processed reference stimuli and all test conditions. The correlation was done only with the modulation depth of the reference stimuli and not with the modulation depth of the higher $F0$ comparison stimuli, because of the test procedure used in the present experiment. The subjects were asked to indicate the highest-sounding stimulus and this could be the one with a higher modulation rate or with no modulation at all. This is because for modulated stimuli the perceived pitch is related to the modulation frequency but steady-state (or poorly modulated) stimuli elicit high pitches related to the high overall pulse rate (McKay and Carlyon, 1999). So, clear pitch differences may have been perceived when the comparison contained no modulations at all.

The analysis revealed that the d'_{temp} values are significantly ($p<0.0001$ for both) correlated with the modulation depth of the reference stimulus of the trials for both computational methods ($r=0.4216$; and $r=0.3090$ for with phase and without phase, respectively). Moreover, although both methods of calculating the modulation depth are significantly correlated ($r=0.5910$) with each other, the correlation with benefit from inclusion of temporal pitch cues was 36% higher when phase information was included. Using a test comparing the two sets of predictors (Tabachnick and Fidell, 1996), this lead to a significant higher correlation for the ''with phase'' approach ($Z=2.42$; $p=0.0077$). However, both correlations are rather weak and the results of this analysis should be interpreted with some caution.

The poorest-performing filter bank for temporal pitch

cues (BUT) produced the lowest modulation depths averaged over all reference stimuli. The best-performing filter bank for temporal pitch cues (ACE) produced the highest average modulation depth for the reference stimuli, when modulation depths are calculated taking into account the relative phase differences between the channels. When the phase information is discarded, GTF produced the lowest modulation depths and GTFW with the highest modulation depths. So, the ''with phase'' modulation depth present in the processed stimuli can be a predictor of the amount of temporal pitch cues that can be used by the cochlear implantee.

This may mean that out-of-phase modulations on different channels probably allow for less clear pitch percepts than equally large in-phase modulations, at least when the channels are within a relatively narrow region, as is the case in this study. It has been shown before that modulation patterns on two nearby electrodes are integrated, while more widely separated modulation patterns are perceived separately (McKay and McDermott, 1996).

V. CONCLUSIONS

In the present study, properties of filter banks which affect $F0$ discrimination in cochlear implants were examined. Temporal and place pitch cues to $F0$ discrimination were studied separately and for both cues the mechanisms relating the pitch percepts to physical parameters were investigated.

The results of our experiments demonstrate that both place and temporal pitch cues allow the discrimination of the fundamental frequency, as also evidenced by previous research. However, $F0$ discrimination based upon purely place pitch cues is weak and only possible for differences exceeding one octave for the best-performing filter banks. Moreover, the ACE filter bank, as used by most of the Nucleus cochlear implant recipients in their daily life, did not allow for $F0$ discrimination based upon place pitch cues. Our experiments indicate that our new filter banks having the highest frequency resolution in the $F0$ register gave the best results using place pitch cues. The newly proposed filter banks are likely to improve $F0$ discrimination, relative to the currently clinically available filter bank, in subjects having good place sensitivity and poor temporal sensitivity, at least for spectrally stationary signals. A previous study has shown that for some cochlear implantees the place pitch cues are very important for $F0$ discrimination (Geurts and Wouters, 2004).

When temporal pitch cues were added to the stimuli, performance increased significantly and jnd's for $F0$ discrimination ranged from 6% up to 60% depending on subject, processing condition, and reference $F0$. There were no overall significant differences between the performance of the filter banks when both temporal and place pitch cues were presented to the subjects. The performance of cochlear implant subjects is still very poor compared to the performance of normal-hearing subjects where jnd's for $F0$ are typically less than 1%. The largest temporal pitch cues for $F0$ discrimination were possible with the filter banks that did not resolve the harmonics. In these filter banks all output channels beat at the fundamental frequency. When the beatings of the different channels were in phase, the temporal pitch cues seemed to be clearer compared to the case where beatings of neighboring channels are out of phase.

A modeling study of the results with purely place pitch cues indicated that place pitch is related to the centroid of the excitation pattern along the cochlea. As place pitch cues enable subjects to rank stimuli from low to high, it can be considered a pitch percept. However, taking into account the present model, place pitch is more related to the sharpness or brightness of timbre than to the repetition rate of the processed sound in cochlear implants.

The proposed model indicates that the place-pitch-mediated $F0$ cues and the overall spectral shape are closely related. In other words, there might be interference in the perception between $F0$ and formant frequencies when both are coded onto place of stimulation. This might possibly lead to a reduction of the usefulness of the place-pitch-mediated $F0$ cues when formants vary as in running speech, and might possibly also lead to reduced speech perception when $F0$ strongly varies. Although the spectral shape or brightness of timbre also interferes with pitch perception in normal-hearing subjects (Ohgushi, 1978), the effect in cochlear implants appears to be larger, at least when only place pitch cues are present.

With the filter banks studied, optimal transmission of both temporal and place pitch cues was difficult for one and the same filter bank.

ACKNOWLEDGMENTS

We thank the subjects for their enthusiastic cooperation. We also thank Luc Geurts and Astrid van Wieringen for helpful comments. We thank Robert Carlyon, Andrew Faulkner, and Robert Shannon for comments on earlier versions of this manuscript. This study was partly supported by the Flemish Institute for the Promotion of Scientific-Technological Research in Industry (Project IWT 020540), by the Fund for Scientific Research-Flanders/Belgium (Project G.0233.01), and by Cochlear Ltd.

American National Standards Institute (**1994**). ''American National Standard Acoustical Terminology for physiological and psychological acoustics.'' ANSI S1.1-1994 (American Standards Association, New York).

Anantharaman, J. N., Krishnamurthy, A. K., and Feth, L. L. (**1993**). ''Intensity-weighted average of instantaneous frequency as a model for frequency discrimination,'' J. Acoust. Soc. Am. **94**, 723–729.

Barry, J. G., Blamey, P. J., Martin, L. F., Lee, K. Y., Tang, T., Ming, Y. Y., and Van Hasselt, C. A. (**2002**). ''Tone discrimination in Cantonese-speaking children using a cochlear implant,'' Clin. Linguist Phon. **16**, 79–99.

Cochlear Ltd. (**2002**), ''Nucleus Implant Communicator (NIC) System Overview,'' **N95291 Iss.1**

Cohen, L. T., Busby, P. A., and Clark, G. M. (**1996**). ''Cochlear implant place psychophysics. II. Comparison of forward masking and pitch estimation data,'' Audiol. Neuro-Otol. **1**, 278–292.

Dai, H. P., Nguyen, Q., Kidd, G., Feth, L. L., and Green, D. M. (**1996**). ''Phase independence of pitch produced by narrow-band sounds,'' J. Acoust. Soc. Am. **100**, 2349–2351.

Drullman, R., Festen, J. M., and Plomp, R. (**1994**). ''Effect of temporal envelope smearing on speech reception,'' J. Acoust. Soc. Am. **95**, 1053–1064.

Geurts, L., and Wouters, J. (**2001**). ''Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants,'' J. Acoust. Soc. Am. **109**, 713–726.

Geurts, L., and Wouters, J. (**2004**). ''Better place-coding of the fundamental frequency in cochlear implants,'' J. Acoust. Soc. Am. **115**, 844–852.

Gfeller, K., and Lansing, C. (**1992**). ''Musical perception of cochlear implant users as measured by the primary measures of music audiation—An item analysis,'' J. Music. Ther. **29**, 18–39.
Glasberg, B. R., and Moore, B. C. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.
Green, T., Faulkner, A., and Rosen, S. (**2002**). ''Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants,'' J. Acoust. Soc. Am. **112**, 2155–2164.
Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.
Grey, J. M. (**1977**). ''Multidimensional perceptual scaling of musical timbres,'' J. Acoust. Soc. Am. **61**, 1270–1277.
Jones, P. A., McDermott, H. J., Seligman, P. M., and Millar, J. B. (**1995**). ''Coding of voice source information in the Nucleus cochlear implant system,'' Ann. Otol. Rhinol. Laryngol. Suppl. **166**, 363–365.
Klatt, D. H. (**1980**). ''Software for a cascade-parallel formant synthesizer,'' J. Acoust. Soc. Am. **67**, 971–995.
Laneau, J., and Wouters, J. (**2004**). ''Multi-channel place pitch sensitivity in cochlear implant recipients,'' J. Assoc. Res. Otolaryngol. **5**, 285–294.
Macmillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).
McDermott, H. J., and McKay, C. M. (**1994**). ''Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea,'' J. Acoust. Soc. Am. **96**, 155–162.
McKay, C. M., and Carlyon, R. P. (**1999**). ''Dual temporal pitch percepts from acoustic and electric amplitude-modulated pulse trains,'' J. Acoust. Soc. Am. **105**, 347–357.
McKay, C. M., and McDermott, H. J. (**1996**). ''The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites,'' J. Acoust. Soc. Am. **100**, 1081–1092.
McKay, C. M., McDermott, H. J., and Carlyon, R. P. (**2000**). ''Place and temporal cues in pitch perception: Are they truly independent?'' ARLO **1**, 25–30.
McKay, C. M., McDermott, H. J., and Clark, G. M. (**1994**). ''Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees,'' J. Acoust. Soc. Am. **96**, 2664–2673.
McKay, C. M., McDermott, H. J., and Clark, G. M. (**1996**). ''The perceptual dimensions of single-electrode and nonsimultaneous dual-electrode stimuli in cochlear implantees,'' J. Acoust. Soc. Am. **99**, 1079–1090.
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.
Ohgushi, K. (**1978**). ''On the role of spatial and temporal cues in the perception of the pitch of complex tones,'' J. Acoust. Soc. Am. **64**, 764–771.
Patterson, R. D., Allerhand, M. H., and Giguere, C. (**1995**). ''Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform,'' J. Acoust. Soc. Am. **98**, 1890–1894.
Pijl, S. (**1997**). ''Labeling of musical interval size by cochlear implant patients and normally hearing subjects,'' Ear Hear. **18**, 364–372.
Rubinstein, J. T., Wilson, B. S., Finley, C. C., and Abbas, P. J. (**1999**). ''Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation,'' Hear. Res. **127**, 108–118.
Shannon, R. V. (**1983**). ''Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics,'' Hear. Res. **11**, 157–189.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.
Slaney, M. (**1993**), ''An efficient implementation of the Patterson–Holdsworth auditory filter bank,'' Tech. Rep. **35**, Apple Computer.
Smith, Z. M., Delgutte, B., and Oxenham, A. J. (**2002**). ''Chimaeric sounds reveal dichotomies in auditory perception,'' Nature (London) **416**, 87–90.
Stover, L. J., and Feth, L. L. (**1983**). ''Pitch of narrow-band signals,'' J. Acoust. Soc. Am. **73**, 1701–1707.
Tabachnick, B. G., and Fidell, L. S. (**1996**). *Using Multivariate Statistics*, 3rd ed. (Harper Collins College Publishers, Northridge).
Tong, Y. C., Clark, G. M., Blamey, P. J., Busby, P. A., and Dowell, R. C. (**1982**). ''Psychophysical studies for 2 multiple-channel cochlear implant patients,'' J. Acoust. Soc. Am. **71**, 153–160.
Townshend, B., Cotter, N., Van Compernolle, D., and White, R. L. (**1987**). ''Pitch perception by cochlear implant subjects,'' J. Acoust. Soc. Am. **82**, 106–115.
Versfeld, N. J. (**1997**). ''Discrimination of changes in the spectral shape of noise bands,'' J. Acoust. Soc. Am. **102**, 2264–2275.
Wilson, B. S. (**1997**). ''The future of cochlear implants,'' Br. J. Audiol. **31**, 205–225.
Wouters, J., Damman, W., and Bosman, A. J. (**1994**). ''Vlaamse opname van woordenlijsten voor spraakaudiometrie,'' ''Flemish recording of the wordlists for speech audiometry.'' Logopedie **7**, 28–33.
Zwicker, E., and Fastl, H. (**1999**). *Psychoacoustics: Facts and Models*, 2nd ed. (Information Sciences, Springer, Berlin).

Methodological aspects of an adaptive multidirectional pattern search to optimize speech perception using three hearing-aid algorithms

Bas A. M. Franck[a] and Wouter A. Dreschler
Academic Medical Center, Clinical and Experimental Audiology, Meibergdreef 9, 1105 AZ, Amsterdam, The Netherlands

Johannes Lyzenga
Department of Otolaryngology, VU Medical Center, De Boelelaan 1117, 1081 HV, Amsterdam, The Netherlands

(Received 8 October 2003; revised 19 August 2004; accepted 23 August 2004)

In this study we investigated the reliability and convergence characteristics of an adaptive multidirectional pattern search procedure, relative to a nonadaptive multidirectional pattern search procedure. The procedure was designed to optimize three speech-processing strategies. These comprise noise reduction, spectral enhancement, and spectral lift. The search is based on a paired-comparison paradigm, in which subjects evaluated the listening comfort of speech-in-noise fragments. The procedural and nonprocedural factors that influence the reliability and convergence of the procedure are studied using various test conditions. The test conditions combine different tests, initial settings, background noise types, and step size configurations. Seven normal hearing subjects participated in this study. The results indicate that the reliability of the optimization strategy may benefit from the use of an adaptive step size. Decreasing the step size increases accuracy, while increasing the step size can be beneficial to create clear perceptual differences in the comparisons. The reliability also depends on starting point, stop criterion, step size constraints, background noise, algorithms used, as well as the presence of drifting cues and suboptimal settings. There appears to be a trade-off between reliability and convergence, i.e., when the step size is enlarged the reliability improves, but the convergence deteriorates. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1808220]

PACS numbers: 43.66.Yw, 43.66.Ts [DDO] Pages: 3620–3628

I. INTRODUCTION

In previous studies, multi-directional strategies have been proposed to optimize hearing-aid parameters (e.g., Neuman *et al.*, 1987; Kuk and Lau, 1996). In many studies the (modified) Simplex procedure has been promoted. The Simplex procedure is especially advantageous over tournaments with respect to time efficiency. Moreover, the procedure can take into account interactions between parameters by considering multiple responses at the same time. However, many methodological consequences have not been studied systematically.

In order to apply the Simplex procedure in the auditory domain, a paired-comparison paradigm has been used most frequently.[1] There are several reasons for using paired-comparisons. It often appears problematic to use objective scores like speech intelligibility thresholds, because different parameter settings do not always yield different scores. Other important advantages of the paired-comparison paradigm comprise high sensitivity and psychological advantage due to a high motivation by active subject interaction. For audiological purposes it appears especially advantageous to evaluate not only speech intelligibility, but also listening comfort in order to guarantee better satisfaction from using the hearing aid.

In a pilot study we expanded the domain of the modified Simplex procedure of Neuman *et al.* (1987) from two dimensions to three. The dimensions represented different algorithms instead of different parameters of a single algorithm. The results indicated that the test-retest reliability of the optimization procedure is relatively poor due to the presence of settings that were hard to distinguish. Moreover, listening comfort appeared a more reliable evaluation criterion than speech intelligibility. Therefore, in this study we selected the listening comfort criterion to determine subjective preferences.

In order to improve the reliability we modified the procedure in the following ways. First, the step size in the procedure has been made adaptive. That is, the step size for each dimension was governed by the perceptual differences between settings for each listener. Initially, the step size was large. After three rounds of comparisons the step size was halved. The process of halving continued until the listener could not distinguish the settings any more. In that case the step size was enlarged for that particular dimension. Second, boundary effects for the constrained complex of settings have been made less influential. This was achieved on the one hand by extreme algorithm settings at the boundary that

[a]Author to whom correspondence should be addressed. Clinical and Experimental Audiology, Academic Medical Center Amsterdam, P.O. Box 22660, 1100 DD, Amsterdam, The Netherlands. Telephone: +31-20-5663918; Fax: +31-20-5669068. Electronic mail: b.a.franck@amc.uva.nl

are assumed to be unfavorable. On the other hand the number of inside settings relative to boundary settings was increased. This was achieved by increasing the number of settings from five to ten for each dimension. We hypothesize an increased reliability of the procedure as a consequence of these modifications.

Although the necessity of distinguishable settings is assumed to be an important prerequisite for a reliable optimization procedure, other procedural and nonprocedural factors can be of influence as well. Procedural factors that need closer inspection are the initial setting, the stop criterion, and the maximum number of paired comparisons. This latter constraint prevents listeners from losing their attention. When different initial settings yield different final settings (optima), the optimization procedure performs inconsistently. Next to this effect, the reliability might depend on the initial setting for reasons yet unknown. The stop criterion is an important tool for proper convergence. When the stop criterion has been chosen falsely, the procedure might end too quickly. The procedure can also end too late. This can happen when the, a priori defined, maximum number of paired comparisons has been reached. In both cases at best a sub-optimal setting will be found.

Non-procedural factors that can influence the reliability in a negative way can be caused by acoustical constraints and cognitive factors. When the stimulus material has been recorded real-life, a smearing effect can be introduced as a result of reverberation. A shaping effect in the frequency domain can be expected when the stimuli are presented through a hearing aid. Both effects could have detrimental effects on the reliability. Although optimization of speech perception for hearing aids should be performed in realistic acoustical conditions in future studies, in this study we used artificial speech and noises that are presented by headphones. So, both a stationary car noise and a fluctuating speech-shaped noise have been chosen as experimental parameters. The most important cognitive factor is the criterion that is used by the listener to judge which fragment is better. That is, the listener's criterion can drift during an experiment. This aspect needs close inspection.

In this study three experimental questions will be answered. Does the adaptive multidirectional pattern search perform reliably? Or, stated differently, does the adaptive step size result in high reliability of the optimization strategy? Directly related to this, what are the effects of procedural and nonprocedural factors on the reliability of the multidirectional pattern search? Finally, taking all considerations into account, we evaluate how applicable the multidirectional pattern search is to the auditory domain for the selection of an optimal setting of complex hearing aids.

II. METHOD

A. Algorithms

Three experimental auditory signal-processing algorithms are used in the optimization procedure. The individual dimensions are spectral enhancement (SE), spectral lift (SL), and noise reduction (NR), developed and described in detail by Lyzenga *et al.* (2002).

The spectral enhancement algorithm expands the spectral peaks for frequencies between 200 and 5000 Hz. In that region slow spectral fluctuations are slightly compressed, intermediate fluctuations expanded, and fast fluctuations remain intact to avoid statistical variations of noise from being expanded. The spectral enhancement was realized using a power function. Enhancement factor M represents the exponent, the spectral amplitude of the signal a represents the base of the power function.

The spectral lift algorithm is a linear filter that is assumed to reduce upward-spread-of-masking. That is, to try prevent low-frequency high-level formants from masking higher-frequency low-level formants. Therefore, the algorithm is designed to progressively emphasize frequencies between 0.3 and 3 kHz, the region in which the second to fourth formant are dominant. For frequencies higher than 3 kHz the spectral envelope decreases again, so the filter is sawtooth shaped, and the peak of the filter is centered on 3 kHz. Emphasis of high frequencies is avoided, because sharp sounds are assumed to be uncomfortable.

The noise reduction algorithm attempts to suppress noise based on the temporal behavior of the signal, i.e., the phase variations. When the temporal behavior of spectral components is irregular, the signal is assumed to be noisy, and the linear suppression filter is given small amplitudes. The amplitude of the filter is set high when spectral components behave temporally regular, and are assumed to be speech-like signals.

It was found to be difficult to assemble a series of ten easily discernable settings for each algorithm without introducing intolerable distortions. To overcome this difficulty, we included settings that, at first inspection, seemed counter-intuitive. Not only spectral expansion is allowed but also spectral compression, not only spectral lift but also spectral suppression, and finally, not only noise reduction but also noise amplification. The settings are labeled from -3 to 6, such that 0 is the reference setting.

The spectral enhancement factor M ranges between 0.25 and 2.5 in steps of 0.25 dB/dB. For setting 0, $M=1$ and there is no spectral enhancement. The spectral lift settings have values between -9 and 18, in nine 3 dB steps. A lift-value of 0 represents no spectral lift. The noise reduction multiplication factors are represented by dimensionless values between -3 and 6 in steps of 1, and setting 0 corresponds to a value of 0. The algorithm values have been chosen so that steps are approximately perceptually equidistant.

B. Experimental procedures and conditions

1. Multidirectional pattern search with adaptive step size

In order to simultaneously optimize multiple hearing aid algorithms a multidirectional pattern search design is used in this study. We use three dimensions that represent three hearing aid algorithms. Several procedural features characterize the search strategy, which is aimed at finding the optimal combination of all algorithm settings. The search has a single initial setting. After the first comparison, the next setting is assumed to have improved performance, i.e., listening comfort in this study. To search for improved performance, the

setting under focus is compared pair wise with three neighbors (n paired comparisons for n dimensions). The settings that are included for paired comparisons are orthogonally related to the initial setting under focus and remain orthogonally related. So, three orthogonal search directions are used throughout the procedure for this pattern search design. The decision as to what will be the next setting under focus depends on the outcome of all three paired comparisons in the following way. When neither of the neighbors wins, the setting under focus remains the best setting. When one neighbor wins, this neighbor will be the new best setting. When two or three neighbors win, the new setting is estimated from the winning directions. That is, the *assumed* new best setting is found by addition of the vectors that represent the search directions, i.e., the directions of paired comparisons. As a consequence, the progress can be along a diagonal with respect to the orthogonal system of search directions.

The new search directions for the novel best setting depend on the history of search directions. For each dimension the search direction remains the same when the neighbor yields higher listening comfort. The search direction is reversed when the setting at stake performs better than the neighbor in that particular direction. The incorporation of the history of search directions is very important for this search strategy in order to accelerate the optimization process. The settings can never escape a predetermined lattice structure. This feature is necessary for our study due to the fact that we had to record the speech and noise material in advance.

The step size has been made adaptive, and can be characterized by perceptual distinctness. At the start of the search, the step size is equal to four steps of the lattice structure. After one series of three paired comparisons the step size is halved to two steps. After another three rounds the step size is halved to one step, the minimum step size. The halving process has been incorporated to speed convergence. The initial large step size was mainly chosen to ensure audible differences between settings and to make the listeners acquainted with the perceptual dimensions. At any time the fragments to be compared can be repeated. When the listener cannot distinguish between the two settings, the step size for that particular dimension will be increased until it is equal to three. Again, after three series the step size will be halved and rounded, until the step size is equal to one step. It is important to note that the ''no difference'' button can be used only after the ''repeat'' button has been used once, as the listeners should be convinced that the fragments are indistinguishable, instead of just rather difficult to judge in terms of listening comfort. Each pair of fragments can be repeated once.

The multidirectional pattern search procedure terminates when one of the following three criteria is met. When the same setting is frequented three times or when a setting wins a paired comparison more than six times, the procedure will stop. These stop criteria are a compromise between good convergence and limited required average test time. The third stop criterion is the maximum number of paired comparisons in the procedure. This criterion has been included to limit the maximum test time of the procedure. In a pilot experiment a limit of 54 paired comparisons was found to be a reasonable choice, which corresponds to a test duration of fifteen minutes on average.

For four different starting points a test (T) and a retest (R) have been performed. In the multidimensional setup we used combinations $T(-3)=R(-3)=(-3,-3,-3)$, $T(1)=R(1)=(1,1,1)$, $T(3)=R(3)=(3,3,3)$, and the center of gravity of the winners of $T(-3)$, $R(-3)$, $T(1)$, $R(1)$, $T(3)$, and $R(3)$. This starting point, which could be a suboptimal setting instead of an optimal setting is labeled as T(S) or R(S) depending on the test or retest status.

2. Multidirectional pattern search with step size as parameter

The experimental procedure for this experiment is the same as for experiment 1, except for the step size. Using the step size as a parameter, the dependence of the reliability on the step size can be estimated. In this experiment, the step size is either fixed or variable. The fixed step size is either 3 or 4 lattice grid steps. The variable step size has a lower limit of 2 and an upper limit of 3 steps. Therefore, only the lower limit is different from the lower limit of experiment 1, which was 1 step. Hence, in combination with the first experiment there are four step size configurations. The labeling of the configurations follows the possible step sizes in the procedure. So, the procedure of the experiment 1 is step123, the variable step size with elevated minimum step size is labeled step23. The fixed step size configurations are labeled step3 and step4. The first ''orienting'' step of the multidirectional pattern search is twice as large as the fixed step size throughout the procedure, i.e., 6 and 8, respectively. For the variable step size, the first step is equal to 4, comparable to experiment 1. Experiment 2 has been completed for starting point $(-3,-3,-3)$ and continuous noise only. For this starting point three tests were performed instead of a test and a retest only, used in experiment 1.

For experiment 1 the reliability of the multidirectional pattern search was determined. The reliability is the test-retest reproducibility for the pattern search. The reliability was calculated by the three-dimensional (3D) distance between the optima found in the test and in the retest condition. For experiment 2 two measures were determined. First the reliability was calculated for different pattern search designs that comprised different step size configurations. Second a convergence measure was determined. Convergence relates to the ability of the pattern search to find the optimum. The degree of convergence was studied by comparing the optima found for each step size configuration. That is, for each subject the average optimum setting (avg2) was determined for each step size configuration for starting point $(-3,-3,-3)$. After that the results were compared with the average optimum setting for the continuous noise condition and for all starting points of experiment 1 (avg1). Finally, the three-dimensional distance between avg1 and avg2 was calculated.

C. Subjects and test material

Experiment 1 was conducted by seven normal hearing subjects: s1 to s7. Only part of these subjects were available for experiment 2: s1, s3, and s4. Therefore, two other normal

hearing subjects were added, s8 and s9, thus the data of experiment 2 rely on the results of five subjects. We verified that the performance of s8 and s9 was in agreement with the other subjects for a subset of conditions in experiment 1 (for details see the cluster analysis in the results section).

The subjects were listening with their ''better'' ear; three subjects used their left, six subjects their right ear. In the experiments, the subjects listened to speech in background noise. The duration of the fragments in the optimization procedure range between 1.5 and 2.5 seconds. We used standard Dutch sentences (VU98; see Versfeld *et al.*, 2000), uttered by a Dutch female speaker and equated for their root mean square (RMS)-values. Two background noises have been used: Continuous car noise (cont) and fluctuating speech-shaped noise (fluct, Festen and Plomp, 1990). The speech and noise were presented at a signal-to-noise ratio (SNR) of 0 dB. This SNR was assumed to be appropriate for judging the listening comfort of the speech-in-noise fragments used for the multidirectional search.[2]

D. Experimental setup

1. Recording method

The stimulus material was processed off-line. The speech fragments were preceded by a fixed period of 0.5 s of silence and the length of the silence period after the sentence depended on the length of the sentence, creating a fragment of four seconds. Likewise, four-second noise fragments were cut and added to the speech fragments at a signal-to-noise ratio of 0 dB. To achieve this, levels of the speech fragments were compared to the levels of the specific noise fragments in dBA. The A-weighting was used to correct for differences in spectral contents of the two noise types. After the speech and noise had been mixed, the signals were processed by the three algorithms.

The order of processing is quite straightforward. The noise reduction was used first, because spectral peaks are more pronounced for higher signal-to-noise ratios and may be processed more successfully after noise reduction. The spectral lift algorithm is applied last, after the enhancement stage, because the lift will be defined better once the signals are enhanced. When the order of processing is reversed, the amount of lift will depend on the amount of expansion. We carefully avoided clipping of the signals in the recording. To avoid loudness becoming a cue when fragments are compared pair wise in the experiments, the fragments were equalized with respect to their root-mean-square (RMS) values.

2. Playback method

Before running the optimization procedure the 44 100 Hz sampled signals were stored on disk. During the experiments these monophonic signals were transferred to a signal-processing hardware system (stage 1), filtered using an experimental hearing aid (stage 2), transferred back to the signal-processing hardware system (stage 3), and finally presented through headphones (stage 4).

In the first stage the monophonic signal was fed into channel 1 of the signal-processing hardware (Tucker Davis Technologies TDT system). An appropriate DC-voltage was fed into channel 2 of the signal-processing hardware system in order to create positive voltages needed for the experimental hearing aid in stage 2. After that, the signals of both channels were transferred via several hardware modules in different ways. That is, the AC-part, was converted to analog (using a TDT DA3-unit), filtered by an anti-aliasing filter (TDT FT5) with a cut-off frequency of 16 kHz and a slope of 48 dB/oct, and attenuated 18 dB by a programmable attenuator (TDT PA4). The DC-component was also converted to analog via the same converter (TDT DA3) and summed with the attenuated AC-component using a signal mixer (TDT SM3).

In stage 2, the signal with the appropriate DC-offset was transferred to a flexible experimental hearing aid (DASi-system; Rass and Steeger, 2000) to filter the signal following a predetermined frequency shape. We applied the correction factors described by Dillon (1986) to compensate for the presentation of signals by headphones instead of by a behind-the-ear hearing aid. In order to use this study as an appropriate reference for future studies with hearing-impaired subjects the normal hearing subjects in this study receive slight amplification for certain frequencies as well. The NAL-prescription rule was used for this purpose (National Acoustic Laboratory, Byrne and Dillon, 1986). In effect some frequency shaping was applied.

After this filtering, the signal was transferred back to the signal-processing hardware system in stage 3. In this stage the signal is amplified with a microphone amplifier (TDT MA2) and buffered by a headphone-buffer module (TDT HB6). Finally, in stage 4 the signal was presented monophonically by Telephonics TDH-39P headphones to the subjects seated in a sound-attenuating booth.

III. RESULTS

A. Pattern search optimization with adaptive step size

1. Statistics

The optimization procedure ended most often on meeting two stop criteria simultaneously. They were a setting is frequented three times and a setting has won a paired comparison more than six times. This might indicate that there is a close relation between these stop criteria. The total number of paired comparisons depended on subject and on noise type. To determine the significance of these differences, Wilcoxon (matched pairs signed rank sum) tests were performed. For continuous noise the number of paired comparisons was significantly lower (Wilcoxon, $p<0.05$) than for fluctuating speech-shaped noise, on average they are 30 and 38, respectively. In the paired comparisons the second alternative is used not significantly more frequently than the first alternative, although the second alternative tends to be used more often.

The relation between use of the ''repeat'' and ''no difference'' button was also subject dependent. Most subjects did not use the ''no difference'' button frequently. So, these subjects did not seem to have frequent problems in distinguishing the fragments, irrespective the step size. Two subjects used the no difference button relatively often, probably

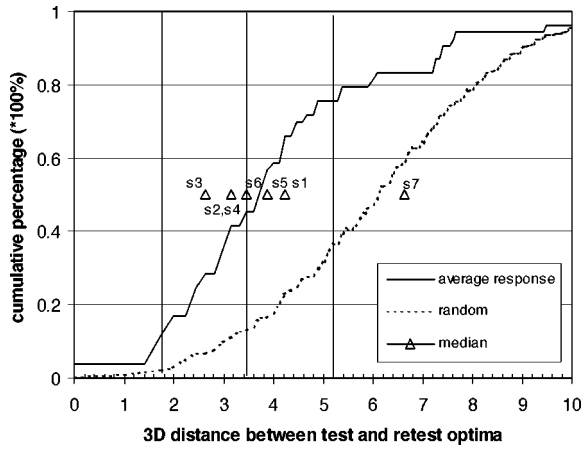


FIG. 1. Cumulative distributions for the calculated 3D distances between the test and retest optima, averaged for two noise types and seven normal hearing subjects (average response). As a reference, the random distribution is also given as calculated based on randomly assigned test and retest optima. The triangles represent the median 3D distance values for each individual (s1 to s7).

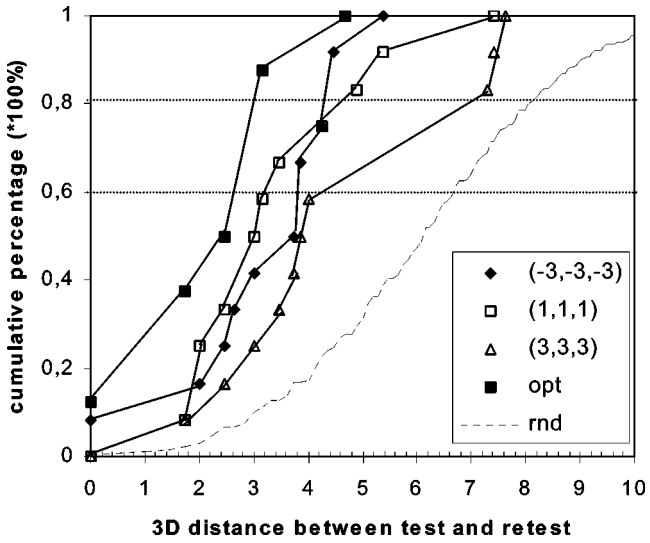


FIG. 2. Cumulative distributions for the calculated 3D distances between the test and retest optima, averaged for six subjects and differentiated by initial setting. The initial setting was $(-3,-3,-3)$, (1,1,1), (3,3,3), or the average optimal setting (opt). As a reference, the random distribution is also given as calculated based on randomly assigned test and retest optima.

because they sometimes had difficulties hearing differences between the fragments. One subject almost always used the repeat button when there was no perceived difference. For the fluctuating noise relative to the continuous noise, the repeat button was used twice as often and the no difference button a factor four more often.

After a startup round with a step size of 4, the step size was halved to 2. From this point on the step size was recorded to determine the minimum step size and maximum step size used in the search. The minimum step size could either be 1 or 2. The maximum step size could be either 2 or 3. The results showed that the minimum step size per dimension was comparable for both noise types and had a value of 1 for approximately 80% of all paired comparisons. However, the maximum step size was noise dependent. A maximum step size of 3 was found in 7% and 27% of the cases, for continuous noise and fluctuating noise, respectively. This large difference was found to be mainly caused in the spectral enhancement dimension. Obviously, spectral enhancement settings may be harder to discriminate in fluctuating noise.

2. Reliability

To acquire information about the reliability we used two strategies. Cumulative distributions were used to determine the factors that influence the reliability. A cluster analysis provided information of the spread and location of optimal and suboptimal settings for each individual.

a. Cumulative distributions. The three-dimensional (3D) distance between the optimum found for the test and retest condition has been calculated in order to determine the reliability of the optimization procedure. In principle, the procedure has maximum reliability when the test and retest results always yield the same optimum settings for all conditions. Cumulative distributions were used in order to analyze the spread of possible test-retest differences. The cumulative percentages were calculated as a function of the maximum distance between test and retest optima. These percentages are given in Fig. 1. From this graph it appears that 11%, 45%, and 76% of the data have a test-retest distance smaller than $\sqrt{3}$, $2\sqrt{3}$, and $3\sqrt{3}$, respectively. These distances represent a maximum step size of 1, 2, or 3, in each dimension. As a reference the random distribution is also given, presented as a dashed line. For this random distribution 1000 test and retest results have been chosen randomly, after which the 3D distances have been calculated. The cumulative distribution of the normal hearing subjects deviates clearly from the random cumulative distribution. The cumulative distribution of the normal hearing subjects contains a few distances that are very large. This is mainly due to the results of subject 7. To illustrate the large test–retest differences, we have plotted the median 3D distance for each individual (see Fig. 1), which shows that the performance of s7 was very close to completely random.

To reliably study the dependence of the reliability on noise type and initial setting the results of s7 have been discarded. For different noise types, the cumulative distributions were found to be approximately equal. For different initial settings, however, differences in the distributions appeared, as shown in Fig. 2. Especially for percentages around 80% differences occur. In general, the test-retest reliability was highest when the average optimal setting (opt) was chosen as the initial setting of the procedure. Initial setting (3,3,3) produced the lowest reliability.

The one-dimensional (1D) distances were calculated for each separate dimension. These results are presented in Fig. 3. It appears that the test-retest reliability depends strongly on the algorithm. The reliability was highest for the noise reduction algorithm and poorest for spectral lift. Additionally, the one-dimensional reliability depended on noise type. The reliability for SE and NR was highest for continuous noise, whereas for SL the reliability was highest for fluctuating speech-shaped noise (not shown in Fig. 3).

b. Cluster analysis. A cluster analysis has been performed to investigate the spread of optima found for differ-

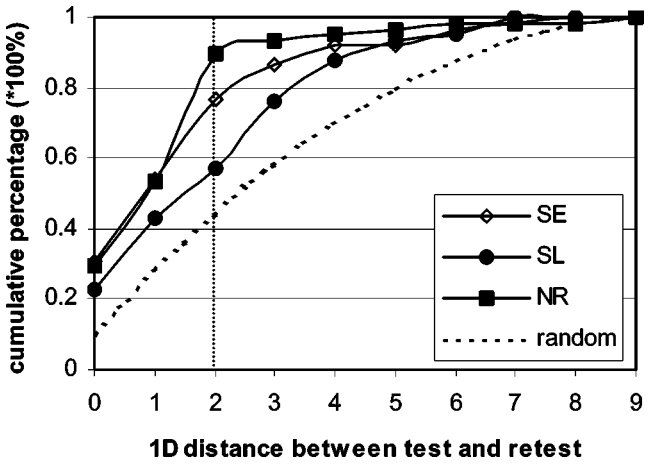


FIG. 3. Cumulative distributions for the calculated 1D distances, i.e., the difference between test and retest for each dimension separately. The dimensions of the optimization procedure represent the algorithms Spectral Enhancement (SE), Spectral Lift (SL), and Noise Reduction (NR). For comparison, the random distribution is also given.

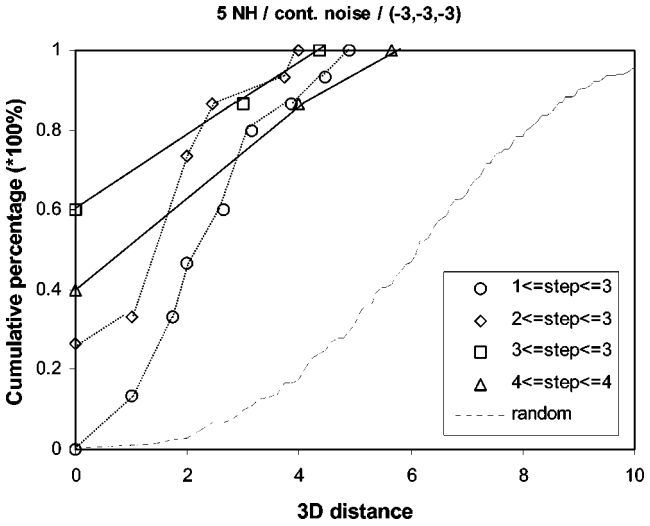


FIG. 4. Cumulative distributions for the calculated 3D distances between the test and retest optima, averaged for five normal hearing subjects (5NH) and differentiated by step-size configuration. The pattern searches were performed for continuous noise (cont. noise) and initial setting $(-3,-3,-3)$ only. The step size can be either adaptive and vary between 1 and 3 ($1 \leqslant \text{step} \leqslant 3$), or between 2 and 3 ($2 \leqslant \text{step} \leqslant 3$), or the step size can be fixed at 3 ($3 \leqslant \text{step} \leqslant 3$) or fixed at 4 ($4 \leqslant \text{step} \leqslant 4$). As a reference, the random distribution is also given as calculated based on randomly assigned test and retest optima.

ent conditions in the three-dimensional space of algorithm settings. The method of hierarchical clustering has been used, based on nearest neighbors and Euclidean distances. The results of the cluster analysis are presented in Table I. The numbers of final settings that are included in the cluster analysis are indicated in the second column (total). The criterion to assign different optima to the same cluster is the average distance between the cluster center and the other points, which should be lower than $\sqrt{3}$. The number of clusters that result from the analysis with this criterion is given in the third column (# clusters). The largest cluster is assumed to be the optimum. The one but largest cluster is considered a suboptimum. The algorithm settings that represent the optimum or suboptimum are rounded values of the cluster center. The number of points (# points) that form the optimum and suboptimum are also included in Table I. Clusters that contain only one or two final settings are considered to be outliers and are given in the last column (# outliers).

For subjects s1 to s6 at least 50% of the final settings is grouped in one cluster. So, for the majority of the subjects the optimization strategy can find an optimal combination of algorithm settings. For s1, s4, s5, and s7 there is also a second cluster of reasonable size. For s1 and s4 this is a true suboptimum, i.e., a serious competitor for being most comfortable. The presence of two large clusters for subjects s5 and s7 appear to be optima that relate specifically to the two noise conditions.[3] The percentage of outliers is 25% or less for all but one subject. Subject s7 has a large scatter among the final settings found by the optimization procedure.

There proved to be a clear relation between the number of paired-comparisons needed and the size of the clusters. For the large and medium optimum and suboptimal clusters the procedure stopped on average after 25 and 30 paired comparisons, respectively. For the outliers, 54% of the procedures that were performed took 45 to 54 paired comparisons, which is close to, or at, the upper limit of paired comparisons needed.

TABLE I. Results of a cluster analysis for seven subjects (s1 to s7). The total number of points (total) included in the analysis and the number of clusters (#clusters) that result from the analysis are given in the second and third column. The two largest clusters are assumed to form the optimum and suboptimum. The algorithm settings of the optimum and suboptimum are indicated. SE, SL, and NR represent algorithms Spectral Enhancement, Spectral Lift, and Noise Reduction. Clusters that contain only one or two final settings are called outliers. The number of points (#points) are given for the optimum cluster, the suboptimum cluster and the clusters of outliers.

			Cluster1: optimum				Cluster2: suboptimum				Outliers
Subject	Total	#Clusters	SE	SL	NR	# points	SE	SL	NR	# points	# points
s1	16	4	−1	5	0	11	0	1	0	3	2
s2	16	4	1	4	0	12					4
s3	16	3	−1	−1	0	14					2
s4	16	6	0	1	0	8	0	−3	1	4	4
s5	14	4	0	3	2	7	−1	−2	1	4	3
s6	14	4	−1	1	1	10					4
s7	16	8	5	−3	5	5	−1	−3	6	4	7

B. Pattern search with step size as parameter

The results of all four step-size configurations (step123, step23, step3, step4) are plotted as cumulative distributions in Fig. 4. For comparison, the random distribution, based on randomly assigned test and retest optima, is plotted in the dashed line.

The results show that the reliability of the procedure tends to increase when the minimum step size increases from one to three (for step123, step23, and step3, respectively). For a fixed step size of four the reliability decreases relative to a step size of 3. Both for the fixed (solid lines) and for the variable step size configurations (dotted lines) there is a parallel shift of the distributions. This shift indicates an increase in reliability for the variable configurations and a decrease for the fixed configurations when the step size is enlarged.

The degree of convergence was studied by comparing the optima found for each step size configuration. The average results indicate that the 3D distance increases with minimum step size, at least for minimum step sizes between 1 and 3. The 3D-distance is 1.0, 1.8, 2.0, and 1.7 for configurations step123, step23, step3, and step4, respectively. There is a significant difference ($p<0.05$, Wilcoxon matched pairs signed rank sum test for the results of five subjects and four configurations) between step123 and the step3 configuration, which indicates that the optimization procedure converges better for the variable step size configuration.

Finally, the number of paired comparisons needed to end the optimization procedure appeared inversely proportional to the minimum step. For step123, step23, step3, and step4 the number of paired comparisons is 22.2, 16.2, 16.2, and 13.8, respectively.

IV. DISCUSSION

A. Adaptive pattern search

An advantage of the pattern search method is that no a priori information is required. For perceptually governed optimization that depends on the listener's capacities and experience this method shows promising capabilities. However, the perceptual capacities of the listeners also constrain the methodological possibilities of the optimization method. The adaptive step size that is used in this study helps to create perceptually distinguishable fragments to be used in the paired comparisons. The results of the pattern search procedure suggest that settings that differ 2 or 3 steps can be judged reliably for our set of stimuli. That is, the reliability of the optimization procedure is high for a variable step size that ranges between 2 and 3, and for a fixed step size of 3.

The optimal settings have been estimated in two different ways considering all pattern searches for each listener. The estimators comprise the average optimum setting (O_{avg}) and the center of the largest cluster (Cl). The three-dimensional distance between the two estimators can be determined. The individual results, average results, and standard deviations are presented in Table II for each background noise type. The results of this study suggest that for both noise types the correspondence between the cluster center and the average optimum setting is close for all but one subject. This can be seen from the small average 3D-distances that are smaller than $\sqrt{3}$ for all but one subject (distances larger than $\sqrt{3}$ are presented in gray in Table II). Although the results presented in Table II cannot simply yield information about the quality of the estimators, the cluster center of the largest cluster is assumed to be the best estimator of the optimum. For in some pattern searches (for different tests and starting points) the ''real'' optimum can be missed when there are grave discontinuities in the three-dimensional perceptual landscape or due to procedural inaccurateness as a result of step sizes used.

TABLE II. Three-dimensional distances between the cluster optimum (Cl) and the average optimum (Oavg) differentiated by subject (s1 to s7) and noise type. The noise type was either continuous car noise (cont) or fluctuating speech-shaped noise (fluct). Additionally, the average three-dimensional distances (avg) and standard deviations (stdev) are indicated. Values larger than $\sqrt{3}$ are presented in italic.

	3D-dist(Cl, Oavg)	
Subject	cont	fluct
s1	1.3	*3.1*
s2	0.2	0.3
s3	0.2	0.3
s4	*3.2*	0.6
s6	0.3	0.5
s7	0.4	0.5
avg	0.9	0.9
stdev	1.2	1.1

B. Reliability and convergence

The cumulative distributions of all subjects except for s7 indicate that the three-dimensional distance between test and retest is lower than $2\sqrt{3}$ for about 50% of the results. The cluster analysis illustrates the presence of large and intermediate clusters around the optimum, small clusters around a suboptimum, and some outliers.

In studies of Neuman *et al.* (1987), Kuk and Pape (1992), and Kuk and Lau (1996) the reliability of the modified ''Simplex'' procedure was higher than in this study. The most important explanation for that may well be connected to the configuration and the stimulus material of the optimization procedure. In our setup the subjects listened to sentences that were between 1.5 and 2.5 s in duration. Moreover, in our experimental paradigm speech-in-noise fragments were presented in succession and could be repeated only once. A problem of this design is that the auditory memory is not sufficient to compare the fragments: At the end of fragment 2, a part of fragment 1 is already forgotten. In the studies mentioned above the paradigm is very different in that continuous discourse is used and listeners are allowed to switch freely between two parameter settings. By switching freely, the effects of auditory memory are less acute, because of the possibility to return to the alternative. In a small parameter or algorithm space such a paradigm appears to be convenient, but for large multidimensional setups, like in this study, such a slow approach is unpractical.

Another, maybe more influential difference is the level equalization of the stimulus material in our study. For the study of Neuman *et al.* (1987) the loudness of low frequen-

cies and high frequencies of the stimulus material was one of the optimization parameters under study. For the studies of Kuk and Pape (1992) and Kuk and Lau (1996), the parameter settings of the two-dimensional optimization procedure were the degree of amplification used for high and low frequencies. However, because the subjects had to adjust the level of the test material to their most comfortable level (MCL), the effective levels can be quite different for different settings. The advantage of these approaches is that the reliability is high because level differences are relevant cues for converging on the optimum setting. An additional advantage is that (parts of) the stimulus material can never be too loud or too soft. Our optimization procedure was intended to focus on the effects of the algorithms themselves and not on level differences. The reason for this is that loudness differences can act as a separate and different evaluation criterion. The procedure used in this study avoids that the optimum setting found is the result of a loudness judgment exclusively.

The choice of the background noise type in combination with the selected signal processing algorithms also affects the reliability. An explanation could be that the differences between settings, in the paired comparisons in the optimization process, are much more difficult to detect for the fluctuating noise than for the continuous noise. Subjects indicated that processing artifacts are much more prominent for the continuous car noise condition relative to the fluctuating speech-shaped noise condition.

When the maximum number of paired comparisons was reached the end point is not likely to be the optimum, and the setting that has won most paired comparisons is in most cases not the optimal setting either. Conversely, when the end point is reached rather quickly, within nine to eighteen paired comparisons, this point is much more likely to be the true optimum. The reliability found for each noise type can be directly related to the stop criterion, if addressed. For fluctuating noise, that had a relatively low reliability, the procedure stopped most often when the maximum number of paired comparisons was reached. For continuous noise, the procedure ended both on three visits and more than six wins. This might imply that combining these both stop criteria yields highest reliability.

The adaptive pattern search of study A (step-size configuration step123) appeared to have the best convergence properties. The unexpected reversal with respect to the degree of convergence as a function of step size might be due to the case that a grid point was accidentally located near avg1. Further inspection of the data revealed that this was the case in only one subject. In three out of five subjects avg1 proved to be closer to a grid point in the step3 configuration than in the step4 configuration. So, this effect is not very likely to be the only explanation.

C. Auditory constraints

In this study we did not record the difference limens for each particular dimension. Although these limens can be measured for each algorithm separately, that would still not produce any knowledge of the limens for combinations of these nonlinearly acting algorithms. The present study shows that the adaptive step size can partly compensate for any unfortunate choices of the step size. The step size configuration appears to affect the reliability. Apart from the fixed step size configurations, we found that the highest reliability was achieved for a minimum step size of two and a maximum step size of three (step23). A variable step size helps to create perceptually distinguishable fragments. In the same line, the elevated minimum step size has an advantageous effect in avoiding the presentation of indistinguishable fragments.

In the studies of Neuman *et al.* (1987) and Kuk and Pape (1992), it seems that the possible existence of multiple optimum settings played a minor role. That is, the combination of the parameters under study yielded a response surface that curved smoothly around a single optimum. Yet, in the study of Kuk and Lau (1996) category rating results showed a multimodal response pattern for three out of seven subjects, when evaluated for clarity for SNR=0 and SNR=5 dB. Also the results of a previous pilot study indicate that the response surface can sometimes be multimodal. In this study, the cluster analyses imply the presence of multiple optima.

V. CONCLUSIONS

Inclusion of an adaptive step size has two advantages. The perceptibility between fragments can be controlled. That is, an increase in the step size can be beneficial in order to create perceptual differences between settings that are compared. As a result the steps in the procedure can be estimated more reliably. The gradual decrease of the step size helps the procedure to converge towards the optimum setting. The reliability, however, depends on many factors, especially on initial value, stop criterion, step size constraints, background noise, algorithms used, and the possible presence of suboptimal settings. Reliability and convergence appear to have a trade-off effect. Specifically, when the step size is enlarged the reliability improves, but the convergence deteriorates.

The results of the cluster analysis and of the cumulative distributions correspond closely, especially for continuous car noise, with the exception of one subject. The cumulative distributions indicate that for the pattern search about 50% of the test-retest optima have a three-dimensional distance lower than $2\sqrt{3}$. The cluster analyses show that large clusters can be formed that cover many end points of the pattern search found for different conditions. For four subjects a second cluster could be identified that either represented the optimum for the second background noise, or a suboptimum.

The presence of a suboptimum can cause the optimization strategy to stop prematurely. Another drawback of the present pattern search procedure is the problematic division between ''not distinguishable'' and ''no preference.'' Nevertheless, the simplicity and efficiency of the search and the assimilation of algorithm interactions make the optimization strategy suitable for fine-tuning of auditory signal-processing algorithms.

ACKNOWLEDGMENTS

This study was supported by the Heinsius-Houbold Fund. We like to thank László Körössy of the Academic Medical Center Amsterdam for his technical support.

[1]Using the paired-comparison paradigm the expression ''Simplex method'' might be confusing. A ''Simplex'' designates a Euclidean geometrical spatial element bounded by a minimum number of points. For two dimensions a Simplex is a triangle, in three dimensions a tetrahedron. So for n dimensions a Simplex contains $n+1$ vertices (boundary points). As such, the expression Simplex can be used, because the settings that are compared pair wise correspond to vertices of a Simplex configuration. In the procedure new Simplices are formed by reflecting one point in the center of gravity (centroid) of the remaining points. This so-called Simplicial feature of the Simplex method is, among others, fundamentally different from the (modified) Simplex procedure of Neuman *et al.* (1987). When using that procedure in a paired-comparison paradigm, it is impossible to decide what vertex should be reflected. This meant that different procedural rules had to be incorporated. Instead of reflection in one point, as is the case in the original procedure, the new Simplex is created by a combination of translations and line-reflections. As a consequence, the new Simplex can share at most 1 point with the old Simplex, while in the procedure of Spendley *et al.* (1962) the new Simplex always shares $n-1$ points with the old Simplex. This procedural difference indicates that the original elegant movement of the Simplex was abandoned in the modified procedure of Neuman. To avoid ambiguous terminology, the most important features of our (and Neuman's) procedure have been selected and compared with other procedures that can be classified as direct search methods (see e.g., Torczon, 1989; Lewis *et al.*, 1998, 2000). To appreciate the lattice structure and the constrained search directions, it is convenient to denote our procedure ''multidirectional pattern search.''

[2]In order to check this assumption, the Speech Reception Thresholds (SRT) of six subjects were measured. This threshold represents the signal-to-noise ratio at which 50% of sentences in noise are repeated entirely correctly (Plomp and Mimpen, 1979). On average, the subjects' SRT is -3.6 (1.6) and -9.1 (2.5) dB for two standard noises, continuous speech-shaped noise and fluctuating speech-shaped noise, respectively. This corresponds well with results of Festen and Plomp (1990) and is substantially lower than the SNR of 0 dB for the speech-in noise fragments used in this study. The SNR for continuous car noise, which was not measured here, can be expected to be lower than the SNR for continuous speech-shaped noise due to the fact that continuous speech-shaped noise will mask speech more efficiently than continuous car noise. So, the paired comparisons were performed for speech in noise presented at levels for which the percentage correct score was well over 50%.

[3]When for each noise type cluster analyses were carried out for each individual, the same conclusion could be drawn. Moreover, it appeared that the number of optima that form the largest cluster and the number of outliers were comparable for subjects s8 and s9 and the other seven subjects. Hence, it is assumed that the results of the five subjects in experiment 2 can be compared safely amongst each other and with the results of experiment 1.

Byrne, D., and Dillon, H. (**1986**). ''The National Acoustic Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid,'' Ear Hear. **7**(4), 257–265.

Dillon, H. (**1997**). ''Converting insertion gain to and from headphone coupler responses,'' Ear Hear. **18**(4), 346–348.

Festen, J. M., and Plomp, R. (**1990**). ''Effect of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing,'' J. Acoust. Soc. Am. **88**(4), 1725–1736.

Kuk, F. K., and Pape, N. C. M. (**1992**). ''The reliability of a modified simplex procedure in hearing aid frequency-response selection,'' J. Speech Hear. Res. **35**, 418–429.

Kuk, F. K., and Lau, C. (**1996**). ''Comparison of preferred frequency gain settings obtained with category rating and modified Simplex procedure,'' J. Am. Acad. Audiol **7**, 322–331.

Lewis, R. M., Torczon, V., and Trosset, M. W. (**1998**). ''Why pattern search works,'' Optima (Mathematical Programming Society Newsletter), 1–7.

Lewis, R. M., Torczon, V., and Trosset, W. (**2000**). ''Direct search methods: then and now,'' J. Comput. Appl. Math. **124**, 191–207.

Lyzenga, J., Festen, J. M., and Houtgast, T. (**2002**). ''Speech enhancement scheme incorporating spectral expansion evaluated with simulated loss of frequency selectivity,'' J. Acoust. Soc. Am. **112**(3), 1145–1157.

Neuman, A. C., Levitt, H., Mills, R., and Schwander, T. (**1987**). ''An evaluation of three adaptive hearing aid selection strategies,'' J. Acoust. Soc. Am. **82**(6), 1967–1976.

Plomp, P., and Mimpen, A. M. (**1979**). ''Improving the reliability of testing the speech reception threshold for sentences,'' Audiology **18**, 43–52.

Rass, U., and Steeger, G. H. (**2000**). ''A high performance Pocket-size System for Evaluations in Acoustic Signal Processing,'' Acta Acust. (Beijing) **86**, 374–375.

Spendley, W., Hext, G. R., and Himsworth, F. R. (**1962**). ''Sequential Application of Simplex Designs in Optimisation and Evolutionary operation,'' Technometrics **4**, 441–461.

Torczon, V. J. (**1989**). ''Multi-directional search: a direct search algorithm for parallel machines,'' Thesis for philosophy doctorate, Rice University, Houston, Texas.

Versfeld, N. J., Daalder, L., Festen, J. M., and Houtgast, T. (**2000**). ''Method for the selection of sentence materials for efficient measurement of the speech reception threshold,'' J. Acoust. Soc. Am. **107**(3), 1671–1684.

Auditory processing of real and illusory changes in frequency modulation (FM) phase

Robert P. Carlyon,[a] Christophe Micheyl,[b] and John M. Deeks
MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, United Kingdom

Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 4 March 2004; revised 22 July 2004; accepted 9 September 2004)

Auditory processing of frequency modulation (FM) was explored. In experiment 1, detection of a π-radians modulator phase shift deteriorated as modulation rate increased from 2.5 to 20 Hz, for 1- and 6-kHz carriers. In experiment 2, listeners discriminated between two 1-kHz carriers, where, mid-way through, the 10-Hz frequency modulator had either a phase shift or increased in depth by ΔD% for half a modulator period. Discrimination was poorer for $\Delta D = 4\%$ than for smaller or larger increases. These results are consistent with instantaneous frequency being smoothed by a time window with a total duration of about 110 ms. In experiment 3, the central 200-ms of a 1-s 1-kHz carrier modulated at 5 Hz was replaced by noise, or by a faster FM applied to a more intense 1-kHz carrier. Listeners heard the 5-Hz FM continue at the same depth throughout the stimulus. Experiments 4 and 5 showed that, after an FM tone had been interrupted by a 200-ms noise, listeners were insensitive to the phase at which the FM resumed. It is argued that the auditory system explicitly encodes the presence, and possibly the rate and depth, of FM in a way that does not preserve information on FM phase.
[DOI: 10.1121/1.1811474]

PACS numbers: 43.66.Ba, 43.66.Nm, 43.64.Bt [GDK] Pages: 3629–3639

I. INTRODUCTION

In this article we investigate how the auditory system encodes changes in the frequency of a sound over time. One simple scheme would be for the system to take a series of ''snapshots'' of the stimulus (Hartmann and Klein, 1980; Dooley and Moore, 1988a, b; Demany and Semal, 1989; Plack and Carlyon, 1995). According to this scheme, when frequency modulation (FM) is imposed on a carrier, the frequency estimated from successive brief temporal windows changes; if this change exceeds a criterion value, then the FM can be detected. A version of this general idea, in which the listener combines multiple ''looks'' obtained throughout the stimulus (Hartmann and Klein, 1980), allows for the decrease in FM detection thresholds (''FMDTs'') with increasing duration. The snapshot hypothesis also predicts that FMDTs should increase with modulation rate, because a window of a given duration will, at higher rates, average the instantaneous frequency over a greater proportion of the modulator period. This prediction has been confirmed for sinusoidal carriers having frequencies up to 4 kHz (Demany and Semal, 1989). Demany and Semal's data indicate that, for these frequencies, the ''snapshot'' hypothesis does a good job of accounting for the detection of sinusoidal FM. Sek and Moore (1995) have argued that, for low-to-medium carrier frequencies, FM is encoded by changes in the pattern of phase locking to the carrier.

An alternative viewpoint is that dynamic changes in the frequency of a sound are explicitly encoded by the auditory system in some way. Unlike the snapshot hypothesis, this view asserts that there is something about the way that the system encodes, say, sinusoidal FM, that is qualitatively different from the processing of a series of brief but separate tones of differing frequency. Early physiological evidence for this ''explicit coding'' of FM came from Rees and Moller (1983), who found units in the rat inferior colliculus (IC) whose firing rates varied systematically as a function of FM rate. Psychophysical evidence comes from a number of paradigms. Dooley and Moore (1988a) measured duration discrimination for tones that were either fixed in frequency and level, or that glided in frequency and/or level over a fixed extent. For the gliding stimuli, the rate of change of frequency and/or level co-varied with duration. Dooley and Moore found that duration discrimination was better for the gliding stimuli than for the steady stimuli, and argued that this indicated a sensitivity to the rate of change of frequency and/or level.

More recently, Cusack and Carlyon (2003) presented subjects with a sequence of tones that were quasi-randomly distributed in frequency and time, and which, on 50% of trials, contained a ''target'' tone. Performance was much better when the target was frequency modulated and the other tones (''interferers'') were steady than for the converse condition where the target was steady and the interferers were all modulated. They interpreted this finding as evidence for a perceptual asymmetry in hearing, and argued that it reflected the presence of neural feature detectors sensitive to the presence of FM. A related finding was reported by Gockel and

[a]Electronic mail: bob.carlyon@mrc-cbu.cam.ac.uk
[b]Present address: Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139-4307.

Carlyon (2000), who investigated FM detection interference (FMDI) in conditions where the (unmodulated) standard and the (modulated) signal were each preceded and followed by an interfering tone in a remote frequency region. When the interferers were unmodulated, the signal interval could be identified by spotting a modulated tone in a background of five steady tones (four interferers plus the standard). FM detection thresholds (FMDTs) were lower in this condition than when the interferers were modulated, and where the task can be thought of as detecting a single unmodulated standard from a sequence of five modulated tones (four interferers plus the signal). They also concluded that their results were consistent with a perceptual asymmetry between the processing of modulated and of steady sounds. Indeed, Carlyon (2000) has argued that this is true for more ''conventional'' FMDI paradigms, in which the interferers and targets are presented simultaneously (Wilson *et al.*, 1990; Moore and Shailer, 1992; Carlyon, 1994). Finally, Sek and Moore (1999) required subjects to discriminate between a pair of pure tones that had the same frequency and a pair in which the second tone was higher than the first. Performance was improved when the silent gap between the two tones in each pair was replaced by a linear frequency glide. Again, the results were interpreted as evidence for a mechanism sensitive to dynamic frequency changes.

Although there is a substantial amount of evidence for mechanisms sensitive to the *existence* of FM, evidence for mechanisms sensitive to FM *phase* is much weaker. Indeed, the behavioral evidence for FM-specific mechanisms derives entirely from paradigms in which subjects were not required to discriminate the phase of FM. This was true for Cusack and Carlyon's work on perceptual asymmetries, and for experiments on FMDI, which is unaffected by the relative phase of the target and interferer FM (Moore *et al.*, 1991; Carlyon, 1994). It was also the case for Sek and Moore's study on the detection of frequency differences between tones that could be joined by linear frequency glides, because subjects were not required to report the direction of the frequency change. Indeed, a recent replication and extension of their study showed that, when a direction judgment was required, the addition of a frequency glide did not help (Lyzenga *et al.*, 2004). Finally, it is worth noting that subjects cannot detect differences in FM phase between two inharmonically related tones that occupy separate frequency regions, when cues associated with distortion products and within-channel interactions are minimized (Carlyon, 1991, 1994, 2000).

The above considerations lead to the interesting hypothesis that the auditory system may encode the presence (and perhaps the rate and depth) of FM in a way that does not preserve information on the modulator phase. This idea can be reconciled with the snapshot hypothesis if one assumes that, at an early stage of processing, changes in instantaneous frequency over time are smoothed with a form of temporal window. To the extent that the modulator period exceeds the window duration, the modulator phase will be preserved in the window output. The output of the window may be used for detecting FM, but, once the FM depth exceeds threshold, an additional form of encoding can be used. This latter form of encoding may register the presence (and perhaps the depth and rate) of FM, but not its phase.

According to this idea, for supra-threshold FM depths, the form of processing that is used depends on the task. We have described some paradigms in which FM phase is not important, and, in these cases, it may be that FM is processed using the additional FM-specific coding mechanism described above. In other tasks, where FM phase information is crucial, listeners will be forced to rely on the output of the temporal window. Under such conditions, we would expect good performance to occur only when the changes in instantaneous frequency are slow enough to be preserved in the output of the temporal window. The experiments reported here evaluate this scheme, by measuring the processing of changes in FM phase both mid-way through a modulated tone, and in a stimulus giving rise to the continuity illusion (Warren *et al.*, 1972).

II. DETECTION OF AN FM PHASE REVERSAL

A. Rationale

Experiment 1 had two aims. First, because later experiments investigated the detection of ''illusory'' changes in modulator phase, it was necessary to identify those conditions under which physical phase shifts could be detected. Second, by measuring detection of phase shifts as a function of FM rate, we could gain some insight into the extent of any ''temporal sluggishness'' in a task using easily detectable FM depths. Evidence for this sluggishness has been previously reported for FM detection thresholds by Sek and Moore (1995), who noted that it is greatest at low-to-moderate frequencies, where, they argue, FM is detected by virtue of a phase-locking code [see also Moore and Sek (1996) and Gockel *et al.* (2001)]. We therefore measured detection of FM phase reversals for two carrier frequencies—1 and 6 kHz.

B. Method

In each two-interval two-alternative forced-choice (''2IFC'') trial, listeners were presented with two tone bursts, frequency modulated at a depth of $\pm 10\%$ of the carrier frequency, and separated by a silent interval of 500 ms. The level of each tone was 60 dB SPL. On each trial, the carrier frequency was held constant at either 1 or 6 kHz, and the modulation rate was fixed at either 2.5, 5, 10, or 20 Hz. The starting modulation phase was selected at random from 0 and π radians. Both the standard and signal had a nominal duration of 800 ms. We describe the generation of the standard and signal for this nominal duration first, before describing why and how the duration sometimes differed from this value.

The standard stimulus is illustrated in the left-hand part of Fig 1(a), and was generated according to Eq. (1):

$$y(t) = \sin(2\pi f_c t + \theta + g(t)), \tag{1}$$

where $g(t)$ is a function describing the FM (implemented by phase modulation):

$$g(t) = (D_m / f_m)[\cos(\phi) - \cos(2\pi f_m t + \phi)] \tag{2}$$

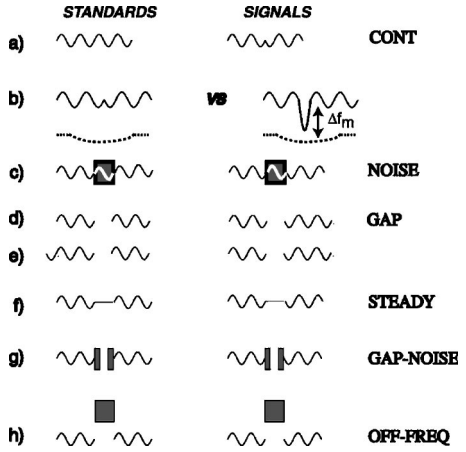


FIG. 1. Schematic spectrograms of some of the stimuli used in the experiments. Standards are shown in the left-hand column and signals in the right. Labels on the far right refer to the names of conditions used in experiments 4 and 5. The dotted lines in part (b) represent the stylized output of a hypothetical temporal window, as discussed in the text. In part (c) the trajectory that the FM tone would have had, if not interrupted by the noise, is shown by the white trace.

and where f_c and f_m are the carrier and modulator frequencies, respectively. The zero-to-peak FM depth is defined as the difference in Hz between f_c and the maximum instantaneous frequency, is denoted by D_m, and is expressed here as a percentage of the carrier frequency ($100\,D_m/f_c$). The initial carrier and modulator phases are given by θ and ϕ, respectively. The value of θ was always zero, and ϕ was equal to zero on half of the trials and π on the other half, selected at random.

The signal differed from the standard in that the FM phase was shifted by π radians mid-way through the stimulus [i.e., after 400 ms; see right-hand part of Fig. 1(a)]. For brevity, we refer to this π-radians phase shift as a "phase reversal." To produce this reversal, the first half of the stimulus was generated according to Eq. (1), again with $\theta=0$ and $\phi=0$ or π. The second half was generated with a new value of ϕ set to differ by π radians from the value it would have had in the absence of a reversal. To avoid any discontinuities in the fine structure associated with this change in the modulator phase, the value of θ for the second half of the signal was set to the value of $g(t)$ at the end of the first half.

Two precautions had to be taken to prevent listeners from using cues other than the FM phase reversal to perform the task. First, as mentioned above, ϕ was randomized from presentation to presentation. If this had not been done, listeners might have identified the signal from the FM phase at the end of the stimulus—as shown in Fig. 1(a), which illustrates a trial with $f_m=5$ Hz and $\phi=0$ for both the standard and signal. The difference in the final phase of the FM is rendered useless as a cue by the randomization of ϕ. However, even with this randomization, if the stimulus duration were fixed, there would be other ways of identifying the signal. For example, the direction of the change in instantaneous frequency just after the start of the signal stimulus is the same as that just before its end, and this is not the case for the standard [Fig. 1(a)]. To prevent this being used as a cue, a half-cycle of FM was added to the start of 50% of all stimuli, and, independently, a half-cycle of FM was appended to the end of 50% of stimuli. These manipulations, which were applied at random from presentation to presentation, increased the signal duration by up to $1/f_m$ s, and forced listeners to base their judgments on the FM phase reversal near the middle of the signal.

C. Listeners and procedure

Eight normal-hearing listeners took part in the trials with a carrier frequency of 1 kHz; four of these also participated in trials with a carrier frequency of 6 kHz. Eight blocks of 25 trials were obtained for each listener for each combination of f_m and f_c, so each data point was based on the average of 200 trials per listener. The values of f_m and f_c were fixed within a block of trials, and the blocks were run in an order counterbalanced for both f_m and, for those listeners who were tested at both carrier frequencies, f_c. Listeners were seated individually in a double-walled sound-attenuating booth and listened through one earpiece of a Sennheiser HD250 Linear 2 headset.

D. Results

The symbols connected by solid lines in Fig. 2 show performance averaged across all listeners tested at carrier frequencies of 1 kHz (squares, $N=8$) and 6 kHz (triangles, $N=4$). To show that the differences in performance between the two carrier frequencies were not due to differences in the

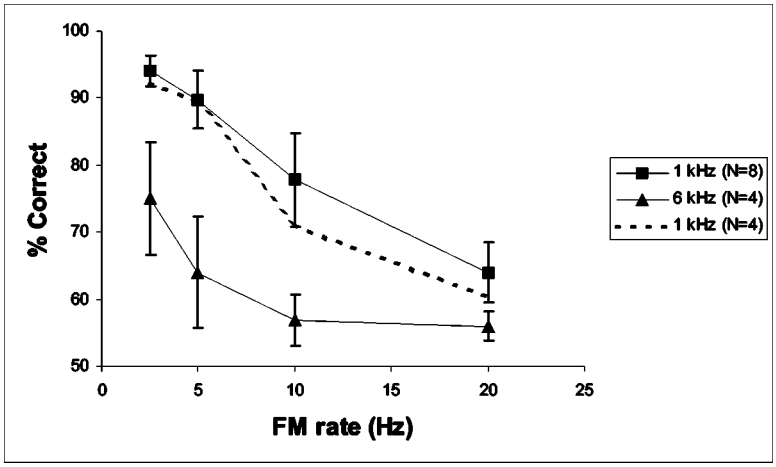


FIG. 2. Mean percent correct detection of an FM phase reversal [cf. Fig. 1(a)] as a function of FM rate, for carrier frequencies of 1 kHz (squares) and 6 kHz (triangles). The dotted line without symbols shows the data for the 1-kHz carrier averaged only over the subset of listeners who were also tested at 6 kHz. Error bars show the standard error across listeners.

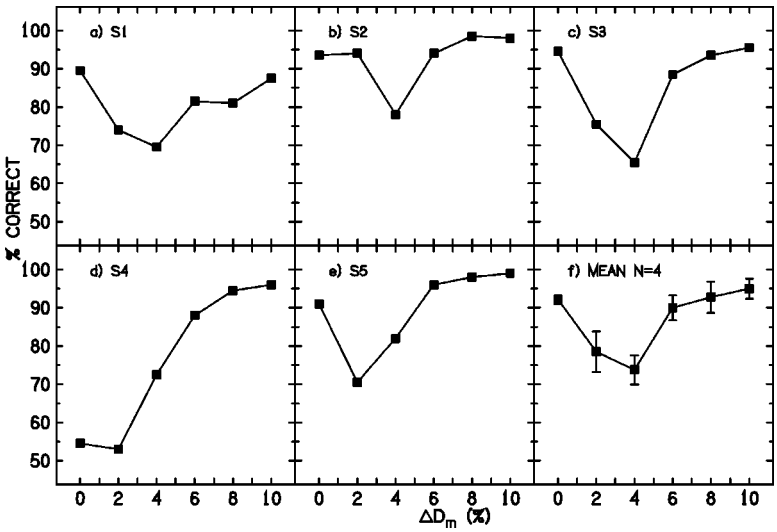


FIG. 3. Percent correct discrimination between an FM phase reversal and an increase in modulator depth over a half-cycle, as a function of the size of that increase (experiment 3). Panels (a)–(e) show data for subjects S1–S5, and panel (f) shows the mean and standard error across all subjects except S4.

subset of listeners tested, mean data for $f_m = 1$ kHz obtained only with those listeners tested at both frequencies are plotted as the dotted line.

For both carrier frequencies, performance decreased markedly as the modulation rate increased. Performance was better for the 1-kHz carrier than for the 6-kHz carrier. One implication of the latter result is that phase reversals were not detected because they caused momentary changes in the frequency spectrum of the signal, because the tones were modulated by the same percentage of each carrier frequency, and because auditory filters are slightly sharper, when expressed as a proportion of their center frequency, at 6 kHz than at 1 kHz (Glasberg and Moore, 1990).[1] The reasons for the difference are not certain, but one possibility is that the most efficient way of tracking changes in instantaneous frequency is via a phase-locking mechanism, which is more effective at low carrier frequencies. A similar explanation has been proposed for the deterioration in pure-tone frequency DLs at high frequencies (Moore, 1973; Micheyl *et al.*, 1998), and for the detection of FM. In the latter case, Sek and Moore (1995) have argued that the increase in FMDTs with increasing modulation rate at low-to-moderate frequencies is due to a ''sluggish'' temporal mechanism, which is not available for high carrier frequencies, where a place-of-excitation cue must be used.

Whatever the reason for the difference in performance between the two carrier frequencies, the deterioration in performance with increasing modulation rate indicates a limitation in the ability of the auditory system to track fast changes in the instantaneous frequency of a sound. Such a limitation might be modeled, at least to a first approximation, by smoothing the function relating instantaneous frequency to time using a sliding temporal integrator (temporal window) of finite duration. As illustrated by the dashed line in the left-hand part of Fig. 1(b), a phase reversal would cause a ''bump'' in the output of such a window. Furthermore, as the modulation rate is increased, more complete cycles of the FM would fall within the window, thereby reducing the size of the bump, and making it harder to detect. This general idea was tested by experiment 2, and a more specific implementation of it is discussed in Sec. VII A.

III. DISCRIMINATING BETWEEN AN FM PHASE REVERSAL AND A BRIEF INCREASE IN FM DEPTH

A. Rationale and method

In experiment 2, listeners discriminated between an FM tone containing a modulator phase reversal, and one containing no such reversal but in which the modulation depth, D_m, was increased by an amount ΔD_m for half a cycle of the modulator [see Fig. 1(b)]. The rationale was that, if instantaneous frequency is averaged by a temporal window, then there should be some value of ΔD_m that produces a bump in the output of that window which is similar to that produced by a modulator phase reversal. If subjects detect the reversal only by virtue of the magnitude of this bump, then they should be poor at discriminating between the two stimuli; this is shown in Fig. 1(b), the dashed lines of which show the output of a hypothetical window.

In each trial, the standard stimulus was a 1-kHz carrier modulated at a rate of 10 Hz and with a depth of 10%, with a π-radians shift in the modulator phase occurring near the middle. This standard was identical to one of the 10-Hz signals in experiment 1; the phase reversal could occur at either a positive or negative zero crossing, and the stimulus had a nominal duration of 800 ms that could be extended by half a modulation cycle at either the beginning, end, neither, or both. This standard was presented in two intervals of each 3I-2AFC trial, but could differ between these two intervals in the starting modulator phase ($\phi = 0$ or π) and/or duration, which was randomized from presentation to presentation. The signal was presented in either the second or third interval of each trial, at random. It was generated in the same way as described by Eq. (1), except that after 400 ms of the nominal duration, the value D_m was increased to $D_m + \Delta D_m$, for one half-cycle of the modulator. To avoid any discontinuities associated with this change in the modulation depth, the value of θ following the momentary increase in modulation depth was set to the value of $g(t)$ at the end of that increase.

B. Results

Panels (a)–(e) of Fig. 3 show percent correct as a func-

tion of ΔD_m for the five listeners tested. Four of these functions are nonmonotonic with a minimum at 4% [panels (a)–(c)] or 2% [panel (e)]. One subject, S4 [panel (d)], shows performance that is near chance for $\Delta D_m \leqslant 2\%$, and which improves monotonically as ΔD_m is raised above 2%. In all cases the minimum level of performance exceeds the chance value of 50%; averaged across listeners, the minimum value was 67% (71% when S4's data were excluded). The final panel (f) shows the functions averaged over all subjects except S4.

The fact that, for most listeners, performance reaches a minimum at intermediate values of ΔD_m is broadly consistent with the idea that instantaneous changes in frequency are averaged using a temporal window with a finite duration. As discussed in Sec. II A [and illustrated in Fig. 1(b)], it is possible that a particular value of ΔD_m will produce a similar change in the output of the temporal window as does an FM phase reversal. Because performance never dropped completely to chance, this class of explanation requires that the way the window output varied over time was never exactly the same for the two types of stimulus. For example, although a phase reversal and a judicious choice of ΔD_m might produce the same maximum excursion in the output of the window, the duration or shape of that excursion might differ. Indeed, Sec. VII A describes a simple model for which this is the case, and which uses a window having a duration of 110 ms. Alternatively, it could be that a finer-grained sampling of values of ΔD_m would reveal a particular value for each subject that yielded chance performance.

IV. PERCEPTION OF FM CONTINUITY

A. Rationale

In a conference presentation, Plomp (1982) reported that when a portion of a 15-Hz FM tone was replaced by a burst of noise, listeners heard the modulation continue throughout the noise burst. However, Plomp did not report any quantitative data on this effect, and the remainder of the experiments described here investigate his finding in more detail. Experiment 3 provides a more formal replication of the original finding (albeit with a slower FM rate), and shows that listeners hear the extent of the FM continue undiminished. This finding, which shows that the *depth* of FM is fully preserved during the illusion, provides a useful contrast to the results of experiments 4 and 5, which show that the *phase* of the FM is not preserved.

B. Methods

The first stimulus presented in each trial consisted of two 400-ms 1-kHz tones, frequency modulated at a depth of $\pm 10\%$ and at a rate of 5 Hz. The tones were turned on and off with 5-ms raised-cosine ramps and were separated by a 200-ms gap (zero-voltage points). They had a level of 60 dB SPL. In one condition, the gap was filled by a digitally generated 220-ms bandpass noise (800–1200 Hz) that was turned on 5 ms before the start of the first tone burst's offset ramp, and ended 5 ms after the second tone burst had reached full amplitude. The noise spectrum level was 52 dB SPL, giving a level at the output of an auditory filter

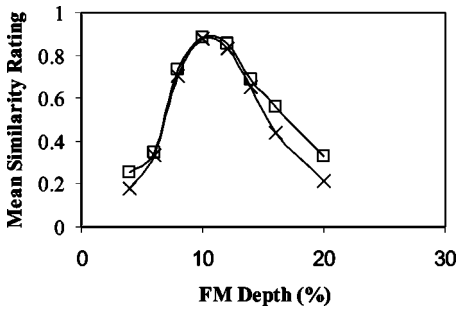


FIG. 4. Results of experiment 3, in which listeners rated the similarity of the illusory portion of an interrupted FM carrier to that of a comparison (uninterrupted) carrier in which the depth of the central portion of the FM was varied; the value of the FM depth of the central portion is shown on the abscissa. For all stimuli, the first and last portions of the stimulus consisted of a 1-kHz tone modulated at a rate of 5 Hz by $\pm 10\%$ of its carrier frequency. The interruption lasted 200 ms and was filled by either a noise burst (squares) or an intense 1-kHz carrier modulated by $\pm 10\%$ at a rate of 13.6 Hz (crosses).

(''ERB,'' Glasberg and Moore, 1990) centered on 1 kHz of 73 dB SPL. In another condition the gap was filled by a 73 dB SPL 1-kHz tone, frequency modulated at a rate of 13.6 Hz and a depth of 10%, and with a modulator starting phase of zero. This FM tone was turned on and off in the same way as the noise in the first condition.

The second stimulus was presented 500 ms later and consisted of a 1-s 5-Hz FM tone, whose modulation depth during the first and last 400 ms was also 10%. During the middle 200 ms, the FM depth was, on separate trials, 4%, 6%, 8%, 10%, 12%, 14%, 16%, or 20%. Discontinuities associated with changes in modulation depth were avoided by generating the stimulus according to Eq. (1) in three sections, corresponding to the first 400 ms, middle 200 ms, and final 400 ms of the stimulus. The value of θ at the start of the second and third sections was then set to the value of $g(t)$ at the end of the previous section.

Four normal-hearing listeners were instructed to ignore the interrupting sound (noise or 13.6-Hz FM tone) during the first stimulus and to focus on the middle portion of the stimulus. They were asked to adjust a horizontal slider to indicate how similar this sounded to the middle part of the second stimulus. The slider readings were digitized, divided by the maximum value used by each listener, and then averaged across listeners.

C. Results

Figure 4 shows the mean similarity rating as a function of the FM depth of the central portion of the comparison stimulus. The function reached a maximum at an FM depth of 10%, both when the inducer was a noise (squares) and when it was a 13.6-Hz FM tone (crosses). In both conditions, then, listeners heard the modulation continue to the same extent during the interruption as was physically present before and after it. The finding is particularly interesting for the case with the modulated inducer, as it was (of course) also possible for subjects to attend to the (louder) 13.6-Hz FM tone and ignore the tone with 5-Hz FM. This means that, depending on the instructions to the subject, s/he can, for the same stimulus, hear either an (illusory) slow or a (''real'')

fast FM applied to a 1-kHz carrier. To our knowledge there is no ''nonillusory'' stimulus that produces this percept; this issue is discussed further in Sec. VII D.

V. EXPERIMENT 4: DETECTION OF REAL AND ILLUSORY REVERSALS IN FM PHASE

A. Method

Experiment 4 investigated the ability of listeners to detect changes in FM phase under the conditions illustrated in Figs. 1(a) and (c)–(f). All of these conditions used a 2IFC task with feedback, a modulation rate of 5 Hz, a carrier frequency of 1 kHz, and a zero-peak modulation depth of 10%. The ''continuous'' condition [Fig. 1(a)] was simply a replication of that part of experiment 1 that used a 1-kHz carrier and a modulation rate of 5 Hz. As in that experiment, ϕ [Eq. (1)] was set to zero on half of all trials and to π on the other half. Subjects had to detect a phase shift of π radians in an otherwise uninterrupted FM. This condition provided a baseline level of performance for comparison with the other conditions. The nominal duration of the tone was 800 ms, as in experiment 1, and again this could be increased by either $0.5/fm$ or $1/fm$ s (100 or 200 ms).

In the remaining conditions, the FM tone was interrupted approximately half-way through for 200 ms. In the ''noise'' condition [Fig. 1(c)], this 200-ms interval was filled with a burst of noise, identical to that used in experiment 2. The durations of the portions of the tone before and after the noise were nominally 400 ms, but these could be increased (independently and at random) by 100 ms. Hence, as in the continuous condition, the total duration of the tone (excluding the 200-ms interruption) was nominally 800 ms, but could be increased by 0, 100 or 200 ms. The rationale behind this noise condition was that, if the phase (as well as the presence and depth) of FM is preserved during the noise, then subjects should be able to tell whether or not the second portion of the tone started in the modulation phase that it would have had if the modulation had continued throughout. This illusory FM is illustrated by the white line continuing through the noise in Fig. 1(c).

The ''gap'' condition [Fig. 1(d)] was similar to the noise condition except that the noise was absent and the reversal in modulator phase occurred after a silent gap. If subjects are not sensitive to FM phase *per se*, then performance in this condition should be at chance. Note that if we had not randomized duration, there would have been an extra cue that subjects could have used in the gap condition. Specifically, subjects could compare the modulator phase at the start of each tone burst within a given interval of the trial: these phases would have been the same for the standard stimuli [Fig. 1(d), left] but different for the signals [Fig. 1(d), right]. This cue was rendered useless by the duration randomization, as illustrated in Fig. 1(e). We will return to this issue in Sec. VI.

Finally, in the ''steady'' condition [Fig. 1(f)], the 200-ms gap was filled by an unmodulated tone. We would also expect performance to be at chance in this condition, unless there were ringing in a frequency modulation filterbank, analogous to that proposed for amplitude modulation (Dau *et al.*, 1996). If that were the case, and if the ringing persisted long enough, subjects might be able to detect a phase reversal between the end of the ringing response to the first burst and the start of the second burst (cf. Shailer and Moore, 1987; Sek and Moore, 2002).

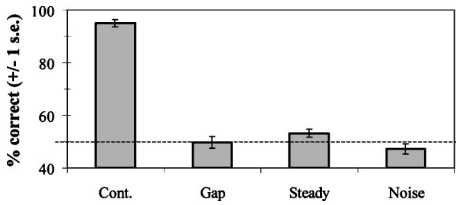


FIG. 5. Results from the four conditions of experiment 4, showing means and standard errors calculated across listeners. The horizontal dashed line indicates chance performance. Schematic spectrograms of the various conditions are shown in Fig. 1.

The method of signal generation and other signal parameters were the same as in experiment 1. Five normal-hearing listeners completed blocks of 25 trials per condition, before a fresh set of blocks was started, in a different order. This was continued until a total of eight blocks per condition had been run, so that each data point reported here was the mean of 200 trials per listener.

B. Results

The results of experiment 4, averaged across listeners, are shown in Fig. 5. Performance was close to perfect in the continuous condition, but was not significantly better than chance in the noise condition. The fact that subjects could not detect an ''illusory'' phase reversal in the noise condition suggests that FM phase is not preserved for 200 ms during the continuity illusion. Combined with the finding of experiment 3 that the FM depth is perceived to continue unabated throughout the noise, this suggests that subjects were experiencing a somewhat paradoxical percept during the signal intervals. That is, they heard the FM continue, but did not notice what would have been an obvious phase shift in that FM. Subjectively, one hears a continuous FM whose extent and rate do not change during the noise, but without the discontinuity that would occur if the modulation really had continued throughout (as in the continuous condition).

Performance was also at chance in the gap and steady conditions, suggesting that, if there is ringing in a modulation filterbank, the duration of the ringing is too short to allow listeners to detect a shift in phase of the modulation when the modulation is interrupted for 200 ms (corresponding to one whole modulator period).

VI. EXPERIMENT 5: IMPAIRMENTS IN PERFORMANCE PRODUCED BY ILLUSORY CONTINUITY

A. Rationale and method

Experiment 5 addressed the same issue as experiment 4 but using a paradigm in which the continuity illusion is predicted to *impair* performance unless information on FM phase is retained. To do this we exploited an alternative strategy, described in Sec. V A, for how listeners might perform the task in the gap condition [Fig. 1(d)]. That section pointed

out that, if overall duration were not randomized, subjects could discriminate between the standard and signal intervals by comparing the starting phase of the two tone bursts within each interval. (These starting phases would be the same for the standard, but not for the signal). In experiment 5, the duration of all stimuli was fixed, and listeners were alerted to this strategy. The reasoning was that, if the stimulus is heard to continue uninterrupted through the noise (experiment 3), then the strategy should not be available—because, perceptually, there are no longer two tone bursts to compare within each interval. Hence subjects should only be able to perform the task in the noise condition if the phase of the FM is preserved in the auditory system during the noise burst.

Because one prediction for experiment 5 is that the noise [Fig. 1(c)] should impair performance relative to that for the gap condition [Fig. 1(d)], due to it inducing the continuity illusion, we introduced two controls for alternative ways in which the noise might impair performance. To control for any distracting effect, the ''off-frequency'' condition [Fig. 1(h)] used a noise burst that was filtered between 2000 and 2400 Hz. This noise should still distract the subject but, because its center frequency was remote from that of the 1-kHz carrier, it should not induce the illusion. The rms level, duration, and timing of the off-frequency noise were identical to those of the on-frequency noise used in the noise condition. Another possibility is that the on-frequency noise would produce some masking of those portions of the tone that were temporally adjacent to the gap, and that this would prevent subjects from comparing the initial modulator phase of the two carrier bursts that comprise each interval. To control for this, the central 100-ms portion of the noise was deleted in the ''gap-noise'' condition [Fig. 1(g)], and 5-ms raised-cosine ramps were applied to the remaining portions of noise that were temporally adjacent to this deleted portion. The reasoning was that the remaining noise bursts would still mask the portions of the carrier adjacent to the gap,[2] but that the interruption in the noise would abolish the continuity illusion. Hence a preserved ability to perform the task in this condition, but not in the noise condition, would provide evidence that the poor performance in the latter was not due to masking of those portions of the tone adjacent to the gap. Such a finding would also argue against interpretations in terms of an auditory analog of the ''change blindness'' phenomenon, in which the transient obliteration of part of a visual scene prevents subjects from noticing an otherwise obvious change (e.g., Rensink *et al.*, 2000).

B. Results

Figure 6 shows that, as in experiment 4, subjects could easily detect a phase reversal in the continuous condition, but did not perform consistently above chance in the noise condition. However, the removal of the duration randomization meant that, unlike the case with experiment 4, all listeners except S2 could now perform the task reliably in the gap condition. Unfortunately, despite the lack of a duration rove and instructions to compare the FM phase at the start of each tone burst, it appears that S2 did not learn to use this strategy consistently. Importantly, the remaining subjects, who did perform well in the gap condition, could also perform substantially above chance in the off-frequency and gap-noise conditions. Hence, subjects could do the task in conditions where the tone was perceived as being interrupted, even when there was a noise present that potentially distracted the subjects and/or masked portions of the tone adjacent to the silent gap. This conclusion should be qualified slightly by the observation that performance in the gap-noise and off-frequency conditions was below that for the gap condition, suggesting that distraction and/or partial masking effects may have had some effect on performance. Nevertheless, the only condition in which a noise reduced performance to chance was the one in which an on-frequency noise completely filled the silent gap between the two tone bursts in each interval. Performance in that condition was significantly worse than in the off-frequency [$t(5)=5.80$, $p=0.002$] and gap-noise [$t(5)=3.74$, $p=0.014$] conditions. We attribute this to the noise inducing a perception of continuity, which prevented subjects from comparing the FM phases at the start of the two tone bursts in each trial.

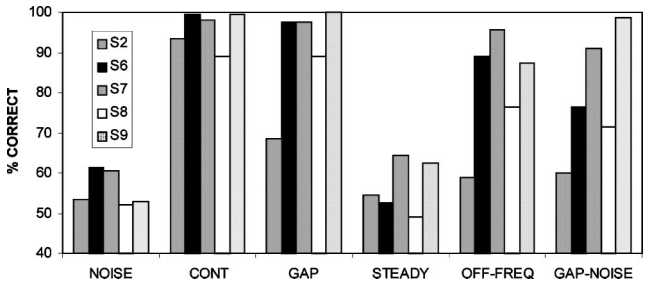


FIG. 6. Data from the six conditions of experiment 5, presented separately for each listener. Schematic spectrograms of the various conditions are shown in Fig. 1.

Further light on the way in which perceived continuity can affect performance comes from the steady condition, in which the carrier was continuous throughout, but where the central 200-ms portion of the carrier was unmodulated. Subjectively, one hears this stimulus veridicaly: first modulated, then steady, and then modulated again. One might therefore expect subjects to be able to perform the task in much the same way as they do in the gap condition—by comparing the phase of FM at the start of each unmodulated portion. However, as shown in Fig. 6, this was not the case, and performance was at chance. One explanation is that listeners may only be able to compare the FM phase at the start of two modulated sections if the time of onset of the modulation is ''labeled'' in some way. In the gap condition this labeling would be provided by the onset of the carrier. Because processing of FM is sluggish, in the steady condition, the FM may have to be present for some time (following the steady portion) before listeners become aware that the FM has resumed. [For comparable arguments about the detection of gaps in amplitude modulation, see Sek and Moore (2002)].

VII. DISCUSSION

A. Discriminating between changes in FM phase and depth

Section I introduced the idea that changes in FM phase might be detected via the output of a sliding temporal win-

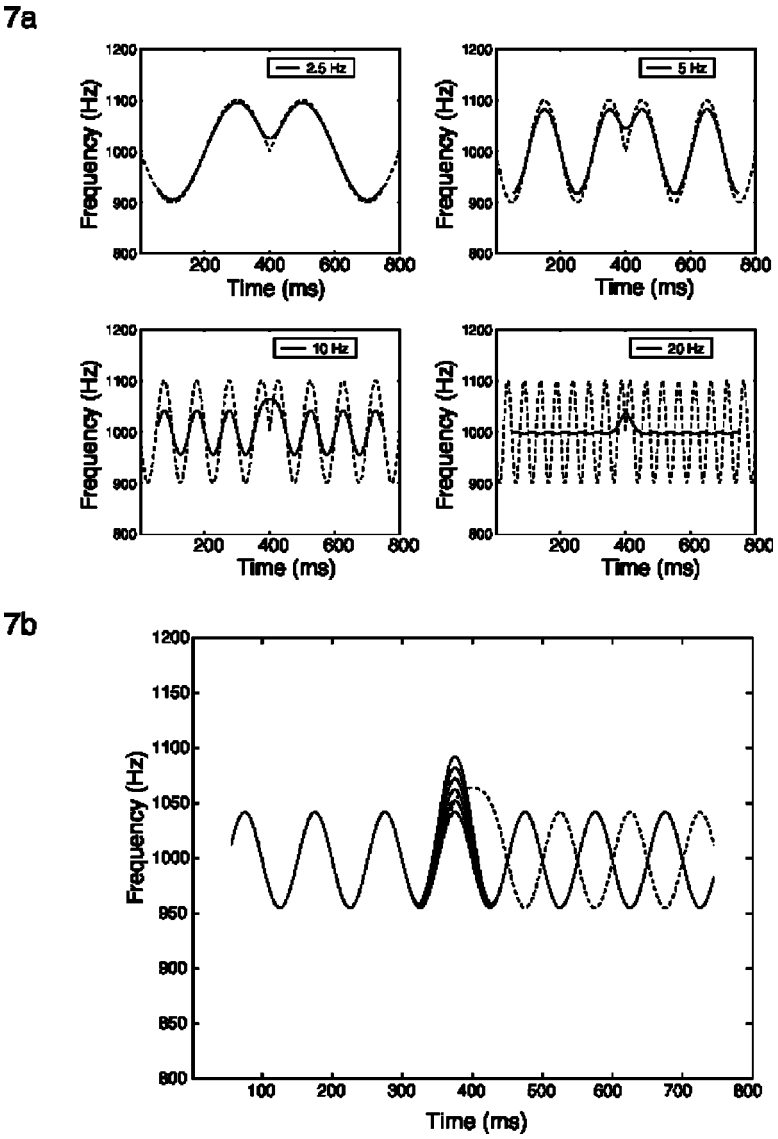


FIG. 7. (a) Dashed lines in each panel show the pattern of instantaneous frequency of a sinusoidal modulator of a given rate, with a phase reversal mid-way through. Superimposed solid lines show the output of a sliding temporal window. The window had a total duration of 110 ms, and so the output is undefined for the first and last 55 ms of the 800-ms stimulus. (b) The solid lines show the outputs in response to stimuli in which the FM depth during one half-cycle was increased by amounts ΔD_m of 0%, 2%, 4%, 6%, 8%, and 10% of f_c. Successively higher values of ΔD_m produce increasingly higher excursions in the middle of the trace. The dashed line shows the window output in response to 5-Hz FM containing a phase reversal mid-way through, with an FM depth of $\pm 10\%$ of a 1000-Hz carrier frequency.

dow, whose effect is to average instantaneous frequency over a finite duration. We argued that such a scheme was consistent with the results of experiment 1, which showed that listeners' sensitivity to an FM phase reversal deteriorated with increasing FM rate. The reasoning is illustrated by the solid lines in Fig. 7(a), which show instantaneous frequency convolved with a Hann window, for FM rates ranging from 2.5 to 20 Hz, superimposed on the patterns of instantaneous frequency (dashed lines). The window had a total duration of 110 ms, and a duration between the 6-dB-down points of 55 ms. For low FM rates, the window output reflects the individual cycles of the sinusoidal FM, and the phase reversal mid-way through the stimulus has a distinct effect. For higher FM rates, the phase reversal produces a small ''bump'' in the window output. As the FM rate increases further, the bump shrinks, consistent with the reduction in performance with increasing FM rate.

In experiment 2, listeners were required to discriminate between a tone in which the FM phase reversed mid-way through, and one in which the FM depth was increased by $\Delta D\%$ for half a modulator period. Performance varied nonmonotonically as a function of ΔD, reaching a minimum for most listeners at a value of around 4%. This result was interpreted as providing some support for the idea that changes in instantaneous frequency are smoothed by a temporal window, and that listeners detect such changes using the output of that window. However, performance never quite dropped to chance, leading us to accept this interpretation only with some caution.

Figure 7(b) shows the output of the 110-ms Hann-shaped temporal window in response to the FM-phase-reversal standard of experiment 2 (dashed line), and to the signals containing FM depth increments (ΔD) between 0% and 10% (solid lines). It can be seen that the signal with $\Delta D = 4\%$ produces a bump having approximately the same height as that produced by the standard. This is consistent with the minimum in the function relating discriminability to ΔD (Fig. 3). It can also be seen, from Fig. 7(b), that the window output in response to a signal with $\Delta D = 4\%$ is not *identical* to that produced by the phase-reversal standard. One difference can be ignored: the bumps produced by the different values of ΔD occur earlier than that produced by the standard, but this would not have consistently aided performance because the durations of both stimuli were roved.

A second difference is more likely to have affected performance: the bumps produced by the FM depth increases are narrower than that produced by the standard, and this could account for the fact that scores did not drop to chance, even at $\Delta D = 4\%$.

We should stress, of course, that the above simulations do not prove that the brain smoothes changes in instantaneous frequency using a Hann temporal window with a total duration of 110 ms. The shape of the window was chosen arbitrarily, and its duration was selected to give the best match to the data of experiment 2, on the assumption that the decision metric is the maximum extent of the deviation in the window output. Other window shapes and decision metrics are clearly possible. Also, additional cues may have come into play in some conditions. For example, wide FM excursions could produce onset responses in auditory nerve fibers tuned to frequencies more than 10% away from the carrier, and listeners might detect these onsets without necessarily tracking any change in pitch. This would have had the effect of elevating the functions shown in Fig. 3 for larger values of ΔD.

The simulation does, however, demonstrate that averaging using a fairly long window can qualitatively account for the results of the first two experiments reported here. It is also worth noting that a duration of 110 ms is roughly consistent with that suggested by studies of temporal overintegration in pitch perception (Carlyon, 1996; Micheyl and Carlyon, 1998; Gockel *et al.*, 1999).

B. "Explicit" encoding of FM phase

In the Introduction, we reviewed several sources of evidence that the presence of FM is ''explicitly encoded'' by the auditory system. We interpreted as evidence for such encoding any finding showing that dynamic changes in a sound's frequency were processed in a manner that was qualitatively different from the processing found for steady sounds. Previous psychophysical findings in this category included experiments on glide detection, on FMDI and on the perceptual asymmetries reported by Cusack and Carlyon (2003). The experiments described here allow a different, perhaps more precise, definition of what is meant by ''explicit encoding,'' which relates to those features of a sound that are encoded, when other aspects appear to be discarded by the system. The fact that listeners hear FM continue throughout a noise, but are insensitive to the phase at which the FM resumes after the noise ends, is consistent with the system encoding the presence of FM in a way that does not preserve phase information. In contrast, when listeners have to detect reversals in FM phase, they may need to use a separate mechanism, which does preserve FM phase. This mechanism is probably sluggish, causing performance to drop dramatically for modulation rates above about 10 Hz, and causing listeners to confuse FM phase reversals with momentary increases in FM depth.

We should stress that this does not mean that changes in FM phase can only affect perception when they are sufficiently slow to be preserved in the output of a fairly long temporal window. Rather, such changes may be perceptible, if they influence the output of some other mechanism. For example, although listeners are insensitive to differences in FM phase between two inharmonically related (simultaneous) pure tones, they can easily detect such differences when the tones are harmonically related, because the FM phase difference causes them to become mistuned (Carlyon, 1991, 1994, 2000).

The ''recoding'' of sinusoidal FM may also have implications for the processing of other forms of dynamic frequency change. Dannenbring (1976) presented subjects with an alternating series of ascending and descending linear frequency glides, so that the instantaneous frequency of the signal underwent an alternating series of peaks and troughs. He replaced the peaks of this FM pattern with bursts of noise, so as to induce an illusion of continuity, and asked subjects to adjust a comparison sound to the maximum frequency that they heard during this illusion. The results showed that the maximum frequency subjects heard corresponded to the maximum instantaneous frequency that was physically present. As with our results, this shows that subjects were not perceiving the pattern of instantaneous frequency change that they ''would have'' heard if the signal had really continued throughout the noise. Our results, however, differ from his conclusion that subjects' percepts correspond to an interpolation between the frequencies of the glide just before and just after the noise. If this had been the case for the sinusoidal FM reported here, subjects would have heard the modulation stop during the noise bursts.

C. A modulation filterbank?

An alternative explanation of the present results comes from the idea that FM may be processed via a modulation filterbank, analogous to that proposed for AM (Dau *et al.*, 1996, 1997). The FM phase reversals used in experiment 1 would produce ''splatter'' in the modulation domain, and, if the bandwidth of modulation filters increases with their center frequency (as is proposed for AM filters), this splatter would be less detectable at faster rates. This in turn could account for the reduction in performance with increasing FM rate. Furthermore, ''modulation splatter'' might not provide a useful cue when the FM phase reversal occurred on either side of a 200-ms noise burst, or on either side of an unmodulated carrier, since the offset and onset of the modulation would result in splatter in the modulation domain in both intervals of a trial. The probable ineffectiveness of the ''modulation splatter'' cue in such a case is consistent with the results of experiments 3 and 4. One argument against an explanation in terms of a modulation filter bank is that, if the FM filters are similar to those proposed for AM, they will be very broadly tuned, with a Q factor of about 1 (Sek and Moore, 2003). This would make splatter in the modulation domain very hard to detect. However, in the absence of definitive information on the shapes of the putative FM filters and on the size of the change in filter output that can be detected, it may be premature to reject this idea outright. If it is correct, then an ''excitation pattern'' in the modulation domain, calculated from the rms outputs of different modulation filters, would provide one way in which the auditory system could explicitly encode the presence, depth, and rate of FM, without explicitly encoding its phase.

D. Paradoxical percepts and informal observations

Two aspects of the results presented here suggest that, during the continuity illusion, listeners perceived a pattern of FM that could not be produced by any ''real'' modulator. One of these comes from the combination of the results of experiment 1, which showed that listeners can easily detect a reversal in the phase of a 5-Hz FM, with those of experiments 3–5. These latter experiments revealed that FM is heard to continue through an interrupting noise, but that listeners are insensitive to the relative phase at which the FM is resumed after the noise ends. In other words, listeners hear the FM continue, but do not notice a phase reversal which, if the FM really had continued, would have been obvious. One way out of this paradox would be to assume that the internal representation of the FM rate changed during the noise, in such a way as to avoid an abrupt change in phase when the real FM tone resumed. For example, a rate change from 5 Hz to either 2.5 or 7.5 Hz would produce a modulator whose phase shifted by π radians over 200 ms, allowing the FM to continue without any sudden phase reversal. Of course, no such rate change should occur in the standard intervals [e.g., left of Fig. 1(c)], and so one would have to assume that a change in the representation of FM rate occurred at some early level of processing, but that listeners could not detect this change. This might be possible, because, as noted in the Introduction, there is some evidence that the perception of FM is ''sluggish'' (Sek and Moore, 1995). We therefore generated versions of our signals in which the 200-ms central portions were replaced, not by a noise, but by FM at either 2.5 or 7.5 Hz. In both cases the rate change was clearly audible, providing further support for the idea that the illusory FM perceived by our subjects did not correspond to any physically present FM.

A second paradoxical percept is suggested by the results of experiment 3, where the central portion of a 1-kHz tone, modulated at 5 Hz, was replaced by a more intense tone with the same carrier frequency but a different modulation rate (13.6 Hz). Listeners perceived the 5-Hz FM continue unabated through the 13.6-Hz FM tone. Although subjects were told to ignore the more intense tone, one can choose to attend to it, and, when one does, the faster FM rate can clearly be heard. This means that the same stimulus can give rise to the percept of two different FM rates on the same carrier frequency. One may not be able to *perceive* the two rates simultaneously, but rather may have to direct attention to one or the other. Nevertheless, we could not think of a nonillusory stimulus where one could choose to attend to one or another sinusoidal FM rate, applied to the same carrier. One candidate would be a 1-kHz carrier modulated by a complex (13.6+5 Hz) tone, but, when we generated this stimulus, the resulting percept had a complex beating quality, and one could not hear the component modulation rates individually.

The existence of paradoxical FM percepts is not something that can be easily accommodated by a simple view in which processing is limited to taking a series of ''snapshots'' of sounds. Rather, it is more consistent with the view that sounds are encoded in terms of auditory features, which can, under specific conditions, be analyzed and recombined in seemingly contradictory ways. In this sense, the paradoxical percepts reported here are analogous to the paradoxical figures, such as the ascending and descending staircases (e.g., Escher, 1960), that are well known in the visual modality. The results presented here add to the growing body of evidence that FM is such a feature (Rees and Moller, 1983; Moore and Shailer, 1992; Sek and Moore, 1999; Cusack and Carlyon, 2003).

VIII. SUMMARY

(i) Experiment 1 showed that detection of a π-radians shift in FM phase deteriorated as the modulation rate increased from 2.5 to 20 Hz, for carrier frequencies of 1 and 6 kHz.

(ii) Experiment 2 required listeners to discriminate between a 1-kHz carrier in which the 10-Hz frequency modulator shifted in phase by π radians mid-way through, and one in which the depth of FM was increased (from $\pm 10\%$) by an amount ΔD for half a modulator period. The function relating performance to ΔD was nonmonotonic, and reached a minimum at $\Delta D = 4\%$.

(iii) Experiment 3 showed that listeners heard FM of a 1-kHz carrier continue at its full depth through an interrupting sound, which was either a burst of noise or a 1-kHz carrier with FM at a 13.6-Hz rate. In the latter case, they heard the 5-Hz FM continue behind the 13.6-Hz FM.

(iv) Experiments 4 and 5 showed that, although listeners heard FM continue through a noise burst, they were insensitive to the phase at which the FM resumed after the noise had ended. Combined with the results of experiment 1, this shows that they are insensitive to a phase shift, which, if applied to a stimulus with uninterrupted FM, would be easily detected.

(v) The results of the first two experiments can be qualitatively explained by a model in which instantaneous frequency is averaged using a sliding temporal window with duration about 110 ms. For low FM rates, the output of this mechanism preserves information on the rate and phase, as well as the presence, of FM

(vi) The results of the last three experiments, together with those of previous studies, indicate that the presence (and perhaps the rate and depth) of FM are represented in the auditory system via a separate code. This code may be useful for faster FM rates, and for more central processing where different features of sounds need to be combined. It is responsible for the perception of FM during the continuity illusion, but does not preserve information on FM phase.

ACKNOWLEDGMENTS

Author CM was supported by Grant No. GR/N64861/01 from the Engineering and Physical Sciences Research Council (U.K.) to authors RPC and BCJM.

[1]The equivalent rectangular bandwidth (ERB) of an auditory filter centered on 1 kHz, according to Glasberg and Moore's (1990) formula, is 133 Hz (13.3% of the center frequency). The ERB of an auditory filter centered on

6 kHz is, according to the same formula, 672 Hz, which is 11.2% of the center frequency.

[2]Although reducing masker duration can lower thresholds in forward masking, the noise used here temporally overlapped with the onsets and offsets of the tones. In such circumstances, thresholds actually *increase* when masker duration is reduced (Carlyon, 1988), presumably for reasons related to the overshoot effect (Zwicker, 1965), in which thresholds for a brief tone are lower at the end of a short than at the end of a long masker. Hence any improvement in performance caused by deleting the central portion of the noise burst (thereby converting it into two shorter bursts) is unlikely to be due to reduced masking of tone onsets and offsets.

Carlyon, R. P. (**1988**). ''The development and decline of forward masking,'' Hear. Res. **32**, 65–80.

Carlyon, R. P. (**1991**). ''Discriminating between coherent and incoherent frequency modulation of complex tones,'' J. Acoust. Soc. Am. **89**, 329–340.

Carlyon, R. P. (**1994**). ''Further evidence against an across-frequency mechanism specific to the detection of FM incoherence between resolved frequency components,'' J. Acoust. Soc. Am. **95**, 949–961.

Carlyon, R. P. (**1996**). ''Masker asynchrony impairs the fundamental-frequency discrimination of unresolved harmonics,'' J. Acoust. Soc. Am. **99**, 525–533.

Carlyon, R. P. (**2000**). ''Detecting coherent and incoherent frequency modulation,'' Hear. Res. **140**, 173–188.

Cusack, R., and Carlyon, R. P. (**2003**). ''Perceptual asymmetries in audition,'' J. Exp. Psychol. Hum. Percept. Perform. **29**, 713–725.

Dannenbring, G. L. (**1976**). ''Perceived auditory continuity of alternately rising and falling FM sweeps,'' Can. J. Psychol. **30**, 99–114.

Dau, T., Kollmeier, B., and Kohlrausch, A. (**1997**). ''Modeling auditory processing of amplitude modulation: I. Detection and masking with narrowband carriers,'' J. Acoust. Soc. Am. **102**, 2892–2905.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). ''A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure,'' J. Acoust. Soc. Am. **99**, 3615–3622.

Demany, L., and Semal, C. (**1989**). ''Detection thresholds for sinusoidal frequency modulation,'' J. Acoust. Soc. Am. **85**, 1295–1301.

Dooley, G. J., and Moore, B. C. J. (**1988a**). ''Detection of linear frequency glides as a function of frequency and duration,'' J. Acoust. Soc. Am. **84**, 2045–2057.

Dooley, G. J., and Moore, B. C. J. (**1988b**). ''Duration discrimination of steady and gliding tones: a new method for estimating sensitivity to rate of change,'' J. Acoust. Soc. Am. **84**, 1332–1337.

Escher, M. C. (**1960**). *Ascending and Descending*. (Escher, B.V., The Netherlands).

Glasberg, B. R., and Moore, B. C. J. (**1990**). ''Derivation of auditory filter shapes from notched-noise data,'' Hear. Res. **47**, 103–138.

Gockel, H., and Carlyon, R. P. (**2000**). ''Frequency modulation detection interference produced by asynchronous and nonsimultaneous interferers,'' J. Acoust. Soc. Am. **108**, 2329–2336.

Gockel, H., Carlyon, R. P., and Micheyl, C. (**1999**). ''Context dependence of fundamental frequency discrimination: Lateralized temporal fringes,'' J. Acoust. Soc. Am. **106**, 3553–3563.

Gockel, H., Moore, B. C. J., and Carlyon, R. P. (**2001**). ''Influence of rate of change of frequency on the overall pitch of frequency modulated tones,'' J. Acoust. Soc. Am. **109**, 701–712.

Hartmann, W. M., and Klein, M. A. (**1980**). ''Theory of frequency modulation detection for low modulation frequencies,'' J. Acoust. Soc. Am. **67**, 935–946.

Lyzenga, J., Carlyon, R. P., and Moore, B. C. J., (**2004**). ''The effects of real and illusory glides on pure-tone frequency discrimination, using tones with and without gliding mid sections,'' J. Acoust. Soc. Am. **116**, 491–501.

Micheyl, C., and Carlyon, R. P. (**1998**). ''Effects of temporal fringes on fundamental-frequency discrimination,'' J. Acoust. Soc. Am. **104**, 3006–3018.

Micheyl, C., Moore, B. C. J., and Carlyon, R. P. (**1998**). ''The role of excitation-pattern cues and temporal cues in the frequency and modulation-rate discrimination of amplitude-modulated tones,'' J. Acoust. Soc. Am. **104**, 1039–1050.

Moore, B. C. J. (**1973**). ''Frequency difference limens for short-duration tones,'' J. Acoust. Soc. Am. **54**, 610–619.

Moore, B. C. J., and Shailer, M. J. (**1992**). ''Modulation discrimination interference and auditory grouping,'' Philos. Trans. R. Soc. London, Ser. B **336**, 339–346.

Moore, B. C. J., and Sek, A. (**1996**). ''Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking.'' J. Acoust. Soc. Am. **100**, 2320–2331.

Moore, B. C. J., Glasberg, B. R., Gaunt, T., and Child, T. (**1991**). ''Across-channel masking of changes in modulation depth for amplitude-and frequency-modulated signals.,'' Q. J. Exp. Psychol. **43A**, 327–348.

Plack, C. J., and Carlyon, R. P. (**1995**). ''Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics,'' J. Acoust. Soc. Am. **98**, 1355–1364.

Plomp, R. (**1982**). ''Continuity effects in the perception of sounds,'' Psychoacoustics of music; Jablonna, Poland.

Rees, A., and Moller, A. R. (**1983**). ''Responses of neurons in the inferior colliculus of the rat to AM and FM tones,'' Hear. Res. **10**, 301–310.

Rensink, R. A., O'Regan, J. K., and Clark, J. J. (**2000**). ''On the failure to detect changes in scenes across brief interruptions,'' Visual Cognit. **7**, 127–145.

Sek, A., and Moore, B. C. J. (**1995**). ''Frequency discrimination as a function of frequency, measured in several ways,'' J. Acoust. Soc. Am. **97**, 2479–2486.

Sek, A., and Moore, B. C. J. (**1999**). ''Discrimination of frequency steps linked by glides of various durations,'' J. Acoust. Soc. Am. **106**, 351–359.

Sek, A., and Moore, B. C. J. (**2002**). ''Mechanisms of modulation gap detection,'' J. Acoust. Soc. Am. **111**, 2783–2792.

Sek, A., and Moore, B. C. J. (**2003**). ''Testing the concept of a modulation filter bank: The audibility of component modulation and detection of phase change in three-component modulators,'' J. Acoust. Soc. Am. **113**, 2801–2811.

Shailer, M. J., and Moore, B. C. J. (**1987**). ''Gap detection and the auditory filter: phase effects using sinusoidal stimuli,'' J. Acoust. Soc. Am. **81**, 1110–1117.

Warren, R. M., Obuseck, C. J., and Ackroff, J. M. (**1972**). ''Auditory induction: perceptual synthesis of absent sounds,'' Percept. Psychophys. **20**, 380–386.

Wilson, A. S., Hall, J. W., and Grose, J. H. (**1990**). ''Detection of frequency modulation (FM) in the presence of a second FM tone,'' J. Acoust. Soc. Am. **88**, 1333–1338.

Zwicker, E. (**1965**). ''Temporal effects in simultaneous masking by white-noise bursts,'' J. Acoust. Soc. Am. **37**, 653–663.

Direct measurement of onset and offset phonation threshold pressure in normal subjects

Randall L. Plant[a] and Gary L. Freed
Department of Otolaryngology/Head and Neck Surgery, Eastern Virginia Medical School, Norfolk, Virginia 23507

Richard E. Plant
Department of Biological and Agricultural Engineering, University of California, Davis, California 95616

(Received 12 January 2003; revised 28 August 2004; accepted 1 September 2004)

Phonation threshold pressures were directly measured in five normal subjects in a variety of voicing conditions. The effects of fundamental frequency, intensity, closure speed of the vocal folds, and laryngeal airway resistance on phonation threshold pressures were determined. Subglottic air pressures were measured using percutaneous puncture of the cricothyroid membrane. Both onset and offset of phonation were studied to see if a hysteresis effect produced lower offset pressures than onset pressures. Univariate analysis showed that phonation threshold pressure was influenced most strongly by fundamental frequency and intensity. Multiple linear regression showed that these two variables, as well as laryngeal airway resistance, most strongly predicted phonation threshold pressure. Two of the five subjects demonstrated a significant hysteresis effect, but one subject actually had higher offset pressures than onset pressures. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1812309]

PACS numbers: 43.70.Aj [AL] Pages: 3640–3646

I. INTRODUCTION

Theoretical studies and limited measurements on laryngeal models, excised larynges, and *in vivo* subjects have indicated that the threshold pressure required to sustain vocal-fold vibration may be less than what is required to initiate vibration. As a result, there may be a hysteresis effect such that the offset pressure for vibration is less than the onset pressure. In this study we have directly measured the onset and offset threshold pressures during vocalization in five normal subjects. The goal of the study was to determine which laryngeal factors influence threshold pressure for vocalization and to determine if a hysteresis effect for threshold pressures was present.

Titze[1] derived the threshold pressure necessary to sustain oscillation using a body-cover model of vocal-fold structure. His theoretical results showed that threshold pressure was inversely related to the vocal-fold thickness and directly related to the tissue damping coefficient, the prephonatory glottal width, and the mucosal wave velocity in the vocal-fold cover. In order to determine the relationship between threshold pressure and fundamental frequency, he reasoned that an increase in fundamental frequency should decrease vocal-fold thickness and increase the mucosal wave velocity. As a result, threshold pressure should increase as the square of fundamental frequency.

Titze *et al.*[2] and Chan *et al.*[3] used a physical model to study phonation threshold pressure. Their model of the vocal folds consisted of a thin silicone membrane, encapsulating fluids of differing viscosity (representing the vocal-fold mucosa), placed within a Plexiglas airflow chamber. Various combinations of membrane thickness, fluid viscosity, and glottal configuration were tested. Phonation threshold pressure was found to be lower with thin membranes and low fluid viscosity. Offset threshold pressures were consistently lower than onset pressures.

Lucero[4] adopted Titze's body-cover model and modified it to include large body oscillations. The lung pressure necessary to sustain vibration at a given amplitude was calculated for several different glottal configurations (e.g., divergent, convergent, rectangular). He found that in all three configurations the pressure necessary to maintain oscillation dropped once vibration began. For a rectangular glottic configuration, the ratio between offset and onset pressure was calculated to be 0.5.

Lucero[5] next examined this hysteresis phenomenon more closely by calculating the amount of energy absorbed and dissipated during each vocal-fold vibration. His analytical results indicated that the ratio of offset/onset pressure was a function of the phase delay (the delay between the movement of the upper and lower portion of the vocal folds) and the airflow separation coefficient (defined as the ratio between the glottal entry area and the area at the point at which the airflow detaches from the glottal wall). Moreover, he found that as the amplitude of vocal-fold oscillation increased, the point of airflow separation shifted in the glottis. His calculations showed that this would decrease the amount of energy absorbed by the vocal folds and therefore enable vocal vibration to continue at pressures below those at phonation onset.

Jiang and Titze measured phonation threshold pressures in excised canine larynges, both in a whole preparation and as a hemilarynx positioned against a glass plate.[6] In both

[a]Current address: Dept. of Otolaryngology/Head and Neck Surgery, Alaska Native Medical Center, 4315 Diplomacy Drive, Anchorage, AK 99508. Electronic mail: rlplant@anmc.org

cases, threshold pressures were measured at approximately 0.9 kPa.

Many studies investigating threshold pressures in human subjects have used a noninvasive indirect technique for measurement of subglottic pressure. This technique measures the peak intraoral pressure as the subject speaks a voiceless plosive consonant followed by a voiced sound (e.g., ''pah''). Although the method has significant limitations in estimating subglottic pressure during prolonged phonation,[7] it has been shown to be fairly accurate in measuring onset pressure in certain circumstances.[8] One major limitation is that it can only be used for vibrations that follow a plosive consonant. Direct techniques allow measurement of onset or offset pressure for any type of vocalization, and in particular can be used for voiced onsets.

Verdolini-Marston *et al.* used this technique to examine the effect of hydration on phonation threshold pressure.[9] The threshold pressure for three different pitches (low, medium, and high) was determined in ''dry'' and ''wet'' hydration conditions. Group average thresholds ranged from 3–5-cm H_2O in low pitches and 4–7-cm H_2O in the high pitches.

Sundberg *et al.* used a similar technique to measure threshold pressure in operatic baritone singers.[10] The actual values for threshold pressure are not listed in their paper, as they were instead interested in the ratio $(P_s - P_{thr})/P_{thr}$, where P_{thr}=threshold pressure and P_s=subglottic pressure.

Jiang *et al.*[11] recognized the shortcomings of techniques based on labial opening and developed a new method that utilized measurement of supraglottal pressure following interruption of airflow during vocalization. Their results showed a slight increase in phonation threshold as sound intensity increased. Threshold pressures ranged from 1- to 7-cm H_2O.

Plant and Hillel[7] directly measured subglottic pressure and laryngeal airflow during speech in normal subjects and subjects with adductor spasmodic dysphonia. Subglottic pressure was measured by percutaneous puncture of the cricothyroid membrane and was therefore not subject to the limitations present with indirect techniques. Their results showed that subglottic air pressure in normal subjects could vary by as much as 30%–50% during vocalization of a single syllable.

Plant and Younger[12] studied the interaction of subglottic air pressure, vocal intensity, and fundamental frequency in seven normal individuals. This interaction was found to be highly variable, particularly with regard to the relationship between subglottic pressure and vocal intensity. In particular, some subjects could increase output intensity at certain frequencies by as much as 50% with minimal changes in subglottic pressure. There was a trend for lower threshold pressure at offset, but other factors such as fundamental frequency were not taken into account in these comparisons.

These previous results imply that threshold pressure may be influenced by, among other things, fundamental frequency and vocal intensity. Rather than determining a single threshold pressure, it may therefore be more appropriate to consider the relationship between onset or offset pressure and other voicing parameters.

One such parameter is the closure force of the glottis. For example, if an individual strongly adducts his folds prior to initiation of airflow, one might expect the phonation threshold pressure to be higher than with a breathy voice onset. During phonation, the force with which the vocal folds are approximated is affected primarily by contraction of the thyroarytenoid and the interarytenoid muscles[13] and tissue properties of the vocal folds. Subglottic pressures are known to be much higher in conditions associated with increased muscle contraction in the larynx, such as adductor spasmodic dysphonia.[7]

Measurements of glottic adduction forces have been made using direct laryngoscopy in anesthetized canines,[14] but there is no easy technique to determine these forces during normal speech in human subjects. However, for constant vocal-fold mass, the acceleration (and therefore the closure speed) of the folds should increase with an increase in the force applied to the fold. Peak closure speed of the vocal fold can be calculated from the first derivative of the EGG signal. Orlikoff has in fact shown that the slope of the increasing EGG contact signal is related to vocal intensity and declination in airflow.[15]

Laryngeal airway resistance is known to be larger in individuals with increased contraction of the vocal-fold adductors.[7] The laryngeal airway resistance may therefore represent the degree of glottalization during vocalization. Holmberg *et al.* found that laryngeal airway resistance decreased as subjects changed from a normal voice to a soft voice.[16]

In the present study, onset and offset threshold pressures were measured in five normal subjects under a variety of voicing conditions. Simultaneous measurements were made of: (1) fundamental frequency; (2) vocal intensity; (3) laryngeal airway resistance; and (4) maximum of the first derivative of the electroglottography signal (dEGG/dt). The goals of the study were (1) to determine if a relationship existed between phonation threshold pressure and these four variables; and (2) to determine if onset threshold pressure for phonation differed from offset pressure.

II. METHODS AND MATERIALS

The study group consisted of five healthy individuals with no previous history of voice disorder. None of the subjects was a serious singer or trained in voice. The design of the study was approved by the Institutional Review Board at Eastern Virginia Medical School. The subjects ranged in age from 26 to 47 (mean 31); there were four men and one woman (subject No. 5).

Airflow, electroglottography, subglottic pressure, and the sound signal were simultaneously measured during vocalization of the sound /i/. The specific tasks given to the subjects were to (a) say three tokens with a deliberately breathy onset; (b) say three tokens at low pitch, the first token at low intensity and the next two at increasing intensity; (c) say three tokens at midpitch, increasing in intensity as in the previous case; (d) say three tokens at high pitch, increasing in intensity as in the previous case; and (e) say three tokens with a glide from low to high pitch. When possible, the process was repeated.

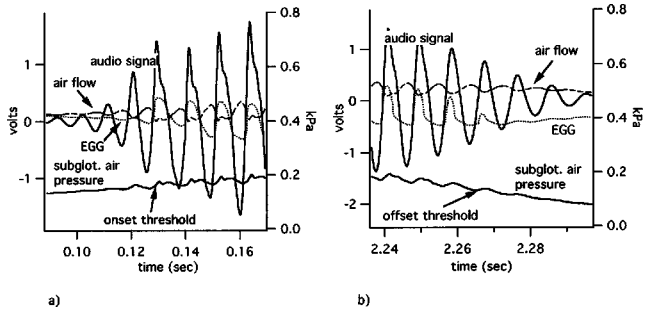


FIG. 1. Signal waveforms and the defined subglottic threshold air pressure at (a) onset and (b) offset of vocal-fold vibration. Axis for subglottic air pressure is on the right.

A Rothenberg-type mask was placed over the subject's face throughout the vocalization. EGG electrodes were positioned on either side of the neck overlying the thyroid cartilage. Subglottic pressure was measured by inserting a 21-gauge needle through the cricothyroid membrane. The needle was attached to a pressure transducer using a short (~10-cm) section of arterial pressure tubing. A condenser microphone was placed within the mask to measure the sound signal.

Airflow measurements were calibrated with a standardized flow meter. Pressure measurements were calibrated using a manometer. The four channels of data were sampled at 6000 Hz and directly stored on a laptop computer for analysis.

The onset of vocal-fold vibration was defined as the first appearance of a periodically varying EGG signal. The actual point of phonation onset or offset was chosen as the peak of the first derivative of the first or last EGG waveform (Fig. 1). This did not necessarily correspond to the initiation of the sound or the airflow fluctuations. In many cases there were one to three cycles in the audio and airflow signals prior to the first periodic EGG signal. These initial vibrations presumably occurred prior to vocal-fold contact, although airflow did drop to zero between the peaks in these early vibrations.

Fundamental frequency at the point of vocal-fold onset and offset was determined by averaging the frequency of the first three and the last three wave cycles. Intensity was determined by calculating the root-mean-square of the first three and last three wave cycles. Laryngeal airway resistance was defined as the quotient of the root-mean-square of the pressure and the airflow over the first three and last three wave cycles. The EGG signal was differentiated and the maxima of the first three and last three peaks were averaged.

During the end of vocalization there often were gradual decreases in the amplitude of the EGG, rather than the abrupt changes seen at onset. As a result, the precise point of vibration offset was occasionally difficult to determine. However, since subglottic pressure was often only slowly changing with time at offset, this had little effect on the determination of the offset threshold pressure.

III. RESULTS

A. Relationships with threshold pressure

The number of tokens available for analysis from each subject ranged from 15 to 36. The variation was due primarily to the technical ability to collect data. (In some cases the needle would become plugged, and a maximum of two punctures was done in most cases.) Both univariate and multiple linear regression analysis of the relationship between threshold pressure and the study variables were performed. The univariate relationships between onset and offset pressure and each of the study variables for each of the subjects are shown in Figs. 2 through 5. A simple linear regression was performed for each of these relationships and is also shown on the graphs. Each fitted regression line was tested to determine if its slope was statistically different from zero ($p < 0.05$). Those fitted lines in which the slope was statistically different from zero are indicated by capitalization and underlining of their identifying annotation (e.g., ONSET or OFFSET).

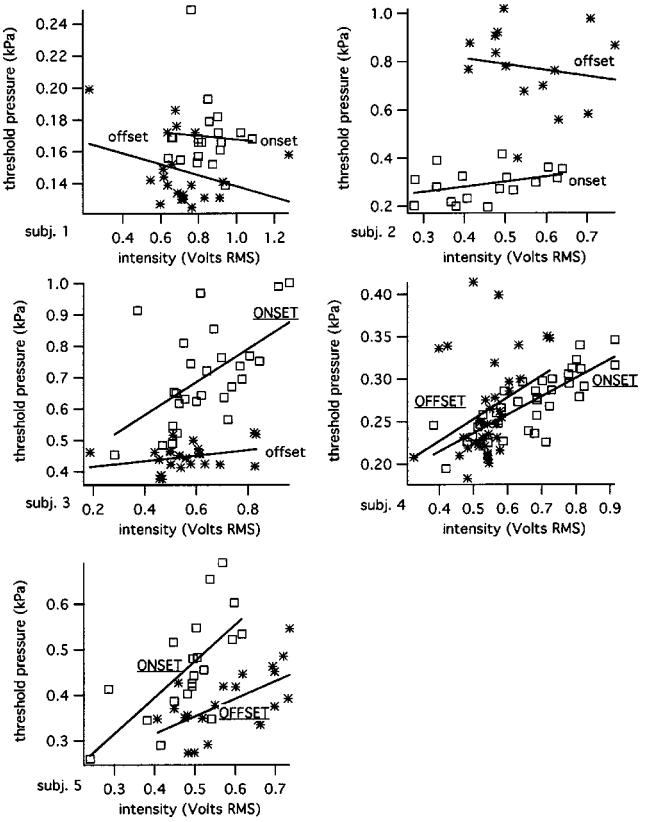


FIG. 2. Scatter plots and fitted regression lines showing the relationship between phonation threshold pressure and sound intensity. Lines whose slope was significantly different from zero are capitalized and underlined (e.g., ONSET and OFFSET).

The largest number of significant nonzero relationships occurred in relationship to intensity (Fig. 2) and fundamental frequency (Fig. 3). Every subject had either an onset pressure or an offset pressure that showed a significant relationship to fundamental frequency, with a total of six significant relationships in all. There were five significant relationships observed with intensity: Two subjects had significant relationships with both onset and offset (subjects 4 and 5), one subject had a significant relationship with onset (subject 3), and the other two subjects showed no relationship between intensity and threshold pressure (subjects 1 and 2).

The slopes for most of the fitted regression curves were non-negative. There were two instances in which the slopes

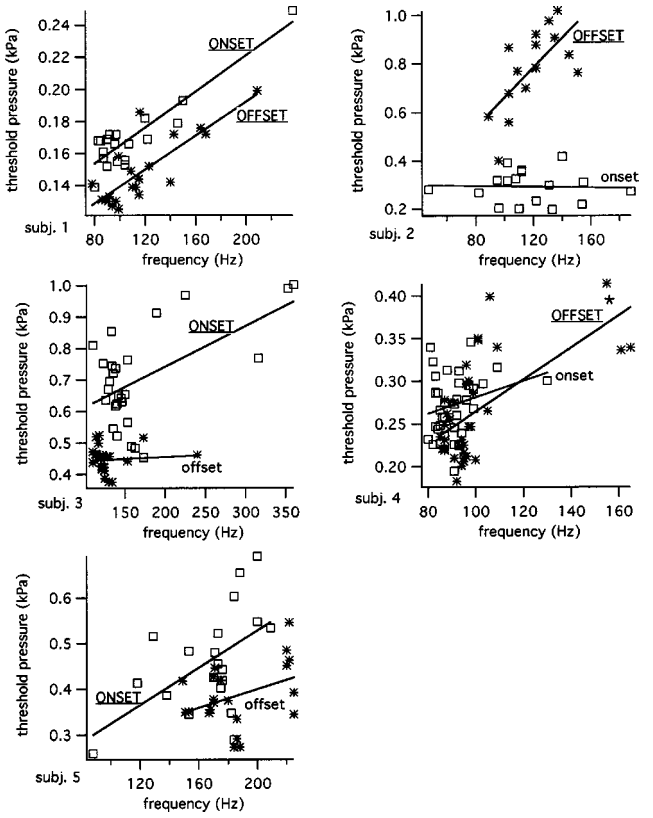


FIG. 3. Scatter plots and fitted regression lines showing the relationship between phonation threshold pressure and fundamental frequency. Lines whose slope was significantly different from zero are capitalized and underlined (e.g., ONSET and OFFSET).

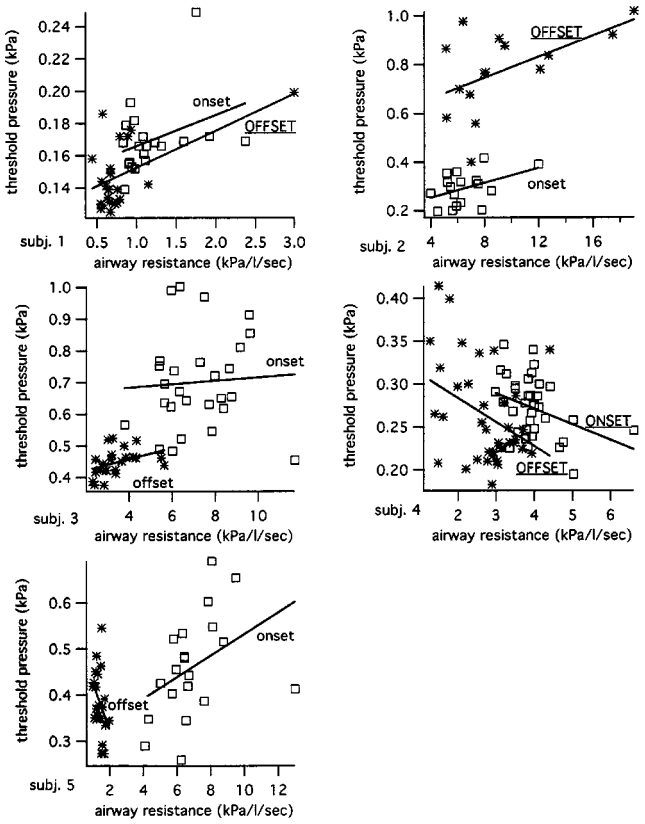


FIG. 4. Scatter plots and fitted regression lines showing the relationship between phonation threshold pressure and laryngeal airway resistance. Lines whose slope was significantly different from zero are capitalized and underlined (e.g., ONSET and OFFSET).

were significantly negative; that behavior was seen in subject 4 with regard to laryngeal airway resistance and threshold pressure (Fig. 4, fourth graph).

Multiple regression analysis of the relationships was also performed for each subject. The Mallow's $C(p)$ statistic was calculated to determine which combination of variables produced a model that best predicted threshold pressure for onset and offset for each subject. The results of this analysis, as well as the value for R^2 for the model, are shown in Table I.

Results of multiple regression analysis showed a more prominent relationship among the study variables than seen in the simple linear regression analysis. Intensity, airway resistance, and frequency were found to significantly contribute to the prediction of threshold pressure in nine, eight, and seven of the combinations, respectively. The first derivative of the EGG signal made a significant contribution in only three of the cases.

B. Comparison between onset and offset threshold pressures

In order to compare onset and offset pressure, 90%-confidence curves were drawn around the fitted simple linear regression lines. Results for each of the subjects and each of the study variables are shown in Figs. 6 through 9.

Two of the subjects (subject 3 and subject 5) had significantly lower offset pressures over most of the range of the study variables. For subject 5, there was an overlap of the confidence curves only in the frequency range between 170 and 200 Hz (Fig. 7, fifth graph). Subjects 1 and 4 had an overlap of the onset and offset confidence curves for most of the study variables, indicating no significant difference between these two conditions. Finally, one subject (subject 2) had an unusual behavior in which offset pressure was significantly higher than onset pressure.

IV. DISCUSSION

In order for vocal-fold vibration to occur, the subglottic air pressure must exceed a certain threshold. This study has analyzed the factors influencing threshold pressure at the onset and offset of phonation. Theoretical work and early measurements in larynges have indicated that the threshold pressure may be influenced by the fundamental frequency of vibration and by certain physical properties of the vocal folds. Moreover, some results indicate that offset pressure is less than onset pressure.

Percutaneous puncture into the airway was used in this study to measure the subglottic pressure in our study subjects. This technique is commonly used for many office-based laryngeal procedures and was well tolerated by all of our subjects.

The influence of four variables on threshold pressure was first studied. Fundamental frequency and sound intensity were chosen because of their known relationship with vocal-fold vibration. Fundamental frequency is directly related to the physical properties of the vibrating folds; frequency rises

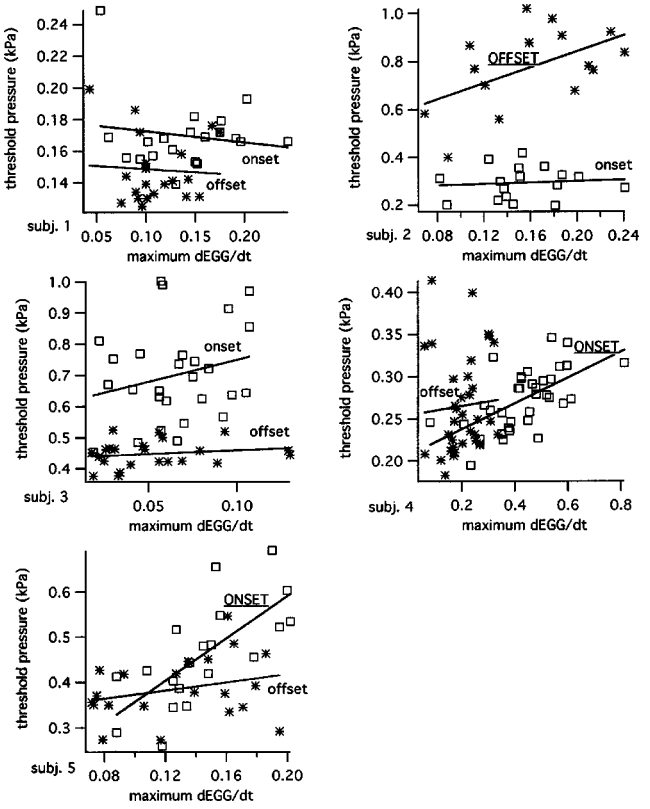


FIG. 5. Scatter plots and fitted regression lines showing the relationship between phonation threshold pressure and maximum dEGG/dt. Lines whose slope was significantly different from zero are capitalized and underlined (e.g., ONSET and OFFSET).

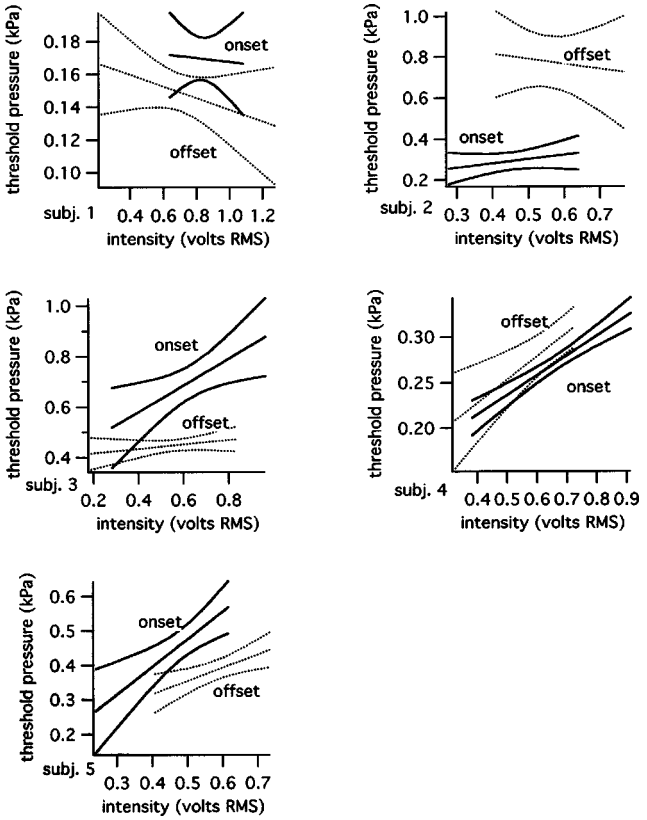


FIG. 6. Fitted regression lines and 90%-confidence bands ($p<0.05$) for onset and offset threshold as a function of sound intensity.

as vocal-fold tension increases.[17] Sound intensity has previously been shown to increase with the amplitude of the mucosal wave and with subglottic pressure.[6,18]

The study results showed substantial variability in the relationship of threshold pressure to acoustic or vibratory properties among and between subjects. These findings, seen in other *in vivo* studies utilizing direct measurement of subglottic pressure,[12] demonstrate that vocal-fold vibration is influenced by a complex combination of aerodynamic and tissue characteristics. Several trends could still be identified. Univariate analysis showed a strong relationship between fundamental frequency and threshold pressure. All of the significant nonzero slopes were positive, indicating that pressure increased with higher frequency. These results agreed with previous theoretical studies and models.[1–3]

Results of multiple linear regression analysis indicated that threshold pressure was most often influenced by vocal intensity, fundamental frequency, and laryngeal airway resistance. The first derivative of the electroglottographic signal made only a small contribution to the threshold pressure. This parameter was chosen as a possible representative of the closure force of the vocal folds, but it either did not represent that force or the closure force had little effect on threshold pressure.

The importance of laryngeal airway resistance is thresh-

TABLE I. Results of multivariable analysis, showing those variables found significantly predictive of phonation threshold pressure.

ONSET subject	R^2	Best predictors in multiple linear regression:
1	0.89	frequency, intensity, airway resistance, dEGG
2	0.83	frequency, intensity, airway resistance
3	0.73	frequency, intensity, airway resistance, dEGG
4	0.74	intensity, airway resistance
5	0.84	intensity, airway resistance
OFFSET subject		
1	0.72	frequency, intensity
2	0.45	frequency, intensity
3	0.55	frequency, airway resistance
4	0.79	frequency, intensity, airway resistance, dEGG
5	0.73	intensity, airway resistance

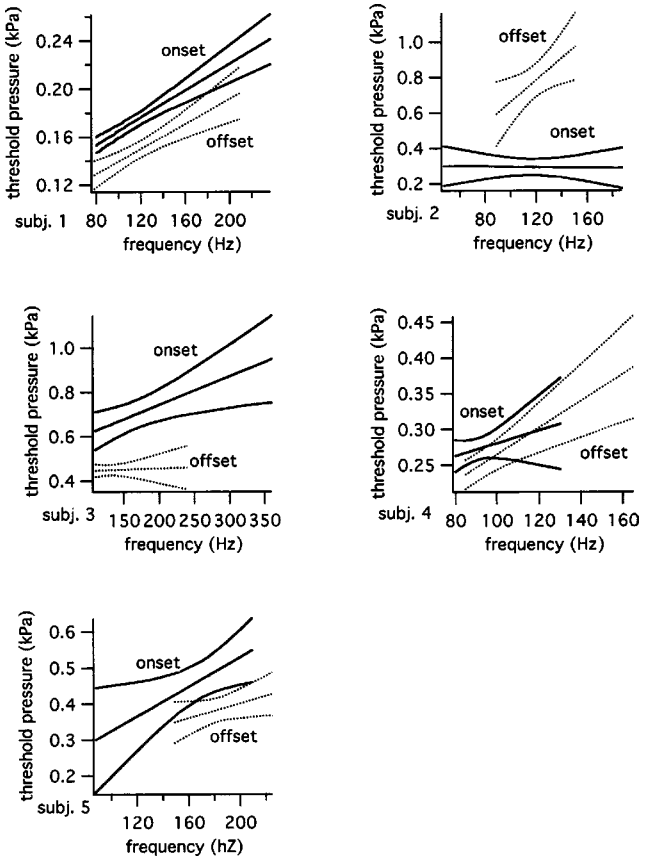


FIG. 7. Fitted regression lines and 90%-confidence bands ($p<0.05$) for onset and offset threshold as a function of fundamental frequency.

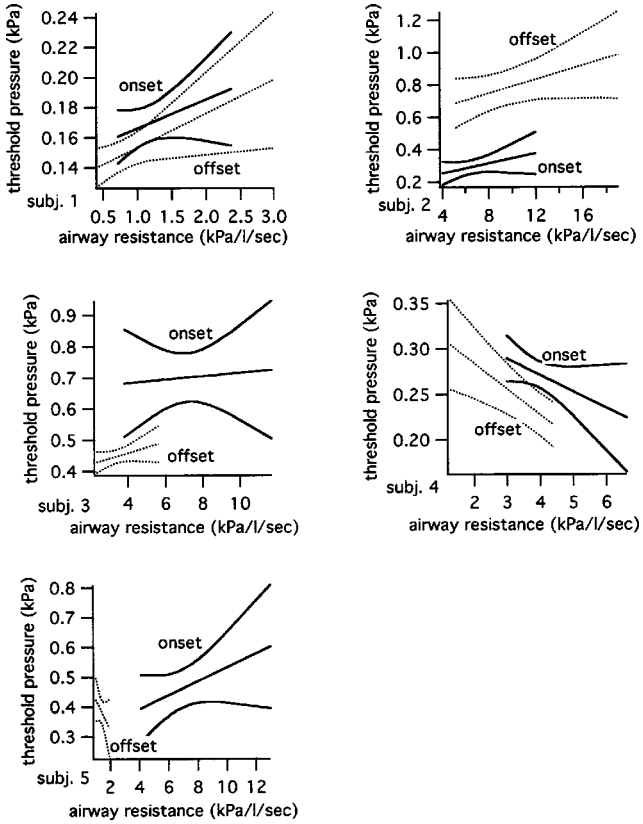


FIG. 8. Fitted regression lines and 90%-confidence bands ($p<0.05$) for onset and offset threshold as a function of laryngeal airway resistance.

old pressure is somewhat misleading. Since subglottic pressure is used to calculate laryngeal airway resistance, correlation between threshold pressure and resistance reflects primarily the influence of airflow on threshold pressure. For example, a strained onset should produce minimal airflow and therefore a high laryngeal airway resistance. There were two unusual instances seen (both in the same subject, number 4) in the univariate analysis in which the threshold pressure actually decreased with increasing airway resistance (Fig. 4, fourth graph). These results would occur if the change in the airflow exceeded the change in the subglottic pressure.

The presence of a hysteresis effect was difficult to uniformly confirm or deny due to the variability in pressure relationships both between and within subjects. Two of the subjects demonstrated such an effect for most test conditions. Two other subjects had little statistical difference between onset and offset pressure. And, subject 2 had offset threshold pressures that were consistently higher than the onset pressure, just the opposite of the predicted behavior. This individual may have had additional ''pressing'' of the vocal folds at offset, although the threshold pressures were higher even at low laryngeal airway resistances.

V. CONCLUSIONS

Direct measurement of subglottic pressure in normal subjects has confirmed that phonation threshold pressure is influenced by fundamental frequency and, to a lesser extent,

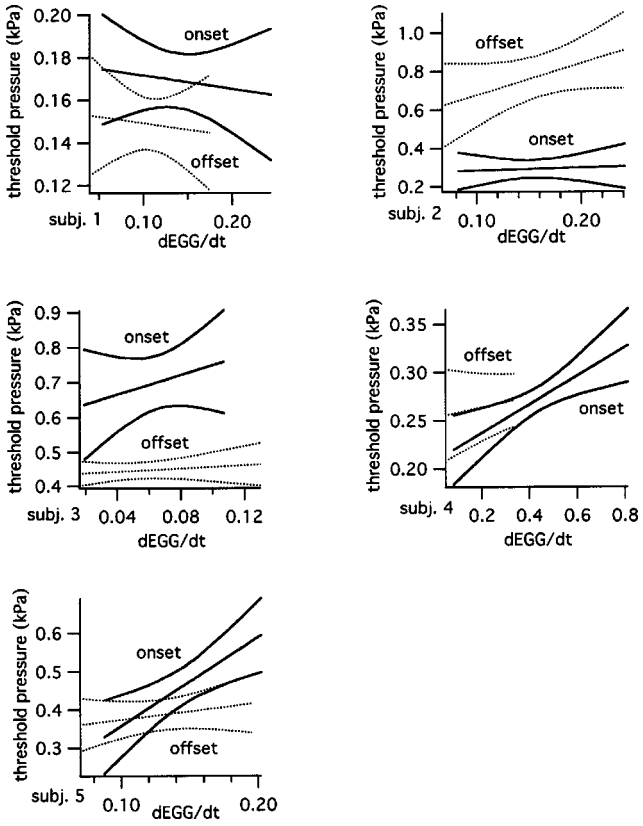


FIG. 9. Fitted regression lines and 90%-confidence bands ($p<0.05$) for onset and offset threshold as a function of maximum dEGG/dt.

laryngeal airway resistance and intensity. However, these results were not seen in all subjects or in all phonation trials. In some cases the threshold pressure for offset was lower than the onset pressure, indicating that the vocal folds develop a hysteresis effect while vibrating.

[1] I. R. Titze, ''The physics of small-amplitude oscillation of the vocal folds,'' J. Acoust. Soc. Am. **83**(4), 1536–1552 (1988).

[2] I. R. Titze, S. S. Schmidt, and M. R. Titze, ''Phonation threshold in a physical model of vocal fold mucosa,'' J. Acoust. Soc. Am. **97**, 3084–3090 (1995).

[3] R. W. Chan, I. R. Titze, and M. R. Titze, ''Further studies of phonation threshold pressure in a physical model of the vocal fold mucosa,'' J. Acoust. Soc. Am. **101**(6), 3722–3727 (1997).

[4] J. C. Lucero, ''The minimum lung pressure to sustain vocal fold oscillation,'' J. Acoust. Soc. Am. **98**(2), 779–784 (1995).

[5] J. C. Lucero, ''A theoretical study of the hysteresis phenomenon at vocal fold oscillation onset–offset,'' J. Acoust. Soc. Am. **105**(1), 423–431 (1999).

[6] J. Jiang and I. R. Titze, ''A methodological study of hemilaryngeal phonation,'' Laryngoscope **103**, 872–882 (1993).

[7] R. L. Plant and A. D. Hillel, ''Direct measurement of subglottic pressure and laryngeal resistance in normal subjects and in spasmodic dysphonia,'' J. Voice **12**(3), 300–314 (1998).

[8] J. R. Smitheran and T. J. Hixon, ''A clinical method for estimating laryngeal airway resistance during vowel production,'' J. Speech Hear. Disord., **46**(2), 138–146 (1981).

[9] K. Verdolini-Marston, I. R. Titze, and D. G. Druker, ''Changes in phonation threshold pressure with induced conditions of hydration,'' J. Voice **4**(2), 142–151 (1990).

[10] J. Sundberg, M. Andersson, and C. Hultqvist, ''Effects of subglottal pressure variation on professional baritone singers' voice sources,'' J. Acoust. Soc. Am. **105**(3), 1965–1971 (1999).

[11] J. Jiang, T. O'Mara, D. Conley, and D. Hanson, ''Phonation threshold pressure measurements during phonation by airflow interruption,'' Laryngoscope **109**, 425–432 (1999).

[12] R. L. Plant and R. M. Younger, ''The interrelationship of subglottic air pressure, fundamental frequency, and vocal intensity during speech,'' J. Voice **14**(2), 170–177 (2000).

[13] A. D. Hillel, ''The study of laryngeal muscle activity in normal human subjects and in patients with laryngeal dystonia using multiple fine-wire electromyography,'' Laryngoscope **111**, 1–47 (2001).

[14] S. Nasri, J. A. Sercarz, B. Azizzadeh, J. Kreiman, and G. S. Berke, ''Measurement of adductory force of individual laryngeal muscles in an *in vivo* canine model,'' Laryngoscope **104**, 1213–1218 (1994).

[15] R. F. Orlikoff, ''Assessment of the dynamics of vocal fold contact from the electroglottogram: Data from normal male subjects,'' J. Speech Hear. Res. **34**, 1066–1072 (1991).

[16] E. B. Holmberg, R. E. Hillman, and J. S. Perkell, ''Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal, and loud voice,'' J. Acoust. Soc. Am. **84**(2), 511–529 (1988).

[17] D. H. Slavit, R. J. Lipton, and T. V. McCaffrey, ''Phonatory vocal fold function in the excised canine larynx,'' Otolaryngol.-Head Neck Surg. **103**, 947–956 (1990).

[18] J. Sundberg, I. Titze, and R. Scherer, ''Phonatory control in male singing: A study of the effects of subglottal pressure, fundamental frequency, and mode of phonation on the voice source,'' J. Voice **7**(1), 15–29 (1993).

Rapid adaptation to foreign-accented English[a)]

Constance M. Clarke[b)] and Merrill F. Garrett
Department of Psychology, University of Arizona, Tucson, Arizona 85721

(Received 13 January 2004; revised 9 September 2004; accepted 14 September 2004)

This study explored the perceptual benefits of brief exposure to non-native speech. Native English listeners were exposed to English sentences produced by non-native speakers. Perceptual processing speed was tracked by measuring reaction times to visual probe words following each sentence. Three experiments using Spanish- and Chinese-accented speech indicate that processing speed is initially slower for accented speech than for native speech but that this deficit diminishes within one minute of exposure. Control conditions rule out explanations for the adaptation effect based on practice with the task and general strategies for dealing with difficult speech. Further results suggest that adaptation can occur within as few as two to four sentence-length utterances. The findings emphasize the flexibility of human speech processing and require models of spoken word recognition that can rapidly accommodate significant acoustic-phonetic deviations from native language speech patterns. [DOI: 10.1121/1.1815131]

PACS numbers: 43.71.Bp, 43.71.Hw [RLD] Pages: 3647–3658

I. INTRODUCTION

Foreign accent is a source of variability in speech that can be particularly detrimental to speech perception. Non-native speech can cause misidentification of words (Lane, 1963; Munro and Derwing, 1995a, b; van Wijngaarden, 2001) and increased processing time (Munro and Derwing, 1995b). However, research shows that experience with accented speech improves perceptual accuracy (Bradlow and Bent, 2003; Clarke, 2000; Weil, 2001; Wingstedt and Schulman, 1987). What is not clear is how much (or little) experience is necessary for changes in perception to occur. Do significant changes require hours, minutes, or only seconds of exposure? Our study addressed this question by recording on-line changes in the ease of processing non-native speech during the first few moments of exposure. A principal assumption was that even a short sample of foreign-accented speech contains phonological regularities that the listener's speech processing system could exploit. Knowledge of these regularities should lead to more efficient decoding of the speech signal, and therefore to improved processing efficiency after exposure to only a few accented utterances.

Traditionally, variability in the acoustic realization of words and phonemes has been viewed as noise and an impediment to language perception. It was assumed that in the initial stages of speech processing, a *normalization* mechanism strips away ''nonlinguistic'' aspects of the signal (such as those due to vocal tract characteristics), revealing invariant acoustic cues to phonetic identity (Liberman *et al.*, 1967; Shankweiler *et al.*, 1977). In compensating for differences in vocal tract characteristics, normalization was thought to involve calibrating vowel perception according to the vowel space dimensions of each speaker (Joos, 1948). There is evidence that vowel identification is indeed evaluated in a relative, rather than absolute, manner. Ladefoged and Broadbent (1957) found that identification of vowels in a /bVt/ context changed according to the frequency range of formants in a carrier sentence. Another finding consistent with the normalization hypothesis is the effect of constantly changing the talker in a series of spoken words. This results in slower processing and increased error rates, presumably due to the need to recalibrate to each talker (Mullennix *et al.*, 1989; Sommers *et al.*, 1994).

However, recent findings indicate that many nonlinguistic aspects of speech are not discarded, as the normalization hypothesis implies, but are retained and affect later speech processing (for a review, see Pisoni, 1997). For example, Palmeri *et al.* (1993) reported faster recognition for previously presented words when the second presentation was in the same voice rather than a new voice. This should not occur if specific voice characteristics are discarded during speech perception. Consequently, many researchers have argued that some forms of variability should be reconceptualized as useful information rather than noise, and that speech perception models must incorporate variability as a fundamental aspect of spoken word recognition (Luce and McLennan, in press; Nygaard and Pisoni, 1995; Pisoni, 1997).

Against this background, foreign accent presents a useful kind of variability for experimentation. Non-native speech contains multiple departures from native norms that can cause perceptual difficulty. These include deviations from native phoneme prototypes as well as non-native phonetic context rules, syllable structure, and prosodic patterns. Foreign-accented speech thus requires adaptation to a range of variability that supplements previous perception research on variability due to vocal tract characteristics, phonetic context, and speaking rate.

[a)]This work contributed to the first author's doctoral dissertation and was conducted under the direction of the second author. Portions of this work were presented in ''Perceptual adjustment to foreign-accented English with short term exposure,'' in *Proceedings of the 7th International Conference on Spoken Language Processing, Denver, CO, 2002*, edited by J. H. L. Hansen and B. Pellom, pp. 253–256.

[b)]Present address: Department of Psychology, University at Buffalo, State University of New York, Buffalo, NY 14260. Electronic-mail: cclarke2@buffalo.edu

This is not to say that the types of phonetic variability observed in accented speech do not occur in native speech. We assume here that the variations of accented speech can be considered an extreme form of that seen among native speakers (Nygaard and Pisoni, 1998). It might be supposed that adapting to a new native voice involves accommodation of new acoustic speech characteristics, such as voice quality, while adapting to an accented voice requires accommodation of new phonetic characteristics. However, it has been shown that familiarity with a native voice is based not only on voice quality but also on pronunciation idiosyncrasies, i.e., phonetic patterns (Remez *et al.*, 1997) typically due to dialect or idiolect. Therefore, it is plausible that similar mechanisms are at work in coping with differences among native speakers and with differences between native and accented speakers. The advantage in studying accented speech is that these processes are exaggerated, whereas they typically occur too quickly for observation with native speech.

Lane (1963) seems to have been the first to establish that word identification is poorer for accented than for native speech. He found that word identification accuracy for Serbian-, Japanese-, and Punjabi-accented English was approximately 36% lower than for native speech in all signal-to-noise ratio and filtering conditions. More recent work has shown lower intelligibility for a Mandarin accent in second language (L2) learners of English (Munro and Derwing, 1995a, b), and an English accent in L2 learners of Dutch (van Wijngaarden, 2001). van Wijngaarden estimated that the reduction in intelligibility for the speech of fluent non-natives was equivalent to lowering native speech by 3 to 4 decibels (dB).

In addition to poorer word identification, accented speech can slow perceptual processing. Schmid and Yeni-Komshian (1999) showed that mispronunciations were detected more slowly for Spanish- and Tamil-accented speech. And, in a speeded true/false sentence verification task, Munro and Derwing (1995b) found that listeners were slower to verify sentences produced by a Mandarin-accented speaker, even when they transcribed them correctly.

However the perceptual system handles variability in speech, some kind of learning must be involved. Normalization as a way to compensate for talker-specific characteristics must be supplemented or supplanted by mechanisms that retain information about those characteristics and apply it to later perceptual processing. Many studies have demonstrated the perceptual benefits of experience with various speech characteristics, including those due to talker differences (Nygaard and Pisoni, 1998; Nygaard *et al.*, 1994), speaking rate (Dupoux and Green, 1997), hearing impairment (McGarr, 1983), and synthetic speech (e.g., Greenspan *et al.*, 1988; for a review, see Duffy and Pisoni, 1992).

The last few years have seen an increased interest in the perceptual learning of accented speech characteristics (Bradlow and Bent, 2003; Clarke, 2000; Weil, 2001). However, the first experiments investigating this topic were in the 1980s. Gass and Varonis (1984) reported improved sentence transcription accuracy following exposure to a story spoken by a non-native speaker. Wingstedt and Schulman (1987) created a ''fake'' accent in Swedish speech and found higher word repetition accuracy for participants who had earlier repeated a set of accented sentences. Because this improvement was attained with novel words, the researchers concluded that the listeners had developed ''phonological perceptual rules'' (p. 339)—rules for translating the acoustic-phonetic input for a particular accent to the native representations of the intended phonemes.

Several recent studies have replicated the findings of perceptual improvement following exposure to accented speech. Clarke (2000) trained native English listeners over three sessions to recognize either several Spanish-accented voices or several Chinese-accented voices. In a subsequent word transcription test that included one of the Spanish-accented voices and one of the Chinese-accented voices from training, listeners were more accurate with the accented voice they were trained with than the other accented voice. In a similar study, Weil (2001) exposed listeners to a Marathi-accented voice using word and sentence transcription tasks over four training sessions. In a following transcription test, trained listeners were more accurate with the Marathi-accented voice than were untrained listeners. Finally, Bradlow and Bent (2003) also found a beneficial effect of familiarity with a Chinese-accented voice in a sentence transcription test after two days of transcription training.

In sum, there is clear evidence of perceptual learning for a variety of nonlinguistic speech characteristics—aspects of speech once thought of as noise to be discarded during language processing. This information is retained and used in a beneficial way. Furthermore, the benefits are integral to the speech system: In various cases it has been shown that task improvement did not rely on the use of semantic context (McGarr, 1983), becoming accustomed to odd physical characteristics of the speech (Greenspan *et al.*, 1988), or memorizing alternative pronunciations for individual lexical items (Nygaard *et al.*, 1994; Wingstedt and Schulman, 1987). Rather, the benefits seem to result from adjustments at a relatively early level of processing, probably the phonological level (Duffy and Pisoni, 1992).

Existing studies demonstrate benefits of several hours experience with accented speech. But an unanswered question is when perceptual learning of accented speech begins. Does the change in processing require prolonged experience, or is the response more rapid? Knowledge of the timeline for perceptual learning will constrain the possible mechanisms underlying this learning. Previous research with talker-specific characteristics (Ladefoged and Broadbent, 1957) and compressed speech (Dupoux and Green, 1997) has shown changes with exposure to as few as one and ten sentence-length utterances, respectively. This indicates that some form of adaptation occurs almost immediately. Moreover, subjective experience suggests that adaptation to accented speech also occurs quickly. Many listeners report that when first listening to a non-native speaker, comprehension can be difficult, but after a few moments of exposure they ''catch on'' to the non-native speech patterns, and comprehension improves. These subjective reports might reflect an increased reliance on utterance or situational context to interpret unintelligible words, or on rapid perceptual learning of phonological patterns, or both. The focus of the research reported

here was the learning of phonological patterns specific to the accented speaker. We attempted to rule out other sources of improvement, such as the use of semantic context or general increases in effort or attention. Thus, a finding of improved processing efficiency would suggest adaptive flexibility within the phonological processing system itself.

We used a cross-modal matching task with a reaction time measure to track processing efficiency during the course of exposure to accented voices. Sentence-length English utterances were presented, each ending in a key word not predictable from the semantic context. Immediately following each auditory sentence, a visual probe word appeared on a computer screen. The task was a speeded yes/no response indicating whether the probe word matched the final word of the sentence. This task had three advantages. First, using sentences rather than isolated words more closely matches conversational language and allows for the influence of all phonological aspects of accented speech, including interword phonetic context effects and prosodic patterns. Second, reaction time (RT) may be a more sensitive measure of subtle changes in processing efficiency during adaptation. RT has already proven sensitive to the perceptual difficulty caused by accented speech even when intelligibility is high (Munro and Derwing, 1995b). Third, processing speed is sampled at the end of every sentence, approximately once every 2 s of speech, providing a relatively fine-grained temporal resolution of the effects of perceptual learning.

II. EXPERIMENT 1

This experiment looked for changes in processing efficiency arising from a short period of exposure to accented speech (under one minute). Three stimulus conditions were tested: 16 sentences produced by a native speaker of Spanish with a moderate accent (*accent* condition); 12 sentences produced by a native speaker of English, followed by four sentences produced by the Spanish-accented speaker (*control* condition); and 16 sentences produced by the native speaker of English (*no accent* condition).

It was predicted that the *accent* group would be slower than the *control* and *no accent* groups at the beginning of exposure, but as the experiment progressed the *accent* groups' RT would decrease due to adaptation. For the last four sentences, the accent group was expected to be faster than the control group because the control group would not have had the previous 12 sentences to adapt to the accented speech. This controlled for the possibility that the accent group's improvement was due solely to practice with the task. Finally, the no accent condition provided a comparison for the level of processing efficiency the accent group reached by the fourth block (i.e., to see whether they reached native-speech processing efficiency).

A. Method

1. Participants

Forty-eight University of Arizona undergraduates (37 females, 11 males) participated and were paid or given partial course credit. All were native speakers of American English and reported no hearing problems at the time of testing. They were not fluent in Spanish and did not have native Spanish-speaking parents. However, 31 participants had other family members or close friends with a Spanish accent. Data from one additional participant were excluded because the participant exceeded a 20% error rate. Participants were randomly assigned to the three conditions (16 each in *accent, control*, and *no accent*).

2. Design

Four blocks of four auditory sentences were presented to each group with no breaks between blocks. In each block, two sentences were followed by matching visual probe words (*yes* trials), and two were followed by mismatching visual probe words (*no* trials). Four block orders were created using a Latin square design such that, across participants, every sentence was presented in every block position. Within each group, an equal number of listeners heard each of the four block orders.

Because between-participants RT comparisons were necessary, and given the typically high RT variability among participants, all experimental RTs were normalized according to a separate measure of each participant's speed at the task: Following the experimental trials, all listeners received eight *baseline* trials. The sentences in these trials were novel and were produced by a different nonaccented speaker. The mean RT for these trials was subtracted from the experimental RTs for each subject, and the difference RTs were the primary dependent measure of processing speed.[1]

3. Stimulus materials

We used 32 low probability (LP) sentences from the Revised Speech Perception In Noise (SPIN-R) test (Bilger, 1984; Kalikow *et al.*, 1977). In LP sentences, the final word is not predictable from the meaning of the sentence (e.g., *Ruth must have known about the* pie.). The 16 experimental sentences were recorded by a female native speaker of American English (age: 31) and by a female native speaker of Mexican Spanish (age: 45; age of English acquisition: 30). In an earlier study, the native Spanish speaker's accent was given a mean rating of 6.1 on a scale of 1 (no foreign accent) to 9 (strong foreign accent). The eight practice and eight baseline sentences were recorded by another female native speaker of American English (age: 19).

Sentences were tape recorded in a quiet room using an Electro-Voice RE16 directional microphone, a DBX 760X microphone preamplifier, and a Tascam 122 mkII tape deck. Stimuli were digitized (Macintosh PowerPC 8100; 22.05 kHz, 16 bits), and each sentence was copied into its own file with silence at the beginning and end trimmed. Each stimulus file was peak normalized to 90% of maximum amplitude resolution.

Because processing efficiency was measured by perceptual identification of the final word of each sentence (the *target* word), the characteristics of these words were carefully controlled. Target words in the experimental sentences were familiar, monosyllabic nouns with a mean frequency of 22 per million (Kucera and Francis, 1967). In both the accented and nonaccented productions, they were each correctly identified in isolation by more than 70% of listeners in

TABLE I. Mean percent error (untransformed; and standard deviations) by experimental block according to voice condition for experiments 1–3.

Condition	Block 1	2	3	4
		Experiment 1		
Accent	14.06 (15.73)	4.69 (10.08)	10.94 (15.73)	4.69 (10.08)
Control	3.13 (8.54)	1.56 (6.25)	0.00 (0.00)	9.38 (15.48)
No accent	1.56 (6.25)	1.56 (6.25)	6.25 (11.18)	0.00 (0.00)
		Experiment 2		
Accent	6.25 (11.18)	4.69 (13.60)	6.25 (14.43)	6.25 (14.43)
Control in noise	14.06 (15.73)	10.94 (12.81)	3.13 (8.54)	4.69 (10.08)
		Experiment 3		
Accent	1.79 (5.25)	2.38 (5.94)	7.14 (11.50)	3.57 (6.96)
Control in clear	2.38 (5.94)	2.38 (5.94)	1.19 (4.37)	4.17 (7.35)
Control in noise	11.90 (12.71)	13.10 (14.60)	11.90 (12.71)	4.17 (7.35)

a prior word intelligibility experiment ($M_{\text{accented}}=92.4\%$; $M_{\text{nonaccented}}=95.8\%$).[2] Target words were never repeated within the experiment.

To control for differences in duration, the accented and nonaccented productions of each target word were digitally compressed or lengthened so they both equaled the mean duration of the two original productions.[3] The accented words were lengthened by a mean of 15% (range: +5% to +28%), and the nonaccented words were compressed by a mean of 11% (range: −5% to −18%). Stimuli were not rescreened for intelligibility, but the duration manipulations were minimal and produced little to no distortion. The mean durations of the preceding portion of each sentence were similar for the two voices ($M_{\text{accented}}=1.29\,\text{s}$; $M_{\text{nonaccented}}=1.36\,\text{s}$) and were not altered.

For half of the experimental trials, the visual probe word did not match the target word but was a phonetic neighbor differing by one phoneme in either the onset (one case), vowel (four cases), or coda (three cases) position. These words were also monosyllabic English nouns and were similar to the target words in mean frequency (20 per million).

4. Procedure

Listeners were tested individually in a quiet room in front of a computer monitor and a two-button response box. They were instructed to respond quickly and accurately and were warned that at some time during the experiment the voice would change. This voice change occurred between blocks 3 and 4 for the control condition, and between block 4 and the baseline trials for the accent and no accent conditions. The experiment began with eight practice trials, followed by four experimental blocks (16 trials), and then eight baseline trials. Trials in the practice and baseline blocks were presented in the same order for every subject. In each experimental block, sentence order was randomized differently for each subject.

Auditory stimuli were presented over headphones at approximately 73 dB(A) sound pressure level. Stimulus presentation and response collection were controlled by an IBM compatible computer using DMDX software (Forster and Forster, 2003). Each trial began with an auditory sentence followed immediately by a probe word presented in capital letters on the computer screen. The probe remained until the listener pressed the *yes* or the *no* button. Accuracy and reaction time feedback were provided on the computer screen after each response. RT measurement began at probe word onset. If a response did not occur within 4 s, it was recorded as *no response*. Participants pressed a foot pedal to begin the next trial.

B. Results and discussion

For all analyses, the between-participants block order counterbalancing variable was included in the analysis of variance (ANOVA) to remove variance due to counterbalancing. If there was no effect involving counterbalancing group by a conservative criterion of $p<0.25$ (see Pollatsek and Well, 1995), the analysis was performed with groups collapsed over this variable. To avoid confusion with the experimental group variable, the counterbalancing group variable will be denoted by cbg. We used an alpha level of 0.05 for all analyses, and the modified Bonferroni correction (Keppel, 1982) to control alpha inflation in planned contrasts.

1. Errors

Incorrect and *no response* trials were counted as errors. Mean error rates for the experimental blocks are shown in Table I. Error data were first transformed to rationalized arcsine units (RAU) to convert percent error to a linear and additive scale (Studebaker, 1985). To test for differences among conditions for the first three blocks, a 3 (group)×3 (block: 1–3)×4 (cbg) mixed design ANOVA was performed. This revealed a significant effect of group, $F(2,36)=5.68$, no effect of block, $F(2,72)=2.33$, ns, and a marginal interaction, $F(4,72)=2.16$, $p=0.08$. Tamhane *post hoc* tests for unequal variances showed the accent group's error rate was higher than the control group's, $t(18)=3.07$, and marginally higher than the no accent group's, $t(23)=2.30$, $p=0.09$. Finally, for block 4, a 3 (group)×4 (cbg) ANOVA showed a main effect of group, $F(2,36)=3.49$. A Tamhane *post hoc*

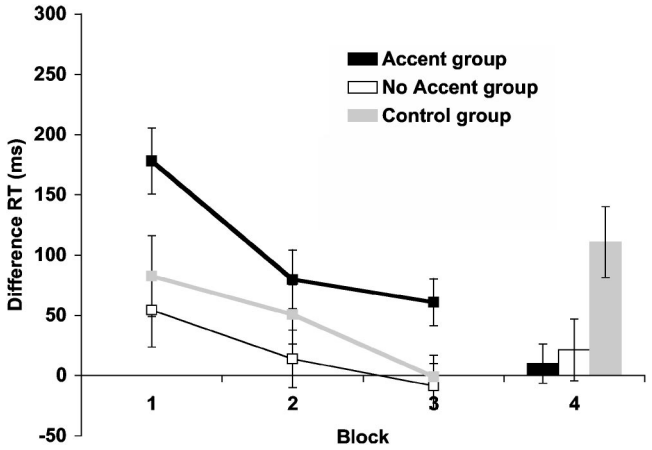


FIG. 1. Experiment 1 mean difference reaction times (experimental −baseline) according to condition. Error bars indicate standard errors.

test for unequal variances showed the control group's error rate was marginally higher than the no accent group's, $t(15)=2.52$, $p=0.07$.

The higher error rate for the accent group in the first three blocks was unexpected because the accented target words had been screened for intelligibility. Although only correct responses were used in the RT analyses, they may have included more guess responses, which do not reflect the time course of accurate phonological analysis and may have inflated the RTs. We note this issue here and return to it in experiment 3.

2. Reaction time

The following data treatment applied to this and all subsequent RT analyses. Only RTs from correct responses were analyzed. RTs less than 200 ms or greater than 2000 ms were excluded, and RTs beyond two standard deviations above or below a given participant's mean for experimental and baseline trials were replaced with that cutoff value. Each participant's mean baseline RT was subtracted from that individual's experimental block mean RTs. These difference RTs were used as the dependent measure.

Mean difference RTs for the experimental blocks are shown in Fig. 1. RTs for the accent group decreased by over 150 ms from block 1 to block 4. This was statistically significant by a one-way ANOVA for the accent condition with block as a repeated measures variable, $F(3,45)=13.24$. Planned contrasts indicated RT decreased significantly between block 1 ($M=178$, $SD=109.62$) and block 2 ($M=80$, $SD=97.41$), $t(15)=3.24$, and between block 3 ($M=61$, $SD=77.66$) and block 4 ($M=10$, $SD=65.34$), $t(15)=2.52$. This decrease in RT is consistent with the hypothesis that processing efficiency increases with brief exposure to accented speech.

To assess the possibility that practice with the task could account for the RT change, the accent and control groups' RTs for block 4 were compared. A 3 (group)×4 (cbg) ANOVA included this as well as the comparison of the accent and no accent conditions. A significant effect of group was found, $F(2,36)=5.50$, and planned contrasts showed that the accent group ($M=10$, $SD=65.34$) was significantly faster than the control group ($M=111$, $SD=117.27$), $t(36)=3.03$, but did not differ from the no accent group ($M=21$, $SD=102.31$), $t(36)=0.34$, *ns*. The control group listeners, who had as much experience with the task, but no prior experience with the accented voice, were significantly slower than the accent group listeners in block 4 when presented with the accented voice. This indicates that the accent group's improvement across the four blocks cannot be explained by practice effects alone. The results also show no difference between the accent and no accent groups in block 4, suggesting that after only 16 sentences of exposure, the group hearing the accented speech was processing it as quickly as the group hearing native speech.

Two effects in the first three blocks were also of interest: (a) whether listeners were slower to respond to the Spanish-accented speech than to the native speech, and (b) whether the control and no accent groups' RTs decreased as they were exposed to the nonaccented voice. A 3 (group)×3 (block: 1–3) mixed ANOVA showed a significant main effect of group, $F(2,45)=4.91$. Planned contrasts indicated that, for the first three blocks, the accent group ($M=106$, $SD=71.35$) was slower than both the control group ($M=44$, $SD=89.68$), $t(45)=2.18$, and the no accent group ($M=20$, $SD=79.20$), $t(45)=3.04$. The slower RT to accented speech is consistent with previous findings of processing difficulty with foreign-accented speech (Munro and Derwing, 1995b). There was also a significant main effect of block, $F(2,90)=17.94$, but no group×block interaction, $F(4,90)=1.20$, *ns*, suggesting all groups improved across the first three blocks.

Overall, the results matched experimental predictions. Listeners were initially slower to respond to the Spanish-accented speech, but this difficulty decreased after a relatively brief period of exposure. At the end of 16 sentences, mean RT had decreased by over 150 ms and was almost identical to the no accent group's RT to native speech. Given that each sentence was approximately 2 seconds long, this exposure comprised less than 1 min of speech. Although practice effects likely played some part in the RT change across blocks, as indicated by the improvement in the first three blocks for the control and no accent groups, it cannot explain the entire effect for the accent group.

III. EXPERIMENT 2

The control condition in experiment 1 ruled out practice effects as an account of the accent group's improvement over the four experimental blocks. However, another possible reason for that improvement is that the listeners in the accent condition developed strategies for understanding difficult speech, such as depending on the fact that the final word was always a noun, or putting more effort or attention into the task.

Experiment 2 was conducted to evaluate these explanations. Noise was added to the nonaccented sentences in the control condition to make them more difficult to understand. A signal-to-noise ratio (SNR) was chosen such that overall RTs to the nonaccented speech in noise were similar to RTs to the accented speech in the clear. Under these new conditions, results for block 4 would remain the same as for ex-

periment 1 if the accent group is learning something specific about how to process accented speech during exposure, rather than learning general strategies or increasing effort. A no accent condition was not included.

A. Method

1. Participants

Thirty-two University of Arizona undergraduates (22 females, 10 males) participated and were paid or given partial course credit. All were native speakers of American English and reported no hearing problems at the time of testing. They were not fluent in Spanish and did not have native Spanish-speaking parents. However, 11 participants had other family members or close friends with a Spanish accent. Data from one additional participant were excluded because the participant exceeded a 20% error rate. Participants were randomly assigned to the two conditions (16 each to *accent* and *control in noise*).

2. Stimulus materials

Materials were identical to those in experiment 1. Using Cool Edit 96 wave-editing software (Syntrillium Software, Phoenix, AZ), the amplitude of the nonaccented sentences was reduced to a mean of approximately 65 dB at the final word, and pink noise at 61 dB was added. The resulting mean SNR at the final word was approximately +4 dB.

3. Procedure

Experimental procedures and block counterbalancing were identical to experiment 1, except that listeners in the control group were warned that the first several sentences were in noise.

B. Results and discussion

1. Errors

Mean error rates for the experimental blocks are shown in Table I. Error rates for blocks 1–3 were higher for the control in noise group (9.38%) than for the accent group (5.73%). However, a 2 (group)×3 (block: 1–3)×4 (cbg) mixed ANOVA on the transformed error percentages (in RAUs) revealed only a marginal effect of group, $F(1,24) = 3.75$, $p = 0.06$. A one-way ANOVA on block 4 showed no significant difference between groups, $F < 1$.

2. Reaction time

Mean difference RTs for the experimental blocks are shown in Fig. 2. To test whether the noise made the control condition more difficult, a 2 (group)×3 (blocks: 1–3)×4 (cbg) mixed ANOVA (Huynh–Feldt corrected because the assumption of sphericity was violated) was performed on the difference RTs for the first three blocks. There was no main effect of group, $F < 1$, suggesting that the two conditions were equated in difficulty for effects on RT. There was a significant main effect of block, $F(2,48) = 5.63$, but no group ×block interaction, $F < 1$, showing that both group's RTs decreased equally across the first three blocks.

Mean RT for the accent group decreased by almost 200 ms from block 1 to block 4. This change was statistically

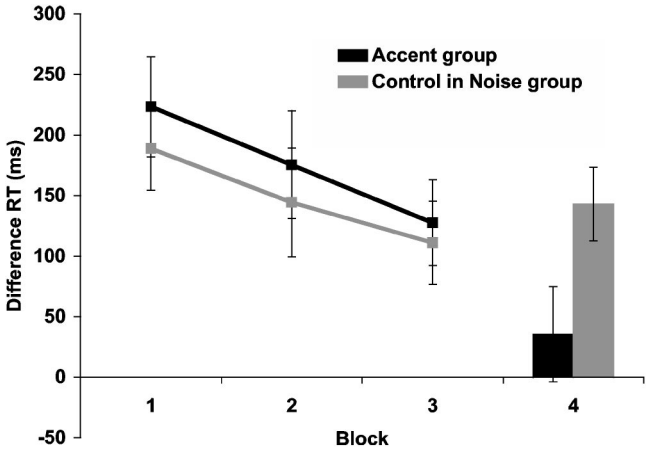


FIG. 2. Experiment 2 mean difference reaction times (experimental −baseline) according to condition. Error bars indicate standard errors.

significant by a 4 (block)×4 (cbg) mixed ANOVA for the accent condition, $F(3,36) = 8.83$. Planned contrasts showed that RT decreased significantly between block 1 ($M = 223$, $SD = 165.55$) and block 4 ($M = 36$, $SD = 157.73$), $t(12) = 3.71$, and between block 3 ($M = 128$, $SD = 141.44$) and block 4, $t(12) = 3.18$. Finally, a 2 (group)×4 (cbg) ANOVA comparing the two voice conditions in block 4 showed that the accent group ($M = 36$, $SD = 157.73$) was significantly faster than the control in noise group ($M = 143$, $SD = 121.69$), $F(1,24) = 5.14$.

Experiment 2 replicated the effect of increased processing efficiency with short-term exposure to accented speech. RTs for the accent group again decreased across the four blocks, and were faster than those of the control in noise group in the last block of the experiment. Because the conditions were equated for difficulty, the assumption that the accent group developed general strategies for coping with difficult speech would also apply to the control group. Yet the control listeners still took longer to process the accented speech in block 4.

An important question regarding rapid adaptation to accented speech is the effect of previous long-term experience with the accent in question. As noted earlier, a large proportion of listeners in experiments 1 and 2 had close family members or friends with a Spanish accent. In order to determine whether the adaptation effect differed based on such experience, the data from the accent groups in the two experiments were combined, and the high experience listeners ($n = 15$) compared with the low experience listeners ($n = 17$). *High experience* listeners were defined as stated above, and *low experience* listeners were those who reported either never having personally known anyone with a Spanish accent or only having Spanish-accented acquaintances with whom they spent little time. Inspection of the data in Table II shows both subgroups improved across the four blocks. Further, in block 1 the low experience subgroup was over 80 ms slower than the high experience subgroup, although this deficit was eliminated by block 4. But the trend indicating an advantage for the high experience subgroup was not statisti-

TABLE II. Mean difference reaction times (in ms; and standard deviations) for accent conditions according to experience with the accent (subgroups are combined for experiments 1 and 2).

	Block			
Experience	1	2	3	4
	Experiments 1 and 2			
Low	239 (137.16)	146(149.90)	103(106.97)	17 (89.59)
High	157 (134.30)	107(151.09)	84(130.96)	29(149.47)
	Experiment 3			
Low	143 (158.21)	70(124.72)	30(110.91)	11(110.64)

cally reliable. A 2 (accent experience)×4 (block)×4 (cbg) mixed ANOVA (Huynh–Feldt corrected for nonsphericity) revealed a main effect of block, $F(2.94,70.54)=17.32$, but no effect of accent experience, $F<1$, or interaction, $F(2.94,70.54)=1.40$, *ns*.

Although previous accent experience was not of central interest in this study, the subgroup analysis suggests that it may affect how efficiently accented speech is perceptually processed, at least initially. However, the important result is that even those listeners with little prior experience with a Spanish accent showed faster RTs with brief exposure. This indicates the processing improvement is due to on-line learning, rather than previous knowledge. Experiment 3 further investigates this hypothesis by testing adaptation effects for a less familiar accent.

IV. EXPERIMENT 3

Experiment 3 explored whether the effect of rapid adaptation would be found with another and less familiar accent, specifically a Chinese accent. This accent is less likely to be encountered by the participant population (living in Tucson, AZ) than the Spanish accent. It is possible that listeners can quickly adapt to accented speech only if they are already familiar with the accent. Otherwise, adaptation may require more experience (possibly on the order of hours as in previous training studies). If adaptation to an unfamiliar accent requires relatively long exposure, it would suggest that the adaptation effects found in experiments 1 and 2 were due to a quick accessing and application of stored knowledge of a generally familiar accent, rather than an on-line adaptation process.

Three conditions were tested: an *accent* condition, in which all four sentence blocks were produced by a Chinese-accented speaker, a *control in clear* condition (similar to the control condition in experiment 1), and a *control in noise* condition (similar to the control in noise condition in experiment 2). In contrast to the first two experiments, 24 experimental sentences (six per block) were used rather than 16 (four per block) because a pilot study showed that RTs to the Chinese-accented voice were slower overall than to the Spanish-accented voice of the previous two experiments. Also, although RTs became faster across the four blocks in the pilot experiment, they were not significantly faster on the fourth block than those of the control in noise group, suggesting that listeners might need more exposure to the Chinese-accented speech to fully adapt to it. It is not clear whether this is because of the Chinese accent itself or the specific voice, which was rated as more accented than the Spanish-accented voice in the previous accent judgment experiment. Using six sentences per block was intended to insure that we could detect adaptation to the Chinese-accented voice.

The control in clear condition was included to test once again the prediction that RTs would initially be slower for accented speech than for nonaccented speech. A retest of this prediction was needed because we discovered that the intelligibility of the Spanish-accented target words in experiments 1 and 2 may have been lower than intended: As noted earlier, the accent group had a higher error rate than the other groups in experiment 1. We also conducted a second intelligibility screening in which the target words were presented in the sentence context used in the experiment (rather than in isolation). With this method, several of the accented words used in experiments 1 and 2 had intelligibility rates less than 70%. Low intelligibility rates raise the possibility that, in experiments 1 and 2, the accent groups' RTs included several guess responses, which might be responsible for the higher mean RTs rather than the hypothesized slower processing of intelligible accented words. For the present experiment, only target words that had greater than 70% intelligibility in the experimental sentence context were selected.

We predicted that RT would decrease across the four blocks for the accent condition and that, in the last block, the accent group would be faster than both control groups. This was based on the assumption that the effects of short-term experience found in experiments 1 and 2 were not due to accessing a stored template for interpreting Spanish-accented speech, but instead resulted from on-line learning of phonological patterns.

A. Method

1. Participants

Eighty-four University of Arizona undergraduates (44 females, 40 males) participated and received partial course credit. All were native speakers of American English who reported no hearing disorders at the time of testing. They were not fluent in Chinese and did not have native Chinese-speaking parents. However, 20 participants had other family members or close friends with a Chinese accent. Data from one additional participant were excluded because the participant reported a strategy of not looking at the screen until after each sentence ended. Participants were randomly assigned to the three conditions (28 each to *accent, control in clear*, and *control in noise*).

2. Stimulus materials

A new set of 40 LP sentences was chosen from the Revised SPIN Test (Bilger, 1984; Kalikow *et al.*, 1977). The 24 experimental sentences (six per block) were selected based on their intelligibility when produced by the Chinese- and nonaccented voices. Each of the sentence final words was identified at greater than 70% accuracy in the LP sentence context by a separate group of listeners ($M_{\text{accented}}=94.32$; $M_{\text{nonaccented}}=99.16$).[4] The target words from the experimental sentences were familiar, monosyllabic nouns with a mean

frequency of 20 per million (Kucera and Francis, 1967). Target words were never repeated within the experiment.

The experimental sentences were recorded by the same female native speaker of American English as in experiments 1 and 2 and by a female native speaker of Mandarin Chinese (age: 24, age of English acquisition: 12). The Chinese speaker's accent was given a mean rating of 7.6 on a scale from 1 (no foreign accent) to 9 (strong foreign accent) in an earlier study. The practice and baseline sentences were produced by the same female native speaker as in experiments 1 and 2. The Chinese-accented sentences were recorded according to the procedures described in experiment 1. In order to match the nonaccented sentences with the highly intelligible Chinese-accented sentences, the native speaker of American English recorded additional sentences. These were recorded onto CD in a WhisperRoom sound isolation booth using a Shure SM57 Dynamic microphone, a Symetrix 302 microphone preamplifier, and an Alesis ML-9600 disc recorder (44.1 kHz, 16 bits).[5] The stimulus file was downsampled to 22.05 kHz, and each sentence was copied into its own file, trimmed, and peak normalized to 90% of maximum amplitude resolution.

In order to control for differences in duration, the accented and nonaccented productions of each target word were digitally compressed or lengthened so they both equaled the mean duration of the two original productions. The same procedure was used to equate the sentence material up to the final word because this portion was produced approximately 250 ms slower on average by the accented voice. For the accented voice, the final word durations were modified by a mean of +4% (range: −10% to +19%), and the precursor portions by a mean of −7% (range: −16% to +3%). For the nonaccented voice, the final word durations were modified by a mean of −3% (range: −14% to +13%), and the precursor portions by a mean of +10% (range: −3% to +23%). Again, the duration manipulations produced little to no distortion, and the stimuli were not rescreened for intelligibility.

The mean sound level measured at the final word was approximately 74 dB for the Chinese-accented sentences and approximately 69 dB for the nonaccented sentences in the control in clear condition.[6] For the control in noise condition, the mean amplitude was lowered to approximately 63 dB at the final word, and pink noise at 62 dB was added to each file, for a mean SNR of +1 dB. In a pilot study, this SNR resulted in RTs similar to those in the accent condition. Finally, for experimental trials in which the visual probe word did not match the corresponding target word, the probes differed from their targets by one phoneme in either the onset (four cases), vowel (four cases), or coda (four cases) position, and were similar to the target words in mean frequency (21 per million).

3. Procedure

Experimental procedures and counterbalancing were identical to the previous experiments except that experimental blocks consisted of six sentences (three *yes* trials, three *no* trials) instead of four.

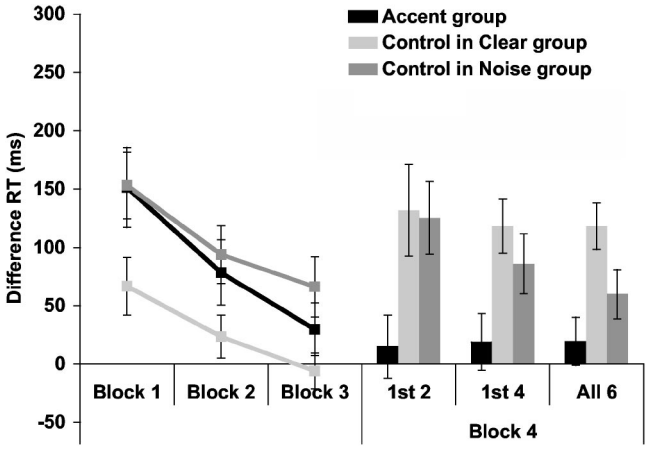


FIG. 3. Experiment 3 mean difference reaction times (experimental −baseline) according to condition. For blocks 1–3, each mean is based on six trials. For block 4, means are based on two, four, and six trials. Error bars indicate standard errors.

B. Results and discussion

1. Errors

Mean error rates for the experimental blocks are shown in Table I. A 3 (group)×3 (block: 1–3)×4 (cbg) mixed ANOVA on the percent error data in RAUs revealed a main effect of group, $F(2,72)=45.86$, and no effect of block or interaction. Planned contrasts showed a higher error rate for the control in noise group than for the accent group, $t(72)=7.42$, and no difference between the accent and control in clear groups, $t(72)=1.53$, *ns*. A 3 (group)×4 (cbg) ANOVA on block 4 showed no significant difference among groups, $F<1$.

2. Reaction time

Mean RTs for the experimental blocks are shown in Fig. 3. To verify that the reduction in SNR in the control in noise condition equated the RTs in the first three blocks with those of the accent group, and to test whether listeners were slower to respond to the accented voice than to the native voice (in the clear), a 3 (group)×3 (blocks: 1–3)×4 (cbg) mixed ANOVA was performed. There were significant main effects of block, $F(2,144)=36.92$, and group, $F(2,72)=3.26$, but no interaction, $F<1$. Planned contrasts showed that the addition of noise in the control condition was successful in equating RTs for the control in noise ($M=104$, $SD=119.59$) and accent ($M=87$, $SD=137.42$) conditions, $t(72)=0.57$, *ns*. In addition, the accent group was marginally slower than the control in clear group ($M=28$, $SD=93.34$), $t(72)=1.87$, $p=0.06$. The difference between the accent and control in clear groups may have been smaller than expected because the accent group adapted to the Chinese-accented voice within the 18 sentences of blocks 1–3. To equate this analysis with the corresponding analysis in experiment 1, the group difference for blocks 1 and 2 only (12 sentences) was tested. For the first two blocks, the accent group ($M=115$, $SD=160.15$) was significantly slower than the control in clear group ($M=45$, $SD=106.28$), $t(72)=2.01$. RTs to the Chinese-accented speech were initially slower than to the native speech, despite the fact that the

intelligibility of the Chinese-accented speech was high, and the error rates were similar for the accent and control in clear conditions. The results are again consistent with Munro and Derwing's (1995b) findings that accented speech is processed more slowly than native speech even when it is highly intelligible.

RTs for the accent group decreased by just over 130 ms from block 1 to block 4. This change was statistically significant by a 4 (block)×4 (cbg) mixed ANOVA (Huynh–Feldt corrected for nonsphericity), $F(2.53,60.71)=16.23$. Planned contrasts showed that RT decreased significantly between block 1 ($M=151$, $SD=181.45$) and block 2 ($M=79$, $SD=148.30$), $t(24)=4.47$, and between block 2 and block 3 ($M=30$, $SD=119.60$), $t(24)=2.43$. Listeners responded faster to the Chinese-accented voice over the course of 24 sentences.

A one-way ANOVA on block 4 RTs tested whether the accent group was faster than the control groups. A significant effect of group was found, $F(2,81)=5.82$, and planned contrasts showed the accent group ($M=20$, $SD=109.60$) was significantly faster than the control in clear group ($M=118$, $SD=105.39$), $t(81)=3.39$, but was not significantly faster than the control in noise group ($M=60$, $SD=111.86$), $t(81)=1.39$, *ns*. The predicted advantage for the accent group in block 4 was borne out in the comparison with the control in clear group but, unexpectedly, not in the comparison with the control in noise group.

One difference between this experiment and the previous experiments was that six sentences were included in block 4, rather than four. Given that the accent groups' RTs decreased significantly after only a few sentences, specifically from block 1 to block 2, in both experiment 1 and the current experiment, the control groups in the current experiment may have adapted to the Chinese-accented speech within block 4. If so, averaging across more sentences in block 4 would attenuate the difference between the control and accent conditions. To investigate this possibility, the mean RTs for block 4 were recomputed based on only the first four sentences, making them equivalent to the means in the first two experiments. A one-way ANOVA on the four-trial means for block 4 revealed an effect of group, $F(2,81)=4.29$. Planned contrasts again showed significantly faster times for the accent group ($M=19$, $SD=129.02$) than for the control in clear group ($M=118$, $SD=122.85$), $t(81)=2.87$. In addition, the accent group was now marginally faster than the control in noise group ($M=86$, $SD=135.54$), $t(81)=1.94$, $p=0.056$.

Inspection of the mean RTs for each sentence in block 4 provides further insight into the situation. As shown in the bottom graph of Fig. 4, RTs for the accent and control in clear groups are relatively constant throughout the block, but the control in noise group seems to have adapted to the Chinese-accented voice within two or three sentences. Consistent with this hypothesis, a one-way ANOVA for means based on the first two sentences of block 4 again showed a significant effect of group, $F(2,80)=3.99$, and planned contrasts now showed that the accent group ($M=15$, $SD=142.95$) was significantly faster than both the control in clear group ($M=132$, $SD=207.43$), $t(80)=2.50$, and the control in noise group ($M=125$, $SD=164.94$), $t(80)=2.38$. When only the first two sentences of block 4 were considered, the accent group was significantly faster than both control groups.

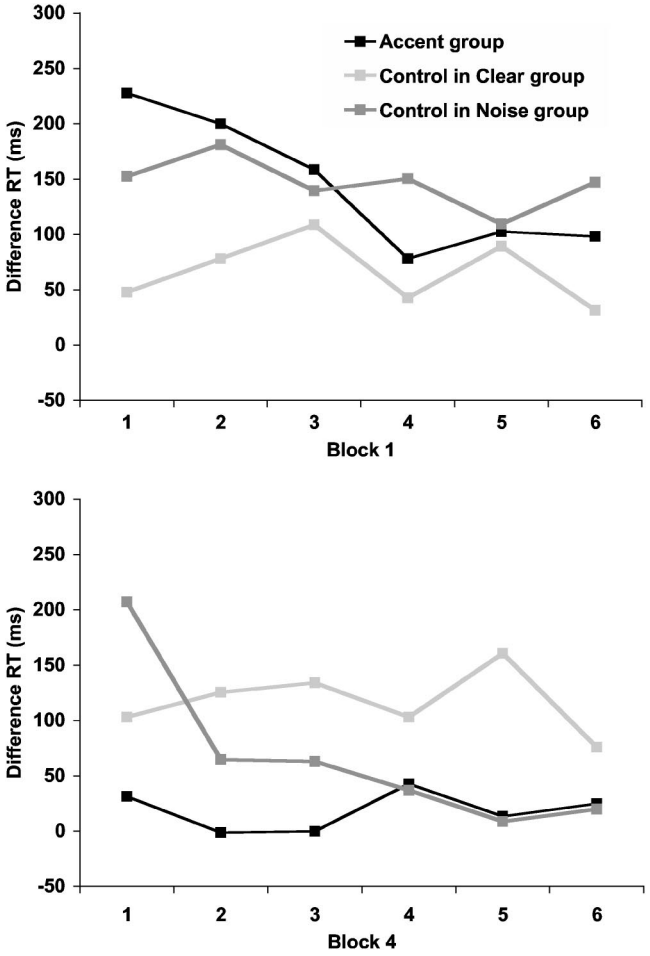


FIG. 4. Single-trial means of difference reaction times (experimental −baseline) in blocks 1 (top) and 4 (bottom) of experiment 3.

If adaptation could occur for the control in noise group within only a few sentences of exposure to the accented voice, the same pattern might also be seen for the accent group in the first block, when they were first exposed to the accented voice. Examination of the mean RTs for each sentence in block 1, shown in the top graph of Fig. 4, indeed suggests a similar pattern. The accent group appears to improve dramatically within the first three or four sentences.

Experiment 3 indicates that the adaptation effect found in experiments 1 and 2 for Spanish-accented speech also occurs during perception of a Chinese accent. The population sampled in this experiment have generally less opportunity to hear Chinese-accented speech in the ambient environment, in contrast to Spanish-accented speech, and few (24%) reported having a close friend or family member with a Chinese accent. To determine whether the adaptation effect in experiment 3 occurred for those listeners least familiar with the accent, a reanalysis was performed including only participants who reported either never having known anyone with a Chinese accent or only having Chinese-accented acquaintances with whom they spent little time (accent: $n=21$; control in clear: $n=21$; control in noise: $n=22$; see

TABLE III. Mean difference reaction times (in ms; and standard deviations) in block 4 based on two-trial means. [n=number of participants. Means in the same row with different subscripts differ significantly (see text for type of analysis for each experiment).]

		Block 4 trials		
Condition	n	1 & 2	3 & 4	5 & 6
Accent				
Experiment 1	16	−17 (59.27)	47 (131.23)	
Experiment 2	15	78 (154.34)	21 (197.55)	
Experiment 3	28	15 (142.95)	25 (148.76)	19 (131.42)
Control in clear				
Experiment 1	16	114 (142.70)	118 (165.34)	
Experiment 3	27	132 (207.43)	130 (171.28)	137 (169.39)
Control in noise				
Experiment 2	16	203_a (184.23)	90_b (120.29)	
Experiment 3	28	125_a (164.94)	48_b (152.88)	10_b (124.69)

Table II). The same pattern of effects was found: the accent group sped up across the four blocks, $F(2.89,49.08)=8.78$ (with Huynh–Feldt correction for nonsphericity), the control in clear group was slower than the accent group in block 4, $t(61)=3.22$, and the control in noise group was slower than the accent group when the first two trials of block 4 were considered, $t(61)=2.03$.

The analysis of only the less experienced listeners should be interpreted somewhat cautiously because the block order counterbalancing was not maintained when the more experienced participants were excluded. Nevertheless, it indicates that the effect was not specific to the experienced listeners alone. The findings with Chinese-accented speech are consistent with the idea that processing improves due to on-line perceptual learning of phonological patterns, rather than a precompiled phonological template. However, Chinese accents are not entirely novel to this participant population, so this possibility cannot be completely ruled out.

The control in noise group's rapid RT decrease within block 4 indicates that adaptation to accented speech may occur even more quickly than the initial results of this study suggested. To test this more sharply, *post hoc* analyses were done for block 4 RT data from all three experiments using two-trial means (see Table III). For experiment 1, *t*-tests showed that there was no change in RT from the first two trials to the second two trials of block 4 for either the accent or control (in clear) conditions, $t(15)=1.76_{\text{accent}}$ and 0.09_{control}, ns. For experiment 2, RTs for the accent group did not change, $t(14)=1.32$, ns, but the control in noise group got significantly faster from the first two to the second two trials of block 4, $t(15)=2.63$. In experiment 3, ANOVAs showed that neither the accent group's nor the control in clear group's RTs changed during block 4, F's<1. However, the control in noise group's RTs decreased significantly, $F(2,54)=7.18$. Planned contrasts indicated a significant decrease by the second two trials, $t(27)=2.51$.

The results are generally consistent. For the accent groups in all three experiments, RT did not change during block 4, as expected given that listeners in this condition had already adapted to the accented voice. For the two control in clear conditions (experiments 1 and 3), RT did not change during block 4, but for the two control in noise conditions (experiments 2 and 3), RT significantly decreased during block 4. It is not clear why the control in clear groups did not adapt to the accented voice within the fourth block as the control in noise groups did. One possibility relates to task difficulty. Because the first three blocks were likely easier for the control in clear groups, they may have been putting less effort into the task. When the accented voice was introduced, they may have been caught off guard by the new difficulty. In contrast, listeners in the control in noise conditions may have already been making substantial effort and were readier to deal with the new perceptual challenge. Although this interpretation is *post hoc*, it suggests a role for attention in the process of adaptation to accented speech and is consistent with evidence that normalization to native voice characteristics consumes processing capacity (e.g., Sommers *et al.*, 1994; for a review, see Nusbaum and Magnuson, 1997). On a methodological note, this finding also shows that measures of processing speed over time must be fine-grained enough to detect very rapid changes.

V. GENERAL DISCUSSION

This study demonstrated that listeners adapt very quickly to accented speech. Initial processing speed is slower for accented speech, but in all three experiments this deficit attenuated with less than one minute of experience. In some circumstances, adaptation required exposure to only two to four sentence-length utterances. The inclusion of control groups discounted the possibility that the effect was purely due to practice with the task or general strategies for handling difficult speech. Further, the extension of the adaptation effect to a less familiar accent is consistent with the hypothesis that the listeners learned the characteristics of the accented speech on-line.

One limitation of this study is that, in the control conditions, there was a change in voice as well as a change in accent, whereas the accent conditions had the same voice throughout. This confound is difficult to avoid. Ideally the same speaker would produce both the native and accented samples, but it is questionable whether one speaker can produce both ''dialects'' authentically. It is possible that a change in voice alone, apart from a change in accent, could induce a normalization process that could account for the slower RTs for the control groups. The only estimate of the magnitude of RT effects due to voice change using natural speech we are aware of is Mullennix *et al.*'s (1989) naming study. Mean naming latencies were 34 and 70 ms slower in mixed-talker conditions than in single-talker conditions. However, the usefulness of this estimate is questionable because the stimuli included male and female voices, and naming times are not directly comparable to the RT data of the present study. We see it as unlikely that a change from one female voice to another would cause the approximately 100-ms slow-down found in this study. Nevertheless, effects of this factor should be investigated in future work. The voice change could also cause a more global ''shock'' reaction, which might slow response times. Although listeners were warned that the voice would change at some point during the experiment, this explanation cannot be ruled out.

Another method for removing the confound might be to add a voice change to the accent condition by presenting a different voice with the same accent in block 4. This experiment, however, would test a subtly different question. A finding of generalization of learning from one voice to another requires not only that the listener adjusts phonological processing to match the characteristics of the current speaker, but that those adjustments are sufficiently abstract to be applied in a new context (taking voice as context). The question of generalization is very interesting, but we are currently satisfied to establish only the first requirement. That is, it is acceptable to us that the current data demonstrate only that the listeners adapted to the specific accented voice they heard. We suspect that adaptation to one voice would indeed generalize to a new voice. But adaptation to a single accented voice is no less impressive even if generalization is not found. The listener is still initially confronted with speech patterns that deviate from native norms and adjusts his or her perceptual processing to more effectively decode them.

The findings of rapid adaptation to foreign-accented speech provide new evidence for the type of normalization first shown by Ladefoged and Broadbent (1957), in which the acoustic-phonetic criteria for a vowel category were altered based on the characteristics of a preceding sentence. We believe that a similar kind of ''extrinsic'' normalization occurred in the present experiments (Nearey, 1989). Speech that deviated from native norms was evaluated more efficiently when recent experience provided information about the systematic ways in which it departed from those norms.

A known problem for the extrinsic normalization hypothesis is explaining how the initial segments or words are correctly evaluated when there is no previous input upon which to calibrate. The approach suggested by Nearey (1989) in his discussion of native speech is the same one offered here. There are usually enough cues within a segment or syllable that identification can be quite accurate even with no previous experience with the speaker (Shankweiler *et al.*, 1977). This is probably largely true even for accented speech. For example, within-category deviations from native phoneme prototypes can be identified and noted without any additional reference information. In addition, there are likely certain kinds of information that can be extracted from the speech signal at a low level, such as general vocal tract posture and speech rhythm. Once some of these properties of the speech are learned, they themselves can improve processing efficiency as well as bootstrap the learning of more complex phonological patterns. Finally, in most real world situations, higher level knowledge of the lexical, semantic, syntactic, and situational context can contribute to the perceptual learning of accented speech.

Rapid adaptation to foreign-accented speech is a clear demonstration of the remarkable flexibility of spoken word recognition. Traditional models of spoken word recognition do not address how this flexibility is achieved. Most assume phonetic context effects can be predicted and hard-wired, and that interspeaker variability can be solved at an early stage with intrinsic or extrinsic vocal-tract normalization. The models then typically focus on an architecture based on abstract linguistic units [e.g., PARSYN (Luce *et al.*, 2000), Shortlist (Norris, 1994), TRACE (McClelland and Elman, 1986)].

However, interspeaker variability includes more than differences in vowel space dimensions. In particular, it extends to higher levels of phonological representation, including phonetic context rules, syllable structure, and prosodic patterns. This is clearly true for accented speech, but also applies to idiolect and dialect differences among native speakers (Klatt, 1988). Therefore, the traditional concept of speaker normalization should be expanded to include interspeaker variability in complex phonological regularities, in addition to simple acoustic properties, and must be integrated as a critical and foundational aspect of spoken word recognition.

Recent attention to perceptual learning in speech perception (e.g., Norris *et al.*, 2003; Pisoni, 1997) suggests some promising lines for theory development. Theories that account for the adaptability of spoken word recognition have potentially strong affinities with enriched conceptions of basic human learning capacities. A striking challenge in the integration of learning theories and language processing performance will be the need to accommodate such theory to the rapidity with which effective change occurs. Further research is needed to explore the characteristics of this remarkable adaptability and uncover its boundary conditions. What are the consequences of rapid adaptation for long-term linguistic representations? Are the various aspects of speech learned with a single learning mechanism or a variety? How do the physical input and previous linguistic knowledge interact to drive adaptation? Answers to these and related questions will form the foundation for better accounts of speech perception as well as accounts of how the speech domain relates to other cognitive capacities.

ACKNOWLEDGMENTS

This work was supported by an NSF Graduate Research Fellowship to CMC and by the Cognitive Science Program at the University of Arizona. We are grateful to Georgine Speranzo for assisting with data collection. We also thank Ken Forster, Rebecca Gomez, Mike Hammond, Paul Luce, David Pisoni, and Natasha Warner for helpful discussion of the ideas presented here and Paul Luce for useful comments on this manuscript.

[1]To verify that the experimental groups did not differ systematically on the baseline measure, the baseline data in each of the three experiments were analyzed. No significant differences among groups were found.

[2]Four of the native English speaker's target words were not included in the earlier word identification test. These missing values were replaced by her mean for the purpose of a paired-items *t*-test comparing the percent correct identification for the two voices. The difference between voices was not statistically significant, $t(15) = 1.33$, *ns*.

[3]Duration was manipulated using the pitch-synchronous overlap and add (PSOLA) algorithm (Moulines and Charpentier, 1990) provided in the Praat wave-editing program (Paul Boersma and David Weenink, University of Amsterdam). This algorithm uniformly modifies the duration of a waveform with minimal change in its pitch or spectral characteristics. For voiced portions of the signal, each pitch period is multiplied by a bell-shaped window and adjacent windows are overlapped and added according to the compression factor. For voiceless portions, the windows are spaced equally. The technique is widely used and results in high quality speech.

[4]A paired-items *t*-test comparing the percent correct identification for the two voices was statistically significant, $t(23)=2.42$, $p<0.05$, indicating the nonaccented words were more intelligible. It is difficult to exactly equate a non-native speaker's intelligibility with a native speaker's. However, the Chinese-accented target words were still highly intelligible: 20 of the 24 words (83%) were identified correctly by over 90% of listeners in the screening test.

[5]Due to different recording circumstances, the new stimuli had a more tinny quality than the original sentence stimuli resulting from greater amplitude of the frequencies above approximately 3 kHz. In order to equate the sound quality, the new stimulus files were filtered using the Cool Edit 2000 software package. Amplitudes of the frequencies above 3 kHz were linearly reduced from 100% at 3 kHz to approximately 5% at 9.5 kHz to 0% at 10.5 kHz using a passive filter with an FFT size of 8192 and a Blackman windowing function. This produced a similar sound quality to the original recordings as judged by the first author.

[6]Although the stimuli in the accent and control in clear conditions differed by 5 dB, the fact that the accent condition was louder predicts a perceptual advantage for the accented voice, which is opposite to the experimental prediction for the first three blocks.

Bilger, R. C. (**1984**). *Manual for the Clinical Use of the Revised SPIN Test* (Univ. of Illinois, Champaign, IL).

Bradlow, A. R., and Bent, T. (**2003**). ''Listener adaptation to foreign-accented English,'' in *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona, Spain, 2003*, edited by M. J. Sole, D. Recasens, and J. Romero, pp. 2881–2884.

Clarke, C. M. (**2000**). ''Perceptual adjustment to foreign-accented English,'' J. Acoust. Soc. Am. **107**, 2856(A).

Duffy, S. A., and Pisoni, D. B. (**1992**). ''Comprehension of synthetic speech produced by rule: A review and theoretical interpretation,'' Lang. Speech **35**, 351–389.

Dupoux, E., and Green, K. P. (**1997**). ''Perceptual adjustment to highly compressed speech: Effects of talker and rate changes,'' J. Exp. Psychol. Hum. Percept. Perform. **23**, 914–927.

Forster, K. I., and Forster, J. C. (**2003**). ''DMDX: A Windows display program with millisecond accuracy,'' Behav. Res. Methods Instrum. Comput. **35**, 116–124.

Gass, S., and Varonis, E. (**1984**). ''The effect of familiarity on the comprehensibility of nonnative speech,'' Lang. Learn. **34**, 65–89.

Greenspan, S. L., Nusbaum, H. C., and Pisoni, D. B. (**1988**). ''Perceptual learning of synthetic speech produced by rule,'' J. Exp. Psychol. Learn. Mem. Cogn. **14**, 421–433.

Joos, M. (**1948**). ''Acoustic phonetics,'' Lang. Suppl. **24**(2), 1–136.

Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). ''Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability,'' J. Acoust. Soc. Am. **61**, 1337–1351.

Keppel, G. (**1982**). *Design and Analysis: A Researcher's Handbook* (Prentice–Hall, London), Chap. 8, pp. 144–168.

Klatt, D. H. (**1988**). ''Review of selected models of speech perception,'' in *Lexical Representation and Process*, edited by W. D. Marslen-Wilson (MIT, Cambridge), pp. 201–262.

Kucera, F., and Francis, W. (**1967**). *Computational Analysis of Present-Day American English* (Brown U.P., Providence, RI).

Ladefoged, P., and Broadbent, D. E. (**1957**). ''Information conveyed by vowels,'' J. Acoust. Soc. Am. **29**, 98–104.

Lane, H. (**1963**). ''Foreign accent and speech distortion,'' J. Acoust. Soc. Am. **35**, 451–453.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (**1967**). ''Perception of the speech code,'' Psychol. Rev. **74**, 431–461.

Luce, P. A., Goldinger, S. D., Auer, E. T., and Vitevitch, M. S. (**2000**). ''Phonetic priming, neighborhood activation and PARSYN,'' Percept. Psychophys. **62**, 615–625.

Luce, P. A., and McLennan, C. T. (**in press**). ''Spoken word recognition: The challenge of variation,'' in *The Handbook of Speech Perception*, edited by D. B. Pisoni and R. E. Remez (Blackwell, Malden, MA).

McClelland, J. L., and Elman, J. L. (**1986**). ''The TRACE model of speech perception,'' Cognit Psychol. **18**, 1–86.

McGarr, N. S. (**1983**). ''The intelligibility of deaf speech to experienced and inexperienced listeners,'' J. Speech Hear. Res. **26**, 451–458.

Moulines, E., and Charpentier, F. (**1990**). ''Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones,'' Speech Commun. **9**, 453–467.

Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (**1989**). ''Some effects of talker variability on spoken word recognition,'' J. Acoust. Soc. Am. **85**, 365–378.

Munro, M. J., and Derwing, T. M. (**1995a**). ''Foreign accent, comprehensibility, and intelligibility in the speech of second language learners,'' Lang. Learn. **45**, 73–97.

Munro, M. J., and Derwing, T. M. (**1995b**). ''Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech,'' Lang. Speech **38**, 289–306.

Nearey, T. M. (**1989**). ''Static, dynamic, and relational properties in vowel perception,'' J. Acoust. Soc. Am. **85**, 2088–2113.

Norris, D. (**1994**). ''Shortlist: a connectionist model of continuous speech recognition,'' Cognition **52**, 189–234.

Norris, D., McQueen, J. M., and Cutler, A. (**2003**). ''Perceptual learning in speech,'' Cognit. Psychol. **47**, 204–238.

Nusbaum, H. C., and Magnuson, J. (**1997**). ''Talker normalization: Phonetic constancy as a cognitive process,'' in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego, CA), pp. 109–132.

Nygaard, L. C., and Pisoni, D. B. (**1995**). ''Speech perception: New directions in research and theory,'' in *Handbook of Perception and Cognition: Vol. 11. Speech, Language, and Communication, 2nd ed.*, edited by J. L. Miller and P. D. Eimas (Academic, San Diego, CA), pp. 63–96.

Nygaard, L. C., and Pisoni, D. B. (**1998**). ''Talker-specific learning in speech perception,'' Percept. Psychophys. **60**, 355–376.

Nygaard, L. C., Sommers, M. S., and Pisoni, D. B. (**1994**). ''Speech perception as a talker-contingent process,'' Psych. Sci. **5**, 42–46.

Palmeri, T. J., Goldinger, S. D., and Pisoni, D. B. (**1993**). ''Episodic encoding of voice attributes and recognition memory for spoken words,'' J. Exp. Psychol. Learn. Mem. Cogn. **19**, 309–328.

Pisoni, D. B. (**1997**). ''Some thoughts on 'normalization' in speech perception,'' in *Talker Variability in Speech Processing*, edited by K. Johnson and J. W. Mullennix (Academic, San Diego, CA), pp. 9–32.

Pollatsek, A., and Well, A. D. (**1995**). ''On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis,'' J. Exp. Psychol. Learn. Mem. Cogn. **21**, 785–794.

Remez, R. E., Fellowes, J. M., and Rubin, P. E. (**1997**). ''Talker identification based on phonetic information,'' J. Exp. Psychol. Hum. Percept. Perform. **23**, 651–666.

Schmid, P. M., and Yeni-Komshian, G. H. (**1999**). ''The effects of speaker accent and target predictability on perception of mispronunciations,'' J. Speech Lang. Hear. Res. **42**, 56–64.

Shankweiler, D., Strange, W., and Verbrugge, R. (**1977**). ''Speech and the problem of perceptual constancy,'' in *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*, edited by R. Shaw and J. Bransford (Erlbaum, Hillsdale, NJ), pp. 315–345.

Sommers, M. S., Nygaard, L. C., and Pisoni, D. B. (**1994**). ''Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude,'' J. Acoust. Soc. Am. **96**, 1314–1324.

Studebaker, G. A. (**1985**). ''A 'rationalized' arcsine transform,'' J. Speech Hear. Res. **28**, 455–462.

van Wijngaarden, S. J. (**2001**). ''Intelligibility of native and non-native Dutch speech,'' Speech Commun. **35**, 103–113.

Weil, S. A. (**2001**). ''Foreign accented speech: Encoding and generalization,'' J. Acoust. Soc. Am. **109**, 2473(A).

Wingstedt, M., and Schulman, R. (**1987**). ''Comprehension of foreign accents,'' in *Phonologica 1984*, edited by W. Dressler, H. Luschutzky, O. Pfeiffer, and J. Rennison (Cambridge U.P., Cambridge), pp. 339–345.

Enhancing Chinese tone recognition by manipulating amplitude envelope: Implications for cochlear implants

Xin Luo[a)]
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057, and Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, Anhui 230027, People's Republic of China

Qian-Jie Fu
Department of Auditory Implants and Perception, House Ear Institute, 2100 West Third Street, Los Angeles, California 90057

(Received 20 June 2003; revised 1 June 2004; accepted 18 June 2004)

Tone recognition is important for speech understanding in tonal languages such as Mandarin Chinese. Cochlear implant patients are able to perceive some tonal information by using temporal cues such as periodicity-related amplitude fluctuations and similarities between the fundamental frequency ($F0$) contour and the amplitude envelope. The present study investigates whether modifying the amplitude envelope to better resemble the $F0$ contour can further improve tone recognition in multichannel cochlear implants. Chinese tone and vowel recognition were measured for six native Chinese normal-hearing subjects listening to a simulation of a four-channel cochlear implant speech processor with and without amplitude envelope enhancement. Two algorithms were proposed to modify the amplitude envelope to more closely resemble the $F0$ contour. In the first algorithm, the amplitude envelope as well as the modulation depth of periodicity fluctuations was adjusted for each spectral channel. In the second algorithm, the overall amplitude envelope was adjusted before multichannel speech processing, thus reducing any local distortions to the speech spectral envelope. The results showed that both algorithms significantly improved Chinese tone recognition. By adjusting the overall amplitude envelope to match the $F0$ contour before multichannel processing, vowel recognition was better preserved and less speech-processing computation was required. The results suggest that modifying the amplitude envelope to more closely resemble the $F0$ contour may be a useful approach toward improving Chinese-speaking cochlear implant patients' tone recognition.
[DOI: 10.1121/1.1783352]

PACS numbers: 43.71.Es, 43.71.Ky, 43.66.Ts [PFA] Pages: 3659–3667

I. INTRODUCTION

In tonal languages such as Mandarin Chinese, the tonality of a syllable is lexically important; syllables containing the same combination of vowels and consonants but different tones have completely different meanings (e.g., Liang, 1963; Lin, 1988; Wang, 1989). There are four lexical tones in Mandarin Chinese, characterized by different patterns of fundamental frequency ($F0$) variation during voiced speech (e.g., Howie, 1976; Wang, 1989). Tone 1 has a flat $F0$ pattern, tone 2 has a rising $F0$ pattern, tone 3 has a falling–rising $F0$ pattern, and tone 4 has a falling $F0$ pattern.

Although the $F0$ contour is the principal feature for tone recognition (e.g., Liang, 1963; Lin, 1988), other temporal cues that covary with the tonal patterns also contribute to tone recognition. These temporal cues include vowel duration, amplitude envelope, and periodicity fluctuations (Liang, 1963; Whalen and Xu, 1992; Fu *et al.*, 1998; Fu and Zeng, 2000). Vowel duration and amplitude envelope cues preserve tonal information effectively when the $F0$ component and its harmonic structure are partially or totally removed. Whalen and Xu (1992) used signal-correlated-noise stimuli to eliminate $F0$ information and other spectral cues and found that, using only vowel duration and amplitude envelope cues, normal-hearing (NH) listeners could correctly identify about 70% of Mandarin Chinese tones. Fu *et al.* (1998) similarly found 80%-correct tone recognition, but only 39%-correct vowel recognition by NH listeners listening to the output of a single-channel noise vocoder in which only the temporal envelope information below 500 Hz was preserved (for both tests, chance performance level was 25% correct). This high level of tone recognition in the absence of spectral cues also provided some Chinese sentence recognition (about 11% correct).

Multichannel cochlear implant (CI) speech processors generally divide acoustic signals into several frequency bands, extract the temporal envelope information from each band, convert the temporal envelope amplitudes into electric currents, and deliver the electric currents to appropriate electrodes situated within the cochlea (e.g., Wilson *et al.*, 1991; Skinner *et al.*, 1994). To recreate the tonotopic distribution of activity within the normal cochlea, the amplitude envelopes from low-frequency bands are delivered to the electrodes located in the apical region, and the amplitude envelopes from high-frequency bands are delivered to the electrodes located in the basal region. While this frequency-

a)Electronic mail: xluo@hei.org

to-electrode mapping provides gross spectral cues for speech understanding, many other cues are not well transmitted, including $F0$ and the temporal fine structure above 500 Hz within each spectral channel.

Several studies have simulated the effects of various CI stimulation parameters on Chinese tone recognition using a noise-band speech processor—an acoustic simulation of the speech-processing strategy in cochlear implants. Fu *et al.* (1998) found that when the number of frequency bands was increased from one to four, Chinese vowel, consonant, and sentence recognition improved dramatically, similar to the improvement observed in English speech recognition (Shannon *et al.*, 1995). However, Chinese tone recognition remained nearly constant, independent of the number of bands (up to the four bands that were tested in the study). Conversely, improving the temporal resolution by changing the cutoff frequency of the envelope filter from 50 to 500 Hz had a significant effect on Chinese tone and sentence recognition, but no effect on vowel and consonant recognition. Xu *et al.* (2002) extended these observations to include a greater range of spectral (1 to 12 frequency bands) and temporal resolution (envelope filter cutoff frequencies from 1 to 512 Hz). Results showed that while tone recognition somewhat depended on the number of bands (up to 12), the dependence was not as strong as that for vowel and consonant recognition; tone recognition significantly improved with increased envelope filter cutoff frequencies (up to 512 Hz), consistent with previous studies (Fu *et al.*, 1998).

Chinese tone recognition has also been directly measured in Chinese-speaking CI patients with different speech-processing strategies (e.g., Liu *et al.*, 1997; Hsu *et al.*, 2000; Fu *et al.*, 2003). Speech-processing strategies in cochlear implants can be roughly categorized into two types: feature-extraction (e.g., $F0/F1/F2$, MPEAK) and filter-bank approaches (e.g., SPEAK, CIS). In the multipeak (MPEAK) strategy (Skinner *et al.*, 1991), important speech features are extracted from the acoustic signal, including the voice $F0$ (which determines the stimulation rate), the amplitude and frequency of peak sound energy in the region of the first ($F1$) and second formants ($F2$), and the amplitude in three high-frequency bands; the extracted speech information is then delivered to the appropriate electrodes. In the spectral peak (SPEAK) strategy (Skinner *et al.*, 1994), the acoustic input signal is filtered into 20 frequency bands; the temporal envelopes extracted from the 6–10 bands containing the highest amplitudes are used to modulate pulse trains delivered to the appropriate electrodes. Although MPEAK explicitly encodes changes in $F0$ by varying the stimulation rate while SPEAK does not, studies have shown that both processing strategies transmitted similar amounts of $F0$ information to Chinese-speaking CI patients (Liu *et al.*, 1997). Compared with MPEAK, the relatively higher stimulation rate of SPEAK (250 pulses per second per electrode) might be beneficial for transmission of temporal envelopes, and therefore benefit transmission of tonal information. Recently, Fu *et al.* (2003) measured tone recognition in nine Chinese-speaking CI patients (all Nucleus-24 users with a minimum of 3 years experience with the device), as a function of different speech-processing strategies: SPEAK, advanced combination encoding (ACE), and continuous interleaved sampling (CIS; Wilson *et al.*, 1991). The results showed that, with the latest-generation speech-processing strategies, CI patients were able to achieve moderate levels of tone recognition, and that speech processors which employed relatively high stimulation rates and increased numbers of channels provided better tone recognition. However, even with higher stimulation rates and a greater number of channels, CI users' tone recognition performance was comparable to that of NH subjects listening to four-channel noise-band speech. A speech-processing strategy specifically designed to improve tone recognition may be necessary to enhance sentence recognition by Chinese-speaking CI users, as Chinese sentence recognition is highly correlated with tone recognition (Fu *et al.*, 1998, 2003).

The results of these experiments with NH and CI listeners suggest that, while increased spectral cues did not necessarily improve tone recognition, temporal envelope cues are important for tone recognition. In attempt to enhance CI users' pitch perception, several researchers have tried manipulating the periodicity fluctuations in a signal's temporal envelope. Geurts and Wouters (2001) reported that larger modulation depths produced better single-electrode pitch discrimination. However, when the larger modulation depths were applied to multiple channels stimulated with synchronous $F0$ patterns, pitch discrimination did not significantly improve. These results indicate that the contribution of periodicity cues to tone recognition may be limited for multichannel cochlear implants.

The correlation between the amplitude envelope and the $F0$ contour of Chinese syllables has been shown in several studies (e.g., Garding *et al.*, 1986; Sagart, 1986; Whalen and Xu, 1992). Fu and Zeng (2000) found that tone recognition with only temporal envelope cues was highly variable across speakers and vowel contexts; acoustic analysis of temporal envelope cues revealed that this variability was directly related to the degree of similarity between the amplitude envelope and the $F0$ contour. Based on these observations, the present study investigated whether modifying the amplitude envelope to better resemble the $F0$ contour can further improve tone recognition in multichannel cochlear implants. Two algorithms were proposed. The first algorithm adjusted the amplitude envelope as well as the modulation depth of periodicity fluctuations for each spectral channel. The modulation depth was increased so that the effects of stronger periodicity fluctuations on tone recognition could be investigated and compared with those of amplitude envelope modification. The second algorithm adjusted the overall amplitude envelope before multichannel speech processing in order to reduce the changes to the signal's spectral envelope caused by modifying individual channels' amplitude envelope. Both algorithms were evaluated by measuring Chinese tone and vowel recognition scores for six Chinese-speaking NH subjects listening to four-channel noise-band speech with and without amplitude modification.

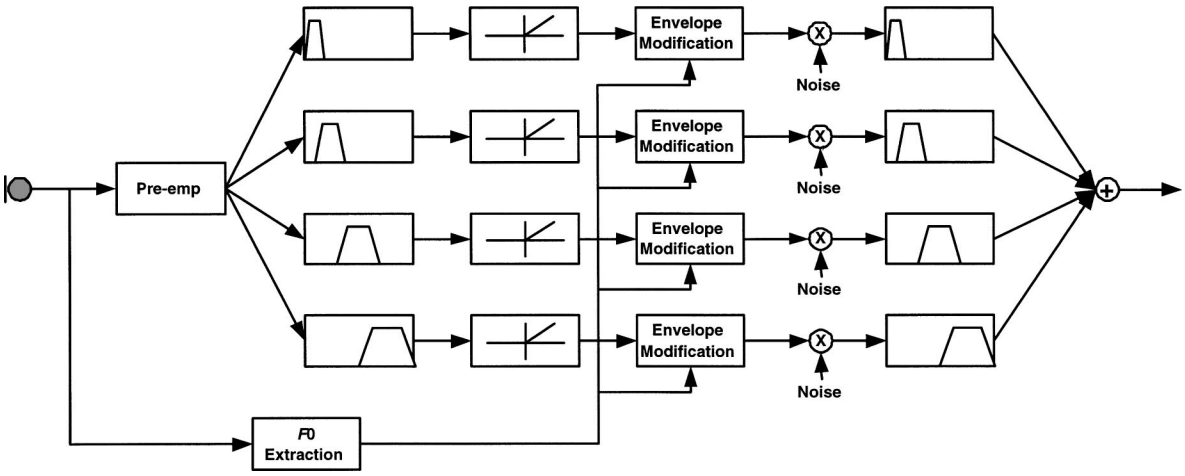


FIG. 1. Block diagram of the individual channel envelope enhancement speech processor.

II. METHODS

A. Subjects

Six young adult native Chinese-speaking listeners (three male and three female) participated in this study and were paid for their services. All subjects were normal hearing and had pure-tone thresholds better than 20 dB HL at octave frequencies from 125 to 8000 Hz in both ears.

B. Stimuli and speech processing

The Chinese vowel stimuli were derived from the Chinese Standard Database (Wang, 1993). Five male and five female speakers each produced four tones for six Mandarin Chinese single-vowel syllables (/a/, /o/, /e/, /i/, /u/, /ü/), resulting in a total of 240 isolated vowel tokens; in Chinese language these isolated vowels are lexically meaningful. These stimuli were digitized using a 16-bit A/D converter at a 16-kHz sampling rate without high frequency preemphasis.

All speech stimuli were processed using a four-channel noise-band acoustic simulation of a cochlear implant fitted with the CIS strategy (Wilson *et al.*, 1991). After preemphasis (first-order Butterworth high-pass filter at 1200 Hz), the input speech signal was divided into four frequency bands (overall frequency range was between 100 and 6000 Hz). The corner frequencies of the analysis bands were determined according to Greenwood's (1990) formula; all analysis filters were fourth-order Butterworth bandpass filters. The temporal envelope from each analysis band was extracted by half-wave rectification and low-pass filtering (fourth-order Butterworth low-pass filter at 500 Hz), and was used to modulate wideband noise, which was then bandpass filtered by filters with the same passbands as the analysis filters. The output speech was the sum of these modulated noise bands.

1. Algorithm 1: Adjusting the temporal envelope of individual channels

The first envelope enhancement algorithm involved modifying the temporal envelope of each bandpass filter output with an amplitude envelope related to the $F0$ contour of the speech signal, as well as increasing the modulation depth of the periodicity fluctuations in the temporal envelope. Figure 1 shows a block diagram of the first envelope enhancement speech processor.

To obtain the modification target of the amplitude envelope, the $F0$ contour of the speech signal was extracted before multichannel processing using an autocorrelation method on a frame-by-frame basis. The $F0$ extraction method used in this study was based on the algorithm proposed by Markel (1972), with some simplification. Figure 2 shows a block diagram detailing the $F0$ extraction. To remove the influence of formant frequencies on $F0$ extraction, 12th-order linear predictive (LP) analysis was performed for each frame using the Levinson–Durbin algorithm (Makhoul, 1975). The speech signal was inverse filtered to give the prediction residual, a signal with an approximately flat spectrum. The autocorrelation of the residual signal was calculated and the location of the autocorrelation peak within an appropriate range (2–20 ms) was chosen as the pitch period of the frame. The voiced/unvoiced (VUV) quality of the frame was chosen as voiced if the normalized level of the peak was beyond an empirical threshold (0.2 in this study); otherwise, it was chosen as unvoiced. The $F0$ analysis frame size was adaptive; if the normal 30-ms frame size was less than 3 times the previous pitch period estimate, the analysis frame size was increased to be 4 times the previous pitch period estimate. The analysis frame shift was fixed to be 10 ms.

Because changes in $F0$ over time tend to be within a limited range, several postprocessing methods were employed to fix errors such as VUV confusion, pitch doubling, and pitch halving. For example, if a voiced frame was found to be between two unvoiced frames (or vice versa), the VUV decision of the intermediate frame was reversed (Markel, 1972). Similarly, when pitch doubling or pitch halving occurred for two successive voiced frames, pitch values were corrected to be the one with higher voicing degree. Finally, a five-point median filter was used to smooth the extracted $F0$ contour (Rabiner *et al.*, 1975).

To extract the amplitude envelope and the periodicity fluctuations in each channel, the half-wave rectified bandpass signals were low-pass filtered at 50 and 500 Hz. The output

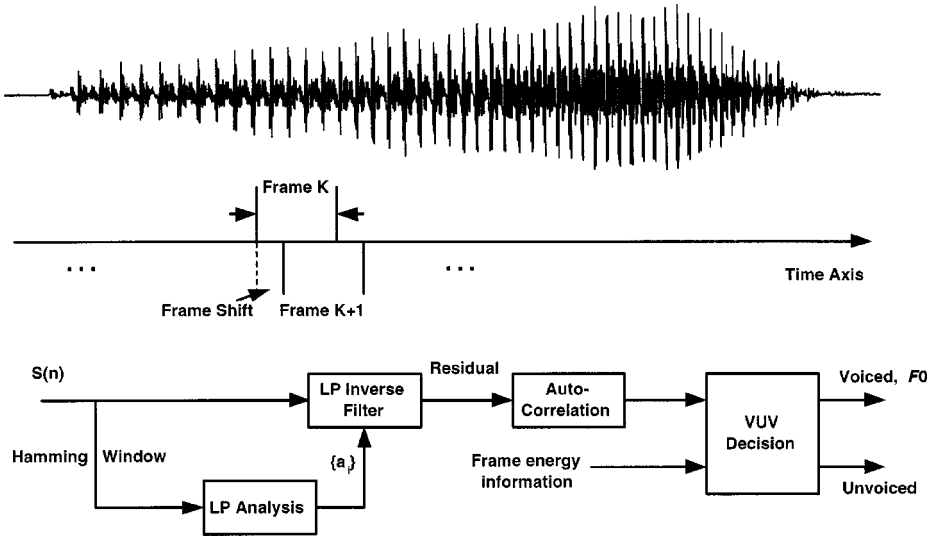


FIG. 2. Block diagram of the $F0$ extraction algorithm.

of the 50-Hz filter produced the low-frequency amplitude envelope; subtracting the output of the 50-Hz filter from the output of the 500-Hz filter produced the periodicity fluctuations corresponding to the human pitch range (50–500 Hz).

To make the shape of the amplitude envelope more similar to that of the $F0$ contour, the root-mean-square (rms) amplitudes first were calculated on a frame-by-frame basis to match the discrete format of $F0$ data. The F_{0}_rms amplitude sequence that was most similar to the $F0$ sequence in shape was selected as the modification target of the rms amplitude sequence, as shown in Eq. (1)

$$F_{0}\_\mathrm{rms}(i)=\left(\frac{\Sigma_{\mathrm{VUV}(k)==\mathrm{Voiced}}\mathrm{rms}(k)}{\Sigma_{\mathrm{VUV}(k)==\mathrm{Voiced}}F_0(k)}\right)F_0(i)$$
$$\mathrm{VUV}(i)==\mathrm{Voiced}, \qquad (1)$$

where k is the frame index of the summations and i is the frame index of the F_{0}_rms amplitude sequence and the $F0$ sequence. Note that amplitude envelope modifications were made only for voiced frames [$\mathrm{VUV}(i)==\mathrm{Voiced}$]. The rms amplitude sequence was then modified according to shape factor s, as shown in Eq. (2)

$$\mathrm{Mod\_rms}(i)=\mathrm{rms}(i)+s\times(F_{0}\_\mathrm{rms}(i)-\mathrm{rms}(i)), \qquad (2)$$

where Mod_rms is the modified rms amplitude sequence. When shape factor $s=0$, there was no modification to the amplitude envelope; $s=1$ produced the largest amplitude envelope modification, making the modified rms amplitude sequence directly proportional to the $F0$ sequence. Finally, modifications of individual amplitude envelope samples were calculated to be the linear interpolation of the frame-by-frame rms amplitude modifications. Figure 3 shows an example of a single channel's amplitude envelope modification. The top panel shows the original waveform of the Chinese vowel /a/ (tone 2), the middle panel shows the extracted $F0$ contour, and the bottom panel shows the original and modified low-frequency amplitude envelope for channel 1.

To enhance the extracted periodicity cues, the modulation depth of the periodicity fluctuations (Ped) were in-

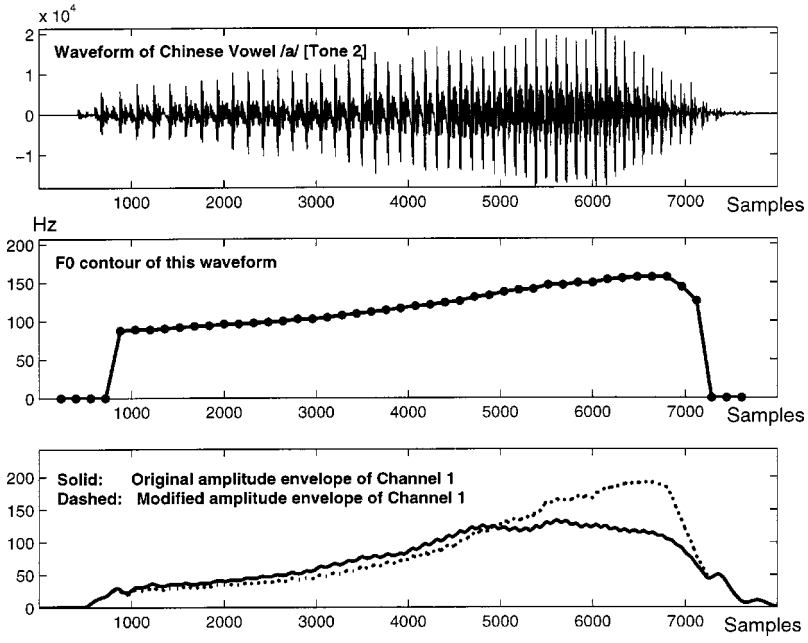


FIG. 3. Example of amplitude envelope modification for one channel of the envelope enhancement speech processor. The top panel shows the original waveform of a Chinese vowel /a/ (tone 2). The middle panel shows the extracted $F0$ contour. The bottom panel shows the original and modified amplitude envelope for channel 1.

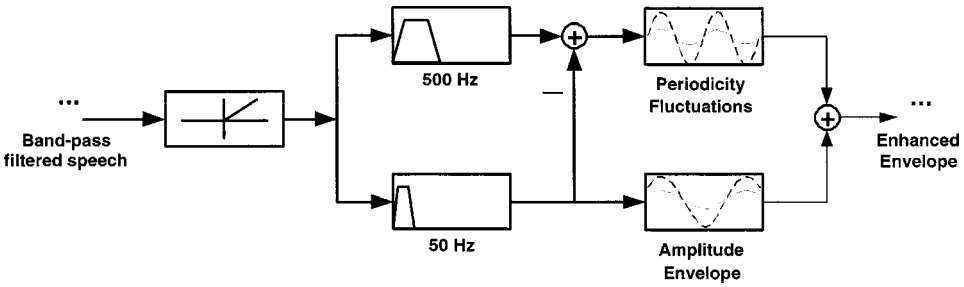


FIG. 4. Block diagram of envelope extraction and modification for one channel of the envelope enhancement speech processor.

creased according to depth factor d, as shown in Eq. (3)

$$\text{Mod\_Ped} = \text{Ped} \times d, \tag{3}$$

where Mod_Ped is the modified periodicity fluctuations. When depth factor $d=1$, there was no modification to the modulation depth; when $d>1$, there was an increase in the modulation depth.

The modified amplitude envelope and the modified periodicity fluctuations were added to construct each channel's enhanced temporal envelope. This enhanced temporal envelope was then used to modulate the appropriate noise carrier band. Figure 4 shows a block diagram of the extraction of the amplitude envelope and the periodicity fluctuations, as well as the combined modifications to the temporal envelope for one channel of the first envelope enhancement speech processor.

2. Algorithm 2: Adjusting the overall amplitude envelope before multichannel speech processing

In the first envelope enhancement algorithm, the amplitude envelopes of individual channels were modified separately. These ''local'' temporal envelope enhancements, when combined, may cause a change to the signal's overall spectral envelope that, in turn, could cause a change in vowel recognition performance. Also, modifying the amplitude envelopes of individual channels required a large amount of computation.

In the second envelope enhancement algorithm, the overall amplitude envelope was extracted from the speech signal and modified before bandpass filtering, thereby largely preserving the signal's spectral envelope. Figure 5 shows a block diagram of the second envelope enhancement speech processor.

The $F0$ extraction and the amplitude envelope modification were the same as described in the first enhancement algorithm. Using the modified overall amplitude envelope (Mod_Envelope), the envelope-enhanced speech (Mod_Speech) was calculated as shown in Eq. (4), and then replaced the original speech before multichannel speech processing

$$\text{Mod\_Speech} = \text{Speech} \times \left(\frac{\text{Mod\_Envelope}}{\text{Envelope}} \right). \tag{4}$$

Figure 6 shows an example of the modification to the overall amplitude envelope of the Chinese vowel /a/ (tone 2). The first panel shows the original waveform of the vowel and the second panel shows the extracted $F0$ contour of the vowel. The third panel shows the original and modified overall amplitude envelope and the fourth panel shows the envelope-enhanced speech waveform.

C. Procedure

Before formal testing began, subjects were asked to preview sample sets of processed speech to gain familiarity with

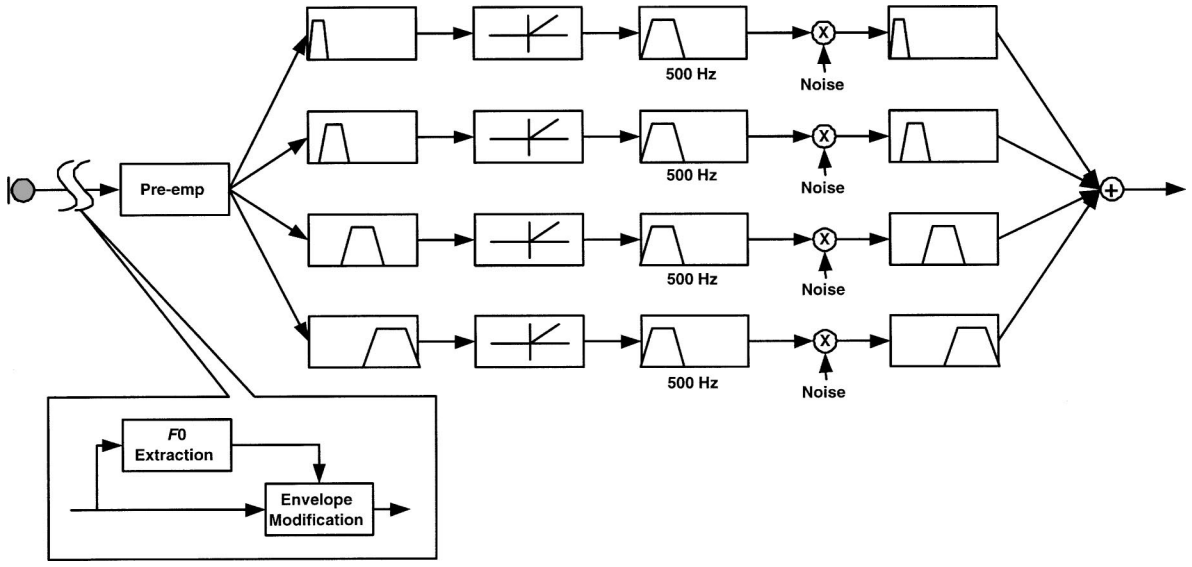


FIG. 5. Block diagram of the overall amplitude envelope enhancement before multichannel speech processing.

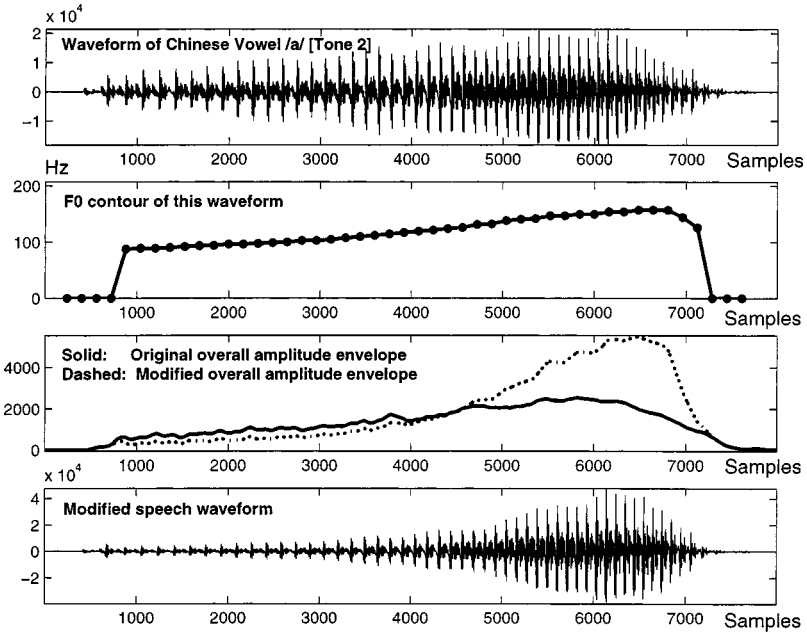


FIG. 6. Example of the modification to the overall amplitude envelope. The first panel shows the original waveform of the Chinese vowel /a/ (tone 2). The second panel shows the extracted $F0$ contour. The third panel shows the original and modified overall amplitude envelope. The fourth panel shows the envelope-enhanced speech waveform.

the noise-band speech processing. For each recognition task, the test order of speech-processing conditions was randomized and counterbalanced across subjects. Subjects were seated in a double-walled sound-treated booth and listened to the stimuli presented in free field over a single loudspeaker (Tannoy Reveal) at 65 dbA. Tone recognition tests were conducted using a four-alternative, forced-choice (4-AFC) task, while vowel recognition tests were conducted using a 6-AFC task; no feedback was provided.

For the first envelope enhancement algorithm, Chinese tone and vowel recognition were tested under different parametric combinations of the shape and depth factors. Shape factors between 0.0 (original amplitude envelope) and 1.0 (largest amplitude envelope modification) were tested, in steps of 0.2. Depth factors between 1.0 (original modulation depth) and 2.0 (largest modulation depth) were tested ($d=1.0,1.2,1.5,2.0$). Note that shape factor $s=0.0$ and depth factor $d=1.0$ were the speech-processing parameters used in the baseline measures.

For the second envelope enhancement algorithm, Chinese tone and vowel recognition were tested under only two processing conditions: shape factor $s=0.0$ (original amplitude envelope, which served as the baseline measurement) and shape factor $s=1.0$ (largest amplitude envelope modification).

III. RESULTS

A. Effects of modifying individual channels' amplitude envelope

Figure 7 shows the speech recognition scores obtained with modification to amplitude envelope of individual channels. The average tone recognition score increased from 67.08% to 77.43% correct as the shape factor was increased from 0.0 to 1.0 (the depth factor was fixed at 1.0). Tone recognition reached asymptotic performance when shape factor $s=0.6$. A one-way analysis of variance (ANOVA) showed that tone recognition was significantly affected by the shape factor [$F(5,30)=2.91$, $P=0.03$]. On the other hand, the average vowel recognition score decreased from 49.5% to 45.9% correct. However, a one-way ANOVA showed that vowel recognition was not significantly affected by the shape factor [$F(5,30)=0.14$, $P=0.98$], most likely due to the high variation across subjects. (These relatively lower tone and vowel recognition scores, compared with those reported by Fu *et al.*, 1998, were due to the more difficult testing materials. In the present study, ten speakers were used instead of one speaker, and a 6-AFC testing procedure was used for vowel recognition tests instead of 4-AFC.)

Figure 8 shows detailed tone recognition scores as a function of the shape factor. Modifying individual channels' amplitude envelope greatly improved recognition of tones 1, 2, and 4, but did not significantly affect tone 3. In the baseline condition (shape factor $s=0.0$, depth factor $d=1.0$),

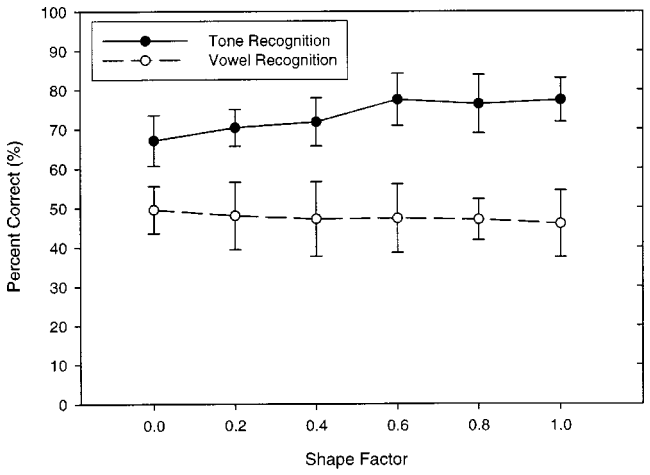


FIG. 7. Chinese tone recognition and vowel recognition as functions of the shape factor in the individual channel amplitude envelope modification; data were collected with fixed depth factor $d=1.0$, i.e., no modulation depth increase. The symbols represent the mean values, and the error bars represent the standard deviations.

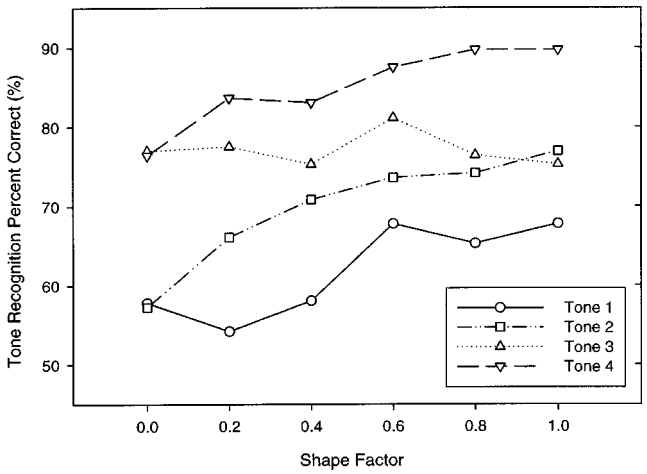


FIG. 8. Recognition performance for individual Chinese tones as a function of the shape factor in the individual channel amplitude envelope modification; data were collected with fixed depth factor $d=1.0$, i.e., no modulation depth increase. The symbols represent the mean values.

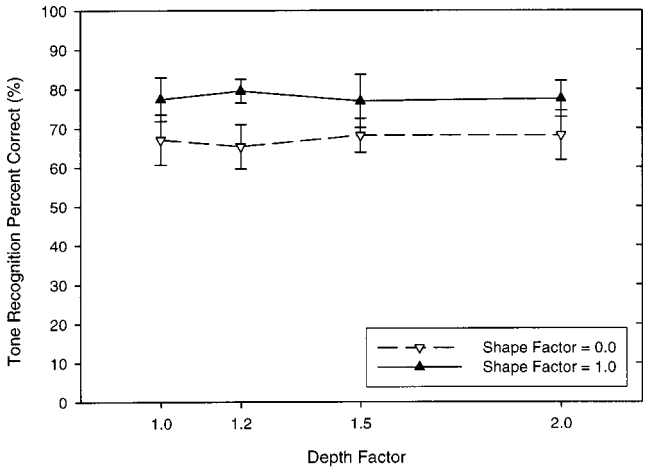


FIG. 9. Chinese tone recognition as a function of the depth factor in increasing the modulation depth for individual channels; data were collected when shape factor $s=0.0$ (▽) and 1.0 (▲). The symbols represent the mean values, and the error bars represent the standard deviations.

recognition scores of tones 3 and 4 were much higher than those of tones 1 and 2, consistent with previous studies (Fu *et al.*, 1998; Fu and Zeng, 2000). With the largest amplitude envelope modification (shape factor $s=1.0$), recognition scores of tones 1, 2, and 4 increased 10, 20, and 13 percentage points, respectively. However, one-way ANOVAs revealed that only tone 2 recognition was significantly enhanced by modifying amplitude envelope of individual channels [$F(5,30)=3.95$, $P=0.01$]. Tone 3 recognition was not improved by the enhancement algorithm, most likely because tone 3's amplitude envelope and $F0$ contour were already quite similar, even before processing. Also, tone 3 recognition is strongly linked to vowel duration cues (Fu and Zeng, 2000), which would not be affected by amplitude envelope modification.

The effects of modifying individual channels' amplitude envelope on tone recognition were somewhat speaker- and stimulus dependent. For example, tone recognition scores showed greater improvement for speakers 2 and 3 than for other speakers. Similarly, the enhancement of tone recognition was much higher for vowels /a/ and /o/ than for other vowels. In general, greater enhancements were found for those speakers and vowels for which tone recognition scores were relatively lower in the baseline condition.

B. Effects of increasing individual channels' modulation depth

In contrast to the effects of modifying individual channels' amplitude envelope, increasing each channel's modulation depth had little effect on tone recognition. Figure 9 shows the tone recognition scores obtained with increased modulation depth of individual channels. One-way ANOVAs showed that tone recognition was not affected by increasing the modulation depth of the periodicity cues, whether with the minimal or the maximal amplitude envelope enhancement (shape factor $s=0.0$: [$F(3,20)=0.33$, $P=0.80$]; shape factor $s=1.0$: [$F(3,20)=0.29$, $P=0.83$]). These results were similar to those of Geurts *et al.* (2001), who found no significant increase in pitch discrimination using larger modulation depths when multiple channels were stimulated with synchronous $F0$ fluctuations.

A two-way ANOVA comparing the interactive effects between the shape and depth factors showed no significant interaction [$F(3,40)=0.60$, $P=0.62$] and confirmed that the shape factor had a significant effect on tone recognition [$F(1,40)=45.51$, $P<0.001$], while the depth factor had no significant effect [$F(3,40)=0.02$, $P=0.99$]. The results suggest that amplitude envelope cues are more salient than periodicity fluctuation cues for Chinese tone recognition in multichannel cochlear implants.

C. Effects of modifying the overall amplitude envelope before multichannel speech processing

Figure 10 shows the speech recognition scores obtained when the overall amplitude envelope was adjusted to better match the $F0$ contour, before multichannel speech processing. The average tone recognition score increased from

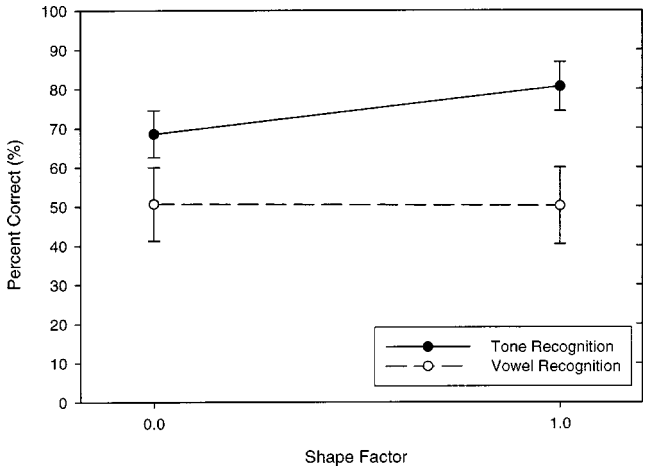


FIG. 10. Chinese tone recognition and vowel recognition as functions of the shape factor when the overall amplitude envelope was modified before multichannel speech processing. The symbols represent the mean values, and the error bars represent the standard deviations.

68.54% to 80.56% correct as the shape factor s was increased from 0.0 to 1.0, while the average vowel recognition score was unchanged. Student's t-tests showed that tone recognition was significantly affected by the shape factor [$t(10) = 3.41$, $p = 0.01$], while vowel recognition was not [$t(10) = 0.09$, $p = 0.93$]. For the largest degree of amplitude envelope modification (shape factor $s = 1.0$), modifying the overall amplitude envelope before multichannel speech processing provided better tone recognition than modifying individual channels' amplitude envelope; vowel recognition was not significantly affected by either enhancement algorithm. These results indicate that modifying the overall amplitude envelope prior to multichannel speech processing can greatly enhance tone recognition without adversely affecting vowel recognition.

The recognition scores for each Chinese tone showed that modifying the overall amplitude envelope prior to multichannel speech processing improved recognition of all four tones. For the largest degree of amplitude envelope modification (shape factor $s = 1.0$), recognition scores of tones 1, 2, 3, and 4 increased 11, 17, 11, and 9 percentage points, respectively. However, Student's t-tests revealed that only recognition of tones 2 and 3 was significantly improved (tone 2: [$t(10) = 3.74$, $p = 0.004$]; tone 3: [$t(10) = 3.65$, $p = 0.004$]). Similar to the first envelope enhancement algorithm, the effects of modifying the overall amplitude envelope on tone recognition were highly variable across different speakers and vowels.

IV. DISCUSSION

The present results demonstrate that, under conditions of reduced spectral resolution comparable to those experienced by cochlear implant listeners, Chinese tone recognition can be improved by manipulating the amplitude envelope. These results further confirm the importance of amplitude envelope cues in Chinese tone recognition when $F0$ information is not salient, consistent with previous studies (e.g., Fu *et al.*, 1998). Tone recognition was significantly improved when either the temporal envelope of individual channels or the overall amplitude envelope was modified. However, adjusting the overall amplitude envelope seemed to better preserve vowel recognition accuracy and required less speech-processing computation. Compared with individual channel amplitude envelope modification, overall amplitude envelope modification produced larger changes in the speech temporal envelope but smaller distortions in the speech spectral envelope, which might explain its better performance in tone and vowel recognition.

Interestingly, no improvement of tone recognition was observed when the modulation depth of the periodicity fluctuations was increased. There are two possible explanations for this lack of effect. One is that the contribution of periodicity fluctuations to Chinese tone recognition is relatively weak when the amplitude envelope is available. Another potential explanation is that even for unprocessed speech, the mean modulation depth is sufficient for listeners to detect the periodicity fluctuations. Therefore, increasing the modulation depth of periodicity fluctuations may not provide additional information for Chinese tone recognition.

The effects of amplitude envelope modification on tone recognition were also highly variable. The amount of tone recognition improvement was dependent on the speakers, vowels, as well as tones. The variability in improvement among the tones was likely due to differences in the degree of similarity between the amplitude envelope and the $F0$ contour in each tone. For example, the amplitude envelope and $F0$ contour of tone 3 are inherently similar, even before speech processing. Syllables produced by different speakers also have different degrees of similarity between the amplitude envelope and the $F0$ contour. With the original temporal envelope cues, tone recognition scores were lower for speakers 2 and 3, indicating that the amplitude envelope of their speech might be less similar to the $F0$ contour. After amplitude envelope modification, greater enhancements in tone recognition were generally found for these speakers because their speech had greater potential for increasing the similarity between the amplitude envelope and the $F0$ contour.

The amount of improvement achieved by enhancing the amplitude envelope may be affected by several other factors. One factor may be the accuracy of the extracted $F0$ contour. Errors in $F0$ extraction may produce distortions of the modified amplitude envelope. For example, if an error of pitch halving happens during the $F0$ extraction of a syllable in tone 2, the enhanced amplitude envelope will have an undesired drop in its general trend of rising. The SIFT algorithm to which our $F0$ extraction method referred had a moderate number of gross pitch estimation errors in a comparative performance study of several pitch detection algorithms (Rabiner *et al.*, 1976). A more accurate and robust $F0$ extraction algorithm may be necessary to avoid distortions in amplitude envelope modification and improve the enhancement algorithms. Another factor may be the degree of manipulation to the amplitude envelope. A large, instantaneous increase in amplitude based on the $F0$ contour may produce a ''popping'' sound, while a large decrease in amplitude based on the $F0$ contour may result in inaudible portions of the speech signal and ''click'' sound artifacts. Different shape factors selected properly for individual speech segments are expected to make the proper degree of amplitude envelope modification.

V. CONCLUSIONS

Modifying the amplitude envelope to more closely resemble the $F0$ contour was effective in enhancing Chinese tone recognition in multichannel cochlear implant simulations. Chinese tone recognition was significantly improved by modifying individual channels' amplitude envelope to better resemble the $F0$ contour, or by modifying the overall amplitude envelope prior to multichannel speech processing. Modifying the overall amplitude envelope before multichannel processing better preserved vowel recognition accuracy and required less speech-processing computation. Both amplitude envelope modification algorithms enhanced recognition of all four Chinese tones, especially that of tone 2. The effects of amplitude envelope modification on tone recognition were somewhat speaker- and stimulus dependent; greater enhancements were generally found for those speakers and vowels for which tone recognition scores were rela-

tively lower in the baseline condition. Increasing the modulation depth of the periodicity fluctuations within each channel did not significantly affect Chinese tone recognition. These results obtained with normal-hearing subjects (listening to acoustic simulations of a four-channel cochlear implant) suggest that cochlear implant users' tone recognition could be significantly improved by enhancing the amplitude envelope of the speech signal to better match the $F0$ contour.

ACKNOWLEDGMENTS

We are grateful to John J. Galvin III for assistance in editing the manuscript. We also thank Andrew Faulkner and an anonymous reviewer for their constructive comments. The research was supported by Grant R01-DC04993 from NIDCD.

Fu, Q.-J., Horng, M.-J., and Hsu, C.-J. (**2003**). ''Effects of speech processing strategy on Chinese tone recognition by Nucleus-24 cochlear implant patients,'' Ear Hear. **25**(5), 501–508.

Fu, Q.-J., and Zeng, F.-G. (**2000**). ''Identification of temporal envelope cues in Chinese tone recognition,'' Asia Pac. J. Speech, Lang. Hear. **5**, 45–57.

Fu, Q.-J., Zeng, F.-G., Shannon, R. V., and Soli, S. D. (**1998**). ''Importance of tonal envelope cues in Chinese speech recognition,'' J. Acoust. Soc. Am. **104**, 505–510.

Garding, E., Kratochvil, P., Svantesson, J. O., and Zhang, J. (**1986**). ''Tone 4 and tone 3 discrimination in modern Standard Chinese,'' Lang Speech **29**, 281–293.

Geurts, L., and Wouters, J. (**2001**). ''Coding of the fundamental frequency in continuous interleaved sampling processors for cochlear implants,'' J. Acoust. Soc. Am. **109**, 713–726.

Greenwood, D. D. (**1990**). ''A cochlear frequency-position function for several species—29 years later,'' J. Acoust. Soc. Am. **87**, 2592–2605.

Howie, J. M. (**1976**). *Acoustical studies of Mandarin Vowels and Tones* (Cambridge University Press, Cambridge, England).

Hsu, C.-J., Horng, M.-J., and Fu, Q.-J. (**2000**). ''Effects of the number of active electrodes on tone and speech perception by Nucleus 22 cochlear implant users with SPEAK strategy,'' Adv. Oto-Rhino-Laryngol. **57**, 257–259.

Liang, Z.-A. (**1963**). ''The auditory perception of Mandarin tones,'' Acta Phys. Sin. **26**, 85–91.

Lin, M.-C. (**1988**). ''The acoustic characteristics and perceptual cues of tones in Standard Chinese,'' Chinese Lang. Writings **204**, 182–193.

Liu, S.-Y., Huang, T.-S., and Follent, M. (**1997**). ''The field trial of the SPEAK versus MPEAK speech coding strategies in Madarin Chinese,'' Adv. Oto-Rhino-Laryngol. **52**, 113–116.

Makhoul, J. (**1975**). ''Linear prediction: A tutorial review,'' Proc. IEEE **63**, 561–580.

Markel, J. D. (**1972**). ''The SIFT algorithm for fundamental frequency estimation,'' IEEE Trans. Audio Electroacoust. **AU-20**, 367–377.

Rabiner, L. R., Sambur, M. R., and Schmidt, C. E. (**1975**). ''Applications of a nonlinear smoothing algorithm to speech processing,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-23**, 552–557.

Rabiner, L. R., Cheng, M. J., Rosenberg, A. E., and McGonegal, C. A. (**1976**). ''A comparative performance study of several pitch detection algorithms,'' IEEE Trans. Acoust., Speech, Signal Process. **ASSP-24**, No. 5, 399–418.

Sagart, L. (**1986**). ''Tone production in modern standard Chinese: An electromyographic investigation,'' Cahiers de Linguistique, Asie Orientale, Paris, 205–221.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). ''Speech recognition with primarily temporal cues,'' Science **270**, 303–304.

Skinner, M. W., Clark, G. M., and Whitford, L. A. (**1994**). ''Evaluation of a new Spectral Peak coding strategy for the Nucleus 22 Channel Cochlear Implant System,'' Am. J. Otol. **15**, (Suppl. 2), 15–27.

Skinner, M. W., Holden, L. K., Holden, T. A., Dowell, R. C., Seligman, P. M., Brimacombe, J. A., and Beiter, A. L. (**1991**). ''Performance of postlinguistically deaf adults with the Wearable Speech Processor (WSP III) and Mini Speech Processor (MSP) of the Nucleus multi-electrode cochlear implant,'' Ear Hear. **12**, 3–22.

Wang, R.-H. (**1989**). ''Chinese phonetics,'' in *Speech Signal Processing*, edited by Y.-B. Chen and R.-H. Wang (University of Science and Technology of China Press), Chap. 3, pp. 37–64.

Wang, R.-H. (**1993**). ''The standard Chinese database,'' University of Science and Technology of China, internal materials.

Whalen, D. H., and Xu, Y. (**1992**). ''Information for Mandarin tones in the amplitude contour and in brief segments,'' Phonetica **49**, 25–47.

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (**1991**). ''Better speech recognition with cochlear implants,'' Nature (London) **352**, 236–238.

Xu, L., Tsai, Y., and Pfingst, B. E. (**2002**). ''Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses,'' J. Acoust. Soc. Am. **112**, 247–258.

Patterns of English phoneme confusions by native and non-native listeners

Anne Cutler[a)]
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Andrea Weber
University of the Saarland, Saarbrücken, Germany

Roel Smits and Nicole Cooper
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

(Received 15 May 2004; revised 3 September 2004; accepted 7 September 2004)

Native American English and non-native (Dutch) listeners identified either the consonant or the vowel in all possible American English CV and VC syllables. The syllables were embedded in multispeaker babble at three signal-to-noise ratios (0, 8, and 16 dB). The phoneme identification performance of the non-native listeners was less accurate than that of the native listeners. All listeners were adversely affected by noise. With these isolated syllables, initial segments were harder to identify than final segments. Crucially, the effects of language background and noise did not interact; the performance asymmetry between the native and non-native groups was not significantly different across signal-to-noise ratios. It is concluded that the frequently reported disproportionate difficulty of non-native listening under disadvantageous conditions is not due to a disproportionate increase in phoneme misidentifications. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1810292]

PACS numbers: 43.71.Es [RLD] Pages: 3668–3678

I. INTRODUCTION

All four authors of this paper fluently speak and understand both English and Dutch; for each of us, at least one of these languages is not the native language. As non-native listeners, we are all too familiar with the phenomenon that listening to non-native language seems disproportionately difficult under disadvantageous listening conditions, such as against a noisy background.

Despite the very large literature on phoneme perception in non-native languages [see, e.g., Strange (1995) for overview papers], the evidence concerning the effects of noise and other distortions on non-native versus native perception remains relatively scant. A series of studies by Nábělek and colleagues (Nábělek and Donahue, 1984; Takata and Nábělek, 1990) demonstrated that speech stimuli which native and non-native listeners reported equally accurately in the clear were reported significantly less accurately by the non-native listeners against a noisy or reverberant background. The stimuli in question were the sentences of the Modified Rhyme Test (MRT; Kreul *et al.*, 1968), English words in the context *Say the word—again*. Gat and Keith (1978) had found the same result with similar materials. The MRT stimuli presented as synthetic speech to non-native listeners by Greene *et al.* (1985) produced a greater performance decrease compared to natural productions than the decrease observed with native listeners (see Pisoni, 1987). Van Wijngaarden *et al.* (2002) found that German and English sentences were perceived significantly better under noise by native than by Dutch listeners. Florentine (1985a, b) and Mayo *et al.* (1997), using the Speech Perception in Noise test (Kalikow *et al.*, 1977), found greater relative effects of noise on non-native than on native reports of high-predictability sentences (e.g., *The boat sailed across the bay*). Conrad (1989) reported that the greater the rate of compression applied to simple sentences (e.g., *The traveler saw a lighthouse in the distance*), the larger were the differences in listening accuracy between native and non-native listeners.

These results confirm what non-native listeners so frequently report: disadvantageous conditions affect listening to a greater degree in the non-native than in the native language. However, they do not uniquely indicate the source of this disproportionate effect. One obvious possibility is, of course, gross disruption of phonetic processing. Where the phoneme categories of the non-native language fail to match those of the native language, phonetic decisions can be influenced by the native repertoire (Best, 1995; Strange, 1995); it may be that this influence becomes stronger when stimuli are harder to perceive. Interestingly, though, a number of results suggest that the difficulty may not be (exclusively) located at the phonetic processing level. When semantically anomalous sentences (e.g., *A jaunty fork raised a vacant cow*) were presented as natural or synthetic speech to native and non-native listeners by Mack (1988), it was the native listeners who showed the proportionally greater increase in errors from the natural to the synthetic condition. Hazan and Simpson (2000) studied the effects of cue enhancement (selective amplification of the acoustic cues critical for particular contrasts) on phoneme perception in noise; their investigation revealed that all listeners' identification performance

[a)]Address for correspondence: Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. Electronic mail: anne.cutler@mpi.nl

benefited from such enhancement, but the benefit for non-native listeners was never greater than, and sometimes less than, that for native listeners. In the studies of Florentine and her colleagues (Florentine, 1985a; Mayo *et al.*, 1997), it was mainly in the effective use of contextual predictability that native listeners outstripped non-native listeners when perceiving speech in noise. Together, these studies suggest that non-native listening difficulty may have a more complex etiology than disruption of phonetic processing.

The previous literature does not, however, motivate a definitive conclusion. The aim of the present study was, therefore, to provide a new data set of phonetic identification in noise by native and non-native listeners, using materials for which higher-level factors such as lexical frequency or contextual plausibility were irrelevant, and covering almost the entire phoneme inventory of a language. The usefulness to speech perception research of very large data sets, ideally covering a complete phoneme inventory, needs no special advocacy—for American English, the classic studies of Peterson and Barney (1952), Miller and Nicely (1955), and Wang and Bilger (1973) remain valuable, now supplemented by the more recent studies of Hillenbrand *et al.* (1995) and Benkí (2003a). Smits *et al.* (2003) similarly reported perceptual data on the complete diphone set of Dutch. On the basis of such sets, it is possible to estimate the contribution of phoneme perceptibility to recognition of any spoken word of the language; our present aim was to provide such a necessary basis for understanding Dutch listeners' recognition of American English, under differing listening conditions.

The noise masking which we used was, as in the studies of Takata and Nábělek (1990), Florentine (1985a, b), and Mayo *et al.* (1997), multi-speaker babble, which best mimics difficult listening conditions in the natural experience of non-native listeners. The stimuli were CV and VC syllables covering almost all such possible sequences in American English. The native listeners were speakers of American English; the non-native listeners were Dutch. These non-native listeners were fluent users of English, but dominant in their native language. Where the phoneme categories of Dutch [16 vowels, 19 consonants (Booij, 1995; Gussenhoven, 1999)] fail to match those of American English, misidentifications are expected in the non-native responses; English contains a number of consonants with no Dutch counterpart (the final consonants of *path*, *smooth*, *edge*, and *egg*) and several vowel contrasts which collapse to a single near category in Dutch (e.g., the contrast in *bat–bet* in any variety of English, and the contrast *cot–cut* in American English). The question particularly at issue here, however, is whether under noisier listening conditions relatively more such misidentifications are observed.

We might further expect our non-native listener population to experience difficulty with syllable-final consonants, since in Dutch all consonants in syllable-final position are voiceless; English final voicing contrasts such as *at* versus *ad* should therefore prove difficult. Several recent studies of native listening in English have in fact reported better recognition of consonants presented in noise in syllable-initial than in syllable-final position (Redford and Diehl, 1999; Benkí, 2003a). Again, however, the question of principal interest here is not only whether non-native listeners experience special problems with perceiving voice in final position, but whether these problems become disproportionately more marked under increasing levels of noise.

In summary: If the repeated demonstrations of greater difficulty of non-native than of native listening under noise reflect disproportionate effects of noise on phoneme identification, then we will observe a phoneme identification difference between native and non-native listeners which increases with increasing noise, as the sentence score differences collected by Mayo *et al.* (1997) did. However, if the extra difficulties of non-native listening under noise are not exclusively, or not at all, due to problems at the level of phoneme identification, then we may observe some other pattern of results: insignificant increase in the difference between native and non-native scores, a constant difference between native and non-native scores across noise conditions, or even a decrease in the native versus non-native difference with increasing noise.

II. METHOD

A. Participants

Sixteen native listeners of American English, mostly students at the University of South Florida, participated in the experiment; they received either course credit or a small monetary compensation. Sixteen Dutch-native listeners, students at the University of Nijmegen, also participated; they received a small monetary compensation. In all cases listeners were rewarded per session and additionally with a bonus upon completion of the eight-session experiment. The Dutch native listeners had all completed their school education in the Netherlands, involving 7 to 8 years of English instruction beginning on average at age 11. All were fluent in English but none had lived for longer periods in an English-speaking country.

B. Materials

Twenty-four consonants and 15 vowels were combined to form all possible standard American English CV and VC sequences, excluding those with schwa. All vowels (12 monophthongs and three diphthongs) occurred either in initial or in final position; thus lax vowels were allowed in syllable-final position although stand-alone syllables ending in lax vowels do not occur in the language. Twenty-two consonants (not /ŋ/ or /ʒ/) occurred in initial position, and 21 consonants (not /h/, /w/, /j/) in final position. The full phoneme set can be found in Appendix A. The complete set of stimuli comprised 645 (330 CV and 315 VC) syllables.

The 645 syllables were transcribed phonemically. A phonetically trained female native speaker of American English (born in the Mid West, who had lived as child and teenager in four different states) read these transcriptions in a quiet room via a Sennheiser ME64 microphone to Digital Audio Tape. The sampling rate at recording was 44.1 kHz, later downsampled to 16 kHz. Stop consonants in final position were released.

Each syllable was centrally embedded in 1 s of multi-speaker babble noise. The babble was constructed from a

recording of three male and three female speakers having a conversation in English in a quiet room. The recording was made directly onto DAT tape using a Sennheiser microphone placed in the middle of a table around which the speakers were seated. For each speaker, a 1-s stretch was selected during which no background noises were present and he or she was speaking alone at a normal (i.e., not too loud or soft or excited) tone. These six stretches of speech were then equalized for rms amplitude and added together. The test syllables were normalized for rms amplitude of the vowel and were then combined with the babble noise at three signal-to-noise ratios (SNRs): 0, 8, and 16 dB (normalized vowel amplitude/babble amplitude). These SNRs were chosen on the basis of a pretest to yield difficult, intermediate, and easy English phoneme perception for the Dutch non-native listeners. The whole stimulus set thus comprised 1935 tokens (645 syllables×3 SNRs).

C. Procedure

Each listener participated in eight testing sessions, made up of 3870 trials in total. Each of the 1935 tokens was presented twice (always in separate sessions), once with the listener's task being to identify the consonant and once to identify the vowel. In each session, listeners received two stimulus blocks, one for consonant and one for vowel identification; the blocks consisted solely of CV or solely of VC stimuli. Every listener received the items of a block in a different pseudo-random order. SNRs were mixed within blocks.

The presentation of items was self-paced. If the listener did not respond by 15 s after stimulus offset, the trial was recorded as a miss. Listeners signaled responses by clicking on a word exemplifying the appropriate sound on a computer screen. Prior to the experiment they were familiarized with these example words: they saw the display screen and heard the same speaker as in the experiment pronounce each alternative, e.g., *v as in very*. Different words were used for vowels, initial consonants, and final consonants; the words are listed in Appendix A.

III. RESULTS

No response (''miss'') was registered on in total 0.64% of trials (less than 0.1% for the non-native listeners; just over 1% for the native listeners); these trials were discarded from the data set.

The principal findings of our study can be derived at a glance from Fig. 1, which shows the overall proportions of correct responses for the two listener groups for vowels and for consonants at the three signal-to-noise ratios. First, it can be seen that the identification performance of the non-native listeners is significantly and consistently worse than that of the native listeners, but that this performance disadvantage is unaffected by signal-to-noise ratio. Second, it can be seen that an increase in signal-to-noise ratio results in a clear improvement in performance (for both listener groups) in the identification of consonants, but has very little effect on the identification of vowels, which (again for both listener groups) is at a relatively high level even at 0 dB SNR. The

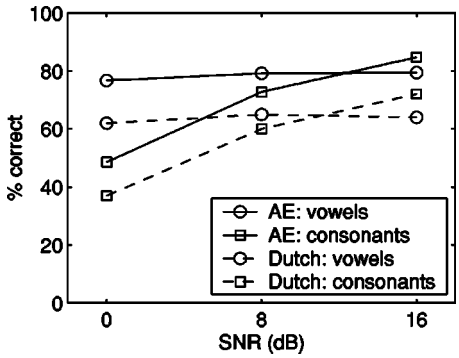


FIG. 1. Percentages of correctly recognized vowels and consonants as a function of SNR, separately by language group (''AE''=American English). Data have been pooled across initial and final positions, phonetic contexts, and subjects.

pattern of results thus strongly suggests that the greater difficulty of non-native than of native listening under noise is not due to disproportionate effects of noise on phoneme identification.

Analyses of variance across subjects confirmed that overall performance was better for native (grand mean 73.6% correct) than for non-native listeners (grand mean 60%; $F[1,30]=21.1$, $p<0.001$) and for vowels (grand mean 71.1%) than for consonants (62.5%; $F[1,30]=44.66$, $p<0.001$); further, performance was strongly affected by SNR (grand mean of 75% at 16 dB SNR, 69.3% at 8 dB SNR, and 56.1% at 0 dB SNR; $F[2,60]=2191$, $p<0.001$). The latter two effects interacted significantly—consonant performance showed significantly more effect of SNR than did vowel performance ($F[2,60]=2066$, $p<0.001$)—but the native versus non-native comparison interacted neither with SNR nor with the vowel/consonant factor. *Posthoc* analyses revealed that for both listener groups there was a significant difference in performance between 0 and 8 dB SNR for both vowels and consonants (vowel difference 2% for each group, consonant difference 24% for native and 23% for non-native), but a significant difference between 8 and 16 dB SNR for consonants only (vowel difference 0% for native and 1% for non-native, consonant difference 12% for each group).

Figure 2, which presents overall identification in initial versus final position in the carrier syllable, shows a further clear effect: for native listeners, identification of both consonants and vowels is better in final than in initial position. Non-native listeners show the same final-position advantage

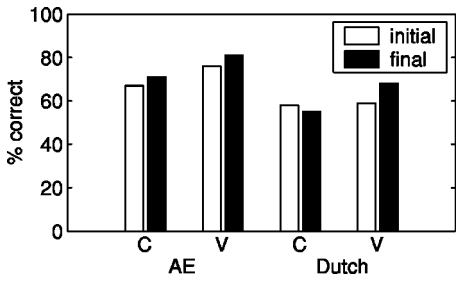


FIG. 2. Percentages of correctly recognized consonants (C) and vowels (V) in initial and final positions, separately by language group (''AE''=American English). Data have been pooled across SNRs, phonetic contexts, and subjects.

TABLE I. Confusion matrix for initial consonants at 0 dB SNR categorized by the American English listeners. Percentages of correct responses have been pooled over participants and vowel contexts.

	Response																					
Stimulus	pie p	tie t	car k	far f	thin θ	see s	she ʃ	chin tʃ	hi h	be b	do d	go g	very v	there ð	zoo z	joke dʒ	yell j	my m	no n	lie l	row r	win w
p	**15.4**	2.9	4.6	10.4	3.3				39.2	6.7	1.3	2.1	2.5	1.3		0.8	1.3	1.3	0.4	0.4	0.8	0.8
t	10.4	**19.6**	9.6	5.4	6.3	1.3	0.8	0.4	27.9	2.9	1.3	3.3	0.4	3.3	1.3	0.4	0.8	0.4	0.8	0.4		0.8
k	11.7	14.6	**25.8**	1.7	2.1		0.4	0.8	27.9	2.1	0.8	2.1	0.4	0.8			0.4	0.4	2.1			1.7
f	22.9	2.1	3.8	**19.2**	7.5	0.4			14.2	8.8	1.3	0.8	3.8	5.8		0.4	1.7					3.8
θ	12.5	5.4	3.8	13.3	**18.3**			0.4	10.4	7.5	2.1	1.3	3.8	14.6	0.8		0.4	0.8	0.4	0.4	0.4	1.7
s	0.4	2.1	0.4	9.2	10.0	**51.7**	2.1	0.4	2.1	2.5				9.6	8.8							
ʃ	0.4		0.4			0.8	**76.7**	19.6				0.4	0.4	0.4		0.4						
tʃ		5.0	0.8	1.7	0.8		1.3	**83.8**	0.4			0.4		1.7		2.9	0.4		0.4			
h	14.6	5.0	4.6	9.6	4.6	0.4		0.4	**36.7**	7.1	0.4	1.7	2.9	2.1		0.4		1.7	0.4	0.4	0.4	1.7
b	2.1		1.3	5.8	5.8				15.0	**19.6**	1.3	1.3	5.0	8.3	0.4	0.8	3.8	10.8	0.4	4.2	1.7	3.8
d		2.9		1.3	7.9	0.4		0.8	4.6	7.9	**14.6**	2.9	0.8	14.6	0.4	0.8	7.1	3.3	19.6	6.7		0.4
g	1.7	0.8	1.3	2.9	2.5	0.4			10.4	3.8	2.5	**29.6**	4.2	2.1	0.4	0.8	19.2	1.3	7.9	2.9	1.3	1.3
v	2.9	1.7	0.4	5.8	4.2	0.8		0.4	8.8	18.3	0.4	3.3	**17.5**	14.6		0.8	2.1	4.2	1.7	0.8	1.7	5.0
ð		1.3		1.3	14.6	0.8		0.4	1.7	9.6	4.2	3.3	5.8	**30.4**	4.2	2.1	1.3	1.3	4.6	10.0		0.8
z		1.7			9.2	2.5			0.8	1.7	1.7	2.5	8.3	21.3	**31.3**	2.1	0.4	0.8	3.3	1.3	1.7	7.5
dʒ	0.4	0.4	0.4	0.4	2.5	0.4		4.6	1.3	1.7	4.2	2.9	0.4	8.3		**68.8**	0.8	0.4	0.4	1.3		
j		0.8			0.4				2.9	3.3	5.4	3.3	1.3	1.3	1.7	2.9	**65.8**	2.5	2.5	1.7		2.1
m	0.4		0.4	1.7	0.4				3.3	3.8	0.8	1.3	6.3	0.4	0.4		0.8	**63.8**	5.8	5.8	1.3	1.7
n					0.8				0.4	0.4	0.4		0.4	0.4	0.4		0.4	12.5	**77.9**	4.2	0.8	0.4
l	0.4				2.5			0.4	0.8	5.0	1.7	2.1	3.3	4.2	0.4		2.1	12.5	5.0	**54.2**	2.1	0.8
r	0.8	0.4	1.3	1.3					7.1	5.4	0.4	2.1	5.0	0.8				2.5	0.4		**68.8**	2.9
w	0.8	0.4							1.7	4.2		0.8	2.9	0.4			4.2	5.8		2.9	0.4	**73.3**

for vowels; for consonants, however, their performance is worse in VC than in CV. Analyses of variance confirmed that the overall advantage for final position was significant ($F[1,30]=27.34$, $p<0.001$), but this effect interacted with the vowel/consonant factor ($F[1,30]=20.4$, $p<0.001$), the final advantage being larger for vowels than for consonants. Moreover, the three-way interaction of these two factors with listener language was also significant ($F[1,30]=11.44$, $p<0.002$); *posthoc* tests revealed the source of this interaction to be a significant interaction (reversal of the position effect) of initial/final and vowel/consonant for non-native listeners ($F[1,15]=26.91$, $p<0.001$), but no significant interaction for native listeners ($F<1$).

Since the overall advantage for final position which we observed contrasts with previous findings of Redford and Diehl (1999) and Benkí (2003a), we conducted further analyses in direct comparison with these earlier studies; these analyses are described in Appendix B.

Detailed results are presented as confusion matrices (separately for native and non-native listeners and for consonants and vowels in initial versus final position) in Tables I–VIII. These tables show the identification results at 0 dB (the more accurate results at better SNRs are available at http://www.mpi.nl/world/persons/private/anne/materials.html). Where rows do not sum to 100%, the remainder was missing data.

It can readily be seen from the confusion matrices that the phonemes which were most difficult for non-native listeners were also difficult for native listeners. Thus although the Dutch listeners had difficulty identifying the English consonants without counterpart in Dutch, these consonants were also difficult for the native listeners; and although the Dutch listeners confused the vowel sounds which share one near Dutch category, native listeners made such confusions, too. Characteristics of the masking babble noise presumably influence these patterns. In fact, the percent correct identification rate of the two listener groups across phonemes was very highly correlated: at 16 dB SNR $r=0.83$, at 8 dB 0.87 and at 0 dB 0.91, in all cases $p<0.001$. It can also be seen from the matrices that there were no strong effects of phonotactic legality of syllable-final lax vowels; errors on these vowels, for both listener groups, tended to be other lax vowels. The one clear effect of native phonology on non-native listening appeared in the Dutch listeners responses to syllable-final consonants; as described above, Dutch phonotactics prohibit voicing contrasts in final position, and the Dutch listeners made many more voicing errors on final consonants such as /b, d, g/ than native listeners did.

Since Miller and Nicely (1955), it has been customary to view perceptual data of the present type in terms of percentage of information transmitted for broad feature classes. In contrast to raw percent correct, transmitted information (TI) takes account of response biases, and, regardless of the number of response alternatives, gives a result of zero when subjects guess randomly. The number of response alternatives varies across features, so only TI measures allow direct comparisons of the accuracy with which different features can be recognized. Thus only TI allows us to compare native and non-native featural sensitivity. Smits (2000) further explains these advantages, and equations for TI calculation are presented by Miller and Nicely (1955).

Figure 3 presents TI analyses of our data set, and Table IX shows the phonemes associated with the featural values we used. We considered the broad features of consonants to be place and manner of articulation, and voicing, rather than the more detailed feature systems (coronal, anterior, conso-

TABLE II. Confusion matrix for final consonants at 0 dB SNR categorized by the American English listeners. Percentages of correct responses have been pooled over participants and vowel contexts.

	Response																				
Stimulus	lip p	hot t	sick k	off f	path θ	pass s	fish ʃ	such tʃ	grab b	odd d	egg g	love v	smooth ð	buzz z	beige ʒ	edge dʒ	am m	on n	ring ŋ	ill l	far r
p	**50.0**	16.3	14.2	5.8	5.8				0.8	0.4	0.4		2.1				0.4				0.8
t	5.0	**77.1**	5.8	0.8	4.6			2.1		0.4		0.4	2.5	0.4							
k	11.3	12.5	**63.3**	0.8	5.0			2.1	0.4	0.4	0.4	0.8	2.1								
f	10.0	10.0	6.7	**45.0**	12.9		0.8		0.8			1.3	5.4		0.8		0.4			0.4	1.7
θ	9.2	17.9	4.2	30.8	**19.2**	0.8	0.4	0.8	0.4	0.4	0.4	2.5	7.5		0.4			0.4	0.8		0.4
s	0.8	2.9	0.8	12.9	8.8	**65.4**	2.9	0.4			0.4		1.7	0.8	0.4						
ʃ					0.4	1.3	**80.8**	14.2					0.4		2.5	0.4					
tʃ	0.4	3.8	0.4					**89.6**							1.7	4.2					
b	1.3	1.3	4.2	3.8	2.5	0.4		0.4	**35.0**	10.4	9.2	15.4	4.6	0.4	2.5	1.3	2.1	1.7		0.4	1.3
d		3.3	0.4	3.8	2.5	1.7			3.8	**42.9**	4.6	6.7	5.8	1.7	5.4	5.8	0.4	5.8	2.9		0.4
g	0.4	3.3	1.3	2.9	5.4	0.4		0.8	5.4	9.2	**35.4**	14.2	5.4	0.8	2.1	1.7	1.3	2.1	2.5	0.4	0.8
v	0.4	0.8	1.3	9.2	2.5			0.4	2.9	4.6	7.9	**47.5**	5.8	0.4	3.8	1.7	2.9	1.3	0.8	0.8	1.7
ð		2.1		1.7	4.2	2.5	0.4	0.4	2.9	22.5	5.0	17.5	**16.7**	5.8	5.4	7.5		1.3	0.8	0.4	0.4
z	0.4				2.5	7.5	0.4	0.4	1.3	10.0	1.3	12.5	9.6	**37.1**	5.4	4.6	0.4	2.5		0.4	0.8
ʒ				0.4		0.8	2.5	2.1		2.5	1.3	4.2	4.6	3.8	**51.7**	23.3	0.4	1.7		0.4	0.4
dʒ	0.4	0.8			0.4		0.4	3.3	0.4	5.8	2.5	0.8	0.4	0.4	17.9	**64.6**	0.4				0.8
m		0.4		1.3	0.4				1.7	1.3	2.1	7.1	0.4			0.4	**56.3**	12.1	14.2	0.4	0.8
n					0.4				0.4	5.4	1.3	3.3	1.7	1.3	1.3	0.8	12.5	**59.6**	10.4		0.4
ŋ		0.4	0.4	0.8				0.4	0.4	1.3	9.2	5.8	1.3		0.4		15.4	25.4	**35.0**	0.8	1.7
l	0.4	0.8	0.8	7.9	0.4					0.8	0.8	6.7	3.3		0.4	0.4	1.3	0.4		**70.8**	0.8
r		0.4	0.4	1.3	0.8				0.4	0.8	0.4	3.8	1.3			1.3	1.3	0.4		0.8	**84.2**

nantal, sonorant, continuant, etc.) used in formal phonology (e.g., Kenstowicz, 1994). The values of place of articulation were held to be labial, dental, alveolar, palatal, velar, and glottal; this classification strikes a balance between a very detailed phonetic inventory of places within the English consonant inventory, which would have very few consonants at many places, and a gross classification into only labial, coronal, and dorsal. As values of manner of articulation we used stop, affricate, fricative, liquid, glide, and nasal. Voicing had two values, voiced and voiceless. The features used for vowels were height (three values: high, mid, and low), backness (three values: front, central, and back) and tenseness (two values: tense and lax). Because the three diphthongs always change value on height and tenseness, and two of the three also change value on backness, we excluded them from the vowel calculations in Fig. 3; for the TI calculations it was therefore also necessary to discard diphthong responses to monophthongal stimuli, a total of 915 cases (1.85% of the total monophthongal vowel dataset).

Statistical analyses of the comparisons in Fig. 3 showed a significant improvement in percentage of transmitted featural information with increasing SNR, for five of the six broad feature classes (all comparisons $p<0.001$; vowel tenseness insignificant). For all three vowel features, and for consonant manner, information was transmitted more efficiently in final position in the syllable than in initial position

TABLE III. Confusion matrix for initial vowels at 0 dB SNR categorized by the American English listeners. Percentages of correct responses have been pooled over participants and consonant contexts.

	Response														
Stimulus	beat i	bit ɪ	wait eɪ	bet ɛ	bat æ	hot ɑ	cut ʌ	caught ɔ	boat oʊ	cook ʊ	boot u	buy aɪ	boy ɔɪ	shout aʊ	bird ɝ
i	**78.9**	8.3	0.3	2.7				0.3		0.3	1.8	1.2			3.9
ɪ	1.5	**81.8**	0.9	8.0	0.9		1.2	0.3		0.3	1.5	0.9			1.8
eɪ	5.7	5.4	**74.4**	4.5	5.7				0.3	0.3		0.3		0.3	1.2
ɛ	0.6	4.2	2.4	**84.2**	2.7		1.2					0.3			3.0
æ		1.2	6.5	3.9	**78.3**	0.6		1.2				0.3	0.3	4.8	2.1
ɑ		0.6	1.2	0.3	9.8	**42.3**	12.5	26.8	0.9			0.6		0.9	1.8
ʌ			0.3		1.2	12.5	**64.9**	8.3	1.8	1.2		1.2		4.2	3.0
ɔ		0.3	0.3		0.6	36.3	4.5	**47.3**	3.9	1.2		0.6	2.1	0.9	
oʊ		0.3				4.8	1.2	0.9	**69.6**	8.9	6.5		3.3	2.1	0.3
ʊ						2.1	14.0	2.1	0.9	**63.7**	6.8	0.3	3.0	2.1	0.9
u	3.6	1.5	0.3			0.6	3.0	1.5	1.8	19.3	**62.5**	0.3	1.2	1.8	1.2
aɪ		8.3	2.1				0.3					**87.2**	0.6		0.6
ɔɪ	0.3		0.6	0.3			0.3	0.6	1.2	0.6	0.3		**92.9**	3.0	
aʊ	0.3		0.3	2.1	0.6	3.3		7.1	2.4	0.3	0.3		0.9	**81.5**	0.3
ɝ	0.6	0.3		1.5			1.5	0.3							**95.5**

TABLE IV. Confusion matrix for final vowels at 0 dB SNR categorized by the American English listeners. Percentages of correct responses have been pooled over participants and consonant contexts.

	Response														
Stimulus	beat i	bit ɪ	wait eɪ	bet ɛ	bat æ	hot ɑ	cut ʌ	caught ɔ	boat oʊ	cook ʊ	boot u	buy aɪ	boy ɔɪ	shout aʊ	bird ɝ
i	**93.5**	0.3		3.7				0.3			1.4			0.3	
ɪ	0.9	**84.4**	0.3	10.5	0.3		2.0		0.3			0.9		0.3	
eɪ	0.6	2.0	**91.5**	2.0	2.8										
ɛ	0.6	6.3	2.3	**73.6**	8.5		2.3	2.3		0.3		0.3		0.3	0.3
æ		0.6	1.1	12.2	**82.7**			1.1			0.3	0.3			0.3
ɑ			1.1	0.9	8.2	**33.5**	24.4	27.0	0.6	0.3		0.3		1.1	
ʌ			0.9	2.3	6.0	11.4	**65.3**	11.1	0.3	0.9	0.3		0.3	0.3	
ɔ			0.9		2.6	23.9	3.7	**65.3**	0.9	0.6				0.9	
oʊ	0.3	0.3		0.3		1.7		0.6	**90.6**	0.3	0.9		1.4	2.3	
ʊ	0.3			0.6		2.0	21.6	0.6	0.6	**68.2**	2.6	0.3	0.3		
u	3.7			0.9		0.3	0.9	0.6	0.3	6.8	**81.8**		0.3	2.6	
aɪ		6.0	0.9	0.3		0.3		0.3	0.3			**91.5**		0.3	
ɔɪ						0.6	0.3	0.9	1.4	0.9	0.3	0.3	**92.0**	2.3	
aʊ	0.3			0.9	0.6	2.0		1.7	8.2	0.6			1.7	**82.4**	0.6
ɝ	0.3	0.3		1.1			0.3								**97.7**

(three comparisons $p<0.001$, vowel backness $p<0.05$). Place and voicing showed no significant main effect of position in the syllable. There was an interaction between SNR and position within the syllable for all feature classes, reflecting in each case a greater improvement with increasing SNR for phonemes in syllable-initial position than for phonemes in syllable-final position (five comparisons $p<0.001$, consonant manner $p<0.05$). For all types of phonetic information, the masking effects of noise (especially at 0 dB SNR) are thus greatest in syllable-initial position. Listener group language did not interact with SNR for any feature comparison, but interacted with syllable position for consonant voicing ($p<0.001$) and vowel height ($p<0.05$). For the native listeners, voicing information was perceived better in final position, but for the Dutch listeners, as expected, voice was much less well perceived in final position [Fig. 3(e)]. The vowel height interaction was due to the advantage of final position over initial position being larger for non-native than for native listeners [Fig. 3(b)].

For each listener group separately, we compared the relative informativeness of types of featural information. There were significant differences in informativeness among the consonant features for the native ($F[2,30]=16.94$, $p<0.001$) and non-native listeners ($F[2,30]=57.69$, p

TABLE V. Confusion matrix for initial consonants at 0 dB SNR categorized by the Dutch listeners. Percentages of correct responses have been pooled over participants and vowel contexts.

	Response																					
Stimulus	pie p	tie t	car k	far f	thin θ	see s	she ʃ	chin tʃ	hi h	be b	do d	go g	very v	there ð	zoo z	joke dʒ	yell j	my m	no n	lie l	row r	win w
p	**30.8**	3.3	9.2	9.6	2.9			0.4	19.2	11.7	1.3	1.3	2.9	1.7	0.4		0.8	0.8	1.3	1.3		1.3
t	24.6	**14.2**	12.5	7.5	7.9	0.8		2.9	11.3	7.1	0.4	2.1	1.3	1.7				2.1	3.3	0.4		
k	25.0	7.9	**25.8**	3.8	4.2	0.4	0.8	0.4	13.8	4.2	1.3	3.8	1.3	0.8	0.4	0.4	1.3	0.4	1.7	1.3	0.4	0.8
f	24.6	2.1	9.2	**15.0**	7.1		0.4	0.4	9.2	15.0	1.7	2.9	5.4	4.2		0.4	0.4	0.4	0.4		0.4	0.8
θ	18.8	6.3	3.8	13.3	**12.1**	0.4	0.4	0.4	7.1	14.2	2.5	1.7	2.9	7.5			0.4	1.3	2.9	2.9		0.8
s	0.4	2.5	0.4	12.5	24.6	**30.4**	0.8	0.4		0.8	1.3		3.3	7.9	14.6							
ʃ		0.4			1.3	6.7	**72.5**	18.3									0.8					
tʃ	3.3	4.2	1.3	2.1	2.5	1.3	4.6	**70.8**	1.3	1.3	0.4		0.4	0.8		5.4	0.4					
h	26.3	4.6	12.1	11.3	5.0	0.4	0.4	0.8	**17.9**	8.3	1.3	0.4	4.6	1.7	0.4		0.8		0.8	1.7	0.8	0.4
b	7.5	0.4	5.8	9.2	1.7	0.4		0.4	12.5	**28.3**	2.5	0.4	4.6	2.1			2.9	7.1	2.9	5.0	1.3	5.0
d	2.5	2.1	1.3	1.3	5.4				8.8	12.1	**10.8**	2.5	2.1	12.9	0.4	0.4	6.3	4.2	12.5	12.5		1.7
g	3.3	1.3	9.2	2.9	2.5		0.4	1.3	9.2	10.0	5.0	**17.1**	1.7	3.3		0.8	24.2	0.8	2.5	2.5	0.4	1.7
v	7.5	2.9	2.5	8.8	5.0	0.4			6.7	30.0	1.3	1.7	**9.6**	7.9			2.1	3.8	1.7	0.4	2.5	5.4
ð	2.5	1.3	2.5	1.7	14.6	2.1	0.4	0.8	2.1	17.1	10.0		1.3	**18.8**	1.3	1.7	1.3	1.3	3.3	12.1	0.8	2.9
z		0.8	1.3	1.3	9.6	3.3	0.4	1.3		7.9	5.0		2.5	23.8	**27.1**	1.3	2.5	2.1	5.0	0.4	0.8	3.8
dʒ	4.2	0.4	2.5	0.4	2.1		0.8	18.3	2.1	2.1	6.7	2.1		7.5		**40.4**	5.8	0.4	1.3	2.9		
j	1.3		0.8	0.8	0.8		0.8	0.4	2.1	4.2	2.5	1.3	0.4	1.3	0.4	4.6	**69.6**	2.5	4.2	1.3		0.8
m	3.8	0.8	2.5	2.9	0.4				2.1	9.6	0.8		2.5					**50.0**	10.0	5.0	5.4	4.2
n					0.4				2.1	1.7	1.3					0.8	0.4	12.9	**73.8**	4.6	0.4	1.7
l	5.8	1.3	1.7	1.3	1.3			0.4	2.1	8.3	1.7	0.8	2.5	3.3			1.3	10.0	4.2	**46.7**	2.1	5.4
r	2.5	1.3	1.7	2.1				0.4	5.8	14.6	0.8	2.1	1.3	0.8				0.8	0.4	0.4	**58.3**	6.7
w	1.7		0.4	0.4	0.4	0.4			1.7	5.8	0.8		2.1	0.4				5.8	0.8	2.5	1.7	**75.0**

TABLE VI. Confusion matrix for final consonants at 0 dB SNR categorized by the Dutch listeners. Percentages of correct responses have been pooled over participants and vowel contexts.

	Response																				
Stimulus	lip p	hot t	sick k	off f	path θ	pass s	fish ʃ	such tʃ	grab b	odd d	egg g	love v	smooth ð	buzz z	beige ʒ	Edge dʒ	am m	on n	ring ŋ	ill l	far r
p	**24.2**	13.8	11.7	5.8	8.8			0.4	21.3	5.8	2.1		3.3		0.8		0.4	0.4		0.4	0.8
t	4.6	**45.0**	5.4	1.7	9.2	1.7		1.3	1.7	20.4	0.4	0.4	5.4	0.4	0.4	0.8				0.4	0.8
k	8.3	12.5	**44.6**	4.2	5.0	0.4	0.4	0.8	2.5	3.8	12.5	0.4	1.7			1.3		0.4			1.3
f	7.5	21.7	6.7	**22.1**	9.6	1.7	1.3	0.4	6.3	10.4	0.4	3.3	5.0	0.4	0.8	0.4	0.8			0.8	0.4
θ	7.9	24.6	3.8	17.9	**17.5**	0.8	0.4	1.7	2.9	7.9	0.4	2.5	9.2		0.8	1.3					0.4
s		3.8		17.1	14.6	**37.5**	5.0	0.8	0.4	0.4		2.5	10.0	4.6	1.7	0.8					0.8
ʃ					0.4	6.7	**66.7**	10.4			0.4		1.3	1.7	10.0	2.1					0.4
tʃ	0.4	0.8		0.4	1.3	0.4	6.7	**42.5**	0.8	3.3	0.4		3.3		5.4	34.2					
b	5.0	7.5	6.7	2.9	5.4			0.4	**30.4**	15.0	3.8	7.9	5.0	1.3	0.4	2.9	2.5	0.8		0.8	1.3
d	1.3	16.3	0.4	2.5	5.8	0.8	0.4		2.1	**39.6**	2.1	3.8	7.9	1.7	2.9	5.8	0.4	1.7	2.5		2.1
g	0.8	12.9	4.6	0.8	7.5			0.8	2.1	20.4	**25.8**	4.2	5.0	0.8	0.8	6.3	2.5	1.7	0.8	0.8	1.3
v	1.3	12.9	3.8	12.1	5.4		1.7	0.4	4.6	15.8	3.3	**15.8**	5.4	1.3	1.3	1.3	2.1	3.3	1.3	4.2	2.9
ð	0.4	11.3	0.8	4.2	8.3	3.3	1.7	1.3	2.5	29.2	2.1	10.0	**8.3**	2.9	1.3	7.9	0.4	1.3		0.8	2.1
z		3.8		1.7	10.0	12.1	2.1	2.5	0.4	8.8	2.1	5.0	10.0	**25.8**	5.0	2.1	2.1	1.7	1.3	1.3	2.5
ʒ		3.3	0.4		2.1	2.5	14.2	4.2	0.4	3.8	0.4	1.3	4.2	6.7	**45.0**	9.2	0.4	0.4		0.8	0.8
dʒ		2.9		1.7	2.9	0.8	2.9	13.3	0.4	8.8			4.2	0.4	9.6	**51.7**					0.4
m		9.2		0.4	4.6	0.8				5.8	1.3	2.1	2.1	0.8			**41.3**	20.0	8.3	2.1	1.3
n	0.4	9.6		1.3	1.3	0.4	0.4	0.4	0.8	8.3		0.4	2.1	2.1	0.4	2.1	8.3	**48.3**	9.2	2.1	2.1
ŋ		6.7	0.4	2.1	2.1	0.4			2.5	5.8	6.3		1.7	1.7			13.8	22.5	**30.4**	2.1	1.7
l	0.4	8.3	1.7	5.8	3.3		0.4			9.2	1.3	2.1	1.7	0.4			0.8	2.5	0.4	**57.5**	4.2
r	0.4	7.9	0.4	0.8	3.8		0.4		1.7	6.7		1.7	2.9	0.4	0.4	0.8		2.1		1.7	**67.9**

<0.001), and among the vowel features also for both the native ($F[2,30]=120.23$, $p<0.001$) and non-native groups ($F[2,30]=186.61$, $p<0.001$). Interfeatural comparisons showed that for native listeners consonantal manner information was transmitted most accurately and place information least accurately; manner and voicing did not differ significantly but each was significantly more accurately perceived than place ($p<0.001$ for manner, $p<0.01$ for voicing). For non-native listeners manner was also transmitted most accurately (significantly more so than place, $p<0.001$), but voicing least accurately (significantly less so than place, $p<0.001$). For both groups the vowel features ordered similarly: backness was more accurately transmitted than height, and height was more accurately transmitted than tenseness (all comparisons $p<0.001$).

These analyses thus further confirm the parallel effects of the noise masking on the performance of the native and non-native listener groups. Although the groups differed overall in one respect, namely in sensitivity to final voicing contrasts, importantly, on no type of information at all did listener group language interact with SNR.

IV. GENERAL DISCUSSION

The identification performance of non-native listeners in our study fell clearly short of the native listeners' perfor-

TABLE VII. Confusion matrix for initial vowels at 0 dB SNR categorized by the Dutch listeners. Percentages of correct responses have been pooled over participants and consonant contexts.

	Response														
Stimulus	beat i	bit ɪ	wait eɪ	bet ɛ	bat æ	hot ɑ	cut ʌ	caught ɔ	boat oʊ	cook ʊ	boot u	buy aɪ	boy ɔɪ	shout aʊ	bird ɝ
i	**75.6**	16.7	1.8	1.8		0.3	0.6			1.5			0.3	0.3	0.9
ɪ	1.5	**86.0**	0.6	5.4	0.3		0.3		0.3	0.3	1.5	1.2			2.4
eɪ	25.0	14.6	**46.7**	6.5	4.2		0.3	0.6	0.3				0.6		1.2
ɛ	0.3	12.5	0.9	**58.3**	25.0		0.3							0.3	2.1
æ		1.8	1.2	33.6	**56.3**	0.3	0.6					0.9		4.2	1.2
ɑ	0.3			0.6	13.4	**29.2**	31.5	15.5	1.2	0.3	0.3	3.9	0.3	1.2	1.8
ʌ			0.3	0.9	4.8	27.4	**44.3**	7.7	3.9	0.3	0.6	1.5	0.3	3.9	4.2
ɔ					0.6	63.4	5.4	**22.0**	2.4	0.6	0.9	0.6	3.0		1.2
oʊ			0.6			8.6	0.9	2.1	**53.0**	14.9	14.9	0.9	1.8	1.2	0.6
ʊ	0.6	0.3			0.3	11.6	4.2	2.4	3.9	**50.9**	17.3	0.6	3.6	2.4	1.5
u	8.6	1.5	0.3		0.3	2.7	0.9	1.8	4.8	46.7	**29.2**	0.3		1.8	1.2
aɪ		3.0	24.7	0.6	2.7		0.3	0.6	0.6			**64.3**	0.9		2.4
ɔɪ			1.5			3.6		0.3	0.9	0.6	0.3	0.9	**90.5**	1.2	0.3
aʊ			0.3	2.7	0.6	2.7		2.7	16.4	0.9	0.9	2.4	0.3	**70.2**	
ɝ	0.9	0.3	0.3	0.9		0.3	19.0	0.9	0.3			0.6		0.6	**75.9**

TABLE VIII. Confusion matrix for final vowels at 0 dB SNR categorized by the Dutch listeners. Percentages of correct responses have been pooled over participants and consonant contexts.

	Response														
Stimulus	beat i	bit ɪ	wait eɪ	bet ɛ	bat æ	hot ɑ	cut ʌ	caught ɔ	boat oʊ	cook ʊ	boot u	buy aɪ	boy ɔɪ	shout aʊ	bird ɚ
i	**97.4**	0.3	0.3	1.7								0.3			
ɪ	0.3	**95.5**		2.3		0.3								0.3	1.4
eɪ	6.0	4.3	**84.1**	3.1	1.1			0.3			0.3	0.9			
ɛ	0.3	15.6	0.3	**60.5**	22.2			0.3						0.3	0.6
æ		0.9	4.8	39.2	**51.7**	0.6	1.1		0.3			0.6			0.9
ɑ	0.3	0.6		0.6	16.8	**23.6**	22.7	28.4	2.6	0.3	0.3	1.7	0.3	2.0	
ʌ		0.3	0.6	1.7	10.2	25.0	**41.5**	16.8	1.4	0.3	0.3	1.4		0.6	
ɔ					4.8	34.7	5.4	**50.0**	3.4		0.3		0.3	0.6	0.3
oʊ		0.6			0.3	11.1	0.6	4.8	**69.6**	2.8	6.0		2.6	1.7	
ʊ		0.6				5.4	4.3	4.5	1.7	**73.3**	6.8	0.6	1.4	0.6	0.9
u	17.9	0.3		0.6		0.6	1.4	1.1	0.9	31.3	**45.2**	0.3		0.6	
aɪ		3.4	14.8	0.3	1.1	0.3	0.6	0.6	0.9			**74.1**	1.4	0.3	2.3
ɔɪ			0.3			2.3		0.6	0.9	0.3		1.7	**93.8**	0.3	
aʊ		0.3	0.3	2.0	0.6	4.0	0.6	5.1	17.0	0.3	0.9	3.1	0.9	**64.5**	0.6
ɚ	0.3	0.3		0.9	0.3	0.6	7.1	0.3				0.6		0.3	**89.5**

mance. Also, all listeners were adversely affected by noise—the higher the SNR, the better the performance, for both native and non-native listeners. Crucially, however, the effects of language background and of noise did not interact. Of the possible patterns of results we listed in the Introduction, the one we have observed is a constant disadvantage for non-native compared with native listeners, irrespective of the degree of noise-masking. Thus our study clearly suggests that the disproportionate effects of noise on listening to non-native, as opposed to native, language are not due to exacerbation of the difficulty of phoneme identification.

We did observe effects of the native inventory on the non-native listeners' identifications. Where the non-native listeners performed as well as the native listeners (for instance, in the vowel confusion matrices, for the vowels of *beat*, *bit*, and *boy*), it was for phonemes which occupy highly similar positions in the two inventories (Gussenhoven, 1999; Ladefoged, 1999). Where the inventories mismatched, non-native performance often fell behind. However, these native-inventory effects were not heightened under noise. If anything, the similarity of non-native and native performance became stronger under noise, as our correlation analyses showed.

Phonological constraints of the native language also affected performance; the absence of a voicing contrast in final position in Dutch was reflected in the non-native listeners' poor performance on voicing decisions in final position [Fig. 3(e)], which led to a reversal, for non-native consonant identifications only, of the otherwise constant advantage of final over initial position with our stimuli (Fig. 2). Again, however, this effect did not significantly interact with the effects of SNR; if anything, the advantage of initial over final voicing decisions for Dutch listeners was actually greater at 16 dB than under more severe noise.

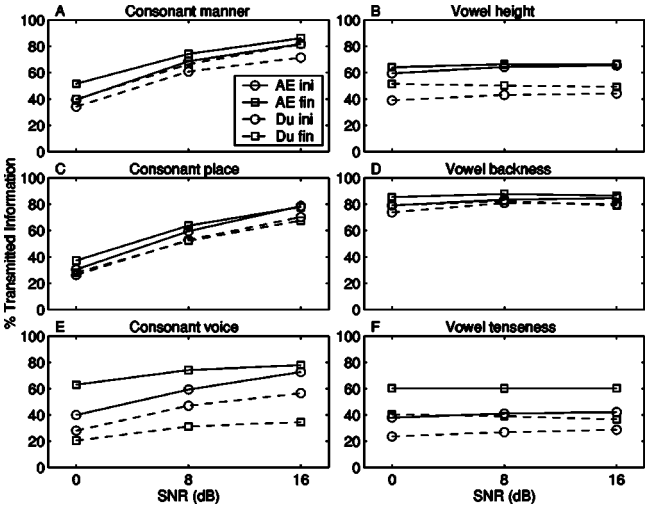


FIG. 3. Percentages of transmitted information for six phonological features as a function of SNR. Data are presented separately by position ("ini" =initial, "fin"=final) and language group ("AE"=American English, "Du"=Dutch), and have been pooled across phonetic contexts and subjects.

TABLE IX. Feature system used for the information-theoretical analyses.

Feature	Values	Phonemes
Consonant manner	stop	/p t k b d g/
	affricate	/tʃ dʒ/
	fricative	/f θ s ʃ h v ð z ʒ/
	liquid	/l r/
	glide	/j w/
	nasal	/m n ŋ/
Consonant place	labial	/p f b v m w/
	dental	/θ ð/
	alveolar	/t s d z n l/
	palatal	/ʃ tʃ ʒ dʒ r/
	velar	/k g ŋ/
	glottal	/h/
Consonant voice	voiced	/b d g v ð z ʒ dʒ j m n ŋ l r w/
	voiceless	/p t k f θ s ʃ tʃ h/
Vowel height	high	/i ɪ ʊ u/
	mid	/eɪ ɛ ʌ oʊ ɚ/
	low	/æ ɑ ɔ/
Vowel backness	front	/i ɪ eɪ ɛ æ/
	central	/ɚ/
	back	/ɑ ʌ ɔ oʊ ʊ u/
Vowel tenseness	tense	/i eɪ ɔ oʊ u ɚ/
	lax	/ɪ ɛ æ ɑ ʌ ʊ/

Thus although the non-native listeners in our study unquestionably performed below native phoneme-identification levels, they did so at more advantageous SNRs as well as under more severe noise, and the degree to which they suffered additional difficulty appeared to remain fairly constant across SNRs within the range tested here. In all our analyses, adverse effects of noise on non-native listening seemed to parallel adverse effects of noise for native listeners. We conclude from these results that it is not disproportionately increasing problems of phoneme identification that underlie the extra difficulty of listening to non-native language in noise.

Instead, we suggest that non-native listening is disproportionately affected by noise because non-native listening is, at all processing levels, slower and less accurate than native listening. Phoneme identification is, as we have seen, less accurate. Phoneme identification problems may be particularly important in that all later levels of processing will be affected by the decisions made at the phonemic level; but at all later levels, non-native listening is also less efficient. Segmentation of continuous speech into words is less efficient, because of interference from native prosodic expectations (Cutler *et al.*, 1986; Cutler and Otake, 1994) and from native phonotactic expectations (Weber, 2001). Lexical recognition is less efficient: phoneme identification problems can cause pseudo-homophony (Japanese listeners may have difficulty distinguishing *right* from *light*, Dutch listeners may confuse *bat* with *bet*), and this can lead to additional competition in the word recognition process (Broersma, 2002; Weber and Cutler, 2004). Spurious competition also arises from the native vocabulary, while native recognition is less likely to be affected by competition from nondominant non-native languages (Weber and Cutler, 2004). Syntactic processing is less efficient, even at high levels of proficiency in the non-native language (Sorace, 1993); prosodic distinctions between idiomatic and literal utterances are less efficiently processed (Vanlancker-Sidtis, 2003); and semantic processing, including the exploitation of prosody for information structure, is less efficient (Akker and Cutler, 2003). The effect of disadvantageous listening conditions, such as a babble of voices, is to slow down the process from the beginning, allowing the cumulative effects of lesser efficiency at all levels to become more noticeable and perhaps to exceed thresholds of auditory memory storage. Compensatory sources of information which all listeners will call upon under difficult listening conditions—knowledge of relative lexical frequencies of occurrence, of transitional probabilities, and of contextual plausibility—will also be less extensive, and less efficiently exploited, in non-native listening [as, indeed, Florentine and her colleagues observed (Florentine, 1985a; Mayo *et al.*, 1997)]. Interestingly, Van Wijngaarden *et al.* (2002) showed that a measure of linguistic entropy (letter-by-letter guessing of visually presented materials) significantly predicted the speech recognition performance of non-native listeners in noise; these authors therefore also concluded that less effective use of context, especially reduced exploitation of semantic redundancy, was a major factor in non-native listening difficulty in noise.

V. CONCLUSION

The identification of phonemes is adversely affected by increasing noise to a similar extent for native and for non-native listeners. Non-native identification scores were around 80% of native identification scores at each of the SNRs used in the present study. This pattern of results suggests that the robustly observed disproportionate difficulty which non-native listeners experience with speech in noisy conditions cannot be simply attributed to exacerbation of phoneme misidentification by noise interference; instead, it may reflect cumulative effects of lesser efficiency at all levels of processing, and lesser ability to exploit contextual redundancy.

ACKNOWLEDGMENTS

Participant testing in Florida was enabled by postdoctoral support to AW from NICHD Grant No. 00323 to Winifred Strange, whose support with this project is gratefully acknowledged. Participant testing in Nijmegen was supported by a research stipend from the Max Planck Society to NC. Further support was provided by a SPINOZA award from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek to AC. The native listening results were reported to the 15th International Congress of Phonetic Sciences, Barcelona (Weber and Smits, 2003), and the non-native listening results to the 8th International Conference on Spoken Language Processing, Jeju, Korea (Cooper and Cutler, 2004). We further thank Natasha Warner for recording the speech materials used in this study, and Randy Diehl, José Benkí, Sander van Wijngaarden, and an anonymous reviewer for comments which helped us improve this report.

APPENDIX A: PHONEMES USED IN THIS STUDY

Phonemes used in the study, with for each phoneme the illustrative word used to guide listeners' responses.

Final consonants		Initial consonants		Vowels	
/p/	liP	/p/	Pie	/i/	bEAt
/t/	hoT	/t/	Tie	/ɚ/	bIRd
/k/	siCK	/k/	Car	/u/	bOOt
/b/	graB	/b/	Be	/ɪ/	bIt
/d/	oDD	/d/	Do	/ʊ/	cOOk
/g/	eGG	/g/	Go	/eɪ/	wAIt
/f/	oFF	/f/	Far	/ɔ/	cAUght
/θ/	paTH	/θ/	THin	/ʌ/	cUt
/s/	paSS	/s/	See	/ɛ/	bEt
/v/	loVE	/v/	Very	/ɑ/	hOt
/ð/	smooTH	/ð/	THere	/aɪ/	bUY
/z/	buZZ	/z/	Zoo	/oʊ/	bOAt
/m/	aM	/m/	My	/æ/	bAt
/tʃ/	suCH	/tʃ/	CHin	/ɔɪ/	bOY
/dʒ/	eDGE	/dʒ/	Joke	/aʊ/	shOUt
/n/	oN	/n/	No		
/ʃ/	fiSH	/ʃ/	SHe		
/ʒ/	beiGE	/w/	Win		
/ŋ/	riNG	/j/	Yell		
/l/	iLL	/h/	Hi		
/r/	faR	/l/	Lie		
		/r/	Row		

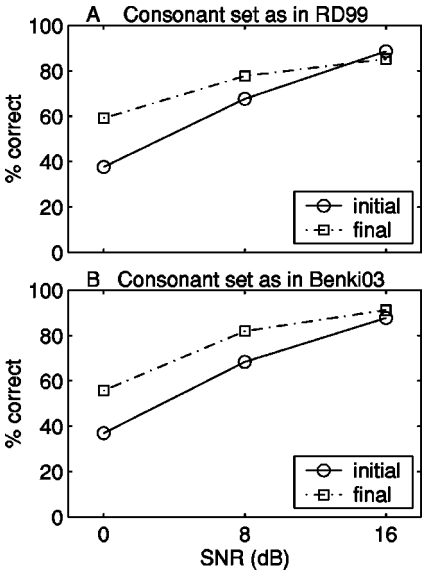


FIG. 4. Percentages of correctly recognized consonants in initial and final position, when consonant sets are restricted to those used by Redford and Diehl (1999), panel (a), and Benkaí (2003a,b), panel (b). Only the data for the American English listeners are given. In the calculations for panel (a), voicing errors have been disregarded, as in Redford and Diehl (1999).

APPENDIX B: ANALYSES OF OUR NATIVE DATA IN COMPARISON WITH REDFORD AND DIEHL (1999) AND BENKI (2003a)

Since we did not find the consistent advantage for initial over final consonants which the listening in noise studies of Redford and Diehl (1999) and Benkí (2003a) had led us to expect, we conducted analyses on our consonant data set in direct comparison with the previous work. These earlier studies differed from ours *inter alia* in that they tested a narrower range of phonemes. We thus extracted from our native listening data set the subset most closely matching the data presented in each of those studies; the relevant subsets are displayed in Fig. 4.

Figure 4(a) shows the American English listeners' identification performance for the voiceless stops and fricatives [p,t,k,f,θ,s,ʃ], with voicing errors ignored [as reported by Redford and Diehl (1999)]. It can be seen that the advantage for final over initial positions holds over 0 and 8 dB SNR in our data set, but disappears—to be replaced by a marginal advantage for initial over final position—at 16 dB SNR. Redford and Diehl used a SNR of 15 dB; thus, at the conditions most closely approximating theirs, we find a result more similar to theirs. At less favorable SNRs, however, the advantage for final position is also robust with this subset.

Figure 4(b) presents the American English listeners' identification performance for the ten initial consonants [b,p, ,d,t,k,s,h,m,l,r] and the ten final consonants [p,d,t,g,k, ,s,z,m,n,l] as used by Benkí (2003a). Again, the final position advantage is stronger at 0 and 8 dB SNR than at 16 dB SNR. However, Benkí used less favorable SNRs (from −14 to −5 dB) than any used in our study, so that our final position advantage with the same subset contrasts with his results.

There were further differences between the studies: we used just one speaker (Redford and Diehl had seven); final stops were released in Benkí's and our stimuli but not in Redford and Diehl's; neither earlier study used multi-speaker babble noise; in both earlier studies, stimuli were CVC syllables in a carrier phrase (respectively *Say—some more* or *Say—again*, and *You will write—please*), while in our study, listeners heard CV or VC syllables in isolation. We chose this format because we were interested in the implications of our findings for natural listening in noise, in which precise onsets are not predictable, and certainly are not accompanied by a constant preceding context. In our study, the syllables varied in length and were centrally embedded in the longer sample of noise, so that the moment of onset of the syllable to be identified was unpredictable and not cued by the preceding context. Under these conditions, the initial vowel or consonant was generally somewhat difficult to identify.

Redford and Diehl (1999) interpreted their positional finding as a result of greater articulatory distinctiveness of initial consonants, a result supported by acoustic evidence that their speakers' initial consonants were longer, louder, and different in fundamental frequency from the final consonants. Benkí (2003a) similarly cited articulatory differences as a likely source of his initial-position advantage. Note also that Benkí (2003b) found that the disadvantage of final consonants largely disappeared when the stimuli presented were words in sparse phonetic neighborhoods, making the final consonant relatively more probable. It seems clear that the relative perceptibility of phonemes as a function of position is not constant, but depends upon the particular characteristics of stimuli and procedure used in a phoneme identification experiment.

Akker, E., and Cutler, A. (**2003**). ''Prosodic cues to semantic structure in native and nonnative listening,'' Bilingualism: Language and Cognition **6**, 81–96.

Benkí, J. R. (**2003a**). ''Analysis of English nonsense syllable recognition in noise,'' Phonetica **60**, 129–157.

Benkí, J. R. (**2003b**). ''Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition,'' J. Acoust. Soc. Am. **113**, 1689–1705.

Best, C. T. (**1995**). ''A direct realist view of cross-language speech perception,'' in *Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Booij, G. (**1995**). *The Phonology of Dutch* (Clarendon P, Oxford).

Broersma, M. (**2002**). ''Comprehension of non-native speech: Inaccurate phoneme processing and activation of lexical competitors,'' in *Proceedings of the 7th International Conference on Spoken Language Processing* (Center for Spoken Language Research, University of Colorado Boulder, Denver) (CD-ROM), pp. 261–264.

Conrad, L. (**1989**). ''The effects of time-compressed speech on native and EFL listening comprehension,'' Stud. Second Language Acquisition **11**, 1–16.

Cooper, N., and Cutler, A. (**2004**). ''Perception of non-native phonemes in noise,'' in *Proceedings of the 8th International Conference on Spoken Language Processing* (Jeju, Korea).

Cutler, A., and Otake, T. (**1994**). ''Mora or phoneme? Further evidence for language-specific listening,'' J. Memory Lang. **33**, 824–844.

Cutler, A., Mehler, J., Norris, D., and Seguí, J. (**1986**). ''The syllable's differing role in the segmentation of French and English,'' J. Memory Lang. **25**, 385–400.

Florentine, M. (**1985a**). ''Non-native listeners' perception of American-English in noise,'' in *Proceedings of Inter-Noise '85*, pp. 1021–1024.

Florentine, M. (**1985b**). ''Speech perception in noise by fluent, non-native listeners,'' Proc. Acoust. Soc. Japan **26**, 1–8.

Gat, I. B., and Keith, R. W. (**1978**). ''An effect of linguistic experience,'' Audiology **17**, 339–345.

Greene, B. G., Pisoni, D. B., and Gradman, H. L. (**1985**). ''Perception of synthetic speech by nonnative speakers of English,'' Speech Research Laboratory, Progress Report 11, Indiana Univ.
Gussenhoven, C. (**1999**). ''Dutch,'' in *Handbook of the International Phonetic Association* (Cambridge U.P., Cambridge, UK), pp. 74–77.
Hazan, V., and Simpson, A. (**2000**). ''The effect of cue-enhancement on consonant intelligibility in noise: Speaker and listener effects,'' Lang. Speech **43**, 273–294.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.
Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). ''Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability,'' J. Acoust. Soc. Am. **61**, 1337–1351.
Kenstowicz, M. (**1994**). *Phonology in Generative Grammar* (Blackwell, Cambridge, MA).
Kreul, E. J., Nixon, N. C., Kryter, K. D., Bell, D. W., Lang, J. S., and Schubert, E. G. (**1968**). ''A proposed clinical test of speech discrimination,'' J. Speech Hear. Res. **11**, 536–552.
Ladefoged, P. (**1999**). ''American English,'' in *Handbook of the International Phonetic Association* (Cambridge U.P., Cambridge, UK), pp. 41–44.
Mack, M. (**1988**). ''Sentence processing by non-native speakers of English: Evidence from the perception of natural and computer-generated anomalous L2 sentences,'' J. Neurolinguist. **3**, 293–316.
Mayo, L. H., Florentine, M., and Buus, S. (**1997**). ''Age of second-language acquisition and perception of speech in noise,'' J. Speech Hear. Res. **40**, 686–693.
Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.
Nábělek, A. K., and Donahue, A. M. (**1984**). ''Perception of consonants in reverberation by native and non-native listeners,'' J. Acoust. Soc. Am. **75**, 632–634.
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Pisoni, D. B. (**1987**). ''Some measures of intelligibility and comprehension,'' in *From Text to Speech: The MITalk System*, edited by J. Allen, D. H. Klatt, and S. Hunnicutt (Cambridge U.P., Cambridge, UK), pp. 151–171.
Redford, M. A., and Diehl, R. L. (**1999**). ''The relative perceptual distinctiveness of initial and final consonants in CVC syllables,'' J. Acoust. Soc. Am. **106**, 1555–1565.
Smits, R. (**2000**). ''Temporal distribution of information for human consonant recognition in VCV utterances,'' J. Phonetics **27**, 111–135.
Smits, R., Warner, N., McQueen, J. M., and Cutler, A. (**2003**). ''Unfolding of phonetic information over time: A database of Dutch diphone perception,'' J. Acoust. Soc. Am. **113**, 563–574.
Sorace, A. (**1993**). ''Unaccusativity and auxiliary choice in non-native grammars of Italian and French: Asymmetries and predictable indeterminacy,'' J. French Lang. Studies **3**, 71–93.
Strange, W. (**1995**). *Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research* (York P, Timonium, MD).
Takata, Y., and Nábělek, A. K. (**1990**). ''English consonant recognition in noise and in reverberation by Japanese and American listeners,'' J. Acoust. Soc. Am. **88**, 663–666.
Van Wijngaarden, S., Steeneken, H., and Houtgast, T. (**2002**). ''Quantifying the intelligibility of speech in noise for non-native listeners,'' J. Acoust. Soc. Am. **111**, 1906–1916.
Vanlancker-Sidtis, D. (**2003**). ''Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one,'' Appl. Psycholinguist. **24**, 45–57.
Wang, D. M., and Bilger, R. C. (**1973**). ''Consonant confusions in noise: A study of perceptual features,'' J. Acoust. Soc. Am. **54**, 1248–1266.
Weber, A. (**2001**). *Language-specific Listening: The Case of Phonetic Sequences*, MPI Series in Psycholinguistics, Vol. 16, Ph.D. dissertation, University of Nijmegen, The Netherlands.
Weber, A., and Cutler, A. (**2004**). ''Lexical competition in non-native spoken-word recognition,'' J. Memory Lang. **50**, 1–25.
Weber, A., and Smits, R. (**2003**). ''Consonant and vowel confusion patterns by American English listeners,'' in *Proceedings of the 15th International Congress of Phonetic Sciences* (Palau de Congressos, Barcelona, Spain), pp. 1437–1440.

Analysis of speech-based speech transmission index methods with implications for nonlinear operations

Ray L. Goldsworthy and Julie E. Greenberg[a)]
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and Harvard-MIT Division of Health Sciences and Technology, Cambridge, Massachusetts 02139

(Received 1 October 2003; revised 16 April 2004; accepted 18 August 2004)

The Speech Transmission Index (STI) is a physical metric that is well correlated with the intelligibility of speech degraded by additive noise and reverberation. The traditional STI uses modulated noise as a probe signal and is valid for assessing degradations that result from linear operations on the speech signal. Researchers have attempted to extend the STI to predict the intelligibility of nonlinearly processed speech by proposing variations that use speech as a probe signal. This work considers four previously proposed speech-based STI methods and four novel methods, studied under conditions of additive noise, reverberation, and two nonlinear operations (envelope thresholding and spectral subtraction). Analyzing intermediate metrics in the STI calculation reveals why some methods fail for nonlinear operations. Results indicate that none of the previously proposed methods is adequate for all of the conditions considered, while four proposed methods produce qualitatively reasonable results and warrant further study. The discussion considers the relevance of this work to predicting the intelligibility of cochlear-implant processed speech. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1804628]

PACS numbers: 43.71.Gv, 43.60Wy, 43.71Ky [KWG] Pages: 3679–3689

I. INTRODUCTION

Early attempts to predict speech intelligibility led to the development of the articulation index (AI) (French and Steinberg, 1947; Kryter, 1962a, 1962b). A fundamental principle of the AI is that the intelligibility of speech depends on a weighted average of the signal to noise ratios (SNRs) in frequency bands spanning the speech spectrum. By accounting for the contribution of different regions of the spectrum to intelligibility, the AI successfully predicts the effects of additive noise and simple linear filters.

The Speech Transmission Index (STI) (Houtgast and Steeneken, 1971; Steeneken and Houtgast, 1980; IEC, 1998) is an intelligibility metric that differs from the AI by using reduction in signal modulation rather than band-specific SNRs. By including modulation reduction in the frequency band analysis, the STI can predict the effects of reverberation as well as additive noise. Calculation of the STI is based on changes in signal modulation when modulated probe stimuli are transmitted through a channel of interest. The responses to probe stimuli are measured in multiple frequency bands for a range of modulation frequencies relevant to speech. The STI successfully quantifies the effects of room acoustics and broadcast channels on speech intelligibility (Steeneken and Houtgast, 1982). The STI can also be adapted for use with hearing-impaired subjects (Humes *et al.*, 1986; Ludvigsen, 1987; Payton *et al.*, 1994).

Steeneken and Houtgast (1980) suggest that applying the STI to nonlinear operations requires more sophisticated probe signals than used in their original procedure. They introduced complex test signals that combine modulated noise with artificial speech-like signals, allowing the STI to predict the effects of automatic gain control and peak clipping. Other researchers have developed variations that use speech, rather than an artificial probe, to investigate nonlinear operations. These speech-based methods have been used to analyze dynamic amplitude compression (Hohmann and Kollmeier, 1995; Payton *et al.*, 2002; Drullman, 1995), spectral subtraction (Ludvigsen *et al.*, 1993), and envelope clipping (Drullman, 1995). In addition, speech-based STI methods have been used to investigate the intelligibility differences between clear and conversational speech (Payton *et al.*, 1994; Payton *et al.*, 1999).

The speech-based STI methods have generally failed to predict performance for nonlinear operations. In some studies, STI intelligibility predictions have been qualitatively inconsistent with performance results. A study of envelope expansion found that ''the prediction from STI is in the wrong direction for the expansion conditions'' (Van Buuren *et al.*, 1998). In an investigation of speech-based STI and spectral subtraction, researchers concluded ''STI, even in its modified version, is an unreliable predictor when non-linear processes are involved.'' (Ludvigsen *et al.*, 1993). Other researchers (Drullman, 1995; Payton *et al.*, 2002; Hohmann and Kollmeier, 1995) have also concluded that speech-based STI methods proposed thus far do not adequately predict the intelligibility of nonlinearly processed speech.

In this work, the various speech-based STI methods are analyzed to determine why they fail to predict intelligibility for nonlinear operations. Simple modifications are proposed to overcome problems with the existing speech-based STI methods. This results in four modified speech-based STI methods that are related to previously proposed methods. These modified STI methods are well correlated with the traditional STI for additive noise and reverberation and also

[a)]Author to whom correspondence should be addressed at: Massachusetts Institute of Technology, Building 36, Room 761, 77 Massachusetts Ave., Cambridge, MA 02139; electronic mail: jgreenbe@mit.edu

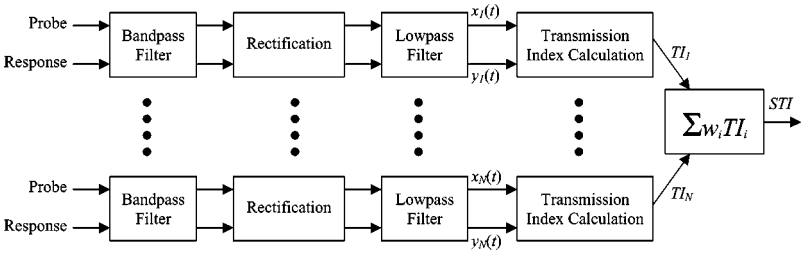


FIG. 1. General form of the STI calculation. For each frequency band, $i=1,...,N$, envelopes of the probe and response signals are compared to determine a transmission index (TI_i). The STI is a weighted average of TI_i values.

exhibit qualitatively reasonable behavior for selected nonlinear operations. As a result, the modified STI methods are promising candidates to predict intelligibility of nonlinearly processed speech.

II. BACKGROUND

Both the traditional and speech-based STI methods employ a frequency band analysis as illustrated in Fig. 1. A bank of bandpass filters splits the probe and response signals into frequency bands, where i indicates the frequency band number. Typically, octave bands with center frequencies from 125 to 8000 Hz are used. For each band, the probe and response envelope signals, $x_i(t)$ and $y_i(t)$, respectively, are computed by rectification and lowpass filtering and then compared to determine a transmission index, TI_i. The TI_i values are combined using a weighted average to determine the STI value. The various STI methods differ in how the envelope signals are computed and in how the TI_i values are computed from the envelopes.

A. Traditional method of computing the STI

For the traditional method (Steeneken and Houtgast, 1980), the TI_i values are determined from an intermediate function called the modulation transfer function (MTF). The MTF is a function of modulation frequency, f, calculated individually for each value of f. For each frequency band, the probe signal consists of speech-shaped noise that has been bandpass filtered and then intensity modulated at a particular modulation frequency. The probe signal is passed through the system to be evaluated. The fractional change in modulation depth between probe and response intensity envelopes is quantified for that value of f, and the process is repeated for other modulation frequencies to determine the complete MTF for one frequency band. The MTF is typically characterized using modulation frequencies ranging from $f=0.63$ Hz to $f=12.7$ Hz in one-third octave intervals. As an alternative to artificial probe signals, Houtgast and Steeneken (1985) proposed determining the MTF for each frequency band from spectra of the intensity envelopes of running speech. Omitting the subscript i to simplify notation, this approach can be described as (Drullman, 1994b)

$$\mathrm{MTF}(f)=\alpha\frac{|Y(f)|}{|X(f)|}=\alpha\sqrt{\frac{S_{yy}(f)}{S_{xx}(f)}}, \tag{1}$$

where $\alpha=\mu_x/\mu_y$, $\mu_x=E\{x(t)\}$, $\mu_y=E\{y(t)\}$, and $E\{\cdot\}$ denotes expected value. $|X(f)|$ and $|Y(f)|$ are magnitude spectra, and $S_{xx}(f)$ and $S_{yy}(f)$ are power spectra, of the probe and response envelope signals, respectively.

The signal-to-noise ratio (SNR) in decibels as a function of f is calculated for each frequency band as

$$\mathrm{SNR}_i(f)=10\log_{10}\left(\frac{\mathrm{MTF}_i(f)}{1-\mathrm{MTF}_i(f)}\right). \tag{2}$$

An overall apparent SNR (aSNR_i) for each frequency band is determined by clipping the $\mathrm{SNR}_i(f)$ values and then averaging across modulation frequencies, that is,

$$\mathrm{cSNR}_i(f)=\begin{cases}-15, & \mathrm{SNR}_i(f)<-15\\ \mathrm{SNR}_i(f), & -15\leqslant\mathrm{SNR}_i(f)\leqslant 15\\ 15, & \mathrm{SNR}_i(f)>15,\end{cases} \tag{3}$$

$$\mathrm{aSNR}_i=\mathrm{mean}(\mathrm{cSNR}_i(f)). \tag{4}$$

The transmission index is a linear function of the apparent SNR for each band, defined to be between zero and one,

$$\mathrm{TI}_i=\frac{\mathrm{aSNR}_i+15}{30}. \tag{5}$$

Finally, the overall STI value is calculated as a weighted average of the TI_i values,

$$\mathrm{STI}=\sum_i w_i\mathrm{TI}_i, \tag{6}$$

where w_i is a psycho-acoustically derived weighting. The weights, w_i, are defined to sum to one, thereby restricting the STI values to a range between zero and one.

B. Speech-based STI methods

This section summarizes four speech-based methods proposed in the literature. The first three speech-based methods use intensity envelopes calculated by squaring and then smoothing, while the fourth uses magnitude envelopes. For each method, the description focuses on the calculation of TI_i for one frequency band. To simplify notation, the subscript i is omitted for intermediate variables such as $\mathrm{MTF}(f)$ and aSNR.

1. Magnitude cross-power spectrum method

Payton and colleagues (2002) proposed a speech-based method where the MTF is based on the magnitude of the cross-power spectra as given by

$$\mathrm{MTF}(f)=\alpha\left|\frac{S_{xy}(f)}{S_{xx}(f)}\right|, \tag{7}$$

where $S_{xy}(f)$ is the cross-power spectrum of the probe and response envelopes. The MTF given by Eq. (7) is used in Eq. (2), and the STI is calculated from Eqs. (2) through (6).

2. Real cross-power spectrum method

Drullman and colleagues (1994b) introduced a phase-locked MTF in order to investigate the effects of reducing low-frequency modulations on the intelligibility of speech. The phase-locked MTF is defined as

$$\mathrm{MTF}(f)=\alpha\,\mathrm{Re}\left(\frac{S_{xy}(f)}{S_{xx}(f)}\right), \tag{8}$$

where $\mathrm{Re}(\cdot)$ denotes taking the real part of the complex-valued function. Although they did not propose a corresponding STI calculation procedure, the MTF in Eq. (8) could be used to calculate the STI in conjunction with Eqs. (2) through (6).

3. Envelope regression method

Ludvigsen and colleagues (1990) proposed a method where the probe envelope signal, $x(t)$, and the response envelope signal, $y(t)$, are compared using linear regression analysis. In this method, the apparent SNR for each frequency band is defined as

$$\mathrm{aSNR}=10\log_{10}\left(\frac{A\mu_x}{B}\right), \tag{9}$$

where A and B are the parameters that produce the best fit for the model $y(t)=Ax(t)+B$. This apparent SNR is clipped to values between ± 15 dB, and the STI is calculated via Eqs. (5) and (6).

4. Normalized covariance method

The normalized covariance method (Koch, 1992; Holube and Kollmeier, 1996) is based on the covariance between the probe and response envelope signals. For each frequency band, the apparent SNR is calculated as

$$\mathrm{aSNR}=10\log_{10}\left(\frac{r^2}{1-r^2}\right), \tag{10}$$

where r is the normalized covariance between $x(t)$ and $y(t)$ given by

$$r^2=\frac{\lambda_{xy}^2}{\lambda_x\lambda_y} \tag{11}$$

with

$$\lambda_{xy}=E\{(x(t)-\mu_x)(y(t)-\mu_y)\} \tag{12}$$

and

$$\lambda_x=E\{(x(t)-\mu_x)^2\}. \tag{13}$$

The apparent SNR of Eq. (10) is clipped to values between ± 15 dB and the STI is calculated via Eqs. (5) and (6).

5. Summary of speech-based methods

The above-described speech-based methods all compute the STI as a weighted sum of TI values determined from the envelopes of the probe and response signals in each frequency band. The key difference among the methods is how the TI values are calculated. Table I summarizes the intermediate modulation metrics used to calculate TI values for the different methods. In the case of the envelope regression method, the modulation metric in Table I is an alternate form that is derived in Appendix A. For the two cross-power spectrum methods, the modulation metric is a function of modulation frequency. For the other two methods there is a single value for each frequency band. The implications of this fundamental difference are discussed in Sec. VI A. In the following sections, these modulation metrics will be used to yield insight into the behavior of the speech-based STI methods.

TABLE I. Intermediate modulation metrics for speech-based STI methods proposed in the literature. These metrics use the normalization term $\alpha=\mu_x/\mu_y$. They are calculated for each frequency band and then combined to produce a single STI value as described in the text.

Magnitude cross-power spectrum	Real cross-power spectrum	Envelope regression	Normalized covariance
$\mathrm{MTF}(f)=\alpha\left\lvert\frac{S_{xy}(f)}{S_{xx}(f)}\right\rvert$	$\mathrm{MTF}(f)=\alpha\,\mathrm{Re}\left(\frac{S_{xy}(f)}{S_{xx}(f)}\right)$	$M=\alpha\frac{\lambda_{xy}}{\lambda_x}$	$r^2=\frac{\lambda_{xy}^2}{\lambda_x\lambda_y}$

III. PROPOSED METRICS

A. Normalization based on noise envelope

Both cross-power spectrum methods [Eqs. (7) and (8)] include the term α, which normalizes the envelopes to account for the power of the probe and response signals. The alternate form of the envelope regression method derived in Appendix A also depends on α; for this method the apparent SNR in Eq. (9) can be expressed as

$$\mathrm{aSNR}=10\log_{10}\left(\frac{M}{1-M}\right), \tag{14}$$

where M is a modulation metric defined as

$$M=\alpha\frac{\lambda_{xy}}{\lambda_x}. \tag{15}$$

Thus, the envelope regression method, as well as the two cross-power spectrum methods, include the normalization term α. This term successfully normalizes the envelopes for the cases of additive noise and reverberation; however, for a large class of operations this normalization ratio is not appropriate. In particular, when the processing reduces the overall amplitude of the response envelope, $y(t)$, α may increase without bound. As shown in Secs. V B and V C, this leads to invalid values of the intermediate modulation metrics listed in Table I.

An alternative normalization term is proposed here. The noise envelope is defined as

$$z(t)=|y(t)-x(t)|, \tag{16}$$

and a new normalization term is defined as

$$\beta=\frac{\mu_x}{\mu_x+\mu_z}. \tag{17}$$

For cases when $y(t)>x(t)$ for all t (as is typically the case for additive noise and reverberation) then $\mu_z=\mu_y-\mu_x$ and, consequently, $\beta=\alpha$. Thus, for certain operations, the pro-

TABLE II. Intermediate modulation metrics for speech-based STI methods proposed in this work. These metrics use the normalization term β as defined in Eq. (17). They are calculated for each frequency band and then combined to produce a single STI value as described in the text.

Magnitude cross-power spectrum	Real cross-power spectrum	Envelope regression	Normalized correlation
$\mathrm{MTF}(f)=\beta\left\lvert\frac{S_{xy}(f)}{S_{xx}(f)}\right\rvert$	$\mathrm{MTF}(f)=\beta\,\mathrm{Re}\left(\frac{S_{xy}(f)}{S_{xx}(f)}\right)$	$M=\beta\frac{\lambda_{xy}}{\lambda_x}$	$\rho^2=\frac{\phi_{xy}^2}{\phi_x\phi_y}$

posed normalization term equals the original.

When the processing reduces the response envelope so that $y(t)<x(t)$ for some values of t, then μ_y decreases, causing α to increase. In some cases, high values of α may result in erroneously high values of apparent SNR for that frequency band. Since $\mu_z+\mu_x$ is always greater than μ_x, β will avoid characterizing reduced response envelopes as improved SNR.

B. Normalized correlation

We hypothesize that the normalized covariance method (Sec. II B 4) is well suited to nonlinear operations. The normalized covariance defined in Eq. (11) is a metric that necessarily falls between zero and one, with a value of unity achieved only when the envelopes are identical. These constraints ensure that the method always produces valid values of the intermediate metric. For the other speech-based methods, the intermediate metrics in Table I are not restricted to values between zero and one, and operations that increase the modulation depth may cause the intermediate metrics to take on invalid values greater than one, as demonstrated in Secs. V B and V C.

As a variation on the normalized covariance method, we consider the normalized correlation,[1] ρ, where

$$\rho^2=\frac{\phi_{xy}^2}{\phi_x\phi_y} \tag{18}$$

with $\phi_{xy}=E\{x(t)y(t)\}$ and $\phi_x=E\{x^2(t)\}$. The STI is subsequently calculated by substituting ρ for r in Eq. (10), clipping to values between ± 15 dB, and applying Eqs. (5) and (6). The normalized correlation method differs from the normalized covariance method only in that the envelope means are included in the correlation terms.

Table II summarizes the intermediate modulation metrics for the proposed speech-based methods. Comparing Table II to Table I reveals the key differences between the methods proposed in this work and those proposed previously.

IV. METHODS

This section describes the calculation of the various speech-based STI methods for three sets of processing conditions: acoustic degradation, envelope thresholding, and spectral subtraction. For the acoustic degradation conditions, speech-based STI values are compared to the traditional STI. For the envelope thresholding and spectral subtraction conditions, the speech-based STI methods are characterized by intermediate modulation metrics for a single frequency band.

A. Common elements

For all speech-based STI methods, the probe stimulus was a 120 s speech signal formed by concatenating 42 phonetically balanced sentences (IEEE, 1969). For the traditional method, the probe stimulus was based on a 60 s noise sequence with the same long-term spectrum as the speech. In both cases the sampling rate was $F_s=22050$ Hz.

The bandpass filters were seven octave-band filters with center frequencies ranging from 125 Hz to 8 kHz. All filters were eighth-order Butterworths. Intensity envelopes were calculated by squaring the bandpass-filtered signals and lowpass filtering. Magnitude envelopes were calculated by full-wave rectification of the bandpass-filtered signals followed by lowpass filtering. In both cases the lowpass filter was an eighth-order Butterworth with 50 Hz cutoff frequency. Envelopes were downsampled to 200 Hz before calculating the various metrics. This resulted in discrete-time probe and response envelope signals, $x[n]$ and $y[n]$, that were $N=24\,000$ samples long for the speech sequence and $N=12\,000$ samples long for the noise sequence.

The octave band weighting function used in Eq. (6) was taken from Houtgast and Steeneken (1985). All processing was performed in MATLAB® on a personal computer with a Pentium III processor.

B. Metric calculation

1. Traditional method

The traditional STI was calculated using fourteen modulation frequencies ranging from $f=0.63$ to 12.7 Hz in one-third-octave increments. Because it requires the use of a probe noise sequence, it was only practical to compute the traditional STI for the acoustic degradation conditions. For each modulation frequency, the noise sequence described in Sec. IV A was amplitude modulated by $\sqrt{1+\cos(2\pi(f/F_s)n)}$ to form the probe signal. The response signal consisted of the probe signal combined with additive noise and/or reverberation. Both the probe and response signals were bandpass filtered into octave bands and intensity envelopes were computed by squaring followed by lowpass filtering. The modulation depth of each envelope was measured as the maximum value of the cross-covariance between the envelope and the function $\cos(2\pi(f/F_s)n)$ normalized by the envelope mean. The MTF value was determined from the ratio of the response envelope's modulation depth to the probe envelope's modulation depth.

2. Cross-power spectrum methods

Both the magnitude cross-power spectrum method and the real cross-power spectrum method use intensity envelopes. Sample envelope means were calculated from the average of the envelope signals. The MTF for the two cross-power spectrum methods requires estimating the auto- and cross-power spectra. This was accomplished using the periodogram method with 4096-point Hanning windows, 4096-point FFTs, and 50% overlap. The resulting 0.05 Hz fre-

quency bins were averaged to produce values in one-third octave intervals (Payton *et al.*, 1999) centered from 0.63 to 12.7 Hz. This resulted in averaging of three bins for the lowest modulation frequency and 60 bins for the highest modulation frequency. These quantities were used in Eqs. (7) and (8) for the original methods, and with β [Eq. (17)] in place of α for the proposed methods. Then STI was calculated via Eqs. (2) through (6).

3. Envelope regression method

The envelope regression method was calculated from the intensity envelopes using the alternate form derived in Appendix A. Sample envelope means were computed from the average of the envelope signals and the covariance was calculated as an unbiased estimate, that is,

$$\lambda_{xy}=E\{(x[n]-\mu_x)(y[n]-\mu_y)\}$$
$$\approx\left(\frac{1}{N-1}\right)\sum_{i=1}^{N}(x[i]-\mu_x)(y[i]-\mu_y). \tag{19}$$

For each frequency band, the modulation metric, M, was calculated using Eq. (15) for the existing method and with β in place of α for the proposed method. The apparent SNR was then calculated from Eq. (14), clipped to values between ± 15 dB, and used in Eqs. (5) and (6).

4. Normalized covariance and normalized correlation methods

The normalized covariance and normalized correlation methods were calculated based on magnitude envelopes. For each frequency band, the normalized covariance, r, was calculated from Eq. (11), with estimates of the variance and covariance calculated as in Eq. (19). The normalized correlation, ρ, was calculated according to Eq. (18) with the correlation estimated as

$$\phi_{xy}=E\{x[n]y[n]\}\approx\left(\frac{1}{N-1}\right)\sum_{i=1}^{N}(x[i]\cdot y[i]). \tag{20}$$

The apparent SNRs were calculated from Eq. (10) (replacing r with ρ for the normalized correlation method), clipped to values between ± 15 dB, and used in Eqs. (5) and (6).

C. Acoustic degradations

For the acoustic degradation conditions, speech-shaped noise was added to the probe stimulus and the resulting signal was convolved with a reverberant impulse response. The speech-shaped noise had the same long-term spectrum as the probe stimulus. Two-second-long reverberant impulse responses were generated using a room simulation based on the image method (Allen and Berkley, 1979). The speech-shaped noise was scaled to produce SNRs between -15 and 30 dB in 3 dB increments as well as a no-noise condition. Reverberation times (T_{60}) ranged from 0 to 1.5 s in 0.3 s increments. The traditional and speech-based STIs were computed for all combinations of SNR and reverberation time.

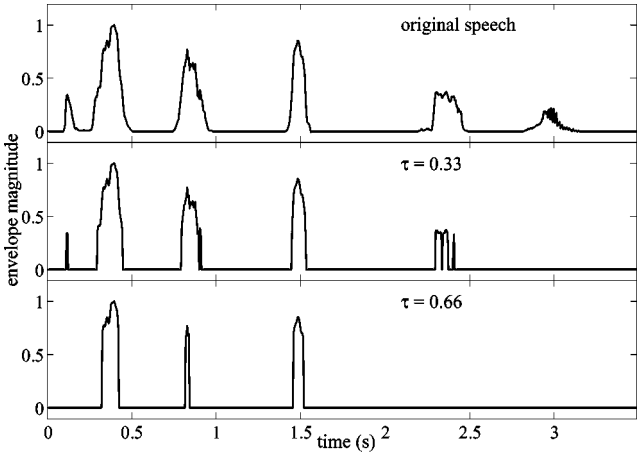


FIG. 2. Effect of envelope thresholding on a speech envelope for the octave band centered at 1 kHz, shown for two values of the fractional threshold, τ.

D. Envelope thresholding

Envelope thresholding is a nonlinear operation that consists of setting to zero any samples of the original envelope that are below a threshold, that is

$$y[n]=\begin{cases} x[n], & x[n]\geqslant\tau\max(|x[n]|) \\ 0, & x[n]<\tau\max(|x[n]|), \end{cases} \tag{21}$$

where $x[n]$ and $y[n]$ are the probe and response envelopes, respectively, and τ is a fractional threshold relative to the maximum value of the probe envelope. Figure 2 illustrates the effect of the envelope thresholding on a speech envelope and shows that increasing the value of the threshold results in greater levels of modulation and increasingly distorted envelopes. Intermediate modulation metrics were calculated for all speech-based STI methods for values of τ ranging from 0 to 1 in increments of 0.02.

E. Spectral subtraction

Spectral subtraction attempts to reduce background noise by subtracting a spectral estimate of the noise from short-time spectra of the noisy signal. Generalized spectral subtraction (Lim and Oppenheim, 1979) scales the noise spectral estimate by a constant factor, that is,

$$|P(F)|=|D(F)|-\kappa|\hat{N}(F)|, \tag{22}$$

where $D(F)$ is a short-time spectrum of the input signal, $\hat{N}(F)$ is the spectral estimate of the noise, $P(F)$ is the processed spectrum, and κ is a parameter that scales the noise estimate. $|P(F)|$ is multiplied by the phase of the original input signal and short-time reconstruction is performed to produce the time-domain output signal.

Figure 3 illustrates the effects of spectral subtraction on speech envelopes. For $\kappa=1$, the noise component of the envelope is suppressed with relatively little effect on the speech envelope. For $\kappa=8$, the noise is suppressed, but the speech envelope is highly distorted. Spectral subtraction with large values of κ is similar to envelope thresholding in that it distorts the envelope and increases the level of modulation.

The speech signal was degraded by noise with the same long-term spectrum as the probe stimulus (0 dB SNR) and

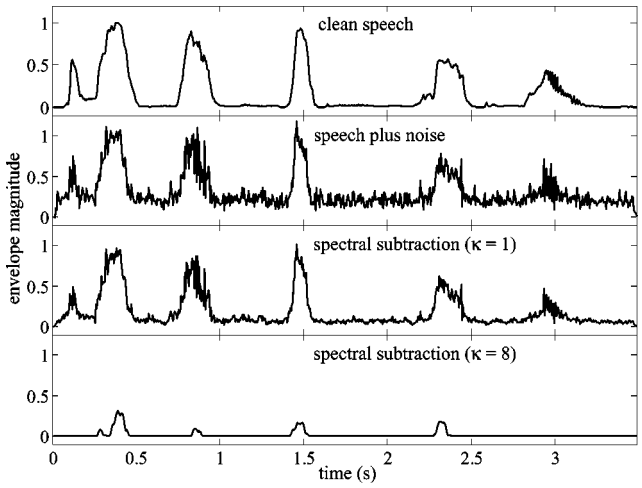


FIG. 3. Effect of spectral subtraction on the envelope of noisy speech for the octave band centered at 1 kHz, shown for two values of the control parameter, κ.

then processed by the spectral subtraction algorithm using the overlap-add method with 25 ms Hamming windows with 50% overlap. Intermediate modulation metrics were calculated for all speech-based STI methods for values of κ ranging from zero to eight in increments of 0.25. A value of $\kappa = 0$ corresponds to no spectral subtraction processing and a value of $\kappa = 1$ corresponds to standard spectral subtraction. A value of $\kappa = 8$ corresponds to an extreme version where the spectral subtraction processing eliminates all but the highest spectral peaks.

V. RESULTS

A. Acoustic degradation

Since the traditional STI method is well established as an accurate predictor of speech intelligibility for additive stationary noise and reverberation, any proposed speech-based method must produce similar values of STI under these conditions. Figure 4 compares the speech-based STI methods to the traditional STI for the acoustic degradation conditions of additive noise and reverberation described in Sec. IV C. Figures 4(a)–(d) show the four previously proposed speech-based methods described in Sec. II B, while Figs. 4(e)–(h) show the methods proposed in Sec. III. Each curve represents STI values calculated over the 45 dB range of SNRs for one level of reverberation.

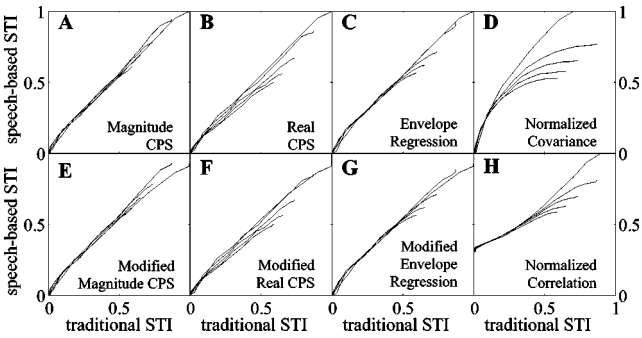


FIG. 4. Comparison of speech-based STI methods to the traditional STI. Each plot shows the relationship between one speech-based method and the traditional STI. Each curve corresponds to the 45 dB range of SNR values for one level of reverberation. More reverberant conditions terminate at lower values of the traditional STI.

In Fig. 4, complete agreement between the traditional STI method and a speech-based STI method would appear as a straight line from the bottom left to the top right of a particular plot. As seen in Figs. 4(a), (b), and (c), the original cross-power spectrum methods and the original envelope regression method all provide a reasonable match to the traditional method, although the real cross-power spectrum method is slightly less well-matched to the traditional than the other two.

Comparing Figs. 4(a), (b), and (c) to Figs. 4(e), (f), and (g) shows that for these acoustic degradation conditions, the modified methods using β as the normalization term are equivalent to the original methods using α. As described in Sec. III A, this equivalence is expected because the acoustic degradations increase the response envelopes relative to the probe envelopes.

The normalized covariance method [Fig. 4(d)] and the proposed normalized correlation method [Fig. 4(h)] are distinctly different from the other speech-based methods. The normalized covariance method does not exhibit a one-to-one relationship with the traditional method. The curves for different levels of reverberation are not superimposed, indicating that the normalized covariance method is not consistent with the traditional method in accounting for reverberation. Given the success of the traditional STI, this implies that the normalized covariance method will not be a good predictor of intelligibility for additive noise and reverberation. The normalized correlation method comes closer to having a one-to-one relationship to the traditional method, with some divergence at high SNRs. This implies that the normalized correlation method may perform poorly when accounting for the effects of reverberation in quiet and low-noise environments.

While the relationship between the normalized correlation method and the traditional STI is approximately one-to-one, they are not equivalent metrics. In other words, some mapping is required to transform the values produced by the normalized correlation method to values corresponding to the traditional STI. To the extent that a unique mapping does exist for these conditions, the new metric will retain the predictive power of the traditional STI for additive noise and reverberation.

B. Envelope thresholding

Figure 5 shows the effect of envelope thresholding on intermediate modulation metrics used to compute the various speech-based STI methods. Investigating these metrics, rather than the final STI values, is necessary to identify methods that produce invalid results. All of the intermediate modulation metrics have a valid range from zero to one, where zero indicates no preservation of the envelope modulations and one indicates perfect preservation. Values of the intermediate metric greater than one indicate a failure of the corresponding method.

Figures 5(a), (b), and (c) reveal that the original cross-power spectrum methods and the original envelope regression method fail for envelope thresholding. In all three plots,

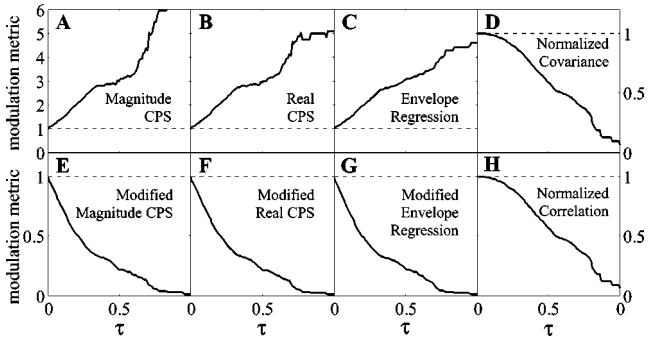


FIG. 5. Intermediate modulation metrics of speech-based STI methods for envelope thresholding as a function of threshold, τ. For the cross-power spectrum (CPS) methods, the intermediate metrics are the MTFs from Eqs. (7) and (8) averaged over modulation frequency. For the envelope regression methods, the intermediate metric is M [Eq. (15)]. For the normalized covariance and normalized correlation methods, the intermediate metrics are r [Eq. (11)] and ρ [Eq. (18)], respectively. All results are for the octave band centered at 1 kHz. The dotted line indicates unity, the maximum valid value for all metrics.

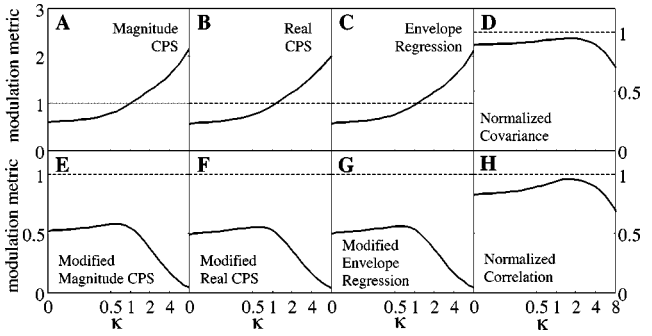


FIG. 6. Intermediate modulation metrics of speech-based STI methods (as in Fig. 5) for spectral subtraction as a function of control parameter, κ. All results are for the octave band centered at 1 kHz. The dotted line indicates unity, the maximum valid value for all metrics.

the modulation metrics increase above one as the threshold increases. These invalid values of the intermediate metrics indicate that these methods are not applicable to the nonlinear operation of envelope thresholding. The remaining five plots reveal that all of the proposed methods [Figs. 5(e)–(h)], as well as the normalized covariance method [Fig. 5(d)], produce valid values of the intermediate metrics. As the threshold increases, all of the intermediate metrics monotonically decrease from an initial value of one.

The general effect of envelope thresholding is to emphasize peaks in the envelope by setting low-amplitude samples of the envelope to zero. As the threshold increases, more samples are set to zero. Because this increases the modulation depth of the envelope, most of the previously proposed speech-based STI methods erroneously interpret this operation as increasing intelligibility beyond the initial value of one for speech in quiet. These methods fail because envelope thresholding reduces the mean of the response envelope, μ_y. Since it is the denominator of the normalization term, α, small values of μ_y can lead to extremely large values of α. Although envelope thresholding also reduces the cross-power spectrum, $S_{xy}(f)$, and cross-covariance, λ_{xy} [which contribute to the numerator of the modulation metrics in Eqs. (7), (8), and (15)], empirical observations indicate that as the threshold increases, these terms decrease more gradually than μ_y, leading to invalid values of the modulation metrics.

The modified methods that use β as the normalization term do not fail in this way because, for envelope thresholding, μ_z varies from zero to μ_x as the threshold goes from 0 to 1, corresponding to values of β ranging from 1 to 0.5 for the full range of envelope thresholding. This causes the intermediate metrics to decrease with increasing threshold.

The results for the three modified methods, as well as the normalized correlation and normalized covariance methods, are qualitatively consistent with the expected effect of envelope thresholding on the intelligibility of speech in quiet. The effect of increasing the threshold is to increase the distortion of the processed signal, thereby making it less intelligible. Increasing the threshold of a slightly different envelope manipulation has been shown to decrease intelligibility (Drullman, 1995). Therefore, the methods that account for envelope thresholding by decreasing as the threshold increases are viable candidates for speech-based STI.

C. Spectral subtraction

Figure 6 shows the effects of spectral subtraction on intermediate modulation metrics used to compute the various speech-based STI methods. Figures 6(a), (b), and (c) reveal that the original cross-power spectrum methods and the original envelope regression method fail for spectral subtraction. In all three plots, the modulation metrics increase monotonically as the control parameter, κ, increases and eventually reach invalid values greater than one. This indicates that these methods are not applicable to spectral subtraction. The remaining five plots reveal that all of the proposed methods [Figs. 6(e)–(h)], as well as the normalized covariance method [Fig. 6(d)], produce valid values of the intermediate metrics. As the control parameter increases, all of the intermediate metrics initially increase to a maximum and then decrease.

The proposed methods as well as the existing normalized covariance method exhibit behavior that is qualitatively consistent with a hypothetical trade-off between noise reduction and signal distortion. For each of these methods, the modulation metric initially increases, predicting slight improvements in intelligibility due to moderate levels of spectral subtraction ($\kappa \approx 1$), and then decreases, predicting degradations in intelligibility for more severe processing ($\kappa > 2$). The modified cross-power spectrum methods and the modified envelope regression method predict the most benefit from spectral subtraction with $\kappa = 0.6$, while the normalized covariance and normalized correlation method predict an optimum value of $\kappa = 1.4$. Further studies are required to determine if the proposed methods predict the intelligibility of speech processed by spectral subtraction and if they account for the effects of musical noise, an unpleasant artifact introduced by spectral subtraction (Goh *et al.*, 1998).

These results imply that spectral subtraction may improve the intelligibility of speech degraded by additive noise. A number of studies have shown that spectral subtraction does not improve the intelligibility of speech for normal-hearing listeners (Lim and Oppenheim, 1979). However,

spectral subtraction has been shown to improve intelligibility for cochlear implant listeners (Weiss, 1993; Hochberg *et al.*, 1992). This is discussed in more detail in Sec. VI B.

VI. DISCUSSION

A. Candidate speech-based STI methods

The results presented in the previous section indicate the suitability of the various speech-based STI methods for predicting intelligibility under conditions of acoustic degradation, envelope thresholding, and spectral subtraction. The long-term goal is to identify and validate a speech-based STI method that accurately predicts intelligibility of speech processed by a wide variety of linear and nonlinear operations. The immediate goal of this study is to identify speech-based STI methods that maintain a one-to-one relationship with the traditional STI for acoustic degradations while also producing qualitatively reasonable results for selected nonlinear operations.

Of the four original methods, only the normalized covariance method exhibited qualitatively reasonable behavior for the nonlinear operations considered in this study. However, this method does not have a one-to-one correspondence to the traditional STI for acoustic degradations. The other three previously proposed methods produce invalid results for the nonlinear operations considered. Therefore, we conclude that none of the four original methods are suitable for both conventional acoustic degradations and nonlinear operations.

The four proposed speech-based STI methods exhibit a one-to-one relationship with the traditional STI for acoustic degradations and produce qualitatively reasonable results for the nonlinear operations. However, the normalized correlation method may be less accurate for predicting the intelligibility of reverberant speech in quiet. Even so, all of the proposed methods are potential candidates to extend the STI to nonlinear operations while retaining its applicability to acoustic degradations. Additional work is required to determine if any of the proposed methods accurately predict speech intelligibility for these and other nonlinear operations.

Substantial differences exist among the four proposed methods. The two cross-power spectrum methods are computed using a modulation transfer function as the intermediate variable for each frequency band, and these MTFs are computed as a function of modulation frequency. In contrast, for the envelope regression and normalized correlation methods, the intermediate metrics consist of a single value for each frequency band and are not computed as functions of modulation frequency. However, it is shown in Appendix B that the normalized correlation method can be expressed as the energy-weighted average of an alternate MTF. The weights applied to the alternate MTF represent the proportion of the total energy in the probe envelope at each modulation frequency. A similar derivation can be performed for the envelope regression method but is complicated by the fact that the intermediate metric is based on covariance rather than correlation.

This interpretation of the normalized correlation and the envelope regression methods as the energy-weighted average of a MTF facilitates comparison with the cross-power spectrum methods. One area of concern relates to nonlinear operations that alter envelope spectra at modulation frequencies above 15 Hz. Such operations will not affect the STI values produced by the cross-power spectrum methods, because those methods only include modulation frequencies up to the one-third-octave band centered at 12.7 Hz. Indeed, there is evidence that modulation frequencies above 16 Hz provide only a marginal contribution to intelligibility (Drullman, 1994a). Because the envelope regression and normalized correlation methods use intermediate metrics that incorporate all frequencies in the envelopes (up to 50 Hz in the current implementation), one might expect these metrics to produce vastly different predictions of intelligibility for alterations in the envelope spectra above 15 Hz. However, since the intermediate metrics can be expressed as the energy-weighted average of a MTF, we must consider how much energy is present at higher modulation frequencies. For typical speech signals, less than 5% of the envelope energy occurs above 15 Hz. As a result, alterations to the envelope spectra above 15 Hz have only minor effects on the STI values produced by the envelope regression and normalized correlation methods.

The normalized correlation method and envelope regression methods can be calculated efficiently because they require estimates of envelope means and variances, which can be computed using running averages or windows of various lengths. The cross-power spectrum methods that calculate the MTF explicitly require at least several seconds of speech in order to estimate power spectra and cross-power spectra with a resolution less than 1 Hz, and calculating these spectra is computationally more intensive than calculating means and variances. Finally, because Figs. 4–6 illustrate that the behavior of the envelope regression method is similar to that of the cross-power spectrum methods, we conclude that the envelope regression method is a more practical choice than the two cross-power spectrum methods.

The normalized correlation method presents a substantial deviation from the traditional STI. The other proposed methods are equivalent to the traditional STI, that is, the speech-base STI values correspond directly to traditional STI values. However, as seen in Fig. 4, the normalized correlation method is not equivalent to the traditional STI, or is it a linear transformation of traditional STI. A (nonlinear) function is required to map the normalized correlation STI values to the traditional STI.

Another difference, illustrated in Figs. 5 and 6, is that the qualitative behavior of the normalized correlation method is substantially different from the other three proposed methods. As mentioned above, additional work is required to determine if any of the proposed methods accurately predict speech intelligibility. Note that although the normalized correlation method uses magnitude envelopes rather than the intensity envelopes used in the other methods, the major differences in qualitative behavior cannot be attributed to this difference in envelope computation. The normalized correlation metric is admittedly a departure from many of the principles of the traditional STI, and it may be preferable to consider it a new intelligibility metric distinct from the STI

except for the common elements of using frequency-band envelopes.

B. Predicting intelligibility of cochlear-implant processed speech

The STI has already been adapted for use with hearing-impaired subjects (Humes *et al.*, 1986; Payton *et al.*, 1994), and it is a good candidate for predicting intelligibility of speech processed by cochlear-implant (CI) speech processors. This expectation is based primarily on similarities between the STI calculation procedure and CI processing strategies; both the STI and conventional CI processing strategies use information from the envelopes in a number of frequency bands and neglect the fine structure. The STI calculation procedures can be tailored to match a particular CI speech processor by matching the frequency bands and method of envelope calculation.

Although the absolute performance of subjects listening to CI-processed speech differs from that of subjects listening to unprocessed speech, additive noise has relatively similar effects in both cases (Hochberg, 1992). Therefore, the STI methods that accurately predict the relative intelligibility among conditions of speech with additive noise (Fig. 4) should also be valid for CI-processed speech with additive noise, although an alternate mapping from STI to percent correct scores may be required for CI-processed speech. It is expected that the same trends will exist for reverberant conditions, although there has been relatively little research assessing the intelligibility of CI-processed speech in reverberation.

The selection of envelope thresholding as a nonlinear operation was guided by our interest in CI-processed speech. Some CI processors use N-of-M processing, coding only a subset (N) of the total (M) frequency-band envelopes during each stimulation cycle (Loizou, 1998). The stimulation cycle is relatively short (a few milliseconds) compared to the STI analysis frame (typically several seconds). The effect of N-of-M processing is comparable to setting the remaining M-N envelopes to zero during intervals when the envelope is not selected. Although this is not identical to envelope thresholding, it has a similar effect on the shape of the envelope, preserving the envelope in intervals where its amplitude is relatively high and eliminating the envelope in intervals where its amplitude is low.

The envelope thresholding results in Fig. 5 indicate that the four proposed methods are potential candidates for predicting the effect of N-of-M processing. If a frequency band is selected all of the time (equivalent to a threshold of 0%), then the intermediate modulation metric is one, contributing a transmission index value (TI_i) of one for that band. If a frequency band is never selected (equivalent to a threshold of 100%), then the intermediate modulation metric is zero and $\mathrm{TI}_i = 0$. If a frequency band is selected intermittently, then the corresponding modulation metric will fall between zero and one, producing a transmission index that reflects that band's partial contribution to intelligibility. While all of the proposed methods are qualitatively correct in that they decrease monotonically from one to zero with increasing threshold, additional work is required to determine which methods, if any, are quantitatively accurate in predicting the effects of envelope thresholding and N-of-M processing on intelligibility.

While research indicates that spectral subtraction does not improve intelligibility for normal-hearing listeners (Lim and Oppenheim, 1979), it has been demonstrated to improve intelligibility for CI users (Weiss, 1993; Hochberg *et al.*, 1992). We hypothesize that this may be related to the effective spectral resolution of the listeners; normal-hearing listeners have relatively fine spectral resolution that permits perception of narrow spectral peaks that rise above the background noise, while CI users are restricted to the relatively broad frequency bands used by their speech processors and therefore cannot perceive spectral peaks within a wider band of noise. As a result, normal-hearing listeners do not benefit from spectral subtraction, since they are already able to listen in relatively narrow bands. On the other hand, CI users benefit from spectral subtraction algorithms that operate in frequency bins substantially narrower than the broader bands used by their speech processors. A related interpretation is that by suppressing narrow frequency bands with low SNR, spectral subtraction removes noise from the broadband temporal envelope, an improvement that provides greater benefit to CI users than to normal hearing listeners. The spectral subtraction results in Fig. 6 indicate that the four proposed methods are potential candidates for predicting the effect of spectral subtraction on CI-processed speech. The intermediate metrics indicate that the proposed STI methods will predict an improvement for speech processed with spectral subtraction algorithms using moderate values of the control parameter, κ. It appears that an appropriate speech-based STI may predict the effect of spectral subtraction on intelligibility more accurately for CI-users than for normal-hearing listeners precisely because it uses a broad frequency-band analysis similar to that used by CI speech processors. In fact, the success of the traditional STI for normal-hearing listeners may be due to the historic focus on broadband distortion such as reverberation and additive broadband noise. For example, consider the case of speech corrupted by a pure tone. This specialized interference would have little or no effect on intelligibility for normal-hearing listeners, but would have a detrimental effect on intelligibility when passed through a CI-speech processor. In computing the STI, the effect of the pure tone would also show up in the apparent SNR for the corresponding frequency band, so that the STI would better predict the effect on intelligibility for CI-processed speech than for a normal-hearing listener.

C. Alternate intelligibility metrics

Because these quantities can be calculated on arbitrarily small speech segments, this raises the possibility of calculating the STI on phoneme-length segments. Traditionally, STI has focused on long-term effects; however, focusing on short segments could prove useful in a number of areas. For example, researchers have studied the effect of mutual independence of adjacent frequency bands based on long-term averages (Steeneken and Houtgast, 1999). However, mutual information may be modeled more accurately using short time segments that carry information concerning the fluctu-

ating short-term SNR. Incorporating short-term averages could potentially lead to speech-based STI metrics that use mutual information from neighboring frequency bands on a phonemic level rather than a global level.

Another approach to combining spectral and temporal information is the physiologically motivated spectro-temporal modulation index (STMI; Elhilali *et al.*, 2003). The STMI is based on an auditory model (Chi *et al.*, 1999) and quantifies the difference in the auditory model output between clean and degraded speech. It operates along spectral and temporal dimensions jointly and explicitly accounts for changes in spectro-temporal modulations. The STMI has been shown to be comparable to the traditional STI for additive noise and reverberation. In addition, for nonlinear distortions consisting of phase jitter or phase shifts, the STMI tracks subject performance on intelligibility tests, while the traditional STI does not. Both the STMI and the methods proposed in this work seek to extend the traditional STI to nonlinear operations. In order to compare these two approaches, future investigations should assess the ability of both the STMI and the proposed metrics to capture the effects of a wide variety of nonlinear operations that includes envelope thresholding, spectral subtraction, phase jitter, and phase shifts.

VII. CONCLUSIONS

The main conclusions of this study follow.

(1) None of the four original speech-based STI methods are suitable for both conventional acoustic degradations and nonlinear operations.
(2) All four of the proposed speech-based STI methods produce reasonable results for conventional acoustic degradations, although preliminary evidence suggests that the normalized correlation method may predict intelligibility less accurately than the other methods for reverberant speech in quiet. All four proposed methods produce qualitatively reasonable results for the nonlinear operations considered in this study. Additional work is required to determine if any of the proposed methods accurately predict speech intelligibility for these and other nonlinear operations.
(3) The normalized correlation and envelope regression methods are computationally less complex than the two cross-power spectrum methods and therefore offer the possibility of computing STI on a short-term (phonemic) level. The envelope regression method is preferred over the two cross-power spectrum methods, because it produces comparable results with less computational complexity.
(4) Of the proposed methods, the normalized correlation method represents the most substantial deviation from the traditional STI. Because it produces results that are qualitatively different from the other methods, it provides an important alternative for fitting data from future speech intelligibility studies.
(5) The proposed speech-based STI methods offer the potential to predict the intelligibility of CI-processed speech.

ACKNOWLEDGMENTS

This work was supported by the National Institute of Deafness and Other Communicative Disorders under Grant Nos. 1-R01-DC00117 and 5-T32-DC0038. The authors are grateful to Karen Payton for her helpful comments on an earlier version of this paper.

APPENDIX A: ALTERNATE FORM OF THE ENVELOPE REGRESSION METHOD

The following is a stochastic reformulation of the envelope regression method (Sec. II B 3) that facilitates comparison with other methods. It begins with the assumption that the linear regression of the sampled response envelope, $y[n]$, onto the sampled probe envelope, $x[n]$, is performed using a minimum mean square error criterion (Ross, 1998). In this case, the optimal fit is

$$y_{\mathrm{MMSE}}[n]=\mu_y+\frac{\lambda_{xy}}{\lambda_x}(x[n]-\mu_x), \tag{A1}$$

where λ_{xy} and λ_x are defined in Eqs. (12) and (13). Thus, the slope (A) and the y-intercept (B) calculated using a minimum mean square error criterion are

$$A=\frac{\lambda_{xy}}{\lambda_x}, \tag{A2}$$

and

$$B=\mu_y-\frac{\lambda_{xy}}{\lambda_x}\mu_x. \tag{A3}$$

Substituting Eqs. (A2) and (A3) into Eq. (9) and rearranging allows the apparent SNR to be expressed as

$$\mathrm{aSNR}=10\log_{10}\left(\frac{M}{1-M}\right), \tag{A4}$$

where M is a modulation metric defined as

$$M=\frac{\mu_x}{\mu_y}\frac{\lambda_{xy}}{\lambda_x}. \tag{A5}$$

APPENDIX B: NORMALIZED CORRELATION METHOD EXPRESSED AS AN ENERGY-WEIGHTED MTF

Equation (18) defines the normalized correlation as

$$\rho=\frac{\phi_{xy}}{\sqrt{\phi_x\phi_y}}. \tag{B1}$$

Using the relationship between the cross-correlation function, $R_{xy}[k]$, and the cross-power spectrum, $S_{xy}(f)$, (Papoulis, 1984) together with the observation that ϕ_{xy} equals the cross-correlation function computed at zero lag, yields

$$\phi_{xy}=R_{xy}[0]=\int_{f=-1/2}^{1/2}S_{xy}(f)\,df, \tag{B2}$$

where $\phi_{xy}\triangleq E\{x[n]y[n]\}$ and $R_{xy}[k]\triangleq E\{x[n]y[n-k]\}$. The normalized correlation can then be expressed as

$$\rho=\frac{1}{\sqrt{\phi_x\phi_y}}\int_{f=-1/2}^{1/2}S_{xy}(f)\,df. \tag{B3}$$

Bringing the denominator inside the integral and multiplying numerator and denominator by the same terms yields

$$\rho = \int_{f=-1/2}^{1/2} \sqrt{\frac{\phi_x}{\phi_y}} \left[\frac{S_{xy}(f)}{S_{xx}(f)} \right] \left[\frac{S_{xx}(f)}{\phi_x} \right] df. \tag{B4}$$

Defining a new MTF,

$$\mathrm{MTF}_\rho(f) \triangleq \sqrt{\frac{\phi_x}{\phi_y}} \frac{S_{xy}(f)}{S_{xx}(f)}, \tag{B5}$$

and a weighting function,

$$W(f) \triangleq \frac{S_{xx}(f)}{\phi_x}, \tag{B6}$$

allows describing ρ as an energy-weighted average of this new MTF, that is,

$$\rho = \int_{f=-1/2}^{1/2} \mathrm{MTF}_\rho(f) W(f) df. \tag{B7}$$

The weighting function, $W(f)$, is the ratio of the power of the probe envelope at each modulation frequency to the total power in the probe envelope.

The MTF defined in Eq. (B5) is similar in form to the MTFs defined for the cross-power spectrum methods in Eqs. (7) and (8). All three MTFs are based on the normalized ratio of the cross-power spectrum between probe and response envelopes to the power spectrum of the probe envelope. The main differences are the factor used for normalization ($\sqrt{\phi_x/\phi_y}$ rather than $\alpha = \mu_x/\mu_y$) and the fact that in Eq. (B5) the MTF is complex-valued. However, since $S_{xx}(f)$ is real and symmetric, and $S_{xy}(f)$ is complex-conjugate symmetric, the integral over equal ranges of positive and negative frequencies will be real-valued.

[1]Motivation for considering the normalized correlation comes in part from studies of binaural detection (Bernstein and Trahiotis, 1996), which have shown that the normalized correlation, ρ, is a better indicator of performance than the normalized covariance, r. By including the envelope means, the metric accounts for the average envelope power as well as the envelope fluctuations. While binaural detection is clearly different than speech intelligibility, it is possible that in both cases the auditory system utilizes the additional information about average envelope power provided by the normalized correlation.

Allen, J. B., and Berkley, D. A. (**1979**). "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943–950.

Bernstein, L. R., and Trahiotis, C. (**1996**). "The normalized correlation: Accounting for binaural detection across center frequency," J. Acoust. Soc. Am. **100**, 3774–3784.

Chi, T., Gao, M., Guyton, M. C., Ru, P., and Shamma, S. A. (**1999**). "Spectro-temporal modulation transfer functions and speech intelligibility," J. Acoust. Soc. Am. **106**, 2719–2732.

Drullman, R. (**1995**). "Temporal envelope and fine structure cues for speech intelligibility," J. Acoust. Soc. Am. **97**, 585–592.

Drullman, R., Festen, J. M., and Plomp, R. (**1994a**). "Effect of temporal envelope smearing on speech reception," J. Acoust. Soc. Am. **95**, 1053–1064.

Drullman, R., Festen, J. M., and Plomp, R. (**1994b**). "Effect of reducing slow temporal modulations on speech reception," J. Acoust. Soc. Am. **95**, 2670–2680.

Elhilali, M., Chi, T., and Shamma, S. A. (**2003**). "A spectro-temporal modulation index (STMI) for assessment of speech intelligibility," Speech Commun. **41**, 331–348.

French, N. R., and Steinberg, J. C. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Goh, Z., Tan, K. C., and Tan, B. T. G. (**1998**). "Postprocessing method for suppressing musical noise generated by spectral subtraction," IEEE Trans. Speech Audio Process **6**, 287–292.

Hochberg, I., Boothroyd, A., Weiss, M., and Hellman, S. (**1992**). "Effects of noise and noise suppression on speech perception by cochlear implant users," Ear Hear. **13**, 263–271.

Hohmann, V., and Kollmeier, B. (**1995**). "The effect of multichannel dynamic compression on speech intelligibility," J. Acoust. Soc. Am. **97**, 1191–1195.

Holube, I., and Kollmeier, K. (**1996**). "Speech intelligibility prediction in hearing-impaired listeners based on a psychoacoustically motivated perception model," J. Acoust. Soc. Am. **100**, 1703–1715.

Houtgast, T., and Steeneken, H. J. M. (**1971**). "Evaluation of speech transmission channels by using artificial signals," Acustica **25**, 355–367.

Houtgast, T., and Steeneken, H. J. M. (**1985**). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Am. **77**, 1069–1077.

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.

IEC (**1998**). *Sound System Equipment—Part 16: Objective rating of speech intelligibility by Speech Transmission Index*, 2nd ed., International Standard No. 60268-16.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," IEEE, NY.

Koch, R. (**1992**). "Gehörgerechte Schallanalyse zur Vorhersage und Verbesserung der Sprachverständlichkeit," ("Auditory sound analysis for the prediction and improvement of speech intelligibility"), Dissertation, Universität Göttingen.

Kryter, K. D. (**1962a**). "Methods for the calculation and use of the articulation index," J. Acoust. Soc. Am. **34**, 1689–1697.

Kryter, K. D. (**1962b**). "Validation of the articulation index," J. Acoust. Soc. Am. **34**, 1698–1706.

Lim, J. S., and Oppenheim, A. V. (**1979**). "Enhancement and bandwidth compression of noisy speech," Proc. IEEE **67**, 1586–1604.

Loizou, P. C. (**1998**). "Mimicking the human ear," IEEE Signal Process. Mag. **15**, 101–130.

Ludvigsen, C. (**1987**). "Prediction of speech intelligibility for normal-hearing and cochlear hearing-impaired listeners," J. Acoust. Soc. Am. **82**, 1162–1171.

Ludvigsen, C., Elberling, C., and Keidser, G. (**1993**). "Evaluation of a noise reduction method—Comparison of observed scores and scores predicted from STI," Scand. Audiol. Suppl. **38**, 50–55.

Ludvigsen, C., Elberling, C., Keidser, G., and Poulsen, T. (**1990**). "Prediction of intelligibility of non-linearly processed speech," Acta Oto-Laryngol., Suppl. **469**, 190–195.

Papoulis, A. (**1984**). *Probability, Random Variables, and Stochastic Processes* (McGraw–Hill, New York), pp. 263–293.

Payton, K. L., and Braida, L. D. (**1999**). "A method to determine the speech transmission index from speech waveforms," J. Acoust. Soc. Am. **106**, 3637–3648.

Payton, K. L., Braida, L. D., Chen, S., Rosengard, P., and Goldsworthy, R. (**2002**). "Computing the STI using speech as a probe stimulus," *Past, Present and Future of the Speech Transmission Index* (TNO Human Factors, Soesterberg, The Netherlands), pp. 125–138.

Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**, 1581–1592.

Ross, S. (**1998**). *A First Course in Probability*, 5th ed. (Prentice Hall, Englewood Cliffs, NJ), pp. 350–354.

Steeneken, H. J. M., and Houtgast, T. (**1980**). "A physical method for measuring speech transmission quality," J. Acoust. Soc. Am. **67**, 318–326.

Steeneken, H. J. M., and Houtgast, T. (**1982**). "Some applications of the speech transmission index (STI) in auditoria," Acustica **51**, 229–234.

Steeneken, H. J. M., and Houtgast, T. (**1999**). "Mutual dependence of the octave-band weights in predicting speech intelligibility," Speech Commun. **28**, 109–123.

Van Buuren, R. A., Festen, J. M., and Houtgast, T. (**1998**). "Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality," J. Acoust. Soc. Am. **105**, 2903–2913.

Weiss, M. R. (**1993**). "Effects of noise and noise reduction processing on the operation of the Nucleus-22 cochlear implant processor," J. Rehabil. R. D**30**, 117–128.

Robust and accurate fundamental frequency estimation based on dominant harmonic components

Tomohiro Nakatani[a)]
NTT Communication Science Laboratories, NTT Corporation, Soraku-gun, Kyoto 619-0237, Japan

Toshio Irino[b)]
Faculty of Systems Engineering, Wakayama University/NTT Communication Science Laboratories, NTT Corporation

(Received 26 August 2002; accepted for publication 6 July 2004)

This paper presents a new method for robust and accurate fundamental frequency (F_0) estimation in the presence of background noise and spectral distortion. *Degree of dominance* and *dominance spectrum* are defined based on instantaneous frequencies. The degree of dominance allows one to evaluate the magnitude of individual harmonic components of the speech signals relative to background noise while reducing the influence of spectral distortion. The fundamental frequency is more accurately estimated from reliable harmonic components which are easy to select given the dominance spectra. Experiments are performed using white and babble background noise with and without spectral distortion as produced by a SRAEN filter. The results show that the present method is better than previously reported methods in terms of both gross and fine F_0 errors. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1787522]

PACS numbers: 43.72.Ar, 43.66.Hg [DOS] Pages: 3690–3700

I. INTRODUCTION

Achieving robust and accurate fundamental frequency (F_0) estimation is one of the most important problems in speech signal processing. F_0 is also cited as a major clue in relation to a person's ability to extract a desired sound from other sounds in the real world.[1] A number of useful speech applications have already been presented based on F_0 estimation. For example, we improved a speech recognition system by employing an F_0-based sound segregation system as a preprocessor.[2] Recently, a very high quality vocoder, STRAIGHT, was developed based on the idea of F_0 adaptive processing.[3] In any application, F_0 accuracy is of the greatest importance since any error in F_0 determination has a detrimental effect on system performance. Many applications also require robust F_0 estimation under adverse noise conditions.

A number of F_0 estimation methods have been proposed.[4,5] Typical methods are categorized into two types, namely temporal and spectral methods. The former is mainly based on the autocorrelation of input signals and includes the maximum likelihood (ML) method. The latter is mainly based on spectral peak extraction and includes the cepstrum method. Recently, several new F_0 estimation methods have been proposed with the goal of providing better performance than the traditional methods.[3,6–10] Some methods are more accurate, while others are more robust in the face of background noise. For example, YIN is a temporal method that is one of the most accurate F_0 estimators when there is no background noise.[9] Shimamura's method is a temporal one that is described as being robust against white Gaussian noise.[7] TEMPO is a newly developed spectral method that enables reliable F_0 estimation in support of a new high quality vocoder, STRAIGHT.[3] Liu and Lin presented a spectral method that is robust against white Gaussian noise.[8] F_0 estimation methods for telephone quality speech have also been proposed recently.[6,10]

Although these methods are superior to conventional methods in certain respects, the limited evaluation experiments undertaken thus far make it impossible to decide which method is the best. For example, robustness and accuracy have not been comprehensively evaluated for real world applications in the presence of both background noise and spectral distortion. The conditions under which speech sounds are recorded are not constant and are usually adversely affected by the acoustic properties of microphones and the environment. It is also necessary for the F_0 estimation methods to support telephone line use.

Furthermore, most methods have not been evaluated with large databases that contain reliable reference F_0 values. There are problems with regard to the definition of the ''correct'' F_0 value. It must be defined without using speech signals for a fair evaluation since speech sounds are time-varying filtered or ''distorted'' versions of glottal pulse sequences. Therefore, we believe this makes it desirable to use databases consisting of simultaneous recordings of speech and electro-glottal graph (EGG) signals because EGG signals are extracted directly from glottal vibrations.

In this paper, we present a new F_0 estimation method, referred to as *DominAnce Spectrum based Harmonics extraction (DASH)*, and show its effectiveness with regard to both background noise and spectral distortion using a comprehensive evaluation method. We define a dominance spectrum to reduce the factors degrading F_0 estimation accuracy. We evaluate the robustness and accuracy of DASH using several large databases containing simultaneous recordings of speech and electro-glottal graph (EGG) signals.[11–13] We use white noise and babble noise as background noise. We

a)Electronic mail: nak@cslab.kecl.ntt.co.jp
b)Electronic mail: irino@sys.wakayama-u.ac.jp

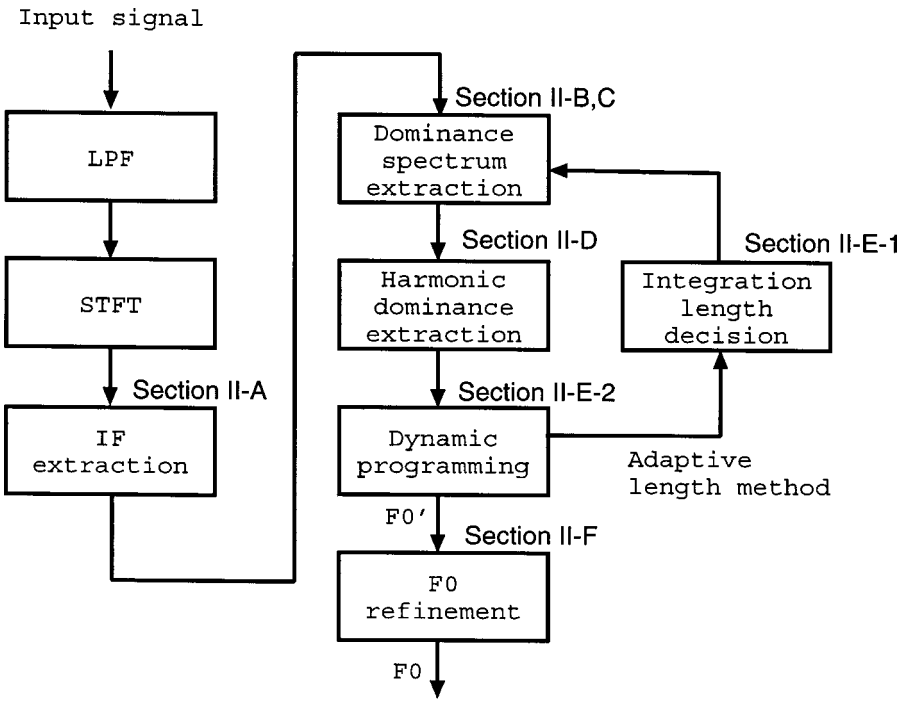


FIG. 1. Flow diagram of F_0 estimation.

also apply spectral distortion to the speech signals using a SRAEN filter. This is a high pass filter, above 300 Hz, recommended by the ITU-T[14] for simulating a telephone handset. As a result, the effectiveness of our F_0 estimation method is reliably examined under various kinds of adverse conditions.

In the rest of this paper, we describe the proposed F_0 estimation method based on dominance in Sec. II and experimental results in Sec. III.

II. METHOD

The F_0 estimation method proposed in this paper is an extension of a spectral method based on instantaneous frequency (IF). It is referred to as *DominAnce Spectrum based Harmonics extraction (DASH)*. The processing flow of DASH is summarized in Fig. 1. By way of preparation, the input signal is first down-sampled (to 4 kHz), and converted into a signal in the frequency domain by short-time Fourier transformation (using 512 points with a 50-m Hanning window and a 1-m frame shift). The IF and the dominance spectrum are then extracted as described in Secs. II A and II B, respectively. The effectiveness of the dominance spectrum for both background noise and spectral distortion is explained in Sec. II C. Section II D describes a method for gathering harmonics information, namely harmonic dominance. Several techniques for increasing the robustness are described in Sec. II E. For example, Sec. II E 1 introduces an adaptive mechanism to decide the integration length for calculating the dominance spectrum. Finally, F_0 estimation is refined based on the degree of dominance of fixed points to provide very accurate results as described in Sec. II F.

A. F_0 estimation based on instantaneous frequencies

If a signal is represented as $s(t)=a(t)e^{j\phi(t)}$, the IF of the signal is defined as $\dot{\phi}=d\phi/dt$. The concept of IF was introduced by Flanagan[15] as one of the speech features for a phase vocoder. The use of IF in F_0 estimation was initiated by Charpentier.[16] Recently, it was reported that IF use is effective in F_0 estimation under noisy conditions.[3,13,17] In these methods, the IF is calculated for each frequency bin of the short-time Fourier transform (STFT) at each time frame. The STFT can be viewed as a set of bandpass filters as follows:

$$S(f,t)=\int_{-\infty}^{\infty} g(\tau-t)x(\tau)e^{-j2\pi f\tau}d\tau, \tag{1}$$

$$=e^{-j2\pi ft}\int_{-\infty}^{\infty} g(\tau-t)e^{-j2\pi f(\tau-t)}x(\tau)d\tau, \tag{2}$$

$$F(f,t)=e^{j2\pi ft}S(f,t), \tag{3}$$

where $x(\tau)$ is the time series of the input signal, f is the center frequency of each STFT bin, $g(\tau)$ is a window function, and t is the center time of each time frame. Equation (2) implies that $S(f,t)$ can be interpreted as the output signal of a bandpass filter, $g(\tau)e^{-j2\pi f(\tau)}$, multiplied by $e^{-j2\pi ft}$. Therefore, signal $F(f,t)$ or $e^{j2\pi ft}S(f,t)$ is the output signal of the bandpass filter. The IF, $\dot{\phi}(f,t)$, is defined as the time derivative of the phase of $F(f,t)$ as in

$$\dot{\phi}(f,t)=\frac{\partial}{\partial t}\arg(F(f,t)). \tag{4}$$

The IF can be calculated according to Flanagan's method[15,17] as follows:

$$\dot{\phi}(f,t)=f+\frac{a(f,t)(\partial/\partial t)b(f,t)-b(f,t)(\partial/\partial t)a(f,t)}{a(f,t)^2+b(f,t)^2}, \tag{5}$$

where

$$a(f,t)=\mathrm{Re}\{S(f,t)\}, \tag{6}$$

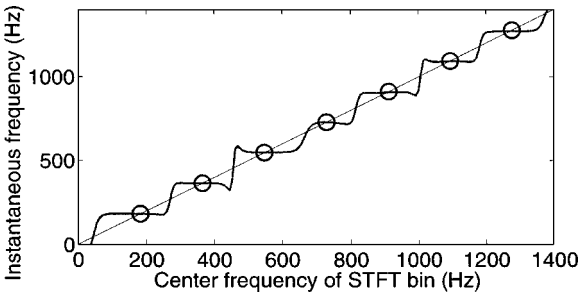


FIG. 2. Instantaneous frequency (thick line) and fixed points (circles) of a speech sound.

$$b(f,t)=\mathrm{Im}\{S(f,t)\}, \tag{7}$$

and time derivatives of $a(f,t)$ and $b(f,t)$ are, respectively, obtained as the real and imaginary parts of

$$\frac{\partial}{\partial t}S(f,t)=\int_{-\infty}^{\infty}-h(\tau-t)x(\tau)e^{-j2\pi f\tau}\,d\tau, \tag{8}$$

where $h(t)$ is the time derivative of $g(t)$.

When we draw a mapping function between the center frequency of the STFT bin, f, and the derived IF, $\dot{\phi}(f,t)$, for a voiced speech signal, it has the appearance of a regularly spaced staircase as shown in Fig. 2.[16] The harmonic components of the fundamental frequency produce the individual stairs. This is because each harmonic component is the only dominant signal included in the frequency bands corresponding to its neighboring frequency bins, and the IFs for the bins coincide with the frequency of the component. Thus the best estimate of the harmonic frequency is the frequency at each *fixed point* where the IF coincides with the center frequencies of the bins, that is, $\dot{\phi}=f$. The fundamental frequency, F_0, is estimated as the difference between the frequencies of adjacent fixed points since the harmonic frequencies are integer multiples of F_0. The use of IF is superior to conventional methods, such as cepstrum methods, because it can provide a precise frequency estimate of the dominant harmonic component included in each STFT bin.

Abe *et al.* proposed the "IF amplitude spectrum,"[17] which is a modified version of the amplitude spectrum, $|S(f,t)|$. The magnitude of the amplitude spectrum is reorganized according to IF, $\dot{\phi}(f,t)$, instead of the STFT center frequency, f. The IF amplitude spectrum, $G(\dot{\phi},t)$, is defined as

$$G(\dot{\phi},t)=\lim_{\Delta\dot{\phi}\to 0}\frac{1}{\Delta\dot{\phi}}\int_{\dot{\phi}<\dot{\phi}(f)<\dot{\phi}+\Delta\dot{\phi}}|S(f,t)|\,df. \tag{9}$$

Since the IF amplitude spectrum enhances the harmonic structure of speech sounds, it was considered applicable to F_0 estimation in the presence of background noise. However, we suspect that the performance would deteriorate in the presence of spectral distortion caused by, for example, telephone handsets. This is because the IF amplitude spectrum is also affected by spectral distortion as it is inherited from the amplitude spectrum. Moreover, the IF amplitude spectrum would not reduce additive noise components since it is defined using a derivative operation on the frequency axis, which usually enhances the noise.

Atake *et al.* recently proposed an F_0 estimation method by defining the stability of fixed points[13] based on Cohen's bandwidth equation.[18] This method makes it possible to estimate F_0 very accurately even under low signal-to-noise ratio (SNR) conditions. However, it requires an initial rough estimate of F_0 to define an adaptive short-time window for precise F_0 estimation. The degree of robustness may largely depend on this initial estimation. Therefore, one of our goals is to develop a method using a fixed window that can provide a robust estimation of F_0.

In order to overcome the problems posed by existing F_0 estimation methods, we propose a new IF-based method that is robust and accurate in the presence of both background noise and spectral distortion.

B. Definition of dominance

We introduce a new measure based on IF to evaluate the magnitude of a harmonic component relative to the other components in each STFT bin. We refer to this measure as "the degree of dominance" and define it as $D(f,t)$ in Eq. (10)[19] for each frequency bin centered at f at a time frame centered at t:

$$D(f,t)=10\log_{10}(1/B(f,t)^2), \tag{10}$$

$$B(f,t)^2=\frac{\int_{f-\Delta f/2}^{f+\Delta f/2}(\dot{\phi}(f',t)-f)^2S(f',t)^2\,df'}{\int_{f-\Delta f/2}^{f+\Delta f/2}S(f',t)^2\,df'}, \tag{11}$$

where $\dot{\phi}(f',t)$ and $S(f',t)^2$ represent the IF and the power spectrum, respectively, for a center frequency f' in a fraquency range neighboring to f. $B(f,t)^2$ is derived as the weighted average of the squared difference between f and $\dot{\phi}(f',t)$. The averaging is performed over the frequency range where $|f'-f|<\Delta f/2$, with the weighting function being the power spectrum, $S(f',t)^2$.

The value of $B(f,t)^2$ reaches a minimum when the dominant frequency component of a signal coincides with the center frequency, f, because the value of $\dot{\phi}(f',t)$ for the adjacent frequency bins approaches f. This is seen as the flat stairs around the fixed points in Fig. 2. Here, the degree of dominance, $D(f,t)$, reaches its maximum since it is defined as the logarithm of the inverse of $B(f,t)^2$ in Eq. (10). By contrast, the dominance value becomes smaller and does not have a sharp peak when the frequency component is greatly affected by noise because $\dot{\phi}(f',t)$ increases as f' increases and thus the difference between $\dot{\phi}(f',t)$ and f becomes large. Consequently, the degree of dominance has sharp peaks only at fixed points corresponding to dominant frequency components.

C. Effectiveness of dominance spectrum

When the degree of dominance, $D(f,t)$, is calculated for all STFT bins, it takes the form of a spectrum [Fig. 3(a)]. This is referred to as a "dominance spectrum." The following sections show that basing F_0 estimation on the dominance spectrum makes it robust with regard to both additive background noise and spectral distortion.

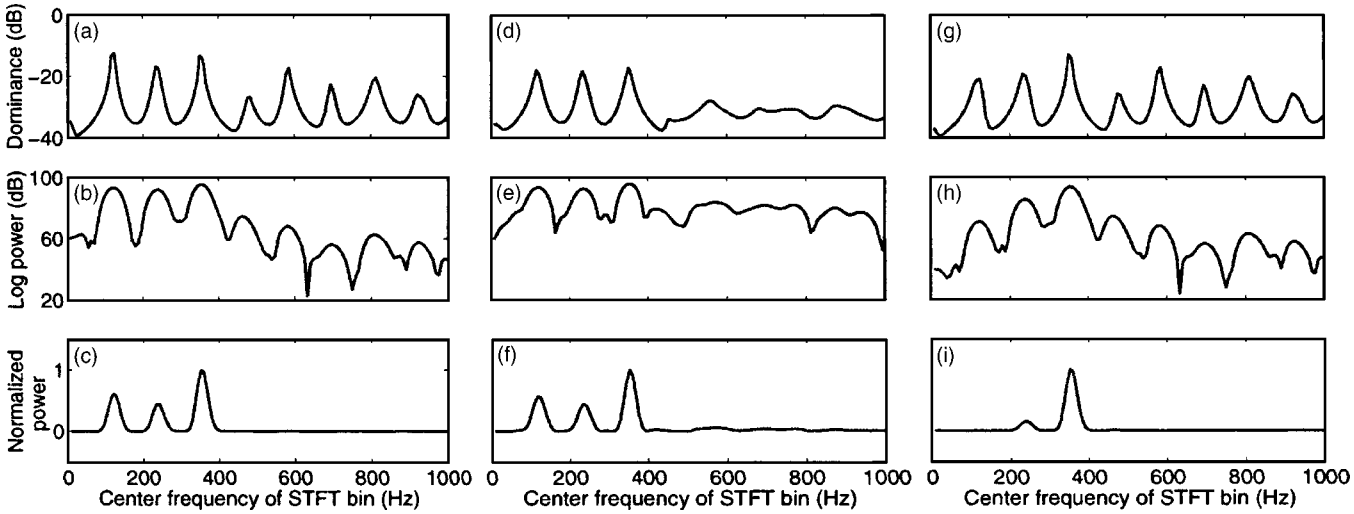


FIG. 3. Dominance spectra [(a), (d), and (g)], logarithmic power spectra [(b), (e), and (h)], and power spectra [(c), (f), and (i)] of clean speech (left three panels), speech with background white noise (middle three panels, SNR: 0 dB), and speech convolved with a SRAEN filter (right three panels).

1. Robustness against additive noise

Figures 3(a)–(c) show, respectively, examples of a dominance spectrum, a logarithmic power spectrum, and the power spectrum of a clean speech signal when there is no background noise. The dominance spectrum [Fig. 3(a)] includes sharp peaks that correspond to the harmonic components. The peaks are much sharper than those of a usual logarithmic power spectrum [Fig. 3(b)]. Figures 3(d)–(f) show the same kinds of spectra when the speech signal is smeared with background white noise with 0 dB SNR. Although neither the dominance spectrum [Fig. 3(d)] nor the logarithmic power spectrum [Fig. 3(e)] have sharp peaks corresponding to the harmonic components above 500 Hz, the peak-to-trough ratio is smaller in the dominance spectrum. By contrast, the peak-to-trough ratio below 500 Hz is greater in the dominance spectrum. That is, the dominance spectrum enhances the peaks of the harmonic components and suppresses the variation produced by noise. This property is particularly useful for robust F_0 estimation. (Further detailed analysis is provided in Appendix A to explain how the dominance spectrum changes with the level of the background white noise.)

The difference between the power spectra of clean speech [Fig. 3(c)] and noisy speech [Fig. 3(f)] is very small. In other words, the proportion of peaks derived from harmonic components and noise components is very large in the power spectrum. Therefore, the power spectrum is also considered to be a good robust measure against additive background noise.

2. Robustness against spectral distortion

The dominance spectrum exhibits the property of whitening the spectral envelope and this reduces the effect of spectral distortion. This can be clearly demonstrated by passing a speech signal through a SRAEN filter to simulate the telephone handset case. Figure 3(g) shows the dominance spectrum for the filtered sound. The first and second peaks of the dominance spectrum in Fig. 3(g) are slightly smaller than those in the original spectrum in Fig. 3(a). But this small decrement does not affect F_0 estimation. Figure 3(i) shows the power spectrum of the filtered sound. The peaks below 300 Hz are greatly suppressed by the filtering, and this shows that the power spectrum is much more sensitive to spectral distortion than the dominance spectrum. By contrast, in the logarithmic power spectrum [Fig. 3(h)], the peaks at low frequencies clearly remain, albeit with a small decrement. This demonstrates that the logarithmic power spectrum is robust with regard to spectral distortion.

Consequently, only the dominance spectrum is robust against both background noise and spectral distortion, and thus must be useful for F_0 estimation under various adverse conditions.

D. Harmonic dominance

We now define a decision measure that summarizes the degree of dominance for all harmonic components at a time frame. This measure is referred to as ''harmonic dominance'' and is defined as $H(f_n,t)$ in Eq. (12):

$$H(f_n,t)=\sum_{l=1}^{lf_n<F_{\max}}(D(r(lf_n),t)-\overline{D(f,t)}), \tag{12}$$

$$\overline{D(f,t)}=\sum_{f} D(f,t)/N, \tag{13}$$

$$F_0=\arg\max_{f_n}\{H(f_n,t)\}, \tag{14}$$

where f_n is one of the quantized F_0 candidates[19] located within a specified F_0 search range, $n(=0,1,2,...)$ is its index, lf_n corresponds to the frequency of its lth harmonic component, and $r(lf_n)$ is a function that transfers lf_n to the center frequency of the nearest STFT bin. Harmonic components up to $F_{\max}(=1500\,\mathrm{Hz})$ are summed as in Eq. (12). $\overline{D(f,t)}$ is the average of $D(f,t)$ over the number of STFT bins, N, where f refers to the center frequency of an STFT bin. $\overline{D(f,t)}$ is a term that ensures that the measure is unbiased. Our preliminary experiment showed that specific errors known as ''double pitch'' and ''half pitch'' errors were effectively re-

duced by this unbiasing (see Appendix B). The frequency that maximizes the harmonic dominance, f_n, becomes the estimate of the fundamental frequency F_0 as in Eq. (14).

E. Robust F_0 estimation based on dominance

We introduce two additional techniques to improve the robustness of our F_0 estimation method based on dominance. One is a method for determining the optimum integration length for calculating the dominance spectrum, and the other is the use of dynamic programming (DP) to reduce discontinuous transition errors in the F_0 trajectory.

1. Frequency range of bandwidth integral

The optimum frequency range, Δf, for the integral in Eq. (11) roughly depends on the F_0 of the input signal. F_0 can be estimated more robustly if the optimum range is given based on a rough estimate of the F_0. In our preliminary experiment, the optimum frequency range, Δf, was about 130 Hz for male speech and 260 Hz for female speech, which approximately coincide with the average F_0 of the target speech. In general, however, no *a priori* information on the target speech is available in advance, so the optimum range cannot be obtained in advance.

To overcome this problem, we introduce an adaptive method to decide the frequency range. With this method, the harmonic dominance is maximized twice, first for rough F_0 estimation, and second for more precise F_0 estimation. In the first estimation, we use a fixed frequency range for the integration that covers both male and female speech, and then we use the roughly estimated F_0 to determine the optimum frequency range for the precise estimation. In our experiment, the 260-Hz range (the same as the optimum value for female speech) was found to be suitable for the initial rough estimate, and about 67% to 110% of the initial rough F_0 estimate was suitable for the precise estimation.

2. The use of dynamic programming

Dynamic programming (DP) has often been used to obtain a smooth F_0 trajectory by tracking continuous peaks in the time series of a vector such as cepstral coefficients. In our F_0 estimation, the harmonic dominance calculated for each frequency bin at each time frame is treated as the vector to which DP is applied, and the frequencies corresponding to the positions of the tracked elements yield the smoothed F_0 trajectory.

F_0 tracking using DP is viewed as a processing approach that finds an optimum F_0 trajectory at the smallest cost overall. With our F_0 estimation, the time series of harmonic dominance are adopted as one cost function for DP. Let $H(f_n,t_k)$ be the harmonic dominance at the kth time frame for the nth quantized F_0 candidate, f_n, searched for in Eq. (12), and $P(t_k)$ be the signal power at the time frame. Then $d(n,t_k)=-P(t_k)H(f_n,t_k)$ is treated as the cost function in our implementation. Another cost function, $p(f_n,f_m)$, is also adopted to determine the cost of the F_0 transition from f_n to f_m between two adjacent time frames. $p(f_n,f_m)=|\log(f_n/f_m)|$ is used in our implementation. This approach means that the F_0 tracking problem can be formalized as finding an F_0 sequence, $\Omega=\{f_{n_0},f_{n_1},\ldots,f_{n_T}\}$, that minimizes the following total cost function, $C(\Omega)$,

$$C(\Omega)=d(n_0,t_0)+\sum_{k=1}^{T}(w\cdot p(f_{n_{k-1}},f_{n_k})+d(n_k,t_k)), \tag{15}$$

where w is a weight to determine the degree to which the F_0 transition cost is weighted in the total cost. Obviously, this cost function can efficiently be minimized using DP.

F. Accurate F_0 estimation based on dominance

To improve the accuracy of the F_0 estimation, we introduce an F_0 refinement method based on the IFs at fixed points. Because fixed points with large dominance values are expected to be derived from dominant harmonic components, the IFs at such fixed points are considered to be good estimates of their harmonic frequencies. Therefore, reliable F_0 candidates can be obtained by dividing their harmonic frequencies by their harmonic numbers. With our method, F_0 is determined as the weighted average of the F_0 candidates derived from fixed points using the degree of dominance as the weight. Because of this weight, F_0 is determined mainly based on the fixed points of dominant harmonic components, so the obtained F_0 is expected to be reliable.

The idea behind F_0 refinement based on fixed points was introduced by Atake *et al.*[13] Cohen's bandwidth equation was then used to evaluate the reliability of fixed points. In this paper, it is modified to use the degree of dominance.[20]

If F_0' is the result of the maximization of the harmonic dominance as discussed in Sec. II E, the refined F_0 is defined as follows:

$$F_0=\frac{\sum_{i=1}^{n}\sum_{\dot{\phi}\in\dot{\Phi}(i\cdot F_0')}(\dot{\phi}/i)(D(r(\dot{\phi}),t)-c)}{\sum_{i=1}^{n}\sum_{\dot{\phi}\in\dot{\Phi}(i\cdot F_0')}(D(r(\dot{\phi}),t)-c)}, \tag{16}$$

$$c=\min_{\dot{\phi}\in\dot{\Phi}(i\cdot F_0'),i=1\sim n}\{D(r(\dot{\phi}),t)\}-\epsilon\quad(\epsilon>0). \tag{17}$$

Here, $\dot{\Phi}(i\cdot F_0')$ is a set of IFs of fixed points that are located within $\pm 10\%$ of the ith multiple of F_0', and n is the number of harmonics. Each $\dot{\phi}$ is a candidate frequency of the ith harmonic component derived from a fixed point, so $\dot{\phi}/i$ is a candidate for F_0, and F_0 is calculated as the average $\dot{\phi}/i$ value weighted by the degree of dominance. "$r(\cdot)$" is used to transform a continuous frequency to its nearest STFT center frequency as used in Eq. (12). This quantization is only used to calculate the reliability of the fixed points, so it does not affect the accuracy of F_0 estimation. Term c enables the weights for all i to be greater than zero.

1. Interpolation for precise IF at a fixed point

IF $\dot{\phi}$ at a fixed point is precisely determined by linear interpolation using the IFs and center frequencies of adjacent STFT bins. Let f_1 and f_2 be the center frequencies and $\dot{\phi}_1$ and $\dot{\phi}_2$ be their IFs. There is a fixed point between these two STFT bins if Eq. (19) holds. The interpolated value of the IF, $\dot{\phi}$, at this fixed point is defined as follows:

$$\dot{\phi}=\frac{(f_2-\dot{\phi}_2)f_1+(\dot{\phi}_1-f_1)f_2}{(\dot{\phi}_1-f_1)+(f_2-\dot{\phi}_2)}, \tag{18}$$

where

$$\dot{\phi}_1>f_1 \quad \text{and} \quad \dot{\phi}_2<f_2. \tag{19}$$

G. The use of the power spectrum

In this section, we propose a simple alternative F_0 estimation method, referred to as the *Ripple-Enhanced Power Spectrum (REPS)*, that is based on the power spectrum instead of the dominance spectrum. In the power spectrum shown in Fig. 3(f), the dominant harmonic components are represented as sharp peaks while the additive noise components are almost negligible. Thus the power spectrum also seems to be applicable to robust F_0 estimation in the presence of additive background noise although its performance would be sensitive to spectral distortions as described in Sec. II C 2 using Fig. 3(i).

Although the power spectrum has been used in some F_0 estimation methods,[2] the robustness of these methods has not yet been well evaluated. We, therefore, designed another method by substituting the power spectrum for the dominance spectrum.

In a preliminary study, however, we found that the direct use of the power spectrum, $S(f,t)^2$, does not necessarily result in the best performance. Instead, we modified the spectrum so as to enhance the spectral ripple corresponding to the glottal pulse. The ripple-enhanced power spectrum, $R(f,t)$, is derived by a method similar to cepstral liftering in the power spectrum (not in the log domain as with usual cepstral liftering), that is, applying inverse discrete Fourier transformation (IDFT) to $S(f,t)^2$, substituting zeros for the lower quefrency components, and applying DFT to the modified coefficients. Finally, we developed our alternative F_0 estimation method by substituting $R(f,t)$ for $D(f,t)$ in Eqs. (12), (13), (16), and (17) without changing the rest of the processing method.

III. EXPERIMENTS

A. Evaluation method

When evaluating an F_0 estimation method, it is of the greatest importance to define the correct F_0 values of the target speech. F_0 values labeled in a speech database have been used in many cases. Unfortunately, such labels are usually attached by hand and thus are not necessarily accurate or large enough for reliable evaluation. For robust evaluation, the correct F_0 values are often determined from clean speech signals using the same algorithm as that being evaluated by assuming that the method is sufficiently good for clean speech. However, we need to be careful about this assumption since nonharmonic components included in certain consonants and the time-varying frequency property of the vocal tract may affect F_0 estimation even in clean speech.

In our evaluation, the correct F_0, F_0^{cor}, is estimated from electro-glottal graph (EGG) signals collected at the same time as the speech signals (see Sec. III B). Since an EGG signal is directly derived from glottal vibration and is largely unaffected by the nonharmonic components of speech, it is an ideal signal for estimating the correct F_0 value.

Since we concentrated on the evaluation of robustness, we calculated two F_0 values, F_0^{cor} and F_0^{est}, from the EGG signal and the noisy speech signal, respectively, using the same F_0 estimation algorithm. We then calculated the difference between them as an error measure for model performance. We defined two measures: gross F_0 error and fine F_0 error. Gross F_0 error is the ratio of the number of frames giving ''incorrect'' F_0^{est} values to the total number of frames. Value F_0^{est} is ''incorrect'' if it falls outside $\pm 5\%$ of the F_0^{cor} value. Fine F_0 error is the normalized root mean square error between the F_0^{cor} value and F_0^{est} value which is not judged as ''incorrect'' in the gross error measurement. The formulation is the root mean square of $(F_0^{\mathrm{cor}}-F_0^{\mathrm{est}})/F_0^{\mathrm{cor}}$.

Through our evaluation, we search for F_0s between 50 to 500 Hz for every 1-ms frame shift, and we evaluate the estimated F_0's only for voiced durations based on voicing labels.

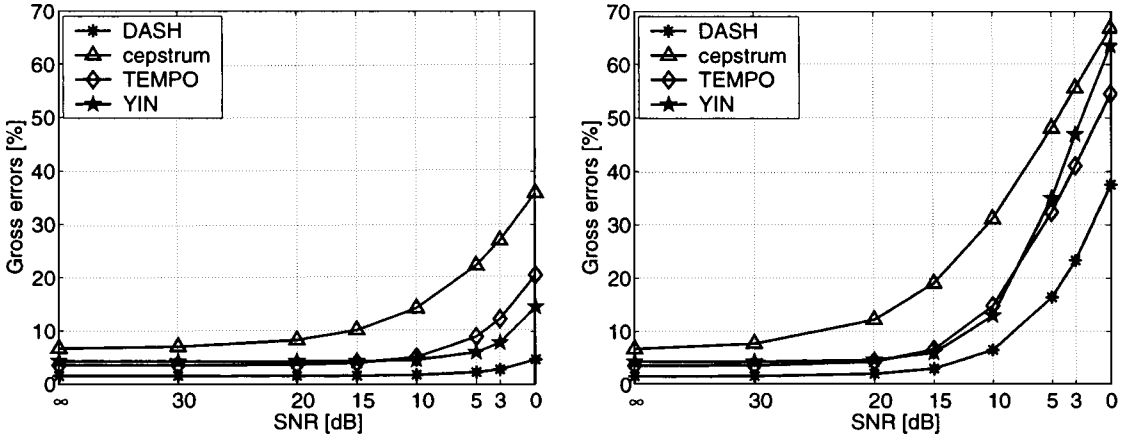


FIG. 4. Gross F_0 errors obtained by DASH, cepstrum, TEMPO, and YIN in the presence of white noise (left panel) and babble noise (right panel).

B. Databases

We used the following three databases of speech and EGG signals recorded simultaneously.

DB1 30 utterances by 14 male and 14 female Japanese speakers (total of 840 utterances, total duration of 40 min, 16-kHz sampling, and 16-bit quantization).[13]

DB2 50 utterances by one male and one female English speaker (total of 100 utterances, total duration of 7 min, 20-kHz sampling, and 16-bit quantization).[11] The database can be downloaded from the following URL, http://www.cstr.ed.ac.uk/~pcb/fda_eval.tar.gz

DB3 Phonetically balanced text utterances by five male and five female English speakers (total duration of 9 min, 20-kHz sampling, and 16-bit quantization).[12] The database can be downloaded from the following URL, ftp://ftp.cs.keele.ac.uk/pub/pitch/Speech

We used the voiced/unvoiced labels included in the three databases, and determined correct F_0's, F_0^{cor}, for each F_0 estimation method from the EGG signals as described in Sec. III A. We did not use the reference F_0 labels present in DB2 and DB3 except as indicated in Appendix C because the reference F_0 labels may have a certain bias depending on the method used for their estimation.

C. Noise and spectral distortion

We used white noise and babble noise as the additive background noise. The SNR in terms of the average power ranged from ∞ to 0 dB. We calculated the average power of a signal as follows: (1) the initial average power of a signal was first obtained by calculating the average power for all frames, (2) frames with less than $\frac{1}{30}$ of the initial average power were discarded, and then (3) the average power was calculated for the remaining frames.

The babble noise was a mixture of the normalized speech signals of ten speakers randomly selected from DB1. In a preliminary listening test involving the speech sounds in the babble noise, the target sounds were almost all distinguishable from the babble noise when the SNR was 0 dB and they were completely indistinguishable when the SNR was −10 dB.

Spectral distortion was introduced by using a SRAEN filter. The SRAEN filter recommended by ITU-T is designed to simulate the ideal frequency property of a telephone handset.[14] It has defined spectral characteristics and is a high-pass filter above 300 Hz. Accordingly, this filter almost completely eliminates the fundamental component of male speech. We designed an FIR filter to fit the spectral characteristics by using the Remez algorithm.

TABLE I. Gross F_0 errors (%) for DB1 with a ±5% error criterion for SNRs of ∞ and 5 dB, with and without a SRAEN filter. The bold font shows the best performance. W: white noise, B: babble noise.

	Without SRAEN			With SRAEN		
Noise		W	B		W	B
SNR (dB)	∞	5	5	∞	5	5
DASH	1.52	2.31	16.4	**2.53**	**3.80**	22.4
$REPS_3$	**0.89**	**1.26**	**8.37**	3.55	4.67	**18.4**
cepstrum	6.69	22.4	48.1	6.64	21.6	46.7
TEMPO	3.53	9.02	32.4	6.43	12.5	30.5
YIN	4.30	6.06	35.0	7.05	11.3	40.9

TABLE II. Gross F_0 errors (%) for DB2 with a ±5% error criterion for SNRs of ∞ and 5 dB, with and without a SRAEN filter. The bold font shows the best performance. W: white noise, B: babble noise.

	Without SRAEN			With SRAEN		
Noise		W	B		W	B
SNR (dB)	∞	5	5	∞	5	5
DASH	2.33	4.51	24.0	**3.47**	**5.03**	**24.9**
$REPS_3$	**2.18**	**3.34**	**15.7**	4.28	6.37	25.1
cepstrum	10.9	29.9	56.0	12.2	30.6	56.2
TEMPO	5.27	17.1	43.1	10.3	25.8	48.0
YIN	3.61	7.15	39.3	7.04	12.7	40.8

D. F_0 estimation methods for comparison

We evaluated the dominance-spectrum-based method with an adaptive integration range (DASH), and the ripple-enhanced power-spectral-based method (REPS). We compared them with a commonly-used cepstrum method (cepstrum), the F_0 estimation method used in STRAIGHT, namely TEMPO,[3] and YIN.[9] TEMPO is an F_0 estimation method based on ''fundamentalness.'' YIN is a modified version of the autocorrelation method. These were recently proposed as improvements to conventional methods.

As a further comparison, we provide several evaluation results in Appendix C using reference F_0 labels and ±20% gross error criterion, both of which are often used as evaluation conditions for recently proposed F_0 estimation methods.[6,9,10]

E. Results

1. Robustness of F_0 estimation

Figure 4 shows the gross F_0 errors for DB1 obtained by DASH, cepstrum, TEMPO, and YIN[21] in the presence of additive white noise (left panel) and babble noise (right panel) for an SNR range of ∞ to 0 dB without any spectral distortion by a SRAEN filter. The precise values for SNRs of ∞ and 5 dB are listed in the second, third, and fourth columns of Table I. In both figure and table, DASH produces smaller errors than cepstrum, TEMPO, and YIN for all SNRs. In Tables II and III, the gross errors obtained for DB2 and DB3 are listed in the same order as in Table I. The

TABLE III. Gross F_0 errors (%) for DB3 with a ±5% error criterion for SNRs of ∞ and 5 dB, with and without a SRAEN filter. The bold font shows the best performance. W: white noise, B: babble noise.

	Without SRAEN			With SRAEN		
Noise		W	B		W	B
SNR (dB)	∞	5	5	∞	5	5
DASH	3.93	5.70	25.2	**5.10**	**6.91**	**28.8**
$REPS_3$	**2.59**	**3.76**	**17.7**	7.37	10.1	31.1
cepstrum	11.0	29.7	52.4	11.7	30.1	51.8
TEMPO	8.07	18.6	39.7	18.9	28.2	46.4
YIN	5.04	10.7	38.4	9.50	17.8	45.7

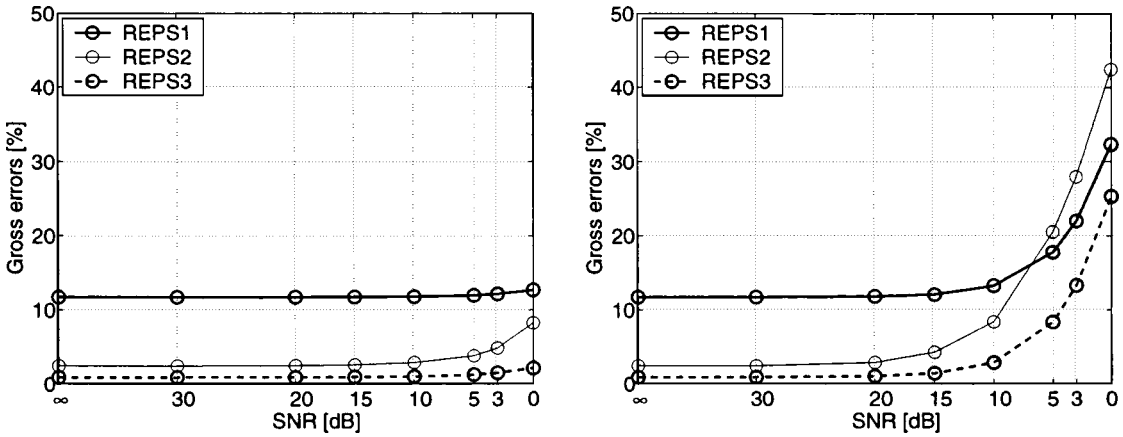


FIG. 5. Gross F_0 errors obtained by REPS_1, REPS_2, and REPS_3 in the presence of white noise (left panel) and babble noise (right panel). See Sec. III E 2 for definitions of REPS_1, REPS_2, and REPS_3.

proposed method, DASH, again gives smaller errors than any other method except for REPS_3 which is also our proposed method (see Sec. III E 2). Since the tendencies of the errors are consistent across the three different databases, the superiority of DASH is clearly demonstrated.

2. The use of the power spectrum

In Sec. II G, we proposed the use of the ripple-enhanced power spectrum as a substitute for the dominance spectrum. We refer to this original method as REPS_1. We evaluated this method under the same conditions as those used in Sec. III E 1. The thick solid lines (REPS_1) in Fig. 5 show the gross F_0 errors for DB1 with white noise (left panel) and babble noise (right panel). The errors were always more than 10% and much worse than those obtained using the dominance spectrum as shown in Fig. 4. We found that the error was very large when estimating F_0 from the EGG signal. This is particularly important since we assumed the estimated values to be the correct F_0 values in our evaluation. A detailed analysis showed that the spectral power at frequencies below F_0 often exceeded the power of the F_0 components in the EGG signal and this precluded exact F_0 estimation. An effective way of eliminating these characteristics in the EGG signal was to use a preemphasis filter (3 dB/oct). We therefore applied a preemphasis filter to both the speech and EGG signals. We refer to this modified method as REPS_2. The results are shown by the thin solid lines (REPS_2) in Fig. 5. The errors were improved at high SNRs but not so much at low SNRs, especially in the presence of babble noise. We also found that the robustness of the F_0 estimation for the speech signals was degraded by the preemphasis. We therefore applied the preemphasis filter to the EGG signals but not to the speech signals. This is referred to as REPS_3. The thick dashed lines (REPS_3) in Fig. 5 show the results. The gross errors became much better than those obtained with REPS_1 and REPS_2. Moreover, the results were even better than those obtained with DASH shown as the solid line in Fig. 4. This is also clearly shown by the gross errors for ∞ and 5 dB in Tables I–III. Thus REPS_3 yielded the best performance if the signal was not smeared by any spectral distortion.

These results show that the ripple-enhanced power spectrum is robust against additive noise provided the spectral distortion is properly compensated. It is, however, difficult to specify an optimum preprocessing filter in advance when sound is recorded in the real world since the impulse response is not necessarily well-defined.

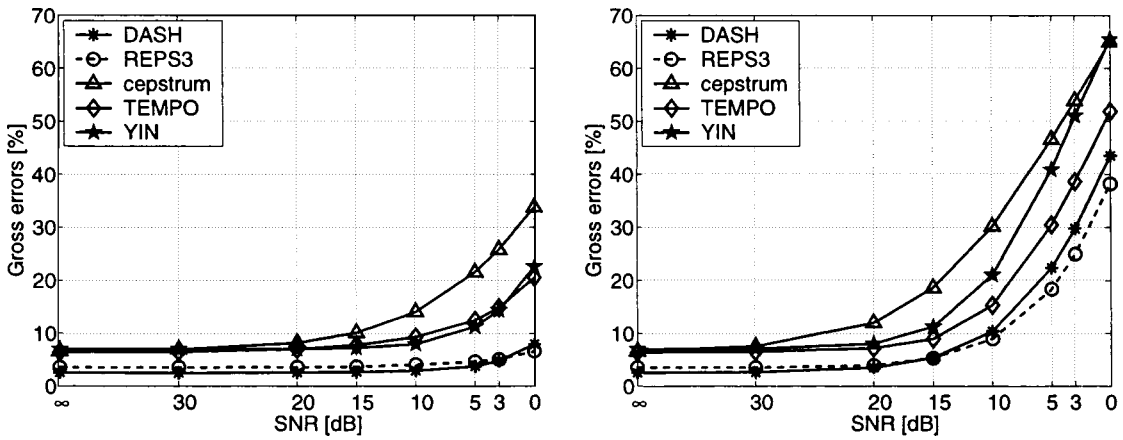


FIG. 6. Gross F_0 errors obtained by DASH, REPS_3, cepstrum, TEMPO, and YIN for speech signals spectrally distorted by SRAEN filtering in the presence of white noise (left panel) and babble noise (right panel).

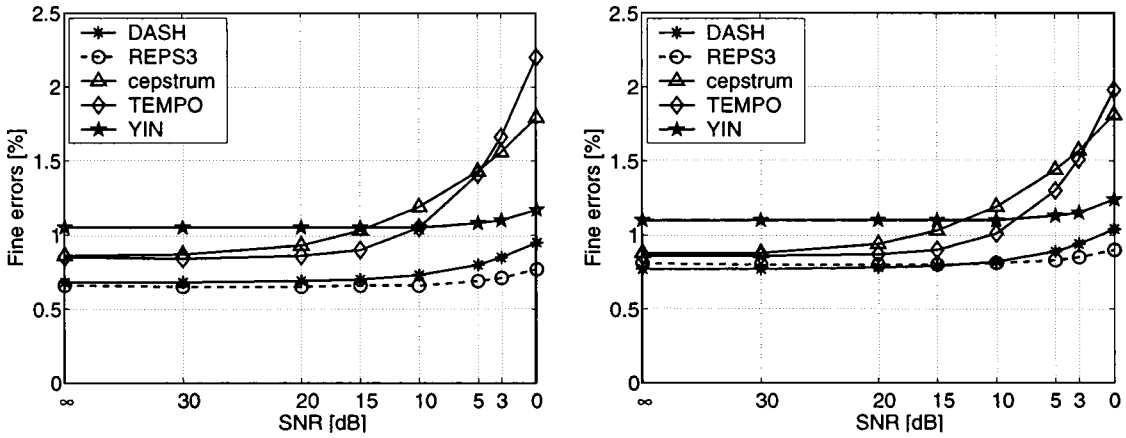


FIG. 7. Fine F_0 errors obtained by DASH, $REPS_3$, cepstrum, TEMPO, and YIN for speech signals without spectral distortion (left panel) and with spectral distortion caused by SRAEN filtering (right panel) in the presence of white noise.

3. Robustness against spectral distortion

To evaluate robustness against spectral distortion, we applied a SRAEN filter to speech signals after adding noise. Figure 6 shows the gross F_0 errors for DB1 obtained with various F_0 estimation methods in the presence of white noise and babble noise. The gross F_0 errors for SNRs of ∞ and 5 dB for DB1 with two types of background noise and a SRAEN filter are also listed in the fifth to seventh columns of Table I. Tables II and III show the gross errors for DB2 and DB3 in the same manner. In these results, the two proposed methods, DASH and $REPS_3$, were always better than any other method. In addition, the errors obtained with DASH were not greatly affected by SRAEN filtering, and were almost always better than the errors obtained with any other method except with DB1 under low SNR conditions. This again implies that the method based on a ripple-enhanced power spectrum is not as robust against spectral distortion as the dominance-spectrum-based method, although it is most robust against background noise. This tendency becomes more obvious when we refer to the results obtained without DP as shown in Sec. III E 5.

From the results presented in Secs. III E 1, III E 2, and III E 3, we can conclude that DASH is recommended for real world applications when the spectral distortion is unknown, unpredictable, or time varying, while REPS is useful when it is possible to compensate the spectral distortion in advance.

4. Accuracy of F_0 estimation

We evaluated the fine F_0 errors of all F_0 estimation methods for DB1 using the two types of background noise and the SRAEN filter. Figure 7 and Table IV show the results. The fine errors were calculated over time frames where all the methods estimated F_0 correctly, i.e., the error was within ±5%, in Secs. III E 1, III E 2, and III E 3. The errors with DASH were almost the same as those with $REPS_3$ and were better than the errors with cepstrum, TEMPO, and YIN. This was because DASH and $REPS_3$ both use F_0 refinement procedures based on fixed points as described in Sec. II F. This shows that the use of fixed points is effective for increasing the accuracy of F_0 estimation.

5. Effect of dynamic programming

Finally, we measured the effect of DP. We compared the gross F_0 errors obtained by the two proposed methods with and without DP using DB1. The results are listed in Table V. Although the gross errors of the methods without DP were worse than with DP, the performance of DASH and that of $REPS_3$ are still generally superior to those of the other methods shown in Table I, except when using $REPS_3$ in the presence of spectral distortion provided by a SRAEN filter. This implies that the F_0 estimation methods based on the dominance spectrum and the ripple-enhanced power spectrum are

TABLE IV. Fine F_0 errors (%) for DB1 for SNRs of ∞ and 5 dB, with and without a SRAEN filter. The bold font shows the best performance. W: white noise, B: babble noise.

	Without SRAEN			With SRAEN		
Noise		W	B		W	B
SNR (dB)	∞	5	5	∞	5	5
DASH	0.68	0.80	1.00	**0.77**	0.89	1.05
$REPS_3$	**0.66**	**0.69**	**0.89**	0.81	**0.83**	**1.00**
cepstrum	0.86	1.43	1.35	0.88	1.44	1.32
TEMPO	0.85	1.41	1.63	0.86	1.30	1.55
YIN	1.05	1.08	1.18	1.10	1.13	1.18

TABLE V. Gross F_0 errors (%) for DB1 obtained by the proposed methods with a ±5% error criterion with/without DP (WDP/WODP) for SNRs of ∞ and 5 dB, with and without a SRAEN filter. The bold font shows the best performance. W: white noise, B: babble noise.

		Without SRAEN			With SRAEN		
	Noise		W	B		W	B
	SNR (dB)	∞	5	5	∞	5	5
DASH	WDP	1.52	2.31	16.4	**2.53**	**3.80**	22.4
	WODP	2.80	6.34	26.9	**5.35**	**7.98**	31.3
$REPS_3$	WDP	**0.89**	**1.26**	**8.37**	3.55	4.67	**18.4**
	WODP	**1.49**	**2.32**	**12.3**	9.07	12.3	**28.2**

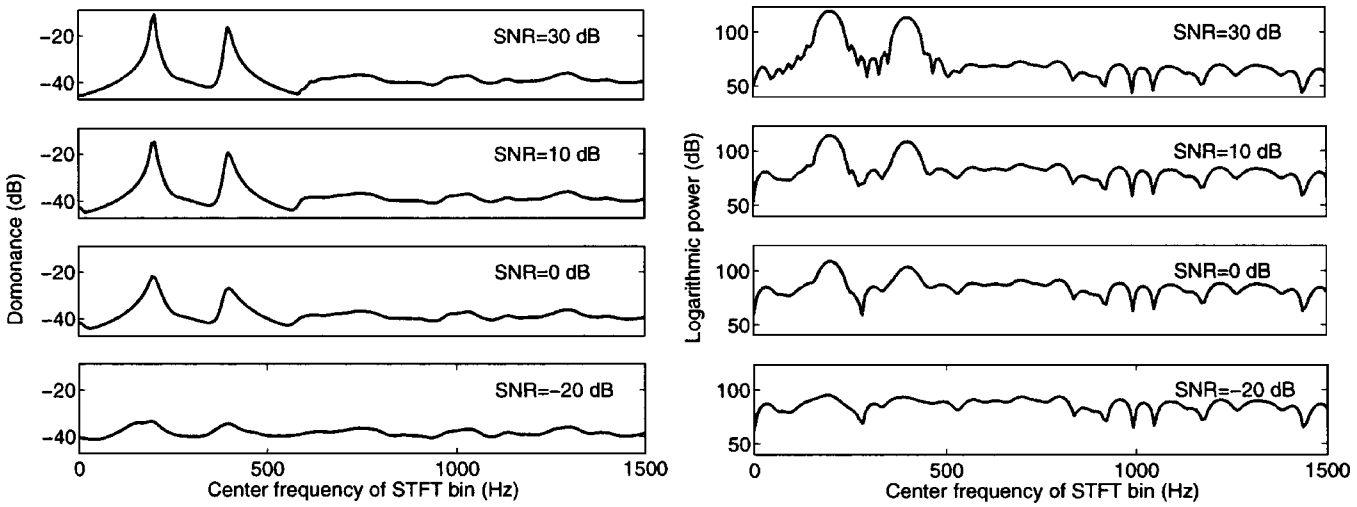


FIG. 8. Dominance spectrum (left panels) and logarithmic power spectrum (right panels) for a mixture of two sinusoidal components and white noise when the SNRs are 30, 10, 0, and −20 dB.

both very robust against background noise, and that the former is also very robust against spectral distortion, even without any F_0 trajectory correction.

IV. CONCLUSION

We proposed a robust and accurate method for F_0 estimation, referred to as DASH, that is based on a new measure, namely the degree of dominance. We defined the degree of dominance to allow us to evaluate the magnitude of the harmonic components of speech signals relative to background noise. DASH was found to be more robust and accurate than conventional methods under background noise conditions of white noise and babble noise. We also demonstrated that DASH was robust against the spectral distortion created by SRAEN filtering. Moreover, we found that a simple method based on a ripple-enhanced power spectrum, referred to as REPS, was a good alternative provided the spectral shape of the input sound was carefully compensated. REPS will be useful when it is possible to compensate the spectral distortion in advance while DASH is recommended for real world applications when the spectral distortion is unknown, unpredictable, or time-varying. Further improvement is required for such applications as sound segregation in the presence of babble noise or under cocktail party conditions. It would be possible to use dual or repetitive F_0 estimation[13] and a method using sound source localization information when a multi-microphone setup is available.[2]

ACKNOWLEDGMENTS

The authors express their gratitude to Professor H. Kawahara of Wakayama Univ. for providing the database of speech with EGG, Dr. A. de Cheveigné of IRCAM for updating his F_0 estimation program, YIN, Dr. S. Katagiri for research support, and members of NTT for helpful discussions. This work was partially supported by CREST of JST.

APPENDIX A: DOMINANCE SPECTRUM DEPENDENCE ON THE BACKGROUND NOISE LEVEL

Figure 8 shows the dominance spectrum and logarithmic power spectrum of a harmonic sound with background white noise. The harmonic sound is composed of two sinusoids of 200 and 400 Hz. White noise is added to the harmonic sound so that the SNR becomes 30, 10, 0, and −20 dB.

The two sinusoidal components have sharper peaks in the dominance spectrum than in the logarithmic power spectrum even when the SNR is low. It is also the case that, independent of the SNR, the logarithmic power spectrum has several clear peaks and deep dips at the frequencies above 800 Hz where only white noise exists, while the dominance spectrum is much smoother in that range. When using F_0-estimation algorithms, such as peak picking or cepstrum methods, the peaks and dips produced by the noise affect the preciseness. It is desirable for the spectral representation to yield distinctive peaks corresponding to harmonics and a flat response in the noise region. This shows that the dominance spectrum is a better representation for F_0 estimation than the logarithmic power spectrum.

APPENDIX B: METHOD FOR AVOIDING DOUBLE/ HALF PITCH ERRORS

The dominance spectrum is always negative when there is no subtraction of $\overline{D(f,t)}$ in Eq. (12). Then, the harmonic dominance of the larger f_n in Eq. (12) tends to become large, resulting in an increase in the double pitch error. This is because harmonic numbers below $F_{\max}$ become smaller as f_n becomes larger, and thus the number of negative values decreases in the harmonic dominance calculation. Conversely, when the dominance spectrum is set positive by adding a large constant value, the half pitch error increases. This is because the harmonic number below $F_{\max}$ becomes larger as f_n becomes smaller, and thus the number of positive values increases in the harmonic dominance calculation. Therefore, it is necessary to normalize the spectrum adequately to avoid these two kinds of errors. For this purpose, we choose the average of the dominance value for all frequencies as the

TABLE VI. Gross F_0 errors (%) using commonly-used F_0 labels with a $\pm 20\%$ error criterion. The bold font shows the best performance. W: white noise, B: babble noise.

		Without SRAEN			With SRAEN		
	Noise		W	B		W	B
	SNR (dB)	∞	5	5	∞	5	5
DB1	DASH	0.30	0.43	8.82	**0.73**	**1.55**	14.1
	REPS	**0.26**	**0.29**	**4.90**	2.11	2.67	**12.7**
	YIN	0.44	2.10	28.4	3.27	7.32	34.6
DB2	DASH	**0.42**	1.34	14.6	**0.63**	**0.97**	**15.1**
	REPS	0.68	**1.05**	**11.1**	2.11	3.25	18.9
	YIN	1.30	4.38	33.5	5.53	10.3	35.0
DB3	DASH	2.81	3.32	16.5	**3.73**	**4.15**	**20.0**
	REPS	2.68	**2.98**	**12.3**	6.91	8.49	26.2
	YIN	**2.57**	7.22	31.0	7.55	14.6	40.0
		Clean			Telephone quality		
	Wang	4.25	···	···	4.34	···	···
	Kasi	(1.7)	···	···	(3.37)	···	···

normalization term. By subtracting the average dominance value from Eq. (12), the dominance values for dominant components become positive while the other dominance values remain negative. We confirmed that this approach was effective in preliminary experiments.

APPENDIX C: GROSS F_0 ERRORS CALCULATED USING COMMONLY USED F_0 LABELS

Although it is very difficult to define universally correct F_0 values, common F_0 labels used for several previous F_0 evaluation reports[6,9,10] may provide better references for a rough comparison of F_0 estimation methods. For this purpose, we evaluated our F_0 estimation methods with the F_0 labels used by de Cheveigné.[9] For DB1 and DB2, we used the labels made by de Cheveigné. For DB3, we used the labels present in the DB itself. All the labels were estimated from EGG signals. In addition, we modified the definition of the gross F_0 errors as the ratios of the estimated F_0 values that fell $\pm 20\%$ beyond the correct F_0 value, in accordance with previous reports.

Table VI shows the results for DB1, DB2, and DB3. These tables also contain the results obtained by YIN in our experiments, and those reported by Wang *et al.*[6] and by Kasi *et al.*,[10] referred to as Wang and Kasi. Note that Wang and Kasi are evaluated under somewhat different conditions. Therefore, direct comparison remains difficult. For example, they did not evaluate their methods under adverse noise conditions. Furthermore, the gross error rates of Kasi were calculated only on frames for which both reference F_0 labels and their voicing decision method indicate voiced frames. The voicing errors of their voicing decision method were reported as being more than 10%, therefore we suspect that their gross error rates would be much worse when they are calculated for all the voicing durations of the reference F_0 labels.

The results in the three tables clearly demonstrate that our proposed methods, DASH and REPS_3, are superior to the other methods.

[1] A. S. Bregman, *Auditory Scene Analysis—The Perceptual Organization of Sound* (MIT, Cambridge, MA, 1990).

[2] T. Nakatani and H. G. Okuno, "Harmonic sound stream segregation using localization and its application to speech stream segregation," Speech Commun. **27**(3–4), 209–222 (1999).

[3] H. Kawahara, I. Masuda-Katsuse, and A. de Cheveigné, "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based f0 extraction: Possible role of a repetitive structure in sounds," Speech Commun. **27**(3–4), 187–207 (1999).

[4] W. Hess, *Pitch Determination of Speech Signals* (Springer-Verlag, Berlin, 1983).

[5] L. R. Rabiner, M. J. Cheng, A. E. Rosenberg, and C. A. McGonegal, "A comparative performance study of several pitch detection algorithms," IEEE Trans. Audio Electroacoust. **ASSP-24**, 399–417 (1976).

[6] C. Wang and S. Seneff, "Robust pitch tracking for prosodic modeling in telephone speech," Proc. ICASSP-2000, pp. 1343–1346 (2000).

[7] T. Shimamura and H. Kobayashi, "Weighted autocorrelation for pitch extraction of noisy speech," IEEE Trans. Speech Audio Process. **9**(7), 727–730 (2001).

[8] D. Liu and C. Lin, "Fundamental frequency estimation based on the joint time-frequency analysis of harmonic spectral structure," IEEE Trans. Speech Audio Process. **9**(6), 609–621 (2001).

[9] A. de Cheveigné and H. Kawahara, "Yin, a fundamental frequency estimator for speech and music," J. Acoust. Soc. Am. **111**, 1917–1930 (2002).

[10] K. Kasi and S. A. Zahorian, "Yet another algorithm for pitch tracking," ICASSP-2002 (2002).

[11] P. C. Bagshaw, S. M. Hiller, and M. A. Jack, "Enhanced pitch tracking and the processing of f0 contours for computer aiding intonation teaching," Proc. EUROSPEECH-93 (1993), Vol. 2, pp. 1000–1003.

[12] F. Plante, W. A. Ainsworth, and G. F. Meyer, "A pitch extraction reference database," Proc. Eurospeech-95 (1995), pp. 837–840.

[13] Y. Atake, T. Irino, H. Kawahara, J. Lu, S. Nakamura, and K. Shikano, "Robust fundamental frequency estimation using instantaneous frequencies of harmonic components," Proc. ICSLP-2000 (2000), Vol. II, pp. 907–910.

[14] ITU-T Recommendation P.11 (1994).

[15] J. L. Flanagan and R. M. Golden, "Phase vocoder," Bell Syst. Tech. J. **45**, 1493–1509 (1966).

[16] F. J. Charpentier, "Pitch detection using the short-term phase spectrum," Proc. ICASSP-86 (1986).

[17] T. Abe, T. Kobayashi, and S. Imai, "The IF spectrogram: A new spectral representation," Proc. ASVA 97 (1997), pp. 423–430.

[18] L. Cohen, *Time-frequency Analysis* (Prentice-Hall, Englewood Cliffs, NJ, 1995).

[19] The integration in Eq. (11) is the numerical integration of the values at discrete points. The frequency quantization is 1 Hz when calculating Eq. (12). Each harmonic frequency, lf_n, in Eq. (12) is transformed to the nearest center frequency by an operator "$r(\cdot)$." Also, in Eq. (16) for F_0 refinement, the operator "$r(\cdot)$" is again used to transform each IF into a discrete center frequency.

[20] The formulation of the degree of dominance is similar in appearance to Cohen's bandwidth equation, i.e., $B^2=\int(f-\bar{f})^2 S(f)^2\,df/\int S(f)^2\,df$, where $\bar{f}$ is an average frequency. However, Cohen's definition does not represent the bandwidth of the instantaneous frequency around a fixed point, but the width of the distribution in the power spectrum around the center of the distribution. The former is considered to be more suitable for evaluating the stability of fixed points. Also, Cohen's definition does not provide a way of localizing integration areas around specified frequencies. Therefore, we use the degree of dominance to improve the accuracy of F_0 estimation.

[21] Under clean speech conditions, the gross pitch error achieved by YIN was 4.3% as shown in Table I, and the result was worse than 0.3% originally reported by de Cheveigné.[9] This is because our evaluation conditions were more severe than theirs. In their work, the F_0 gross errors were calculated according to whether the obtained F_0 was within $\pm 20\%$ of the correct F_0. Moreover, they used their own F_0 labels for the same database. In our preliminary study, we were able to improve the performance of YIN to 0.44% gross errors with their F_0 labels and a $\pm 20\%$ error criterion to simulate their conditions.[9] Note that YIN software was downloaded from de Cheveigné's www page.

A numerical study of the role of the tragus in the big brown bat

Rolf Müller
The Maersk Institute, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

(Received 30 April 2004; revised 15 September 2004; accepted 19 September 2004)

A comprehensive characterization of the spatial sensitivity of an outer ear from a big brown bat (*Eptesicus fuscus*) has been obtained using numerical methods and visualization techniques. Pinna shape information was acquired through x-ray microtomography. It was used to set up a finite-element model of diffraction from which directivities were predicted by virtue of forward wave-field projections based on a Kirchhoff integral formulation. Digital shape manipulation was used to study the role of the tragus in detailed numerical experiments. The relative position between tragus and pinna aperture was found to control the strength of an extensive asymmetric sidelobe which points in a frequency-dependent direction. An upright tragus position resulted in the strongest sidelobe sensitivity. Using a bootstrap validation paradigm, the results were found to be robust against small perturbations of the finite-element mesh boundaries. Furthermore, it was established that a major aspect of the tragus effect (position dependence) can be studied in a simple shape model, an obliquely truncated horn augmented by a flap representing the tragus. In the simulated wave field around the outer-ear structure, strong correlates of the tragus rotation were identified, which provide a direct link to the underlying physical mechanism. [DOI: 10.1121/1.1815133]

PACS numbers: 43.80.Ka, 43.64.Ha [JMS] Pages: 3701–3712

I. INTRODUCTION

Bats have evolved the common anatomical layout of a mammal's outer ear to yield prominent structural features which presumably serve special functions in biosonar sensing. One such feature is the tragus, a frontal prominence, which is highly conspicuous in many bat species. Previous work (Lawrence and Simmons, 1982; Wotton *et al.*, 1995; Wotton and Jenison, 1997) has characterized the sensitivity of the outer ear of the big brown bat (*Eptesicus fuscus*) as a function of direction (directivity) by presenting data on impulse responses or transfer functions. Such results have typically been obtained over a limited angular range and with limited angular resolution. Directivity gain as a function of both of the two angles (azimuth and elevation) which characterize a direction in three-dimensional space has been shown for two frequencies only (Wotton and Jenison, 1997). Similar limitations apply to work performed on other bat species (Grinnell and Schnitzler, 1977; Jen and Chen, 1988; Coles *et al.*, 1989; Obrist *et al.*, 1993; Fuzessery, 1996; Firzlaff and Schuller, 2003). Despite limitations in the scope of the data obtained so far, the research on the big brown bat has revealed interesting functional properties relevant to source localization in elevation. In particular, it has demonstrated the importance of the tragus for elevation discrimination by comparing the performance of animals with natural tragus position to an experimental group for which the tragus had been bent out of the pinna aperture (Lawrence and Simmons, 1982). However, an overview of the spatial distribution of sensitivity with sufficiently high resolution in angle, frequency, and tragus position has not been provided. This prohibits putting features of individual transfer functions obtained for different directions into perspective and hampers understanding the global effect of the tragus underlying these properties. Furthermore, since prior data have been limited to providing black-box system descriptions, they cannot yield definitive insights into the physical mechanisms, which so far have not been observed directly.

Here, a different methodological approach is introduced to this problem, which addresses the above-mentioned shortcomings and hence is capable of yielding a comprehensive picture of the role of the tragus: Based on x-ray computer tomographic cross sections, an accurate high-resolution representation of the shape of an outer ear can be generated. Such a representation can be used for a numerical finite-element simulation study of the diffraction effects taking place around the ear surface. The far-field directional sensitivity arising from the diffraction effects can also be predicted by projecting the wave field outward from the boundaries of the finite-element model's computational domain. This approach offers three key advantages over physical measurements: First, directivity patterns can be estimated with a high angular and frequency resolution, since the time needed to obtain a sensitivity gain estimate is much shorter than in any physical setup with moving parts. Second, the digital representation of the pinna shape can be manipulated in a well-defined way, drawing from the wide range of available computer graphics methods. Because the efficiency advantage of numerical evaluation applies to the manipulated shapes as well, the effects of the manipulations can be studied using high-resolution directivity estimates to describe the functional properties of the outer ear. Third, estimates of the wave field in the volume surrounding the pinna surface, where all diffraction effects take place, are generated as part of the method and can be evaluated to obtain insights into the physical mechanisms behind changes in the directivity pattern. These advantages outweigh disadvantages given by the approximations and simplifying assumptions made by the numerical model. While these approximations represent

sources of error absent in physical measurement, it should be noted that physical measurements have their own sources of error.

In the results presented below, a numerical approach will be used to

(i) obtain a comprehensive spatial picture of the effect the tragus has on the sensitivity of the outer ear;
(ii) understand the role of tragus position (rotation);
(iii) gauge the shape specificity of the spatial sensitivity patterns; and
(iv) characterize the underlying wave-field diffraction effects.

II. METHODS

A. Shape measurement

An outer-ear sample was taken from a big brown bat (*Eptesicus fuscus*) carcass, which had been preserved by cooling. The outer ear was severed from the head with the concha and part of the ear canal left attached. It was mounted (glued to a cylindrical holder) in a life-like manner and subjected to x-ray microtomography in a Skyscan 1072 scanner. X-ray prefiltering was applied to reduce beam-hardening artifacts in the reconstruction. Shadow images of 1024×1024 pixels with a pixel size of 19.35 μm were collected over a half-circle (180°) of target rotation performed with 0.9° resolution. Cross sections of the ear shape were reconstructed using Feldkamp's cone-beam reconstruction algorithm implemented in software provided by Skyscan. Cross-section images (gray scale, 8-bit resolution) were manually postprocessed to remove noise and obvious artifacts which were clearly separated from the ear cross section in the images. A pixel classification into the two categories ''tissue'' and ''air'' was performed by manual adjustment of a gray-value threshold. Stacking of the resulting binary cross-section images produced a three-dimensional voxel-array representation of the shape. It was processed in two ways: for visualization of the shape, a triangular mesh approximation of the boundary surface between ear tissue and air was generated using the marching cubes algorithm (Lorensen and Cline, 1987); these meshes were postprocessed by low-pass filtering and decimation (Schroeder *et al.*, 1992). A rendering of the final shape representation produced by this method is shown in Fig. 1.

Visualization and rendering was performed using functions provided by the VISUALIZATION TOOLKIT software library (VTK, Schroeder *et al.*, 2003). A mesh of finite elements filling the air around the structure was generated by a custom algorithm which evaluated all possible positions for the placement of finite elements in a lattice of cubic elements of uniform size. An element was placed at a given position only if all eight corner voxels for this position had been classified as air. The bias of this decision criterion against placing air elements was adjusted at the voxel classification level, where thresholding was preceded by a three-dimensional smoothing operation using an isotropic Gaussian kernel of adjustable bandwidth.

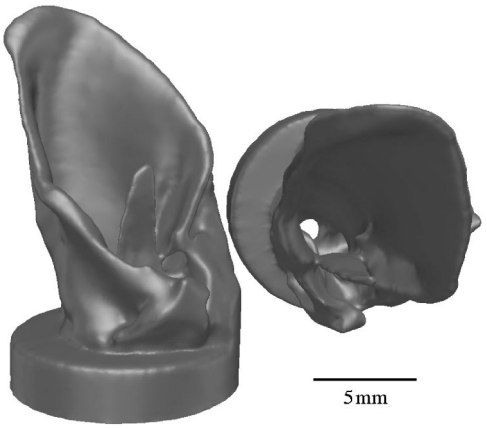


FIG. 1. Tomographic shape reconstruction. Shown is a rendering of a smoothed, triangular surface mesh (119 287 triangles) approximating the boundary between tissue and air.

B. Shape manipulation

The generated ear shape was subjected to two different kinds of manipulation: tragus rotations and shape distortions. Tragus rotation was performed in order to characterize the influence which the relative position of tragus and ear aperture exerts on the directivity. To define the operation, a cuboid was placed manually in the voxel array to enclose the portion of the tragus to be rotated. The cuboid was positioned to include the entire freestanding length of the tragus. The classification of each source voxel (tissue or air) inside the cuboid volume was then transferred to a target voxel, which was found by rounding the rotated position of the source voxel to the resolution of the voxel grid. The classification of all source voxels was set to air, unless they were the target voxel for another portion of the rotated structure. A morphological closing operation was performed after the transfer of the voxel classifications in order to fill any holes produced by the rounding of the target positions in the voxel grid and thereby ensure that the rotated structure was again a homogeneous volume. Rotations were always about an axis placed at the proximal end of the freestanding tragus portion. In the original mounting of the ear, the tragus was placed upright; this position—shown in Fig. 1—is referred to as ''0°.'' Tragus rotation was in the direction outwards from the pinna aperture. It was performed in steps of 15°, spanning a total of 60° from the original position (Fig. 2).

Shape distortions were performed to assess the robustness of the obtained results with regard to any source of shape error or variability by means of a perturbation para-

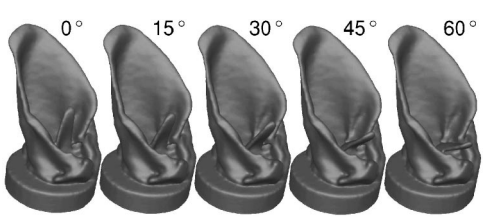


FIG. 2. Shape renderings for tragus rotations performed on the ear shape representation rendered in Fig. 1. The difference in rotation angle between neighboring shapes is 15°; the rotation angle of the upright tragus (left) is defined as 0°.

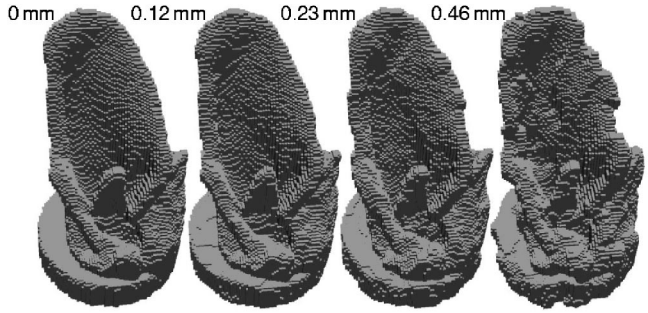


FIG. 3. Example realizations of shapes distorted by random displacement fields. Shown are the boundaries of the finite-element mesh, in which the cubic shape of the individual elements is clearly visible. The parameter (0, 0.12, 0.23, 0.46 mm) is the standard deviation of the displacement magnitude distribution.

digm: the voxel array representation was subjected to random distortions by translating each voxel according to a random displacement vector. The three Cartesian components of the displacement vectors were drawn independently from a zero-mean Gaussian distribution. A strong spatial smoothing of the directions was performed in order to prevent discontinuities in the displacement field which would have resulted in breaking the continuous volume describing the ear tissue. The smoothing operation was carried out using a three-dimensional, isotropic Gaussian smoothing kernel with 15-mm standard deviation applied to each Cartesian vector component. This is equivalent to computing a weighted vector average. The resulting spatially averaged displacement vectors were normalized in length and then scaled with a set of displacement magnitudes drawn independently from a zero-mean Gaussian distribution. The standard deviation of the displacement magnitude distribution was the parameter used to control the amount of shape distortion. The field of displacement magnitudes was spatially smoothed with a kernel identical to the one used for the Cartesian components of the displacement directions. Displacements were applied to the voxel array in the same manner as the tragus rotations (see Fig. 3).

C. Generation of a simple model shape

In order to investigate to what extent the observed properties are specific to the bat ear shape, a simple model shape was studied in parallel with the bat ear. This model shape is based on the obliquely truncated horn introduced by Fletcher and Thwaites (1988) (see also Fletcher, 1992) as an idealized model for the mammalian pinna. In its original form the model does not account for a tragus, however. Therefore, the model was augmented here by addition of a frontal flap intended to mimic the tragus. The overall dimensions of the model shape were adjusted to match the natural ear shape approximately (Fig. 4). The obliquely truncated horn is, however, strictly bilateral symmetric, whereas the natural ear shape strongly deviates from bilateral symmetry.

A detailed description of the augmented obliquely truncated horn model, along with a discussion of other possible features, can be found in Müller and Hallam (2004). Here, the obliquely truncated horn was subjected to the same tragus rotations as the natural ear shape (Fig. 2). The shape

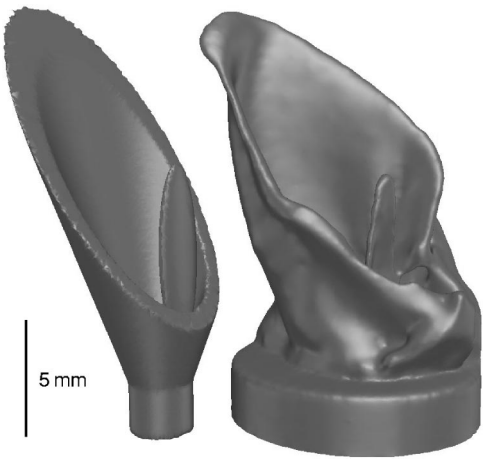


FIG. 4. Comparison of the simple shape model (obliquely truncated horn augmented by a tragus-like flap) with the bat ear shape.

model was implemented using Boolean combinations of implicit functions describing cones (for the inner and outer surface), cutting planes (for the outer-ear aperture and truncation of pinna and ear canal), and cylinders for modeling an ear-canal stump. Binary (1-bit resolution) cross-section images were generated by evaluating the Boolean combinations of the implicit functions and thresholding their function values. The binary cross-section images were then fed into the same processing chain as the tomographic cross sections in order to generate shape visualizations (see Fig. 4) and finite-element meshes.

D. Finite-element model

The finite-element simulation used cube elements (all edges of equal length, eight nodes per element) to approximate the air volume surrounding the outer ear. The air was assigned a density of 1.205 kg/m^3 and a bulk modulus of $1.417\,67\times10^5$ Pa; this corresponds to a sound speed of 343 m/s. The cube mesh was generated from the voxel representation of the shape as described in Sec. II A. Element edge length was set to eight voxels, corresponding to 154.8 μm. Thus, for the shortest considered wavelength ($\sim$4.6 mm at 75 kHz), $\sim$30 elements were placed within one wavelength. A cuboidal computational domain around the ear shape was meshed with finite elements, leaving a minimum distance of several elements between the ear surface and the outer domain boundary on all faces but the bottom face, where the foot plate of the ear shape was placed right on the domain boundary. Where the outer boundary of the computational domain was made up of elements representing air, it was covered with two-dimensional square-shaped absorbing boundary elements, which approximated reflection-free outward propagation (Enquist and Majda, 1977). The ear surface was assumed to be perfectly reflecting.

The finite-element model simulated an acoustical source placed in the proximal ear-canal cross section, treating the ear as a loudspeaker. This approach is valid to determine the ear's properties as a receiver by virtue of the reciprocity principle (Pierce, 1981). It offers the methodological advantage that a single simulation run gives predictions for the field on all faces of the computational boundary, from which

the directional sensitivity of the ear shape can be inferred for all directions as described below. If instead a planar wave impinging from a particular direction were to be simulated, a separate simulation would have to be run for each direction. In order to test the importance of the type of simulated source on the measured system properties, two different sources were tested and the results compared: a point source, where a time-varying boundary condition on pressure was placed on one node near the center of the proximal ear-canal cross section only, and a piston source, where all nodes lying in the cross section were subjected to the same time-varying boundary condition. Taking advantage of the linearity assumption, the signals emitted by the simulated sources were given lattice spectra, which allowed for testing system behavior for an entire set of frequencies of interest in a single simulation. The frequency components present in the signals were spaced 500 Hz apart from 25 to 75 kHz; this covers most of the frequency band of the emission of the big brown bat as reported by Hartley and Suthers (1989) (25 to 90 kHz).

Assembly of the master stiffness matrix from the finite-element mesh and computation of a solution via a preconditioned conjugate gradient method was performed using the COUPLED FIELD SIMULATION (CFS++) program developed by the Department of Sensor Technology at the University of Erlangen-Nürnberg, Germany.

E. Directivity estimation and visualization

The wave fields generated by diffraction from shapes with as little symmetry as the shapes considered here will in general show a joint dependence on all three spatial coordinates and frequency. However, in the far field, the dependencies on distance and angle separate and the wave-field amplitude P can be expressed as

$$P(\phi,\theta,r,\omega) \overset{r\gg 8a^2/\lambda}{\rightarrow} D(\phi,\theta,\omega)\frac{1}{r}e^{-jkr}, \tag{1}$$

where D is the directivity, which depends on direction (two angles, azimuth θ and elevation ϕ) and frequency ω only (Urick, 1983). The second term $[(1/r)e^{-jkr}]$ is the expression for the wave field of a point source; it depends only on distance r and wave number k, which is proportional to frequency. Hence, the directivity D is a complete far-field system description which contains all the information specific to a particular source. In the following, a normalized version of D is presented in order to treat the influence of direction separately from overall gain. The lower boundary for the far field given in Eq. (1) is $r\gg 8(a^2/\lambda)$, where a is the characteristic dimension of the source (here the ear) and λ is the wavelength. The height of the ear shape is $\sim$18 mm; hence, it may be assumed that $a\approx 9$ mm. The smallest wavelength considered here is $\sim$4.6 mm ($\sim$75 kHz). Consequently, the far-field boundary is at $r\gg 14.4$ cm and—at the chosen resolution of the finite element mesh—$\sim$1000 elements would have to placed along each dimension to reach it. In three dimensions, the size of the finite-element problem would become a concern in terms of computation resources and accumulation of numerical errors, if the computation were to be extended significantly beyond the far-field boundary.

To avoid these problems, the wave fields were forward-projected from the boundary faces of the computational domain, which were themselves still inside the near field. Sound propagation into a space free of any diffracting objects (and sources) can be described by a Kirchhoff integral formulation (Jackson, 1999), which was used here to perform the forward projection of the wave field (Ramahi, 1997): In the frequency domain, the wave field $\Psi(\mathbf{x})$ at position $\mathbf{x}$ outside the boundary surface S can be expressed as

$$\Psi(\mathbf{x}) = -\frac{1}{4\pi}\int_S \frac{e^{jkR}}{R}\mathbf{n}\cdot\left[\nabla\Psi + jk\left(1+\frac{j}{kR}\right)\frac{\mathbf{R}}{R}\Psi\right]ds, \tag{2}$$

where $\mathbf{R}$ is the vector between the surface element ds and the position $\mathbf{x}$, $\mathbf{n}$ the surface normal, Ψ in the argument of the integral the field value on S, and k the wave number. The product $\mathbf{n}\cdot\nabla\Psi = \partial\Psi/\partial\mathbf{n}$ is the derivative of the field Ψ with respect to the surface normal $\mathbf{n}$. In order to obtain a prediction of the directivity, the positions $\mathbf{x}$ were chosen on a hemisphere which was placed over the ear. The radius of the hemisphere was chosen large enough to satisfy the condition given in Eq. (1). Because all positions on the surface of the hemisphere differ only in direction and not in distance, the normalized field amplitudes $\Psi(\mathbf{x})$ can be taken as estimates of the directivity D.

The transformation from the time-domain finite-element results to the frequency-domain representation used in Eq. (2) as well as for the directivities was performed using the Goertzel algorithm (Mitra, 2001). Since directivity is a stationary system characterization, data points containing transients generated at the start of the time-domain simulation were discarded prior to this transformation.

The directivity results were visualized using Lambert's azimuthal equal-area projection (Weisstein, 2002). The equal-area property of this projection is desirable for the visualization of directivities, because the relative cross-sectional area of lobes in the beam patterns can be compared in the map. At the same time, the projection keeps distortions of lobe shape close to the theoretical minimum, even if a large portion of the directivity (up to a hemisphere) is shown in a single map.

The combination of directivity and spherical spreading losses, like the geometrical attenuation included in Eq. (1), results in the formation of three-dimensional "sound beams:" in the directions where the directivity has a (local) maximum, sensitivity falls below a given threshold at a larger distance than for the directions of the adjacent (local) minima. These sound beams were visualized by rendering of iso-sensitivity surfaces. Since directivity is also a function of frequency, there is a different isosurface for each frequency. Isosurfaces for a set of linearly spaced frequencies were rendered with different gray levels and overlaid in one image to allow for an easy comparison of the sound beam patterns.

The isosurfaces were generated by evaluating the product of directivity and spherical spreading given in Eq. (1). Absorption as a source of spreading loss was not considered for the sake of simplicity, since only the effect of the frequency-dependent effect of directivity was to be visual-

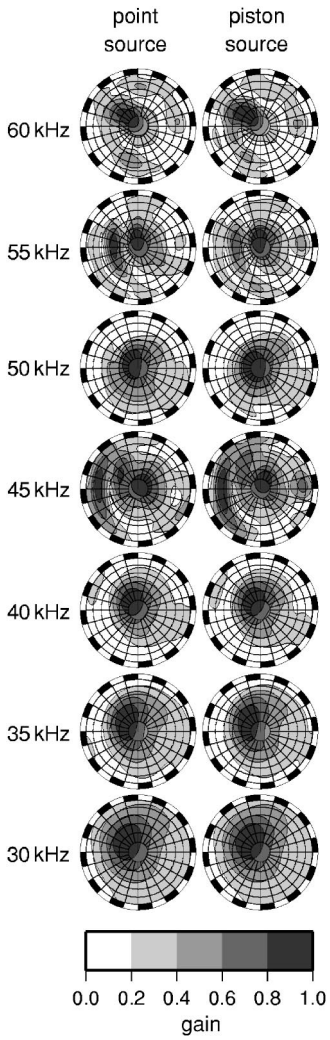


FIG. 5. Comparison of directivity pattern estimates for two different sources of excitation (point and piston source, columns). The directivities are normalized individually for each of the frequencies (rows) shown. The graphs show Lambert's azimuthal equal-area projections of the directivity.

ized in a three-dimensional way. To this end, it suffices that the directivity is multiplied with a term which is a monotone decreasing function of distance and independent of direction.

III. RESULTS

A. Robustness

1. Influence of the excitation source

A comparison of directivity results which were not normalized for each frequency (not shown) indicated a strong effect of the chosen excitation source (piston and point source). In the normalized directivities (Fig. 5), however, only minor differences were found and the overall shape of the directivities was the same for both sources.

Therefore, the choice of the source affected primarily the overall sensitivity as a function of frequency, but it had only a negligible effect on how this sensitivity depended on direction for any given frequency.

2. Influence of random shape distortions

The influence of shape distortion on directivity estimates was tested over a range of values for the standard deviation of the magnitude of the random displacements. The tested values resulted in distortions ranging from very slight to major (Fig. 3). Directivity patterns estimated for example realizations (Fig. 6) were found to be consistent with the notion that—in general—larger shape distortions can be expected to result in larger changes in the directivity patterns. Small shape distortions (standard deviation$<$0.1 mm, i.e., $\sim$2/3 of a finite-element edge length; see Fig. 6) did not alter the pattern of the major lobes in the directivity; the location, shape, and relative weight of these features were largely preserved. Larger shape distortions did change the overall layout of the directivity pattern, although some common features were identifiable even for the largest shape distortion shown here (standard deviation 0.46 mm; see Fig. 3 for a rendering of the tested realization). From the examples examined, it was not apparent that the degree to which shape distortions changed the directivity pattern was dependent on frequency. Hence, it cannot be said that within the studied frequency band (25–75 kHz) low-frequency directivities were more or less susceptible to shape distortion than high frequencies.

B. Tragus rotation and directivities

Rotation of the tragus had a clear, systematic effect on the directivity pattern in the medium frequency range studied. Between $\sim$35 and $\sim$50 kHz, the strength—and to some extent the location and shape—of an extensive, asymmetric (one-sided) sidelobe was dependent on tragus rotation. In the upright tragus position (0°), the sidelobe had the highest relative sensitivity; it was gradually attenuated as the tragus was rotated out of the pinna aperture (see Figs. 7 and 8).

While tragus rotation did result in a large number of small changes to the directivity, the effect on the asymmetric sidelobe stands out by virtue of its magnitude and systematic nature. For frequencies below 35 kHz (e.g., 30 kHz in Fig. 7), no sidelobe is formed, and the effect of tragus rotation on the directivity pattern is minor. For frequencies above 50 kHz, sidelobes are common, but none of them was found to show a systematic dependence on tragus rotation. The results for 53 kHz in Fig. 7 are an example of a strong sidelobe in this frequency band; like the broad main lobe at 30 kHz, it is largely unaffected by the tragus rotation. At a higher frequency resolution (Fig. 8), it is evident that the orientation of the asymmetric sidelobe depends on frequency in a systematic fashion: the sidelobe shifts to higher elevations as frequency increases.

C. Wave field correlates

In the wave fields simulated inside the computational domain of the finite-element model effects were found which showed a systematic dependence on tragus rotation. This dependence paralleled the effects seen in the directivities. The effects were particularly obvious in the phase component of monochromatic wave fields. In the frequency band of the asymmetric sidelobe, rotation of the tragus caused systematic

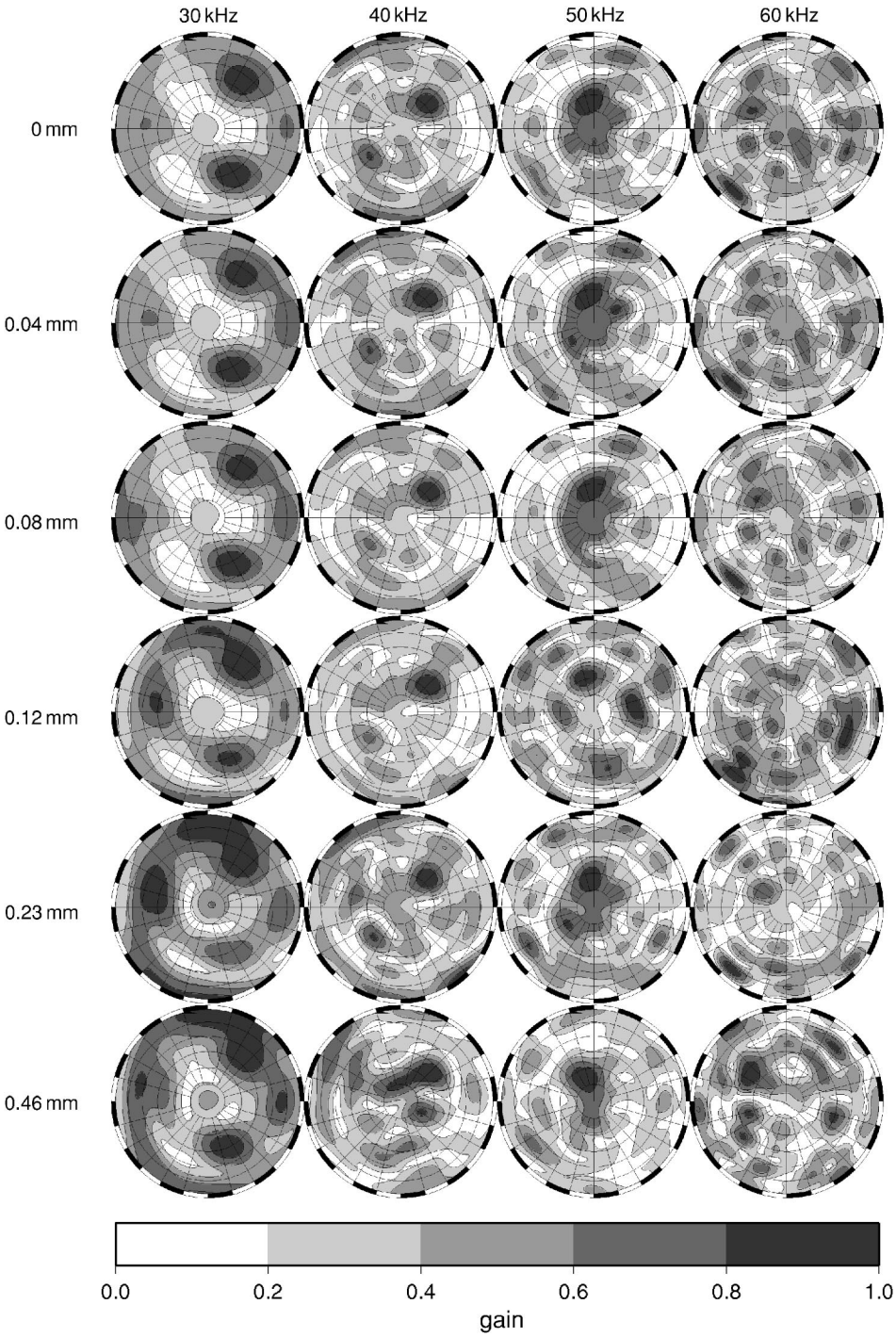


FIG. 6. Directivity estimates for example realizations of distorted shapes. The graphs show Lambert's azimuthal equal-area projections of the normalized directivity. The corresponding shape realizations for 0, 0.12, 0.23, and 0.46 mm are shown in Fig. 3.

shifts in wave field phase in two regions: on the outer side of the tragus, in the area where the tragus obstructed the direct line of sight from the source in the ear canal, and inside and upwards from the tragus (Fig. 9).

Taking the wave field phase values at 60° tragus rotation, where the sidelobe is strongly attenuated, as a reference, it can be seen that phase angles in the former region tend to be larger than the reference, whereas those in the latter tend to be smaller (Fig. 9). The extent of the phase advance region increases as the tragus is rotated upwards. Likewise, the region of phase lag changes shape and the lag magnitude increases (Fig. 9). In a quantitative comparison of phase values along a transect which cuts through both regions, it was found that a strict ordering of phase magnitudes according to tragus rotation exists (Fig. 10). In the phase advance region, upward rotations of the tragus always resulted in larger phase values, whereas in the phase lag region the reverse was the case.

D. Comparison with simple model shape

In general, the directivity patterns predicted for the augmented truncated horn were found to be quite different from those predicted for the bat outer ear. However, despite the overall dissimilarity, an asymmetric sidelobe was found to be

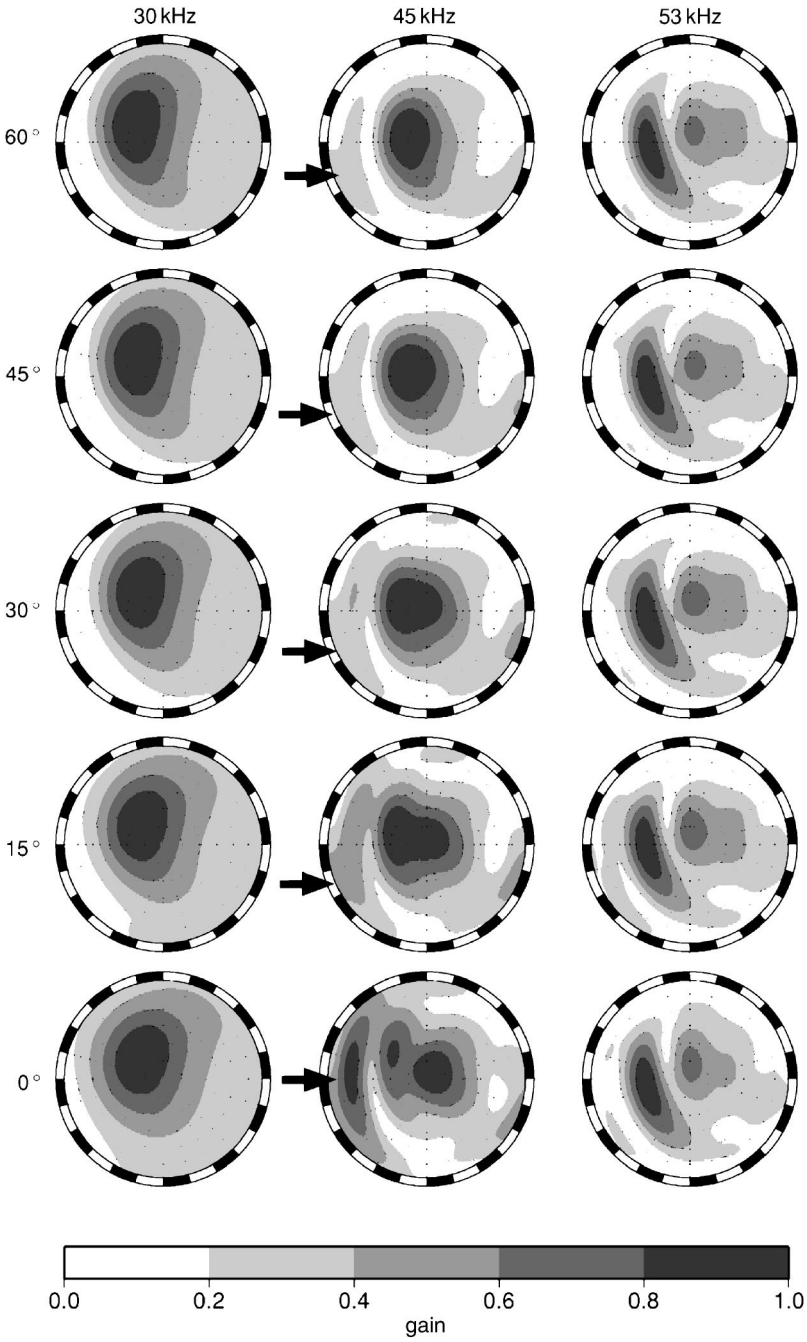


FIG. 7. The effect of tragus rotation (rows) on a large asymmetric sidelobe of the directivity pattern. Shown is an example of a directivity pattern with sidelobe for 45 kHz (center column, approximate sidelobe location marked by an arrow). Examples of directivities for frequencies below (30 kHz) and above (53 kHz) where the effect is absent are shown for comparison. The tragus is pointing approximately to the left in the graphs; the corresponding tragus rotations are shown in Fig. 2. The graphs show Lambert's azimuthal equal-area projections of the normalized directivity.

present at 34.5 kHz which showed the same dependence on tragus/flap rotation as the sidelobe seen in the natural ear shape (Fig. 11).

It was strongest in the upright position of the flap and systematically decreased in sensitivity gain as the flap was rotated downward out of the aperture.

IV. DISCUSSION AND CONCLUSIONS

The directivity pattern of the bat pinna shape studied here arises from sound diffraction at a rather complicated shape devoid of any symmetry which could be used to simplify treatment. Furthermore, since the wavelengths and the size of the outer ear and its structural features—such as the tragus—are similar to each other, the approximations of geometrical acoustics are not valid. The latter insight is not taken into account by ideas which try to explain the acoustical properties of the pinna–tragus combination with distinct ''secondary reflections'' as alluded to in Lawrence and Simmons (1982) and Wotton *et al.* (1995). Instead, an impinging wave is diffracted by the entire shape and the interactions between different parts of the shape (e.g., tragus and pinna) are a continuous process rather than one which could be readily broken up into separate, time-discrete effects. In such a situation, a numerical approach offers the only access to the mechanisms which cause the functionally significant features of the directivity pattern.

A. Robustness and validity of the results

Because no information on the motion pattern of the tympanic membrane in the big brown bat is available, an arbitrary choice for the acoustic excitation source to be used

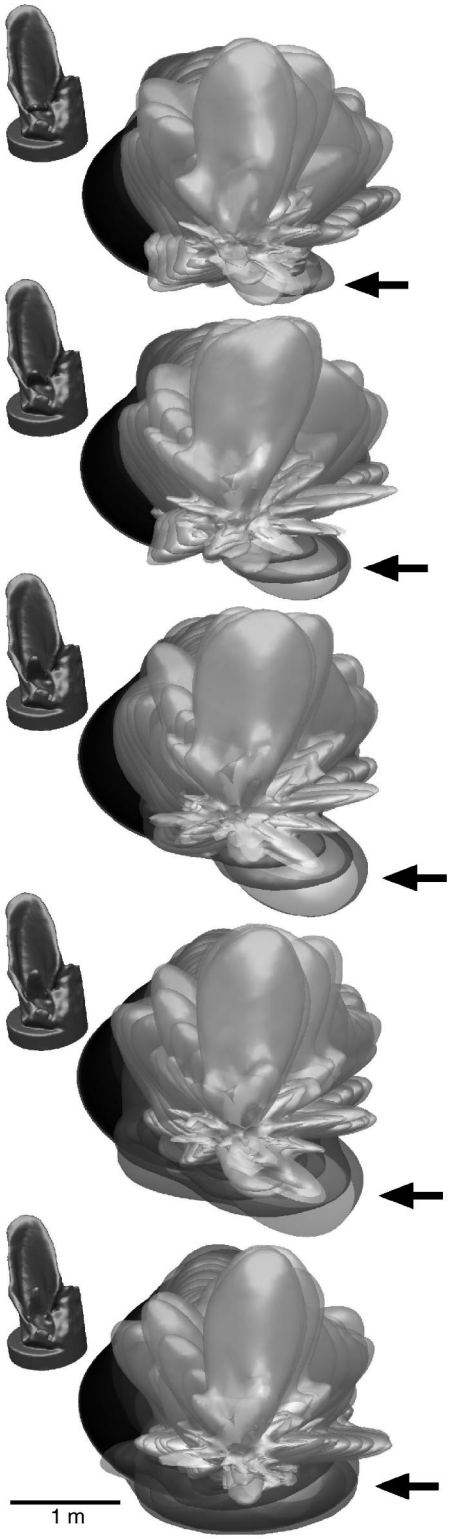


FIG. 8. Three-dimensional iso-sensitivity surfaces for different tragus rotations and frequencies. For each tragus rotation, isosurfaces for frequencies from 25 to 75 kHz in steps of 500 Hz are rendered. Dark-gray values correspond to low frequencies, brighter grays to higher frequencies. The asymmetric sidelobe affected by the tragus rotation is marked by an arrow. The ear shapes are shown as a reference for the orientation of the isosurfaces with respect to the ear. They are 70× magnified with respect to the directivities and displaced outside the isosurfaces to provide an unobstructed view.

in the simulation had to be made. However, the comparison of the results for two very different source choices tested (piston and point source) indicates that this is not a relevant issue for the prediction of directivities: while the nature of the source does affect the overall transfer function of the ear, i.e., the spectral weighting across frequency bands, it does have only a minor effect on normalized directivities, i.e., the spatial weighting within a given frequency band. These experimental results are in agreement with theoretical predictions (Fletcher, 1992): if a source is placed at one end of a pipe which has a diameter smaller than the wavelength, the wavefront at the other end of the pipe will be planar to a very good approximation, irrespective of the source type. The preserved portion of the ear canal represents a pipe for which the scale condition (wavelength larger than pipe diameter) holds for all frequencies considered here. Therefore, both theoretical considerations and experimental results corroborate the notion that directivity patterns can be predicted without knowing the motion pattern of the tympanic membrane. Inaccessibility of the overall frequency transfer function does not affect the objective of the work reported here, namely understanding the spatial sensitivity distribution generated by the pinna. The directivities of the sonar emission and the pinna are the only two points along the sonar communication channel where the spatial sensitivity of the system can be controlled. Understanding the spatial characteristics at these two stages is hence of prime importance. In contrast to this, the overall frequency transfer function is not a spatial property and may be controlled at any internal stage of the channel.

Because only a single ear was used for this study, some sources of variability in ear shape such as individual and gender differences, geographical variation, and postmortem changes cannot be considered. All these factors have the potential for introducing a considerable amount of variability, an assessment of which would most likely require the study of a fairly large sample. However, the shape distortion experiments (see Sec. II B) address this issue by taking a bootstrap approach which creates variability around the sample in order to assess the robustness of the results. Two insights can be gained from the outcome of these experiments: first, the change in directivity as the structure is perturbed is gradual and the basic shape of the directivity is lost only as perturbations get fairly large. This demonstrates that the results are robust against small errors introduced by sample mounting and the shape estimation procedure. Second, larger shape distortions can alter the directivity pattern thoroughly, even if the basic pinna–tragus combination is left intact. This finding offers an experimental corroboration of the fundamental insight that diffraction effects integrated over the entire shape are responsible for the functional (directivity) properties. The specific combination of properties seen in the ear sample, i.e., an asymmetric sidelobe, which occurs over a fairly wide frequency band, has a frequency-dependent axis and a sensitivity controlled by tragus rotation, likely to depend on several aspects of this particular shape. However, the results obtained for the augmented obliquely truncated horn model (see Sec. III D) demonstrate that some individual aspects of the overall behavior can be reproduced by a model

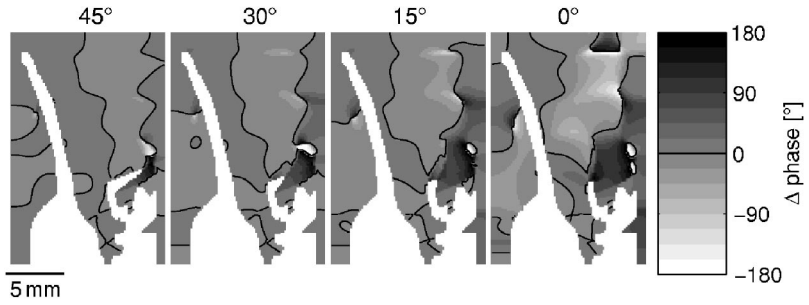


FIG. 9. Difference in wave field phase for a monochromatic field of 45 kHz. The reference is 60° tragus rotation. Depicted is the difference between the phase values of the fields obtained for the other rotations (45° to 0°) and the reference; the corresponding tragus rotations are shown in Fig. 2. Positive values indicate a smaller reference.

which shares very little structural detail with the original ear shape.

Besides the validity issues discussed above, it should be noted that the results presented here solely pertain to the shape of the pinna. In doing so, two simplifying assumptions were made: first, it was assumed that the air–pinna boundary is perfectly reflecting. This simplification is motivated by the large difference between the acoustic impedances of air (~413 rayl under the simulated air properties) and tissue (on the order of magnitude of 10^6 rayl). Second, the shape of the pinna was considered in isolation rather than together with the head. This was necessitated by the limited size of the tomographic imaging volume. While it can be expected that the major shaping influence on the directivity pattern is exerted by the structures which immediately surround the opening of the ear canal, it is possible that parts of the head which are close to the pinna have a modifying influence. This is particularly likely for the lower frequencies where the distance is smaller in terms of the number of wavelengths needed to span it.

B. Directivities and the role of the tragus

The directivity $D(\phi,\theta,\omega)$ is a complete description of the spatial sensitivity of an antenna in the far field. However, it is a function of three variables, two angles (ϕ,θ) and frequency (ω), which poses a challenge for visualization. Head-related transfer functions are one method for making the directivity accessible: for selected combinations of angle values (ϕ and θ), a one-dimensional function of sensitivity gain versus frequency is plotted. Since the spatial variables (ϕ,θ) only appear as labels on a set of one-dimensional functions, head-related transfer functions are a poor choice for making the spatial dimension of the directivity accessible. In the spatial view presented here, a two-dimensional function of the spatial variables is plotted for a set of selected frequency values. If the directivity functions are plotted as a two-dimensional map (as done in Figs. 5, 6, 7, and 11), this necessitates one graph for each frequency. From such maps, the shape of the directivity for a single frequency can be readily assessed, if an adequate map projection is chosen. The choice of the projection is important, since the spatial relations on the surface of a sphere can never be accurately represented in a plane. Therefore, a carefully considered compromise has to be made. Plots using azimuth and elevation as Cartesian coordinates [as presented in Wotton and Jenison (1997)] heavily distort spatial relationships and hence are difficult to read. Irrespective of the chosen projection, comparing the directivities obtained for different, fixed

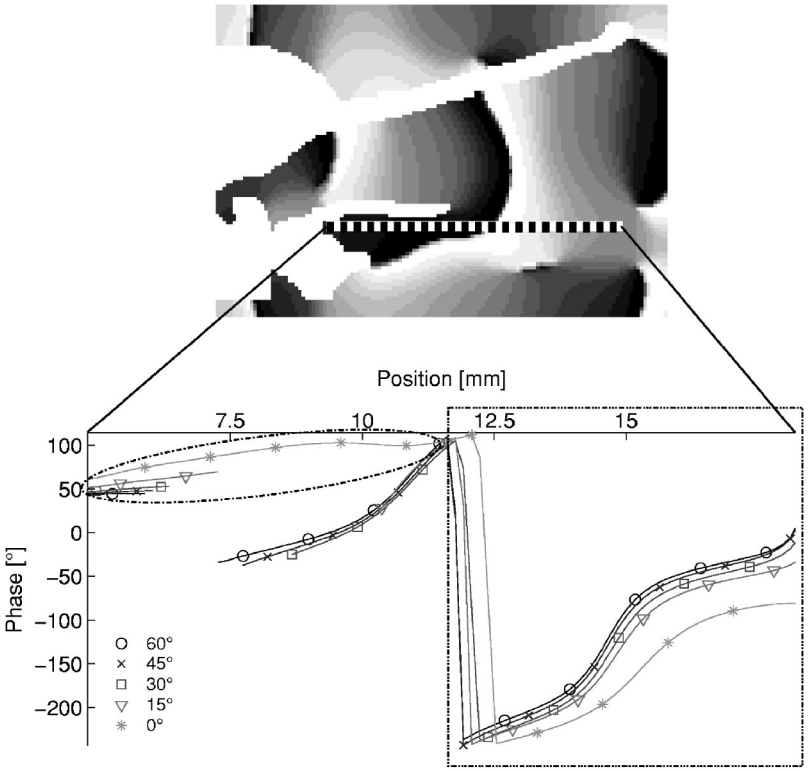


FIG. 10. Wave field phase values along a linear transect through computational domain. The wave field is monochromatic (excitation frequency 45 kHz). Top graph: phase values (coded by gray values) for upright tragus position (0°) and location of the transect (dashed line). Bottom graph: phase values along the transect for all tragus rotations tested. Forward phase shifts with smaller tragus rotation angles are enclosed by the ellipse, phase lags by the square. The interruptions in all curves but the one for 0° at positions in the vicinity of 7.5 mm are due to the transect cutting through the tragus for these shapes. The corresponding tragus rotations are shown in Fig. 2.

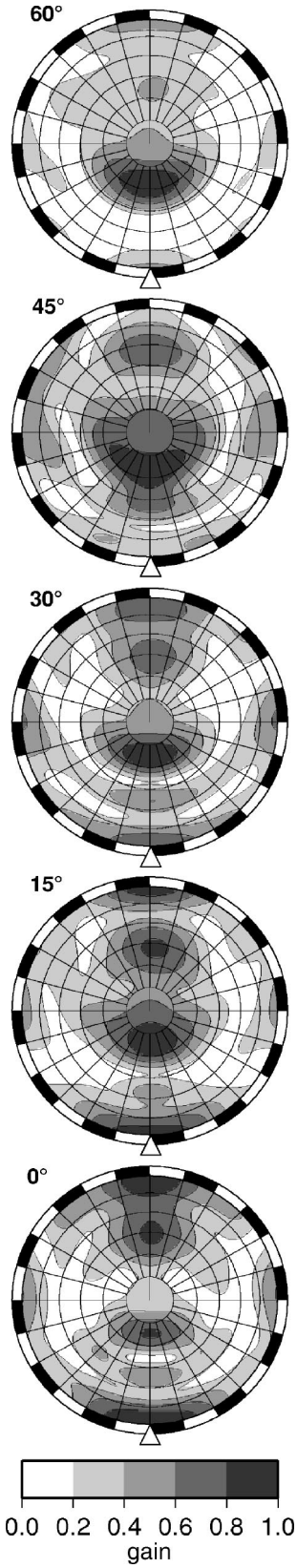


FIG. 11. In the directivity patterns predicted for the augmented obliquely truncated horn (see Fig. 4), an asymmetric sidelobe can be seen for a frequency of 34.5 kHz (approximate location indicated by triangle), which exhibits a similar dependence on tragus rotation as found for the natural bat ear shape (see Fig. 7). The graphs show Lambert's azimuthal equal-area projections of the normalized directivity; the flap modeling the tragus is pointing to the bottom of the graphs.

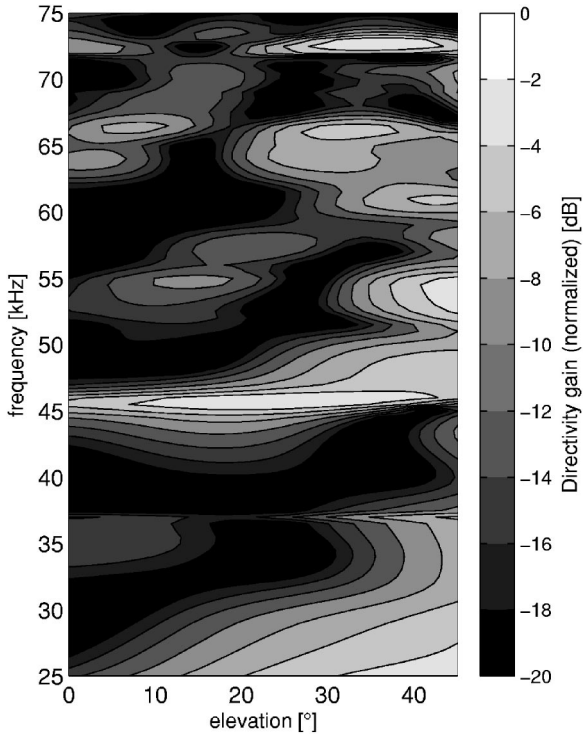


FIG. 12. Example of directivity gains as a function of one angle (elevation) and frequency. Directivity gains are normalized for each elevation value. The foot plate of the shape (see Fig. 1) is aligned with the azimuthal plane and zero elevation is at the equator.

frequencies requires superimposing two or more maps on each other mentally, since a graphical superposition would become cluttered quickly. This encumbers obtaining an understanding of the directivity as a function of all three independent variables. Three-dimensional renderings of sensitivity gain isosurfaces for different frequencies can be superimposed on each other in a single graph (see Fig. 8). While occlusions of one surface by another hide some detail, these renderings provide an intuitive overall view of sensitivity gain as a joint function of spatial variables and frequency.

The numerical simulation and visualization approach taken here has produced a concise description of the impact the tragus has on the spatial sensitivity of the outer ear in the big brown bat: it causes the formation of an asymmetric sidelobe limited to a certain frequency band (here ~35 to ~50 kHz) and the direction the sidelobe is pointing in shifts smoothly upwards with frequency (see Fig. 8). These findings are at least in qualitative agreement with transfer function data presented in Wotton *et al.* (1995), where a spectral notch was observed which shifted upwards in frequency (from ~45 to ~55 kHz) as a function of elevation. An example of directivity gains from the present numerical results arranged as a function of one directional angle (defined as elevation) and frequency shows a similar behavior (Fig. 12).

A precise match between these data and the measurement results of Wotton *et al.* (1995) is not to be expected, because of individual differences between animals, estimation errors associated with both physical measurements and numerical predictions, and—probably most importantly—the arbitrary nature of the choices made in positioning the azi-

muthal plane and the origins of azimuth and elevation. The three-dimensional view presented here puts the results on moving spectral notches into a spatial perspective: occurrence of a spectral notch at a certain frequency in the head-related transfer function for a certain direction means that this direction coincided with the direction of the spatial notch in the directivity for that particular frequency. Moreover, a full view of the directivity can be used as a basis for assessing the role of the outer ear in a different class of biosonar estimation tasks. Head-related transfer functions are useful system descriptions in estimation tasks which can be performed by spectral template matching. Elevation estimation for targets with transfer functions which do not obscure the spectral templates is an example of such a task. However, this estimation approach cannot be extended to biosonar estimation tasks which involve extended targets with many reflecting facets (Müller and Kuc, 2000). For these tasks, the spatial view presented here is more useful because it reveals the spatial weighting imposed on a random spatial distribution of scatters by the outer ear.

C. Wave field correlates

Digital shape manipulations allow well-localized and quantifiable changes to the outer-ear shape to be made. This was used in the tragus rotation experiment and revealed that the strength of the asymmetric sidelobe was strictly decreasing over five successive steps of outward tragus rotation. These findings not only establish a convincing correlation between the tragus and sidelobe formation, they also provide a starting point for the search of further correlates closer to the cause of these effects. Correlates matching the monotone relationship between tragus position and sidelobe strength for the entire sequence of rotation values are likely to be nonrandom and linked to the mechanism by which tragus rotation controls sidelobe strength. The identified changes in the phase of the simulated wave field meet this criterion: the extent of the phase-shift regions as well as the magnitude of the phase shifts showed a monotone dependence on tragus rotation.

Wave field properties in the immediate vicinity of the diffracting structure are interesting, because the causes of sidelobe formation must be localized where diffraction takes place. A hypothesis for a mechanism of sidelobe control must link the changes in shape to changes in the wave field in the vicinity of the structure, and these wave field changes to the far-field directivity. As to the latter step, it may be hypothesized that the two observed phase shifts in the outgoing wavefront—forward on the outside and backwards inside and above the tragus—cause destructive interference in the direction of the notch, which separates main- and sidelobe, and constructive interference in the direction of the sidelobe maximum. As to the link between shape- and wave-field-changes, the phase advance on the outside of the tragus may be attributed to diffraction around the tragus. If a wave has to travel further to a certain point in space, its phase will be advanced at that point. As the tragus is rotated outward and out of the direction of sound propagation, the region accessible only through diffraction around the tragus shrinks, as was observed for the phase advance region. The connection between the observed phase lag and the properties of the shape is not as readily made as for the phase advance. It may depend on less well-localized diffraction effects and hence requires further analysis.

D. Insights from simple shape models

Simplified shape models have the potential to facilitate the analysis of the physical mechanism linking tragus rotation and sidelobe formation. In order to qualify as a comprehensive model, they should reproduce all salient properties of the sidelobe seen in the natural pinna shape, in particular its dependence on tragus rotation and on frequency. The results presented here for the augmented obliquely truncated horn model show only the reproduction of one property, a sidelobe strength which depends on tragus rotation. They did not show the gradual dependence of sidelobe orientation on frequency. Instead, the directivity pattern changed profoundly, even for slight changes in frequency. Therefore, at present, this particular model shape could only be used to study one aspect of the effect. These results do not prove, however, that the general shape model is incapable of reproducing the sidelobe behavior of the bat outer ear in a more comprehensive fashion. This may still be possible for a different set of shape parameters. Further results on the augmented obliquely truncated horn model published elsewhere (Müller and Hallam, 2004) demonstrate that interactions between the tragus and shape features of the pinna (e.g., a surface ripple) take place and can have a profound effect on the directivity. Because this expands the parameter space beyond what has been presented here, finding a suitable combination of parameter values would certainly require a major search effort, which could be undertaken using a numerical searching technique, for instance a genetic algorithm.

E. Tragus rotation as a control mechanism

Tragus rotations were performed here as an experimental method to firmly establish that the tragus is the cause for the large asymmetric sidelobe. However, many bat species, including the big brown bat, show intricate patterns of ear movement (Valentine *et al.*, 2002). If muscular actuation of the ear allowed some of these species to control tragus rotation relative to the pinna, these animals would also have control over certain aspects of their ear directivities. Since only the relative position of tragus and pinna matters, this does not necessarily require that the tragus itself is being moved. A fixed tragus and a mobile pinna or a bit of mobility on both sides would be alternative ways to achieve the same effect. If the pinna were to be moved, however, this would most likely imply a coupling between the overall orientation of the directivity function (controlled by where the pinna is pointing) and its shape. A—hypothetical—bat with a combination of the suitable ear mobility and a tragus with a similar functional significance as in the big brown bat would have control over the extent to which a sidelobe is present in its directivity. Such a bat could extend or retract the sidelobe whenever it facilitates achieving its sensing objectives. A comparison of echoes received with and without the sidelobe being present in the directivity pattern would allow the clas-

sification of echo features as being caused by targets in the direction of the main lobe or by targets in the direction of the sidelobe. This classification could be performed even if main- and sidelobe were both populated with a large number of scattering facets which would prohibit the application of conventional direction estimation techniques. In such a case, the sidelobe could be used like the side view or rear-view mirror of a car, to which the driver can attend at will, whenever deemed necessary.

ACKNOWLEDGMENTS

The ear sample was kindly provided by the Department of Animal Physiology, University of Tübingen, Germany. Joris J. J. Dirckx and Stefan Gea of the Department of Physics, University of Antwerp (RUCA) made their microtomography facilities available and provided essential help with the data collection. Manfred Kaltenbacher and Alexander Streicher of the Department of Sensor Technology, University of Erlangen-Nürnberg shared the CFS++ source code and provided crucial advice on the finite element method. This work was supported by the European Union (IST Program, LPS Initiative, CIRCE Project IST-2001-35144).

Coles, R., Guppy, A., Anderson, M., and Schlegel, P. (**1989**). ''Frequency sensitivity and directional hearing in the gleaning bat, *Plecotus auritus* (Linnaeus 1758),'' J. Comp. Physiol., A **165**, 269–280.

Enquist, and Majda, (**1977**). ''Absorbing boundary conditions for the numerical simulation of waves,'' Math. Comput. **31**, 629–651.

Firzlaff, U., and Schuller, G. (**2003**). ''Spectral directionality of the external ear of the lesser spear-nosed bat, *Phyllostomus discolor*,'' Hear. Res. **181**, 27–39.

Fletcher, N. H. (**1992**). *Acoustic Systems in Biology* (Oxford University Press, Oxford).

Fletcher, N. H., and Thwaites, S. (**1988**). ''Obliquely truncated simple horns: Idealized models for vertebrate pinnae,'' Acustica **65**, 194–204.

Fuzessery, Z. (**1996**). ''Monaural and binaural spectral cues created by the external ears of the pallid bat,'' Hear. Res. **95**, 1–17.

Grinnell, A. D., and Schnitzler, H.-U. (**1977**). ''Directional sensitivity of echolocation in the horseshoe bat, *Rhinolophus ferrumequinum*. II. Behavioral directionality of hearing,'' J. Comp. Physiol. [A] **116**, 63–76.

Hartley, D. J., and Suthers, R. A. (**1989**). ''The sound emission pattern of the echolocating bat, *Eptesicus fuscus*,'' J. Acoust. Soc. Am. **85**, 1348–1351.

Jackson, D. E. (**1999**). *Classical Electrodynamics*, 3rd ed. (Wiley, New York).

Jen, P.-S., and Chen, D. (**1988**). ''Directionality of sound pressure transformation at the pinna of echolocating bats,'' Hear. Res. **34**, 101–118.

Lawrence, B. D., and Simmons, J. A. (**1982**). ''Echolocation in bats: The external ear and perception of the vertical positions of targets,'' Science **218**, 481–483.

Lorensen, W. E., and Cline, H. E. (**1987**). ''Marching cubes: A high resolution 3D surface construction algorithm,'' in Computer Graphics (Proceedings of SIGGRAPH '87), Vol. 21, pp. 163–169.

Mitra, S. K. (**2001**). *Digital Signal Processing: A Computer-based Approach*, McGraw-Hill Series in Electrical and Computer Engineering, 2nd ed. (McGraw-Hill/Irwin, Boston).

Müller, R., and Hallam, J. C. T. (**2004**). ''From bat pinnae to sonar antennae: Augmented obliquely truncated horns as a novel parametric shape model,'' in Proceedings of the Eighth International Conference on the Simulation of Adaptive Behavior, edited by S. Schaal, A. J. Ijspeert, A. Billard, and S. Vijayakumar (to appear).

Müller, R., and Kuc, R. (**2000**). ''Foliage echoes: A probe into the ecological acoustics of bat echolocation,'' J. Acoust. Soc. Am. **108**, 836–845.

Obrist, M., Fenton, M., Elger, J., and Schlegel, P. (**1993**). ''What ears do for bats: A comparative study of pinna sound pressure transformation in chiroptera,'' J. Exp. Biol. **180**, 119–152.

Pierce, A. D. (**1981**). *Acoustics* (McGraw-Hill, New York).

Ramahi, O. M. (**1997**). ''Near- and far-field calculations in fdtd simulations using Kirchhoff surface integral representation,'' IEEE Trans. Antennas Propag. **45**, 753–759.

Schroeder, W., Martin, K., and Lorensen, B. (**2003**). *The* VISUALIZATION TOOLKIT*—An Object-oriented Approach to 3D Graphics*, 3rd ed. (Kitware, Inc.).

Schroeder, W., Zarge, J., and Lorensen, W. (**1992**). ''Decimation of triangle meshes,'' in ''Computer Graphics,'' Vol. 26, pp. 65–70.

Urick, R. J. (**1983**). *Principles of Underwater Sound*, 3rd ed. (McGraw-Hill, New York).

Valentine, D. E., Sinha, S. R., and Moss, C. F. (**2002**). ''Orienting responses and vocalizations produced by microstimulation in the superior colliculus of the echolocating bat, *Eptesicus fuscus*,'' J. Comp. Physiol., A **188**, 89–108.

Weisstein, E. W. (**2002**). *CRC Concise Encyclopedia of Mathematics*, 2nd ed. (CRC Press, Boca Raton).

Wotton, J. M., and Jenison, R. L. (**1997**). ''A backpropagation network model of the monaural localization information available in the bat echolocation system,'' J. Acoust. Soc. Am. **101**, 2964–2972.

Wotton, J. M., Haresign, T., and Simmons, J. A. (**1995**). ''Spatially dependent acoustic cues generated by the external ear of the big brown bat, *Eptesicus fuscus*,'' J. Acoust. Soc. Am. **98**, 1423–1445.

Distortion product otoacoustic emissions provide clues to hearing mechanisms in the frog ear[a)]

Pantelis N. Vassilakis[b)] and Sebastiaan W. F. Meenderink[c)]
Department of Physiological Science, University of California at Los Angeles, Los Angeles, California 90095-1606

Peter M. Narins
Departments of Physiological Science and Organismic Biology, Ecology, and Evolution, University of California at Los Angeles, 621 Charles E. Young Drive, South, Los Angeles, California 90095-1606

(Received 5 March 2004; revised 11 August 2004; accepted 12 September 2004)

$2\,f_1$-f_2 and $2\,f_2$-f_1 distortion product otoacoustic emissions (DPOAEs) were recorded from both ears of male and female *Rana pipiens pipiens* and *Rana catesbeiana*. The input-output (I/O) curves obtained from the amphibian papilla (AP) of both frog species are analogous to I/O curves recorded from mammals suggesting that, similarly to the mammalian cochlea, there may be an amplification process present in the frog AP. DPOAE level dependence on L_1-L_2 is different from that in mammals and consistent with intermodulation distortion expectations. Therefore, if a mechanical structure in the frog inner ear is functioning analogously to the mammalian basilar membrane, it must be more broadly tuned. DPOAE audiograms were obtained for primary frequencies spanning the animals' hearing range and selected stimulus levels. The results confirm that DPOAEs are produced in both papillae, with *R. catesbeiana* producing stronger emissions than *R. p. pipiens*. Consistent with previously reported sexual dimorphism in the mammalian and anuran auditory systems, females of both species produce stronger emissions than males. Moreover, it appears that $2\,f_1$-f_2 in the frog is generated primarily at the DPOAE frequency place, while $2\,f_2$-f_1 is generated primarily at a frequency place around the primaries. Regardless of generation place, both emissions within the AP may be subject to the same filtering mechanism, possibly the tectorial membrane. [DOI: 10.1121/1.1811571]

PACS numbers: 43.80.Lb, 43.64.Jb [JAS] Pages: 3713–3726

I. INTRODUCTION

A. Background—definitions

Distortion product otoacoustic emissions (DPOAEs) arise when the ear is stimulated acoustically by two sinusoidal signals (primaries) with appropriately chosen frequencies (f_1, f_2) and stimulus levels (L_1, L_2) (e.g., Kemp, 1979; Probst *et al.*, 1991). In human subjects, otoacoustic emissions (OAEs) in general are often used for clinical diagnostic screening of frequency-dependent cochlear function. DPOAEs may be used for early screening of outer hair cell (OHC) damage (e.g., Lonsbury-Martin *et al.*, 1993; Stover *et al.*, 1996; Wagner and Plinkert, 1999), for monitoring inner ear function after exposure to ototoxic drugs and/or noise (e.g., Brown *et al.*, 1989; Emmerich *et al.*, 2000), for diagnosis of middle ear damage (e.g., Owens *et al.*, 1992, 1993; Zhang and Abbas, 1997), and may also be used for assessing OHC maturation in premature babies (e.g., Abdala, 2000). In addition, OAEs have provided noninvasive means of examining inner ear mechanisms in mammals (e.g., Maison *et al.*, 1997; Shera and Guinan, 2003) and other vertebrates (e.g., Rosowski *et al.*, 1984), as well as in insects (e.g., Kössl and Boyan, 1998). The present study uses DPOAE recordings from two frog species to examine hearing mechanisms in the frog ear.

The vast majority of physiological and psychophysical studies involving DPOAEs concentrate on the $2\,f_1$-f_2 distortion product, principally because it usually represents the strongest DPOAE and therefore the easiest one to measure. OAEs, including DPOAEs, originate in the inner ear and, in most mammals, have been linked to outer hair cell (OHC) motility and receptor potential (Probst *et al.*, 1991; Robles and Ruggero, 2001). More specifically, the shape of mammalian DPOAE input/output (I/O) curves (i.e., compressive growth at low stimulus levels, although not necessarily as compressive as in the case of basilar membrane and hair cell I/O curves) provides further evidence that may link DPOAE generation to the cochlear amplifier, a process hypothesized to account for the observed increased sensitivity of the mammalian ear at low stimulus levels and for the mammalian ear's compressive and saturating response as stimulus levels increase from low to intermediate. Zheng *et al.* (2000) showed that prestin is the motor protein related to OHC motility, while Liberman *et al.* (2002) demonstrated a direct coupling between electrically stimulated OHC motility and cochlear amplification. In general, DPOAE generation in mammals has also been associated with the overlap of the

[a)]Portions of this work were presented at the 146th meeting of the Acoustical Society of America [P. Vassilakis and P. M. Narins, J. Acoust. Soc. Am. **114**, 2414 (2003)].

[b)]Present address: School of Music, De Paul University, 804 West Belden Avenue, Chicago, Illinois 60614-3296. Electronic mail: pantelis@acousticslab.com

[c)]Present address: Department of Otorhinolaryngology and Head and Neck Surgery, University Hospital Maastricht, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands.

primary tones' basilar membrane (BM) disturbances.

DPOAEs have been recorded from the ears of all terrestrial vertebrate classes, including most frog species tested.[1] The presence of OAEs in general and DPOAEs in particular in the frog raises numerous questions with regards to emission origin, especially due to the fact that the frog inner ear has neither a BM (Capranica, 1976) nor morphologically distinct outer and inner hair cells (Lewis, 1977; Lewis *et al.*, 1982). Based on observations that direct current injection may influence both the frequencies and amplitudes of OAEs (Wit *et al.*, 1989), Probst *et al.* (1991) suggested that frog OAEs may be related to the electrical tuning of hair cells. Others (e.g., Van Dijk *et al.*, 2002) have suggested that OAEs in frogs may be linked to spontaneous hair bundle movement, which may be analogous to the previously mentioned OHC motility in mammals and has been shown to occur in the bullfrog sacculus (Martin and Hudspeth, 1999). Hair bundle movement has been observed *in vitro* on several nonmammalian vestibular organs (review in Hudspeth, 1997) and the hearing organ of turtles (Crawford and Fettiplace, 1985), and *in vivo* on the bobtail lizard (Manley *et al.*, 2001). Recently, hair bundle motility was also observed in rats (Kennedy *et al.*, 2003).

The frog ear responds to airborne sound via two separate hearing organs, the amphibian papilla (AP) and the basilar papilla (BP). The AP nerve fibers are tuned to low and middle frequencies while the BP fibers are tuned to high frequencies (Feng *et al.*, 1975). More specifically, in *R. p. pipiens* the AP responds best to frequencies approximately within the range 100–1250 Hz, while the BP responds best to frequencies approximately within the range 1250–2400 Hz, with a ''break'' in the frequency response of the two papillae at approximately 1250 Hz (Ronken, 1991). In *R. catesbeiana*, the AP range is approximately 100–1100 Hz and the BP range is approximately 1100–1700 Hz, with the frequency response border at approximately 1100 Hz (Feng *et al.*, 1975; Lewis *et al.*, 1982). The present study reveals similarities in the behavior of DPOAEs originating in the frog AP and the mammalian cochlea, supporting analogies between these two organs. Several such analogies and distinctions have already been pointed out (Lewis and Narins, 1999), guiding the present study.

B. Exploration of the stimulus parameter space

The results of parametric studies (e.g., Whitehead *et al.*, 1995a, b; Knight and Kemp, 2000; Kummer *et al.*, 2000; Mills, 2002; Schneider *et al.*, 2003) have revealed a complex interrelationship between DPOAE levels and the four-dimensional primary-tone parameter space and have guided mammalian models of DPOAE generation (e.g., Shera and Guinan, 1999; Talmadge *et al.*, 1998, 1999; Tubis *et al.*, 2000; Knight and Kemp, 2000; Fahey *et al.*, 2000; Lukashkin *et al.*, 2002; Mills, 2002). The precise relationship may not be the same for all mammals, but it exhibits several distinguishing trends. As an example, a generally accepted protocol leading to optimal DPOAE generation within the frequency ranges of interest in clinical DPOAE measurements includes the following

(a) $1.2 \leqslant f_2/f_1 \leqslant 1.25$.[2] Shera and Guinan (1999), Faulstich and Kössl (2000), and others have argued that this value must be at least $f_2/f_1 = 1.15$, to avoid phase complications. In human subjects, $f_2/f_1 = 1.22$ results in the strongest 2 f_1-f_2 emissions (Probst *et al.*, 1991).

(b) $L_1 \leqslant 65$ dB SPL. For most mammals, DPOAE I/O curves show saturation at stimulus levels around 60–70 dB SPL (Probst *et al.*, 1991), followed by a DPOAE level decrease.

(c) $5 \text{ dB} \leqslant L_1 - L_2 \leqslant 10$ dB, for $f_2/f_1 \cong 1.2$ (e.g., Whitehead *et al.*, 1995a, b; Kummer *et al.*, 2000; Fahey *et al.*, 2000). As primary-tone levels become higher and/or frequency ratio becomes smaller, this level difference becomes increasingly unnecessary for the maximization of DPOAE levels.

In frog DPOAE studies, the selection of primary tone parameters has in general been less systematic. For example, Van Dijk *et al.* (2002) used $f_2/f_1 = 1.1$ and $L_1 = L_2 = 90$ dB SPL. Van Dijk and Manley (2001) used $1.05 \leqslant f_2/f_1 \leqslant 1.5$ and $37 \leqslant L_1 = L_2 \leqslant 85$ dB SPL, but only reported the relationship between DPOAE audiograms and f_2/f_1. In the absence of both a full parametric study and a theoretical model of DPOAE generation customized to the amphibian ear, parameter choices have been based on mammalian theoretical grounds, on DPOAE data from studies on other nonmammals, or on a trial-and-error basis.[3]

The present work provides a more systematic exploration of the primary-tone parameter space in frogs. DPOAE amplitudes were recorded from two ranid frog species in a series of experiments that examine the dependence of DPOAE amplitude on the absolute frequencies and on the absolute and relative levels of the primaries. In addition, we examined the effects of frog species and sex on DPOAE levels, controlling for possible effects of the degree of anesthesia.

II. MATERIALS AND METHODS

A. Stimulus parameter space

The dependence of frog DPOAE levels on primary-tone frequency ratio (explored extensively in mammals, e.g., Knight and Kemp, 2000; Schneider *et al.*, 2003) was not addressed in the present study and the primary-tone frequency ratio was set at $f_2/f_1 = 1.15$ for all experiments. In the absence to date of a parametric study examining the relationship between DPOAEs and f_2/f_1 in the frog, this value was chosen based on the previously mentioned recommendation ($f_2/f_1 \geqslant 1.15$) and to facilitate comparison with previous frog research (Van Dijk *et al.*, 2002: $f_2/f_1 = 1.1$).[4] More specifically, the study addresses the following primary-tone parameters:

(a) *Frequency range*: For the DPOAE audiograms, the primary-tone frequencies were selected based on the tuning characteristics of the *R. p. pipiens* and *R. catesbeiana* ears, spanning the frequency range of each papilla (AP and BP) and at frequency steps approximately equal to or less than 10% of f_1.

(b) *Primary-tone absolute levels*: DPOAE I/O curves were obtained for equal-level primaries. Highest DPOAE levels and lowest DPOAE thresholds have in general been measured for primaries within the frequency ranges associated with the most sensitive hearing regions of the species examined (Probst *et al.*, 1991). Additionally, there is a close relationship between auditory threshold and EOAE levels (Wagner and Plinkert, 1999). Therefore, for the DPOAE I/O curves, f_1 was selected so that all frequencies of interest (f_1, f_2, $2f_1$-f_2, $2f_2$-f_1) fell near the center of each papilla's frequency range.

(c) *Primary-tone relative levels* DPOAE audiograms were also obtained for unequal levels of the primaries ($0 \leqslant |L_1\text{-}L_2| \leqslant 10$ dB) from *R. p. pipiens* at $300 \leqslant f_1 \leqslant 1600$. The frequency range examined was selected based on the tuning characteristics of the *R. p. pipiens'* AP.

B. Subjects and procedure

DPOAEs were recorded from both ears of

(a) Ten northern leopard frogs (*R. p. pipiens*), five males (body mass 25.04–54.31 g, snout-vent length 6.08–8.08 cm, cranial width 2.09–3.05 cm, tympanic membrane diameter 0.41–0.53 cm) and five females (body mass 26.59–56.02 g, snout-vent length 6.94–7.61 cm, cranial width 2.32–2.57 cm, tympanic membrane diameter 0.41–0.52 cm) and

(b) Ten bullfrogs (*R. catesbeiana*), five males (body mass 267–370 g, snout-vent length 15.29–17.41 cm, cranial width 5.24–5.95 cm, tympanic membrane diameter 1.63–2.16 cm) and five females (body mass 300–435 g, snout-vent length 14.76–17.55 cm, cranial width 4.82–6.02 cm, tympanic membrane diameter 1.21–1.48 cm).

DPOAE amplitudes are thought to be independent of stimulation side (i.e., left versus right ears) and hearing loss in one ear of a subject does not affect emission levels in the other (Probst *et al.*, 1991; Kastanioudakis *et al.*, 2003). Therefore, each ear tested in our study was considered a separate data point.

The same subjects participated in all experiments, performed in a single session per subject, lasting 4–9 h. Animals were anaesthetized with an intramuscular injection of pentobarbital sodium solution (Nembutal, Abbott Laboratories, 50 mg/ml: ~0.9–1.0 μl/g body mass) and were covered with wet gauze that was regularly moistened with water to facilitate cutaneous respiration. Measurements began between approximately 1.5 and 3 h after Nembutal injection, with the animals deeply anesthetized (no toe-pinch or eye reflex and no acoustic artifacts—i.e., breathing/swallowing noise—in the recorded signal). Some of the DPOAE measurements on *R. p. pipiens* were repeated 5–6 h later under light anesthesia (all reflexes had returned and there were acoustic artifacts present in the microphone signal) to control for possible effects of the degree of anesthesia on emission levels.[5]

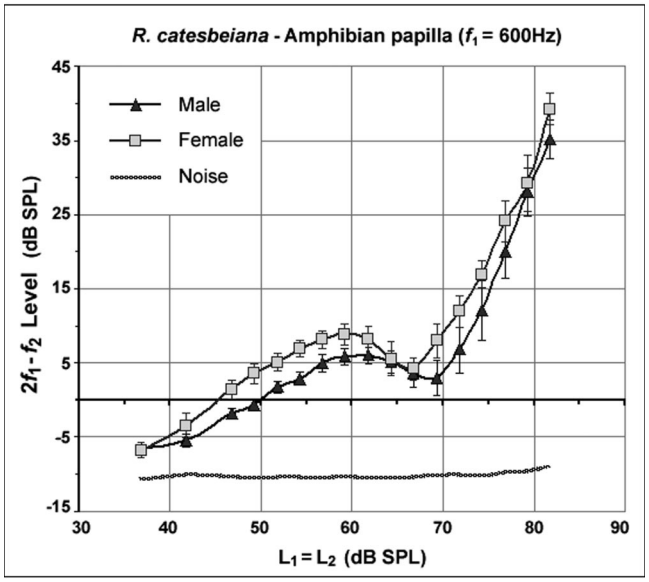


FIG. 1. DPOAE I/O curves ($2f_1$-f_2) from the AP ($f_1 = 600$ Hz) of male (-▲-) and female (-□-) *R. catesbeiana*. Average DPOAE levels and standard errors (ten ears each sex: both ears of five males and five females). Although there are level differences between male and female DPOAEs, the overall shape of the I/O curves is similar regardless of subjects' sex.

C. Equipment and signal analysis parameters

Measurements were performed in a single-wall sound-attenuating chamber with the subjects placed on a vibration isolation table (Newport VH Isostation). Emissions were recorded with a probe assembly (ER-10C, Etymotic Research, Elkgrove, IL) that includes a sensitive microphone connected to a preamplifier (amplification: 20 dB) and two miniature speakers. The ER-10C probe is designed and calibrated for use with human subjects. Since the frog species examined have no ear canals, the probe ear-tip was placed inside a plastic tube, which was coupled to the frog ear by placing its open end adjacent to the skin around the tympanic membrane. The small gap between the plastic tube and the frog's head was sealed using high vacuum grease, a silicone lubricant. The integrity of the acoustic seal was tested using a sweep tone (WG1 waveform generator, Tucker-Davis Technologies, Alachua, FL) to ensure that sound level loss at 250 Hz due to diffraction was $\leqslant 5$ dB, relative to the level measured at 1000 Hz. Because of the large differences in tympanic membrane diameter between *R. p. pipiens* and *R. catesbeiana* and between male and female *R. catesbeiana*, three different ''ear canal'' tubes were used: (1) *R. p. pipiens*: diameter$\cong$0.95 cm, length$\cong$1.20 cm, volume$\cong$0.68 cm^3; (2) male *R. catesbeiana*: diameter$\cong$2.20 cm, length $\cong$0.55 cm, volume$\cong$2.10 cm^3, and (3) female *R. catesbeiana*: diameter$\cong$1.25 cm, length$\cong$1.40 cm, volume $\cong$1.72 cm^3.

Primary frequencies (f_1) were selected within the range 240–3000 Hz[6] and stimuli were generated using a Real Time Processor (TDT-RP2), controlled by software written in Matlab (version 12, MathWorks, Inc., Natick, MA) and RPvds (version 5.0, TDT, Alachua, FL). The stimuli were fed to the two miniature speakers on the probe via two attenuators (TDT-PA4), used for manual control of the signal levels. The

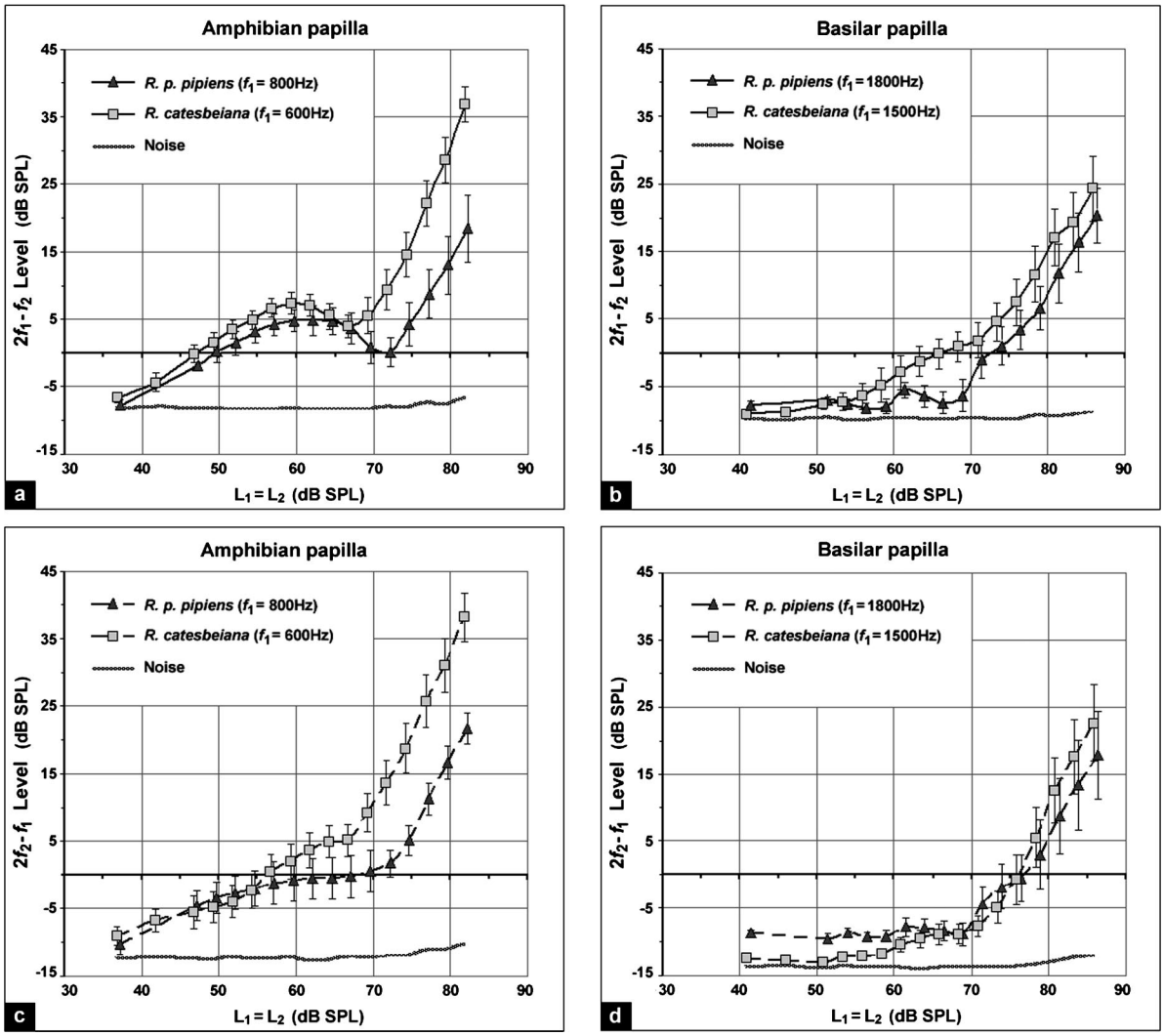


FIG. 2. DPOAE I/O curves for $f_2/f_1 = 1.15$ from *R. p. pipiens* (-▲-, $f_1 = 800$ Hz, 1800 Hz) and *R. catesbeiana* (-□-, $f_1 = 600$ Hz, 1500 Hz). Average DPOAE levels and standard errors (20 ears each species: both ears of five males and five females). (a) and (b) 2 f_1-f_2; (c) and (d) 2 f_2-f_1; (a) and (c) AP; (b) and (d) BP.

probe's microphone signal was fed to a frequency analyzer (SRS-SR770 FFT Network Analyzer, Stanford Research Systems, Sunnyvale, CA).[7] The bandwidth of the Fourier analysis changed as a function of f_1.[8] For each stimulus tone-pair, the DPOAE level reported was the maximum amplitude of five adjacent analysis bands centered at the band containing the DPOAE frequency. The noise level was calculated by averaging the levels of 18 analysis bands surrounding the five bands used to determine the DPOAE level. These DPOAE- and noise-level calculation methods minimized the influence of spectral leakage and gave noise levels comparable to system distortion ($\sim$83 dB below the primaries for $L_1 = L_2 \geqslant 70$ dB SPL and ~ -12 dB SPL for $L_1 = L_2 < 70$ dB SPL), but may have overestimated DPOAE levels that were near the calculated noise floor. Therefore, any emission levels in the results that are <3 dB above the estimated noise-floor may be considered indistinguishable from noise. All levels reported have been corrected to compensate for the frequency response of the ER-10C microphone.

III. RESULTS

A. Experiment 1: DPOAE I/O curves

DPOAE I/O curves at 2 f_1-f_2 and 2 f_2-f_1 were obtained for $f_2/f_1 = 1.15$ and $35 \leqslant L_1 = L_2$ (± 1 dB) $\leqslant 85$ dB SPL (in 2.5-dB steps) with $f_1 = 800$ Hz (AP) and 1800 Hz (BP) for *R. p. pipiens* and $f_1 = 600$ Hz (AP) and 1500 Hz (BP) for *R. catesbeiana*. Although the absolute DPOAE amplitudes were different for male and female subjects, the overall shapes of DPOAE I/O curves within each species were similar between the two sexes (Fig. 1), so data obtained from male and female animals were averaged. Mean levels and standard errors for each distortion product from each papilla and for both species are displayed in Fig. 2.

In Fig. 2(a), the 2 f_1-f_2 DPOAE I/O curves obtained from the AP of both *R. p. pipiens* and *R. catesbeiana* display non-monotonic growth. For low primary levels (<60 dB SPL), the DPOAE level growth rate is <1 dB/dB (compressive), saturating at primary levels ≈ 60 dB SPL, turning negative, and reaching a notch at primary levels ≈ 70 dB SPL (≈ 67 dB SPL for *R. catesbeiana*; ≈ 72 dB SPL for *R. p.*

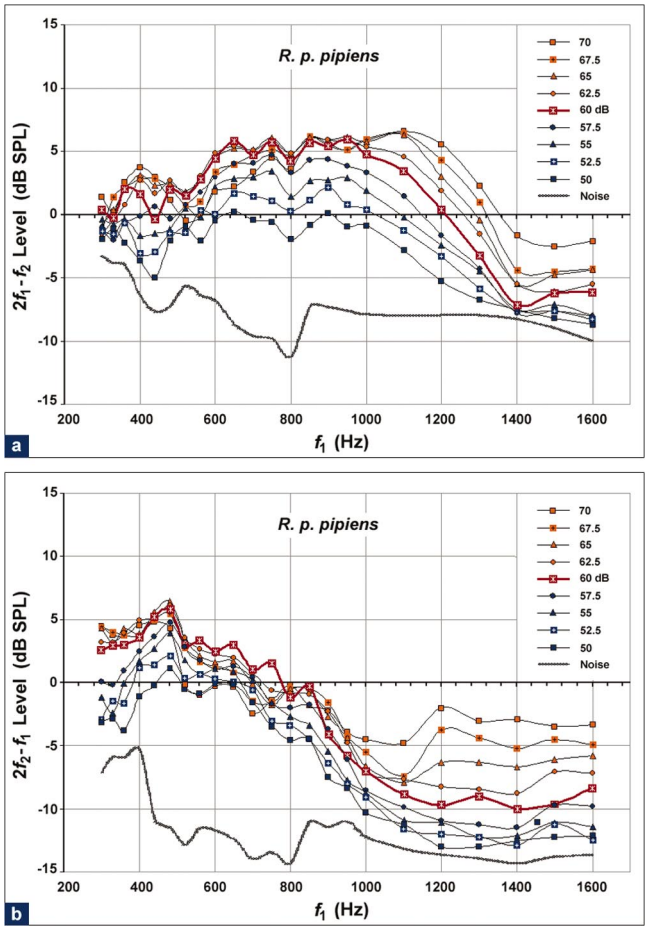


FIG. 3. DPOAE audiograms from *R. p. pipiens* (average DPOAE levels from 20 ears: both ears of five males and five females) for $f_2/f_1=1.15$, $300 \leqslant f_1 \leqslant 1600$ Hz, and $50 \leqslant L_1=L_2 \leqslant 70$ dB SPL. Only the AP may be exhibiting compressive DPOAE level growth. (a) $2f_1$-f_2; (b) $2f_2$-f_1.

pipiens). At higher primary levels, DPOAEs grow linearly at a rate >1 dB/dB ($\approx 1.3-2.2$ dB/dB). The $2f_2$-f_1 DPOAE I/O curves obtained from the AP of both frog species [Fig. 2(c)] show similar trends but with no pronounced notches. For primary frequencies in the BP frequency range [Figs. 2(b) and (d)], DPOAE levels are much lower than those recorded from the AP, with the DPOAE I/O curves for primary levels <70 dB being very close to the noise-floor. For primary levels >70 dB, the slopes of the DPOAE I/O curves from the BP are similar to those from the AP. Also note that, in the AP I/O curves for *R. catesbeiana*, the notch (or deviation from linear growth) is observed at approximately the same primary tone levels for both distortion products [Figs. 2(a) and 2(c); $L_1=L_2 \cong 67$ dB SPL]. This does not seem to be the case in the AP I/O curves for *R. p. pipiens*. Here, the notch in the $2f_1$-f_2 I/O curve occurs at $L_1=L_2 \cong 72$ dB SPL, while the deviation from linear growth in the $2f_2$-f_1 I/O curve is observed at $L_1=L_2 \cong 67$ dB SPL.

A portion of the same results (only *R. p. pipiens* and $50 \leqslant L_1=L_2$ (± 1 dB) $\leqslant 70$ dB SPL) is displayed in Fig. 3 in the form of DP audiograms, for $300 \leqslant f_1 \leqslant 1600$ Hz. All individual DP-audiograms in Fig. 3 show fine-structure, i.e., a periodic variation in DPOAE amplitude as a function of frequency. The frequencies for which respective maxima and minima in DPOAE amplitude were found did not change across different DP-audiograms, indicating that the fine-structure is independent of stimulus level. Similarly to Figs. 2(a) and (b), Fig. 3 supports the observation that, for primary frequencies in the AP and stimulus levels <70 dB there is a compressive growth of DPOAE levels as primary levels increase. This does not appear to be the case for primary frequencies in the BP. More specifically, for both DPOAE components [$2f_1$-f_2: Fig. 3(a); $2f_2$-f_1: Fig. 3(b)], the L_1, $L_2=60$ dB plot is close to the top of the graph for frequencies in the AP ($\sim <900$ Hz) but in between the L_1, $L_2=57.5$ dB and L_1, $L_2=62.5$ dB plots for frequencies in the BP (~ 1000 Hz).

B. Experiment 2: DPOAE level dependence on relative levels of primaries

The relationship between DPOAE level and L_1-L_2 was examined in *R. p. pipiens* (both ears of five males and five females) for $50 \leqslant L_1 \leqslant 60$ dB SPL ($L_2=60$ dB SPL) and $50 \leqslant L_2 \leqslant 60$ dB SPL ($L_1=60$ dB SPL), with $0 \leqslant |L_1-L_2| \leqslant 10$ dB (2.5-dB steps), $300 \leqslant f_1 \leqslant 1600$ Hz, and $f_2/f_1=1.15$. Figure 4 displays the mean DPOAE levels for each L_1, L_2 combination at $f_1=800$ and 1500 Hz. These frequencies were selected to represent DPOAEs recorded from each papilla. Similar responses were observed at other frequencies (Fig. 5). Again, although the absolute DPOAE amplitudes were different for male and female subjects, there were no significant sex differences in the shape/slopes of the curves describing the relationship between L_1-L_2 and DPOAE levels. Therefore, data obtained from male and female animals were averaged.

Independent of species, sex, and DPOAE ($2f_1$-f_2 or $2f_2$-f_1), strongest emissions were recorded for $|L_1-L_2|=0$ or 2.5 dB. As $|L_1-L_2|$ increased, DPOAE amplitudes gradually dropped. Figure 5 displays the same results in the form of DPOAE audiograms, showing maximum DPOAE levels for $L_1-L_2 \cong 0$ for most frequencies tested and both $2f_1$-f_2 and $2f_2$-f_1.

C. Experiment 3: DPOAE audiograms

1. AP versus BP; $2f_1$-f_2 and $2f_2$-f_1

DPOAE audiograms were obtained from both ears of all subjects for $240 \leqslant f_1 \leqslant 3000$ Hz at $\sim 0.1f_1$ steps, with $L_1=L_2=60$ dB SPL, and $f_2/f_1=1.15$. Figure 6 (*R. p. pipiens*) and Fig. 7 (*R. catesbeiana*) display the levels of $2f_1$-f_2 and $2f_2$-f_1 DPOAEs as a function of (a) stimulus frequency f_1, (b) stimulus frequency f_2, and (c) distortion product (DP) frequency. Each plot includes a dashed vertical line that indicates the approximate frequency value separating the frequency ranges of the two papillae. For both species, the frequency separation between the two papillae is based on neural tuning curve characteristics (*R. p. pipiens*: 1250 Hz, Ronken, 1991; *R. catesbaiena*: 1100 Hz, Feng *et al.*, 1975).[9] DPOAE data from both sexes were averaged. The error bars in Figure 9 and 10 are an indication of the variability in the male and female date.

In both figures, two distinct frequency regions can be identified where relative maximum emission amplitudes are found. These frequency regions correspond roughly to the

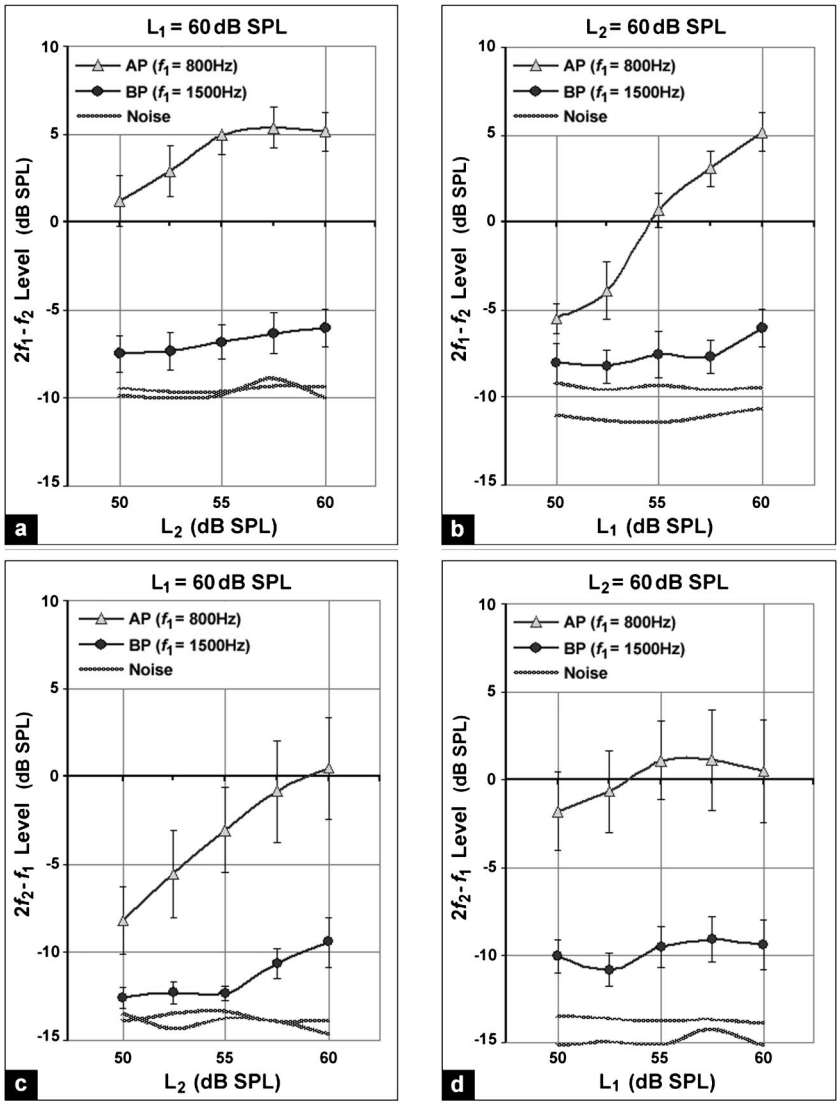


FIG. 4. Average DPOAE levels and standard errors versus L_2, for $L_1 = 60$ dB [(a) and (c)], and versus L_1, for $L_2 = 60$ dB [(b) and (d)] from *R. p. pipiens* (20 ears: both ears of five males and five females). Strongest emissions were generally recorded for $L_1 \approx L_2$. As expected, based on intermodulation distortion properties, the level of each distortion product was affected most by changes in the level of its neighboring (in frequency) primary [(b) and (c)].

frequency tuning of the AP (lower frequency region) and the BP (higher frequency region) of each species, respectively. For the primary level tested (60 dB SPL), the emission levels recorded from the AP were higher than those recorded from the BP.

Figure 6 indicates that, for *R. p. pipiens*, alignment of the $2f_1$-f_2 DPOAE audiogram dip and the break in the frequency coverage between the two papillae (1250 Hz; Ronken, 1991) is best when $2f_1$-f_2 amplitude is plotted as a function of distortion product frequency [Fig. 6(c)]. For the $2f_2$-f_1 DPOAE, on the other hand, closest alignment occurs when $2f_2$-f_1 amplitude is plotted as a function of f_2 [Fig. 6(b)]. Near alignment is also observed when this distortion product is plotted as a function of f_1 [Fig. 6(a)]. Similar results were obtained for *R. catesbeiana* (Fig. 7). All values reported are averages.

To control for possible effects of the degree of anesthesia on DPOAE amplitudes, the experiment was repeated 5–6 h later as the anesthesia was wearing off (eye and toe-pinch reflexes accompanied by acoustic artifacts—i.e., breathing/swallowing noise—in the recorded signal). ''Light'' anesthesia measurements were made on all *R. p. pipiens* for $300 \leqslant f_1 \leqslant 1600$ Hz, $L_1 = L_2 = 60$ dB SPL, and $f_2/f_1 = 1.15$. The mean results are displayed in Fig. 8 and indicate that, for the anesthetic agent used and the frequencies and levels tested, degree of anesthesia did not affect DPOAE amplitudes.

2. Female versus male; R. p. pipiens and R. catesbeiana

Separate analyses of DPOAE data from male and female *R. p. pipiens* and *R. catesbeiana* revealed clear sex differences in DPOAE levels. Figures 9 and 10 compare male and female DPOAE audiograms, demonstrating that, in both species, over most of the frequency range tested, and for the levels of primary tones used, female subjects produce higher emission levels than male subjects. This is more pronounced in the BP.

With the exception of emissions from the high-frequency end of the BP, *R. catesbeiana* demonstrate larger emission level differences between sexes and, as also indicated in Fig. 2, larger emission levels in general.

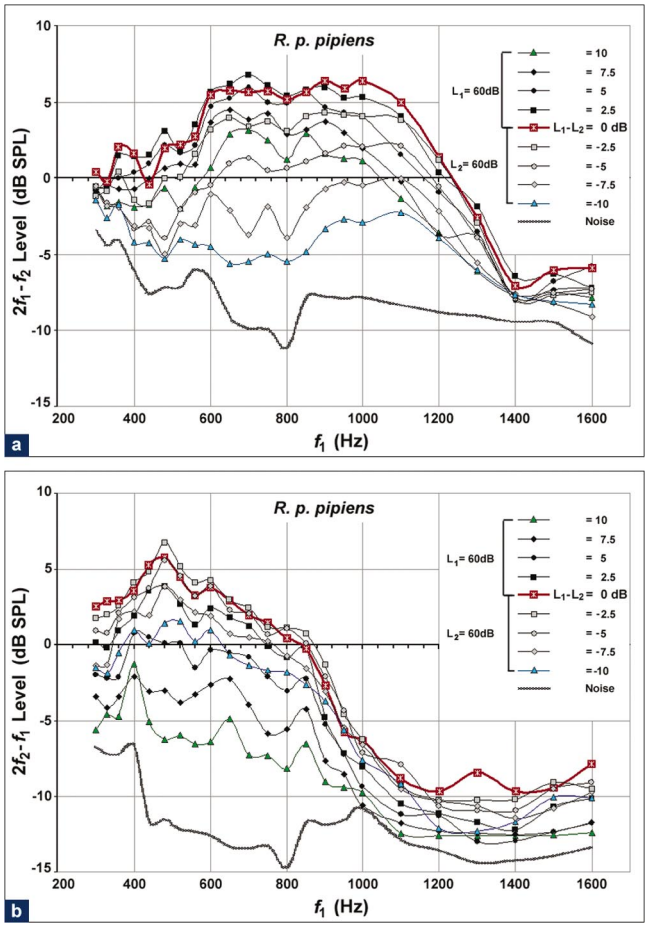


FIG. 5. DPOAE audiograms from *R. p. pipiens* (average DPOAE levels from 20 ears: both ears of five males and five females) for $0 \leqslant |L_1\text{-}L_2| \leqslant 10$ dB, $300 \leqslant f_1 \leqslant 1600$, and $f_2/f_1 = 1.15$. At most frequencies tested, highest DPOAE levels were recorded for $L_1\text{-}L_2 \approx 0$. (a) $2f_1\text{-}f_2$; (b) $2f_2\text{-}f_1$.

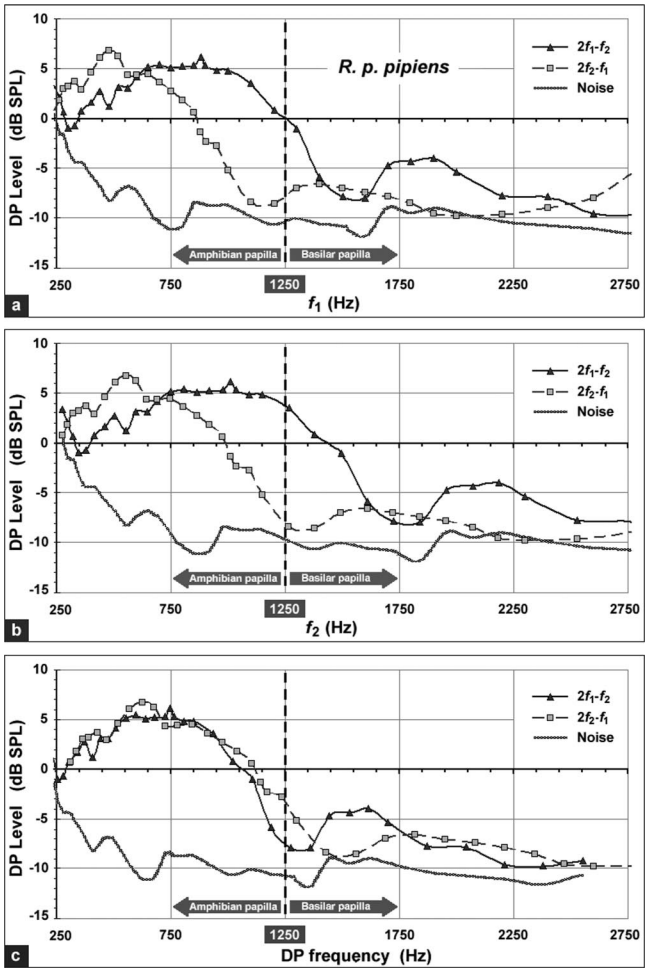


FIG. 6. Average levels of the $2f_1\text{-}f_2$ (-▲-) and $2f_2\text{-}f_1$ (-□-) DPOAEs from *R. p. pipiens* (20 ears: both ears of five males and five females) for $f_2/f_1 = 1.15$, $L_1 = L_2 = 60$ dB SPL (± 2.5 dB), and $240 \leqslant f_1 \leqslant 3000$ Hz as a function of (a) f_1, (b) f_2, and (c) DP frequency. The dashed vertical line marks the "break" in the frequency coverage of the AP and the BP (1250 Hz). For $2f_1\text{-}f_2$, the dip in the plot aligns best with the vertical line when DPOAE level is plotted as a function of DP frequency (c). For $2f_2\text{-}f_1$, the alignment is best when DPOAE level is plotted as a function of f_2 (b), although there is also near alignment in the plot plotted against f_1.

IV. DISCUSSION

A. DPOAE I/O curves provide evidence of an AP amplifier

The $2f_1\text{-}f_2$ I/O curves obtained from the AP of both frog species tested [Figs. 1 and 2(a)] have a salient characteristic in common with the DPOAE I/O curves obtained from mammals: nonlinear, non-monotonic growth. The general shape of I/O curves recorded in mammals is described in a manner strikingly similar to the one used to describe the $2f_1\text{-}f_2$ DPOAE I/O AP curves obtained in our study [Fig. 2(a)] and, with the absence of a clear notch, to the one describing the $2f_2\text{-}f_1$ DPOAE I/O AP curves [Fig. 2(c)]. The slope of mammalian DPOAE I/O curves depends on the stimulus levels used. For primary levels <60 dB SPL slopes are $\leqslant 1$ dB/dB. For intermediate stimulus levels ($L_1 = L_2 \approx 60-70$ dB SPL) the growth saturates, sometimes even resulting in a decrease in DPOAE amplitude (i.e., a notch). For higher stimulus levels ($L_1 = L_2 > 70$ dB SPL) the slope increases again, often to a value >1 dB/dB (e.g., Probst *et al.*, 1991). In mammals, it has been shown that DPOAE levels in general (e.g., Knight and Kemp, 2000) and the exact shape of the DPOAE I/O curves in particular (e.g., Mills, 2002) depend on the distortion product under study, and on the relative levels ($L_1\text{-}L_2$) and frequencies (f_2/f_1) of the primaries. In our DPOAE I/O experiment, $L_1\text{-}L_2$ and f_2/f_1 were not varied. The AP data, however, do exhibit differences in I/O curve shapes between distortion products [Fig. 2(a) versus 2(c)], with only the I/O curve for the $2f_1\text{-}f_2$ DPOAE exhibiting a clear notch. Our results are in general agreement with the DPOAE I/O curve slopes measured by Meenderink and Van dijk (2004) on *R. p. pipiens*, although, in their results, both the low- and high-level I/O slopes obtained from the BP appear relatively steep.

Until recently, the main hypothesis (e.g., in Probst *et al.*, 1991) introduced to explain the DPOAE I/O curve shape (especially the observed notches) in mammals assumes that there is interference between 2 DPOAE components:[10]

(a) a level dependent, nonlinear component, which dominates the recorded DPOAEs when the stimulus levels are low but saturates for higher stimulus levels and
(b) a monotonic component that, as the stimulus levels in-

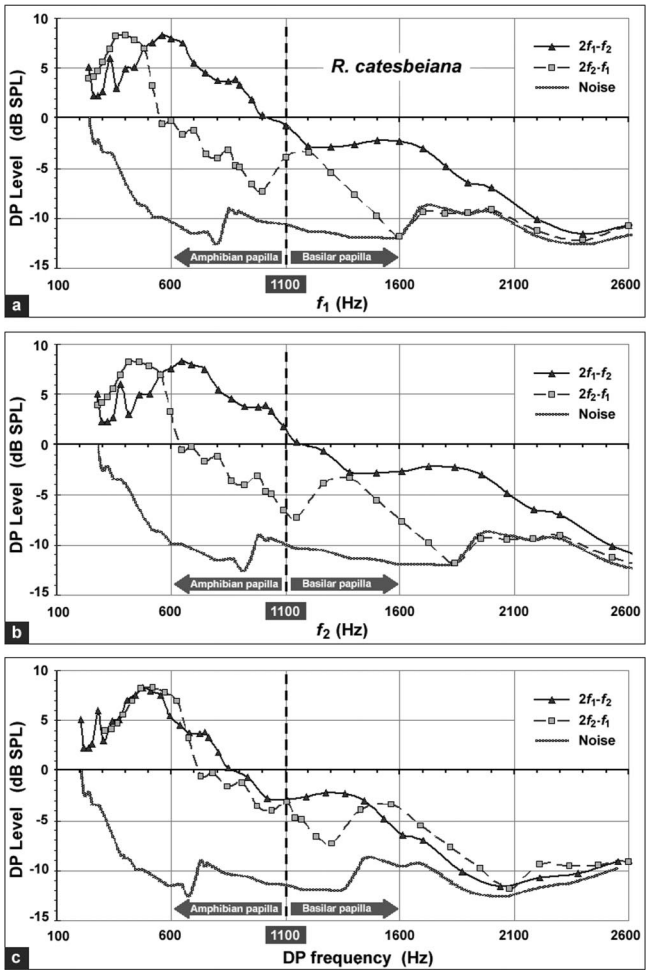


FIG. 7. Average levels of the 2 f_1-f_2 (-▲-) and 2 f_2-f_1 (-□-) DPOAEs from *R. catesbeiana* (20 ears: both ears of five males and five females) for $f_2/f_1=1.15$, $L_1=L_2=60$ dB SPL (± 2.5 dB), and $240 \leqslant f_1 \leqslant 3000$ Hz as a function of (a) f_1, (b) f_2, and (c) DP frequency. The alignment between the dip in the audiograms and the frequency coverage border between the AP and the BP follows the same pattern as in the *R. p. pipiens* data (Fig. 6).

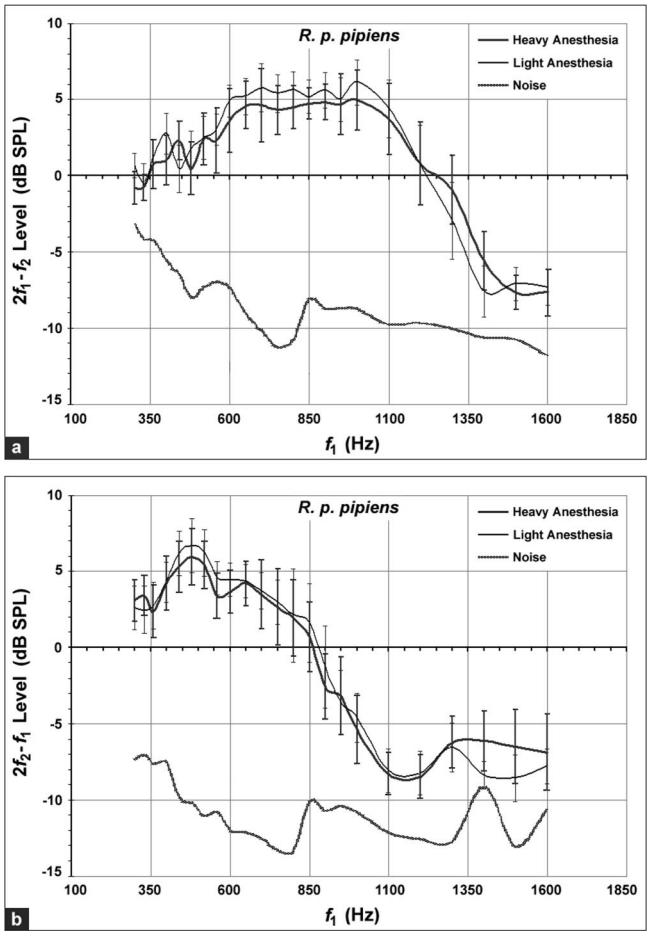


FIG. 8. DPOAE audiograms from *R. p. pipiens* (average DPOAE levels from 20 ears: both ears of five males and five females) for $f_2/f_1=1.15$, $L_1=L_2=60$ dB SPL, and $300 \leqslant f_1 \leqslant 1600$ Hz, recorded from the same animals while heavily and lightly anesthetized. No systematic effect of the degree of anesthesia on DPOAE levels was observed. (a) 2 f_1-f_2; (b) 2 f_2-f_1.

crease and the level-dependent component saturates, becomes dominant in the recorded DPOAEs.

The precise nature of these two components and their place of origin are subjects of ongoing research.

Lukashkin *et al.* (2002) proposed an alternative hypothesis with subtle but important differences to the previous one. Based on the observation that cochlear function impairment results in an upward level shift of the DPOAE I/O curve notch (as opposed to the downward level shift predicted by the earlier hypothesis) they concluded that DPOAE I/O curve shapes can be accounted for by a single source: a nonlinear amplifier with saturating I/O characteristics. A similar hypothesis was introduced by Mom *et al.* (2001) based on their observation that DPOAEs measured during complete cochlear ischemia are highly vulnerable to mild auditory fatigue induced prior to ischemia. Their results indicate that there is no ''passive'' component in the response of the mammalian cochlea and support the single-component model. Whitehead (1998) and clinical data reviewed in Martin (2002) also support such a model.[11]

The common thread in all explanations is the hypothesized presence (whether on its own or along with some other passive nonlinearity) of a saturating nonlinear inner ear amplifier (i.e., cochlear amplifier). The DPOAE I/O curves in Fig. 2(a) of the present study, along with the fact that SOAEs have been recorded from the frog AP, are consistent with the inner ear amplifier hypothesis, despite the fact that the frog ear lacks OHCs, believed to be at the center of the mammalian ear's amplification process.[12] Amplification and OAEs, in the AP of the frog, may be linked to spontaneous movement of the hair cell bundle, which has been shown to occur in the frog sacculus (Martin and Hudspeth, 1999), while there is indirect evidence that it may also occur in the mammalian ear (Kennedy *et al.*, 2003). In other words, hair cell bundle movement in the frog AP may be analogous to OHC and/or hair bundle motility in the mammalian ear.

For primary frequencies in the BP, DPOAE levels were much lower than those recorded from the AP. The portions of the DPOAE I/O curves that are distinguishable from the noise-floor appear monotonic, implying that, in contrast to the AP, an amplification mechanism may not be present in the frog BP. This potential difference is consistent with the absence of SOAEs in the frog BP (Van Dijk *et al.*, 1989) and the differential physiological vulnerability of frog DPOAEs

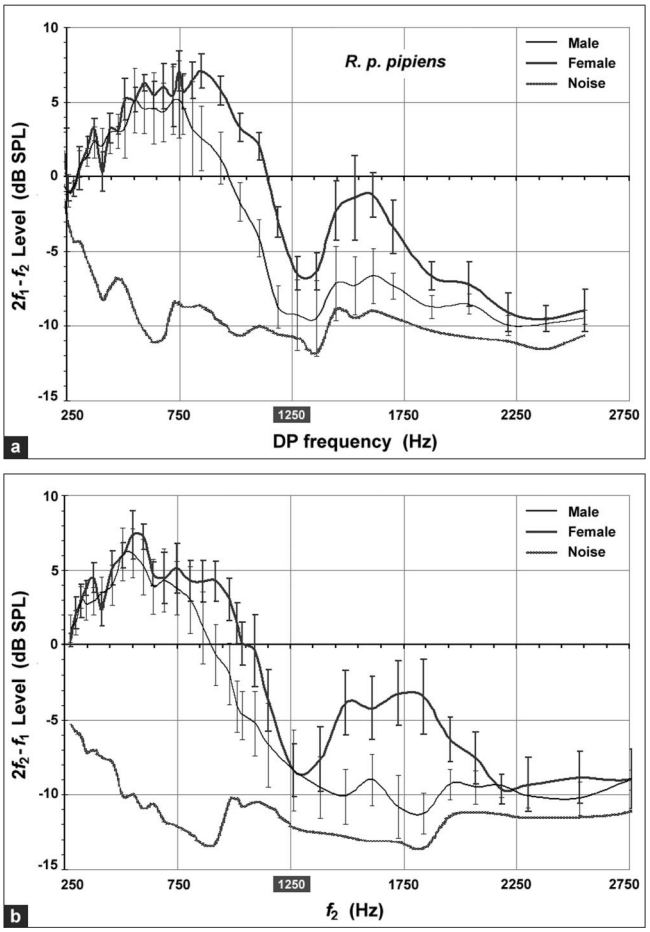


FIG. 9. Comparison between DPOAE audiograms from both ears of five male and five female *R. p. pipiens*. (a) The level of 2 f_1-f_2 is plotted as a function of DP frequency and (b) the level of 2 f_2-f_1 is plotted as a function of f_2 (see Fig. 6). Female subjects exhibited stronger emissions than males, especially from the BP.

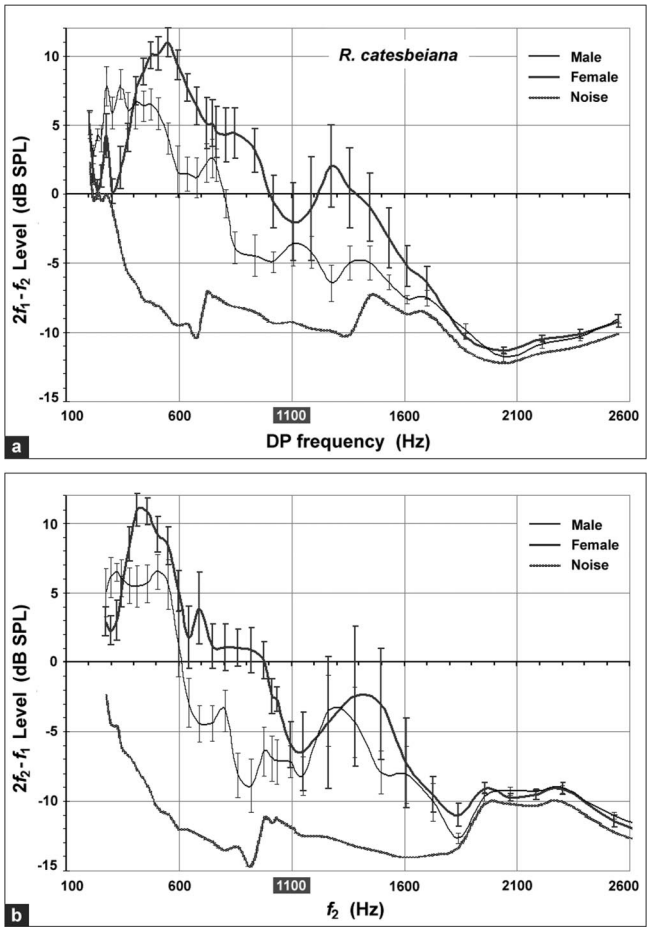


FIG. 10. Comparison between DPOAE audiograms from both ears of five male and five female *R. catesbeiana*. (a) The level of 2 f_1-f_2 is plotted as a function of DP frequency and (b) the level of 2 f_2-f_1 is plotted as a function of f_2 (see Fig. 7). Female subjects exhibited stronger emissions than males.

(Van Dijk *et al.*, 2003), and may echo several physiological differences between the amphibian and basilar papilla. In the bullfrog ear, for example, the AP has approximately 15 times as many hair cells and is innervated by approximately three times as many afferent nerve fibers as the BP (Geisler *et al.*, 1964; Lewis *et al.*, 1985). At the same time, efferent nerve fibers innervate the AP but not the BP (Robbins *et al.*, 1967). The possible relationship between these physiological differences and the observed difference in DPOAE I/O curves, assumed to reflect amplification-process differences, remains to be examined. As a general observation, the fact that only the AP is tonotopically organized (Lewis *et al.*, 1982; Simmons *et al.*, 1992) is consistent with a greater relevance of an amplification mechanism that would increase this papilla's tuning sharpness. In contrast, the BP responds as a band-pass filter with a single characteristic frequency (e.g., Van Dijk and Manley, 2001), rendering the development of an amplification and tuning-sharpening mechanism less essential.

In both *R. p. pipiens* and *R. catesbeiana*, the peak of the compressive-with-saturation portion of the DPOAE I/O curves occurs for $L_1 = L_2 \cong 60$ dB SPL. Since in mammals the amplified or ''active'' portion of the I/O curves is presumed to be dominant only at low stimulus levels, DPOAE audiograms (experiment 3) were obtained for L_1 and L_2 centered at ~60 dB SPL. The relative levels of the primaries (L_1-L_2) corresponding to the highest DPOAE levels were determined in a separate experiment, discussed below.

B. DPOAE amplitude dependence on the relative levels of primaries

In mammals, the amplitude of each distortion product is influenced most by level changes in its neighboring (in frequency) primary tone, consistent with intermodulation distortion expectations. However, and contrary to intermodulation distortion expectations, the highest DPOAE levels are obtained for unequal levels of the primaries (i.e., $L_1 > L_2$; Kummer *et al.*, 2000). This asymmetry in primary tone levels has been linked to the asymmetry of basilar membrane disturbance envelopes and is less critical at low frequencies, high primary stimulus levels, and low frequency ratios of the primaries. Our findings in the frog are more consistent with intermodulation distortion expectations alone, with highest emission levels recorded for equal levels of the primaries (Figs. 4 and 5).[13] The difference between mammalian and frog data regarding the effect of L_1-L_2 on DPOAE levels suggests that, if a filtering mechanism analogous to the mammalian BM is present in the frog (i.e., tectorial membrane), it must be more broadly tuned. This notion is supported by the

shape of frequency threshold curves (i.e., tuning curves) recorded from frog VIIIth nerve auditory fibers [e.g., *R. catesbeiana* in Feng *et al.* (1975); *Eleutherodactylus coqui* and *Bombina orientalis* in Narins and Hillery (1983); *R. p. pipiens* in Ronken (1991)]. In both mammals and frogs, tuning curves display an asymmetry, with relatively shallow low-frequency limbs and high-frequency abrupt cutoffs, which is stronger for high-frequency fibers. However, neural tuning curves recorded from mammals (e.g., Robles and Ruggero, 2001) exhibit narrower tips than those recorded from frogs. In mammals, the above noted asymmetry, along with several other properties of auditory nerve responses including the L_1, L_2 DPOAE asymmetry, is thought to reflect corresponding features of BM vibration (Robles and Ruggero, 2001). Our results (Figs. 4 and 5), along with the difference between mammalian and frog neural tuning curves, suggest that, although the AP and BP of the frog ear lack a BM, some other mechanical structure (possibly the tectorial membrane) may in some respects function analogously to the mammalian BM, while more broadly tuned. This observation is in agreement with phase data from auditory nerve fiber measurements in the frog (Hillery and Narins, 1987), pointing to the tectorial membrane–receptor interface as a possible candidate for the mechanical generation of intermodulation distortion products.[14] Further evidence supporting this suggestion is discussed below. No sex differences in the shape/slopes of the curves describing the relationship between L_1-L_2 and DPOAE levels were observed, suggesting that the mechanisms involved in DPOAE generation may be the same for both sexes.

C. DPOAE audiograms

1. DPOAEs can be evoked in both papillae

The correspondence between frequency ranges to which nerve fibers are tuned and the frequency ranges for which relative maximum DPOAE amplitudes are recorded indicates that DPOAEs are produced in both papillae present in the frog inner ear (Figs. 6 and 7). Our findings are in agreement with previous results on *R. p. pipiens* (Van Dijk *et al.*, 2002; Meenderink and Van Dijk, 2004). For the primary level tested in our DPOAE audiogram experiment (60 dB SPL), the emission levels recorded from the AP were higher than those recorded from the BP. Van Dijk *et al.* (2002) reported almost equal DPOAE levels from both papillae of *R. p. pipiens* for primary levels of 90 dB SPL. The difference in the results of the two studies probably reflects differences in the primary-tone levels used, an explanation supported by our study's DPOAE I/O data (Fig. 2). More specifically, Fig. 2 indicates that, for low primary levels, DPOAE amplitudes recorded from the AP are higher than those recorded from the BP while, for high primary levels, the DPOAE amplitudes recorded from the AP and the BP are comparable.

2. Different generation mechanisms for the $2f_1$-f_2 and $2f_2$-f_1 distortion products

In mammals, the level of the DPOAE at $2f_1$-f_2 is typically plotted as a function of f_2, reflecting very specific theoretical assumptions/hypotheses regarding the generation mechanism and place of origin of the $2f_1$-f_2 DPOAE. In studies on frog DPOAEs (Baker *et al.*, 1989; Van Dijk and Manley, 2001; Van Dijk *et al.*, 2002), DPOAE levels have been plotted as a function of f_1, a choice based on convention (following Köppl *et al.*, 1993). Due to the absence of a theoretical framework for DPOAE generation in the frog ear, we plotted DPOAE levels in three different ways.

The dashed vertical lines in Figs. 6 and 7 mark the frequency separation between the AP and BP for the species tested, as determined by neural tuning curves, and are therefore within a frequency range of low sensitivity. The assumption of lower sensitivity at the edges of a papilla's frequency response versus the center is supported by the results of several studies (e.g., Ehret and Capranica, 1980; Narins and Hillery, 1983; Zelick and Narins, 1985). Since lower DPOAE levels and higher DPOAE thresholds have in general been measured at frequency ranges associated with the least sensitive hearing regions of the species examined (e.g., Probst *et al.* 1991), the degree of alignment between the DP audiogram dips and the vertical lines in Figs. 6 and 7 provides useful information regarding the site of DPOAE generation. For example, if a plot in Figs. 6 and 7 has its dip and vertical line misaligned, the expectation that emission levels should be minimal at the frequency range of minimal response is not satisfied. From this observation it can be deduced that the generation site of the respective distortion product cannot be related to the generation site of the frequency component against which the distortion product levels have been plotted. Figure 6 indicates that, for *R. p. pipiens*, alignment of the $2f_1$-f_2 DPOAE audiogram dip and the break in the frequency coverage between the two papillae (1250 Hz; Ronken, 1991) is best when $2f_1$-f_2 amplitude is plotted as a function of DP frequency [Fig. 6(c)]. This suggests that the generation of the $2f_1$-f_2 distortion product is related to the generation region corresponding to the DP frequency and not to that corresponding to the frequencies of the primaries. For the $2f_2$-f_1 DPOAE, on the other hand, alignment occurs when $2f_2$-f_1 amplitude is plotted as a function of f_2 [Fig. 6(b)]. Near alignment is also observed when this distortion product is plotted as a function of f_1 [Fig. 6(a)]. This suggests that, in contrast to $2f_1$-f_2, the generation of the $2f_2$-f_1 DPOAE is related to the frequencies of the primaries and not to the distortion product frequency. Similar results were obtained for *R. catesbeiana* (Fig. 7), supporting a similar interpretation.[15]

The implication is that the generation of the $2f_1$-f_2 DPOAE in the frog may primarily occur at or near the DPOAE frequency place, while the generation of the $2f_2$-f_1 DPOAE may primarily occur at a frequency place between the two primaries. These results differ from those observed in mammals. As indicated by suppression studies using a variety of methods and confirmed by studies assessing recovery functions following short exposures to pure-tone or noise stimuli, the generation of the $2f_1$-f_2 DPOAE in mammals occurs primarily at a frequency place between the primaries. The distortion product itself then acts as a weak stimulus and may induce stimulus-frequency emissions at the distortion product frequency (Shera and Guinan, 1999). In contrast, the generation of the $2f_2$-f_1 DPOAE occurs at or near the

DPOAE frequency place on the basilar membrane (Probst *et al.*, 1991; Knight and Kemp, 2000). The opposite behavior of the $2f_2$-f_1 and $2f_1$-f_2 DPOAEs observed between mammalian and frog DPOAE data may be related to various anatomical differences between the mammalian and frog ears and needs to be researched further.

Figures 6(c) and 7(c) also reveal that, for frequencies in the AP, the $2f_2$-f_1 and $2f_1$-f_2 curves are very similar in shape and overlaying when DPOAE levels are plotted against emission frequency. This suggests that, regardless of the place of generation and prior to their reemission from the AP, both DPOAEs may be subject to the same mechanical filtering by a structure with broad filter characteristics, possibly the tectorial membrane. Several studies have used DPOAE data to derive filter characteristics (e.g., Brown *et al.*, 1992; Allen and Fahey, 1993; Kössl and Vater, 1996), while a recent study (Lukashkin and Russell, 2003) specifically argues for the existence of a possible tectorial membrane filter in mammals, based on the phase characteristics of different order DPOAEs.

3. Sexual dimorphism manifested in DPOAEs

Figures 9 and 10 indicate that, for the selected primary-tone levels, and over most of the frequency range tested, female subjects produced higher emission levels than male subjects in both species.[16] Overall, *R. catesbeiana* produced higher emission levels than *R. p. pipiens*.[17] In addition, there is a difference in the DPOAE audiogram peaks between males and females, suggesting that there are differences in the frequency tuning between sexes for both species.[18]

Several studies discussed in Probst *et al.* (1991) have shown significantly higher levels of spontaneous OAEs for females in adult, children, and infant humans compared to males. The cause of this sex difference is not clear. In the female organ of Corti, OHCs are arranged much more irregularly than in males, possibly due to the smaller physical dimensions of the female cochlea. It has been argued that such irregularities may be linked to OAE generation (Lonsbury-Martin *et al.*, 1988).[19] McFadden and Pasanen (1998) linked EOAE level differences between males and females to hormonal differences. They reported lower emission levels for homosexual than heterosexual females, and argued that the nature of this difference is consistent with the idea that homosexual females are exposed prenatally to higher levels of androgens than are heterosexual females.

Another possible explanation may lay in sex differences in middle ear transfer functions. In humans, for example, all OAEs can be detected reliably within a range (1–8 kHz) determined to a large extent by the frequency range (1–6 kHz) of effective reverse transmission from the middle ear (Probst *et al.*, 1991). Puria (2003) argues that the forward and reverse transfer characteristics of the mammalian middle ear alter significantly the levels of DPOAEs recorded in the ear canal, in a frequency specific manner (higher frequencies are increasingly attenuated). Johansson and Arlinger (2002) recorded significantly higher DPOAEs from human female subjects and linked high middle-ear compliance present in males to low emission levels. Additionally, clinical OAE studies in infants routinely record higher evoked OAEs emissions from female newborns (e.g., Newmark *et al.*, 1997), with this difference increasing with the frequency of the primaries (e.g., Cassidy and Ditty, 2001). The frequency-dependent change in EOAE levels may therefore be linked to the frequency-dependent transfer characteristics of the middle ear demonstrated by Puria (2003), which, as Johansson and Arlinger (2002) have argued in terms of middle ear compliance, may be sexually dimorphic.

In a recent study of bullfrogs, Mason *et al.* (2003) showed that, according to the current models of middle ear function, there is considerably higher middle ear impedance transform ratio in male bullfrogs compared to females. Although, according to the authors, it is not certain whether this reflects the assumptions of the model or actual physiological differences, such a transform ratio difference between male and female bullfrog middle ears corresponds well with the observed sex differences in the DPOAEs. In addition to middle ear transfer differences, sex differences in the frog have been reported in relation to peripheral auditory response. Narins and Capranica (1976) demonstrated such differences for the neotropical treefrog *Eleutherodactylus coqui*, associating sexual selectivity in the response properties of the BP to sexual selectivity in behavior. Since higher emission levels are thought to correspond to higher auditory sensitivity, frequency-dependent differences between sexes in OAE levels may indeed reflect sex differences in auditory sensitivity, possibly associated with male-female differences in what constitutes a biologically relevant sound.

V. SUMMARY AND CONCLUSIONS

DPOAE measurements were made for the $2f_1$-f_2 and $2f_2$-f_1 distortion products from male and female *R. p. pipiens* and *R. catesbeiana*. DPOAE I/O curves were obtained at frequencies corresponding to the best response in the AP and BP of both species. The $2f_1$-f_2 I/O curves obtained from the AP of both species are very similar to mammalian DPOAE I/O curves, exhibiting a non-monotonic response, with a distinct notch for primary levels $\sim$70 dB SPL. This similarity suggests that, like in the mammalian ear, there may be an amplification process present in the frog AP manifested as a compressive nonlinearity with saturation. In mammals this process is linked to prestin-mediated OHC activity. Since the frog ear lacks OHCs and its hair cells contain no prestin in their basolateral membranes, an alternative amplifier may be based on hair cell bundle spontaneous movement, shown to occur in the frog sacculus. The DPOAE I/O curves obtained from the BP of both species exhibit a rather monotonic response, suggesting that no amplification process may be present in this papilla. The difference in the response between amphibian and basilar papilla may be linked to respective differences in degree and type of innervation. It is also consistent with the fact that only the AP is tonotopically organized and could take better advantage of an amplification mechanism that increases tuning sharpness, as well as with the observation that only this papilla exhibits SOAEs.

The dependence of DPOAE level on the relative levels of primaries (L_1-L_2) was investigated in *R. p. pipiens*, within the range $50 \leqslant L_1$, $L_2 \leqslant 60$ dB SPL. Consistent with intermodulation distortion expectations alone and differently

from mammalian results, highest emissions were generally obtained for $L_1 \cong L_2$. This difference between mammalian and frog data is consistent with a difference in neural tuning-curve shape of mammalian and frog nerve fibers with comparable characteristic frequencies. In the mammalian case, neural tuning curves are thought to reflect BM displacement envelopes, which are in turn considered responsible for the dependence of DPOAE level on the relative levels of the primaries. Since the frog ear lacks a BM, the observed DPOAE level dependence on L_1-L_2 in the frog suggests that, if some mechanical structure (i.e., the tectorial membrane) functions analogously to the mammalian BM, it must be more broadly tuned. Our interpretation is supported by the rest of our results, outlined below, and does not exclude the possibility of the optimal primary-level difference to be L_1-$L_2 \neq 0$ for primary levels outside the level-range examined.

DPOAE audiograms were obtained from males and females of both species for primary frequencies spanning the animals' hearing range and primary levels determined by the present study. The results indicate the following:

(a) In agreement with previous DPOAE frog results (Van Dijk *et al.*, 2002; Meenderink and Van Dijk, 2004), DPOAEs are produced in both frog papillae and, except for primary frequencies at the high-frequency end of the BP, *R. catesbeiana* produce stronger emissions than *R. p. pipiens*.

(b) In contrast to mammalian results, the generation of the $2f_1$-f_2 DPOAE in the frog appears to occur primarily at or near the DPOAE frequency place, while the generation of the $2f_2$-f_1 DPOAE occurs primarily at a frequency place between the primaries. Further study is needed to confirm the observed difference in DPOAE generation sites between mammals and frogs and determine the precise reason behind it. Regardless of DPOAE generation site, both DPOAEs within the AP may be subject to the same filtering mechanism, possibly the tectorial membrane.

(c) For the primary-tone levels used, females of both *R. p. pipiens* and *R. catesbeiana* produce higher emissions than males in both papillae, with the difference being larger in the BP. The observed similarities in male and female frog DPOAE data in terms of I/O curve shapes, dependence on L_1-L_2, etc. suggest that sex differences in overall DPOAE levels are not related to DPOAE generation mechanism differences. DPOAE-based sexual dimorphism, also reported in humans and other mammals, may be linked to inner ear anatomical differences, middle ear transfer impedance differences, hormonal differences, nerve-fiber tuning differences, and/or behavioral and environmental factors.

ACKNOWLEDGMENT

Work supported by NIH Grant No. DC-00222 to PMN.

[1]In frogs, OAEs have been recorded from the ears of *Rana pipiens pipiens* [DPOAEs: Van Dijk *et al.*, 2002; Meenderink and Van Dijk, 2004; spontaneous OAEs (SOAEs): Van Dijk *et al.*, 1996], *Rana catesbeiana* (SOAEs: Gennosa *et al.*, 1989), *Rana esculenta* [SOAEs: Palmer and Wilson, 1982; Van Dijk *et al.*, 1989; Van Dijk and Wit, 1987; stimulus-frequency OAEs (SFOAEs): Palmer and Wilson, 1982], *Rana temporaria* (SOAEs: Baker *et al.*, 1989), *Xenopus laevis* (DPOAEs: Van Dijk *et al.*, 2002), *Hyla cinerea* (DPOAEs: Van Dijk and Manley, 2001; SOAEs: Van Dijk *et al.*, 1996), *Hyla chrysoscelis*, *Hyla versicolor*, and *Leptodactylus albilabris* (SOAEs: Van Dijk *et al.*, 1996). No measurable SOAE levels were obtained from *Xenopus laevis* (Van Dijk *et al.*, 1996) and no measurable DPOAE levels were obtained from *Scaphiopus couchi* (Van Dijk *et al.*, 2002). Neither SOAEs nor DPOAEs could be measured from *Bombina orientalis* (Van Dijk *et al.*, 1996 and Van Dijk *et al.*, 2002 respectively).

[2]E.g., $f_2/f_1 = 1.2$ in Kalluri and Shera (2001); $f_2/f_1 = 1.22$ in Emmerich *et al.* (2000); $f_2/f_1 = 1.25$ in Kemp and Brown (1983) (in Probst *et al.*, 1991).

[3]The only available parametric DPOAE study on the frog to date (Meenderink and Van Dijk, 2004) examines in detail the I/O characteristics of the two papillae in *R. p. pipiens*, offering results that, in general, agree with the results of the present study.

[4]Although exploration of the f_2/f_1 space in mammals has revealed several discontinuities, we assume that the value selected by the present study is close enough to the value used in previous frog DPOAE studies to allow for a meaningful comparison.

[5]The body temperature of the animals was not monitored systematically.

[6]Below 200 Hz the noise floor of the probe/analysis system was too high (>25 dB SPL) and the speaker response too weak (<65 dB SPL), while the hearing range of neither frog species examined exceeds 3000 Hz.

[7]Windowing: Blackman and Harris; averaging: 200 time frames with 90% overlap; input range: adjusted for each primary tone level combination to ensure consistent use of the analyzer's full dynamic range; harmonic/intermodulation distortion of the speaker–microphone–analyzer system $\geqslant 83$ dB below the primaries at all frequencies tested. System distortion was determined by attaching the probe tube system to a hard surface and measuring the level recorded at frequencies corresponding to DPOAE frequencies for selected primary-tone frequency and amplitude values. The dBV reading of the analyzer was converted to dB SPL using a measuring amplifier (B&K 2609, Brüel & Kjær, Nærum, Denmark).

[8]The analysis bandwidth was 0.98 Hz for 240 Hz$\leqslant f_1 \leqslant$800 Hz, 1.95 Hz for 850 Hz$\leqslant f_1 \leqslant$1600 Hz, and 3.9 Hz for 1700 Hz$\leqslant f_1 \leqslant$3000 Hz.

[9]Shofner and Feng (1984) indicate 1000 Hz as the frequency value marking the separation between AP and BP response in *R. catesbeiana*.

[10]One alternative hypothesis involves transition among different vibration modes of the inner ear structures as the levels of the primaries change, claimed to be responsible for the observed DPOAE I/O curve notch. In such an explanation, it is expected that, for a given set of primary tone frequencies, the DPOAE I/O curve notches should be observed at the same stimulus levels, regardless of distortion product. Our data indicate that this may be the case for the *R. catesbeiana* I/O curves but it certainly is not for the *R. p. pipiens* I/O curves [Figs. 2(a) and 2(c)], suggesting that the ''vibration mode transition'' hypothesis may be excluded as an explanation for the DPOAE I/O curve shapes observed. Another explanation for the presence of notches in the DPOAE I/O curves is based on a hypothesized frequency shift of DPOAE fine-structure in human subjects with changing stimulus levels (He and Schmiedt, 1993; Whitehead, 1998). Figure 3 demonstrates that this is not the case for the frog species tested. In mammals, the presence of DP audiogram fine structure has been explained in terms of interference between two emission components (e.g., Talmage *et al.*, 1998, 1999). Further study is needed to determine the origin of the DPOAE fine structure observed in the frog.

[11]In the case of the frog ear, however, the DPOAE phase data obtained and analyzed by Meenderink and Van Dijk (2004) do not appear to support such a model.

[12]The presence of an inner ear amplifier in the echidna, a Monotreme with features common to early mammals, birds, and reptiles and which has no organ of Corti, has also been inferred based on DPOAE data (Mills and Shepherd, 2001).

[13]The range of primary stimulus levels tested (50–60 dB SPL) is below the DPOAE I/O curve saturation point for *R. p. pipiens* (Fig. 2), supporting the suggestion that the observed effect of L_1-L_2 on DPOAE levels is indeed indicative of a difference between mammalian and frog responses as opposed to simply a level saturation response in the frog.

[14]In an alternative interpretation of the data by Hillery and Narins (1987), Pitchford and Ashmore (1987) argue that the large phase accumulations in the phase-locked responses of auditory-nerve fibers in frogs may not reflect phase accumulations of a mechanical traveling wave (e.g., Hillery

and Narins, 1984) but may result from electrical filtering within the hair cells.

[15]The DP audiogram dips for *R. catesbeiana* are shallower than those for *R. p. pipiens*, due to the fact that, in *R. catesbeiana*, there is a slight overlap in the coverage of the two papillae, while in *R. p. pipiens* there is not.

[16]These results are consistent with the results obtained from the first two experiments, examining the relationship of absolute and relative levels of the primaries to emission levels.

[17]With the exception of emissions from the high-frequency end of the BP associated possibly with the reduced response of the male and female bullfrog middle ear at these frequencies ($>\sim$1200 Hz; Mason and Narins, 2002a, b).

[18]Although the dependence of DPOAE levels on animal weight and/or the size of the eardrum was not systematically examined, our data suggest that any DPOAE level difference between species and sexes is most likely not related to such factors.

[19]Sex differences in DPOAE phase delay measures have also been attributed to anatomical differences in cochlear length (Bowman *et al.*, 2000).

Abdala, C. (**2000**). ''Distortion product otoacoustic emission ($2f_1$-f_2) amplitude growth in human adults and neonates,'' J. Acoust. Soc. Am. **107**, 446–456.

Allen, J. B., and Fahey, P. F. (**1993**). ''A second cochlear-frequency map that correlates distortion product, neural tuning measurements,'' J. Acoust. Soc. Am. **94**, 809–816.

Baker, R. J., Wilson, J. P., and Whitehead, M. L. (**1989**). ''Otoacoustic evidence for nonlinear behavior in frog hearing: Suppression but no distortion products,'' in *Cochlear Mechanisms: Structure, Function and Models*, edited by J. Wilson and D. T. Kemp (Plenum, New York), pp. 349–356.

Bowman, D. M., Brown, D. K., and Kimberley, B. P. (**2000**). ''An examination of gender differences in DPOAE phase delay measurements in normal-hearing human adults,'' Hear. Res. **142**, 1–11.

Brown, A. M., Gaskill, S. A., and Williams, D. M. (**1992**). ''Mechanical filtering of sound in the inner ear,'' Proc. R. Soc. London, Ser. B **250**, 29–34.

Brown, A. M., McDowell, B., and Forge, A. (**1989**). ''Effects of chronic gentamicin treatment on hair cells can be monitored using acoustic distortion products,'' Hear. Res. **42**, 143–156.

Capranica, R. R. (**1976**). ''Morphology and physiology of the auditory system,'' in *Frog Neurobiology*, edited by R. Llinas and W. Precht (Springer Verlag, Berlin), pp. 551–575.

Cassidy, J. W., and Ditty, K. M. (**2001**). ''Gender differences among newborns on a transient otoacoustic emissions test for hearing,'' J. Music. Ther. **38**, 28–35.

Crawford, A. C., and Fettiplace, R. (**1985**). ''The mechanical properties of cilliary bundles of turtle cochlear hair cells,'' J. Physiol. (London) **364**, 359–379.

Ehret, G., and Capranica, R. R. (**1980**). ''Masking patterns and filter characteristics of auditory nerve fibers in the green treefrog (*Hyla cinerea*),'' J. Comp. Physiol. **141**, 1–12.

Emmerich, E., Richter, F., Meißner, W., and Dieroff, H. G. (**2000**). ''The effect of impulse noise exposure on distortion product otoacoustic emissions in the awake guinea pig,'' Eur. Arch. Otorhinolaryngol. **257**128–132.

Fahey, P. F., Stagner, B. B., Lonsbury-Martin, B. L., and Martin, G. K. (**2000**). ''Nonlinear interactions that could explain distortion product interference response areas,'' J. Acoust. Soc. Am. **108**, 1786–1802.

Faulstich, M., and Kössl, M. (**2000**). ''Evidence for multiple DPOAE components based upon group delay of the $2f_1$-f_2 distortion in the gerbil,'' Hear. Res. **140**, 99–110.

Feng, A. S., Narins, P. M., and Capranica, R. R. (**1975**). ''Three populations of primary auditory fibers in the bullfrog (*Rana catesbeiana*): Their peripheral origins and frequency sensitivities,'' J. Comp. Physiol. **100**, 221–229.

Geisler, C. D., Van Bergeijk, W. A., and Frishkopf, S. (**1964**). ''The inner ear of the bullfrog,'' J. Morphol. **114**, 43–58.

Gennosa, T. J., Voigt, H. F., and Burkard, R. (**1989**). ''Spontaneous otoacoustic emissions in *Rana catesbeiana*, the American bullfrog,'' J. Acoust. Soc. Am. Suppl. 1 **85**, S35.

He, N., and Schmiedt, R. A. (**1993**). ''Fine structure of the $2f_1$-f_2 acoustic distortion product: Changes with primary level,'' J. Acoust. Soc. Am. **94**, 2659–2669.

Hillery, C. M., and Narins, P. M. (**1984**). ''Neurophysiological evidence for a traveling wave in the amphibian inner ear,'' Science **255**, 1037–1039.

Hillery, C. M., and Narins, P. M. (**1987**). ''Frequency and time domain comparison of low-frequency auditory fibre responses in two anuran amphibians,'' Hear. Res. **25**, 233–248.

Hudspeth, A. (**1997**). ''Mechanical amplification of stimuli by hair cells,'' Curr. Opin. Neurobiol. **7**, 490–496.

Johansson, M., and Arlinger, S. (**2002**). ''Otoacoustic emissions in a general adult population in Sweden,'' Proceedings of Nordic Noise 2002. Stockholm, Sweden.

Kalluri, R., and Shera, C. A. (**2001**). ''Distortion-product source unmixing: A test of the two-mechanism model for DPOAE generation,'' J. Acoust. Soc. Am. **109**, 622–637.

Kastanioudakis, I., Ziavra, N., Anastasopoulos, D., and Skevas, A. (**2003**). ''Measuring of distortion product otoacoustic emissions using multiple tone pairs,'' Eur. Arch. Otorhinolaryngol. **260**, 395–400.

Kemp, D. T. (**1979**). ''Evidence for cochlear nonlinearity and frequency selective wave amplification in the cochlea,'' Arch. Otolaryngol. **224**, 37–45.

Kennedy, H. J., Evans, M. G., Crawford, A. C., and Fettiplace, R. (**2003**). ''Fast adaptation of mechanoelectrical transducer channels in mammalian cochlear hair cells,'' Nat. Neurosci. **6**, 832–836.

Knight, R. D., and Kemp, D. T. (**2000**). ''Indications of different distortion product otoacoustic emission mechanisms from a detailed f_1, f_2 area study,'' J. Acoust. Soc. Am. **107**, 457–473.

Köppl, C., Manley, G. A., and Johnstone, B. M. (**1993**). ''Acoustic distortion products in the ear canal of the bobtail lizard. I: General characteristics,'' J. Acoust. Soc. Am. **93**, 2820–2933.

Kössl, M., and Boyan, G. S. (**1998**). ''Acoustic distortion products from the ear of a grasshopper,'' J. Acoust. Soc. Am. **104**, 326–335.

Kössl, M., and Vater, M. (**1996**). ''Further studies on the mechanics of the cochlear partition in the mustached bat II. A second cochlear frequency map derived from acoustic distortion products,'' Hear. Res. **94**, 78–86.

Kummer, P., Janssen, T., Hulin, P., and Arnold, W. (**2000**). ''Optimal L1-L2 primary tone level separation remains independent of test frequency in humans,'' Hear. Res. **146**, 47–56.

Lewis, E. R. (**1977**). ''Structural correlates of function in the anuran amphibian papilla,'' Scan Electron Microsc. **II**, 429–436.

Lewis, E. R., and Narins, P. M. (**1999**). ''The Acoustic Periphery of Amphibians: Anatomy and Physiology,'' in *Comparative Hearing: Fish and Amphibians*, edited by R. R. Fay and A. N. Popper (Springer-Verlag, New York), pp. 101–154.

Lewis, E. R., Leverenz, E. L., and Bialek, W. S. (**1985**). *The Vertebrate Inner Ear* (CRC, Boca Raton, FL).

Lewis, E. R., Leverenz, E. L., and Koyama, H. (**1982**). ''The tonotopic organization of the bullfrog amphibian papilla, an auditory organ lacking a basilar membrane,'' J. Comp. Physiol. **145**, 437–445.

Liberman, C. M., Gao, J., He, D. Z. Z., Wu, X., Jia, S., and Zuo, J. (**2002**). ''Prestin is required for electromotility of the outer hair cell and for the cochlear amplifier,'' Nature (London) **419**, 300–304.

Lonsbury-Martin, B. L., Martin, G. K., Probst, R., and Coats, A. C. (**1988**). ''Spontaneous otoacoustic emissions in a nonhuman primate: II. Cochlear anatomy,'' Hear. Res. **33**, 69–94.

Lonsbury-Martin, B. L., McCoy, M. J., Whitehead, M. L., and Martin, G. K. (**1993**). ''Clinical testing of distortion-product otoacoustic emissions,'' Ear Hear. **14**, 11–22.

Lukashkin, A. N., and Russell, I. J. (**2003**). ''A second, low-frequency mode of vibration in the intact mammalian cochlea,'' J. Acoust. Soc. Am. **113**, 1544–1550.

Lukashkin, A. N., Lukashkina, V. A., and Russell, I. J. (**2002**). ''One source for distortion product otoacoustic emissions generated by low- and high-level primaries,'' J. Acoust. Soc. Am. **111**, 2740–2748.

Maison, S., Micheyl, C., and Collet, L. (**1997**). ''Medial olivocochlear efferent system in humans studied with amplitude-modulated tones,'' J. Neurophysiol. **77**, 1759–1768.

Manley, G. A., Kirk, D., Köppl, C., and Yates, G. K. (**2001**). ''In-vivo evidence for a cochlear amplifier in the hair-cell bundle of lizards,'' Proc. Natl. Acad. Sci. U.S.A. **98**, 2826–2831.

Martin, G. K. (**2002**). ''DPOAE generation mechanisms,'' OAEs Portal Zone, University of Ferrara, Italy. http://www.otoemissions.org/guest_editorials/editorial_May_June_2002.html. [Last accessed: 10/2004.]

Martin, P., and Hudspeth, A. (**1999**). ''Active hair-bundle movements can amplify a hair cell's response to oscillatory mechanical stimuli,'' Proc. Natl. Acad. Sci. U.S.A. **7**, 14306–14311.

Mason, M. J., and Narins, P. M. (**2002a**). ''Vibrometric studies of the middle ear of the bullfrog *Rana catesbeiana*. I. The extrastapes,'' J. Exp. Biol. **205**, 3153–3165.
Mason, M. J., and Narins, P. M. (**2002b**). ''Vibrometric studies of the middle ear of the bullfrog *Rana catesbeiana*. II. The operculum,'' J. Exp. Biol. **205**, 3167–3176.
Mason, M. J., Lin, C. C., and Narins, P. M. (**2003**). ''Sex differences in the middle ear of the bullfrog (*Rana catesbeiana*),'' Brain Behav. Evol. **61**, 91–101.
McFadden, D., and Pasanen, E. G. (**1998**). ''Comparison of the auditory systems of heterosexuals and homosexuals: Click-evoked otoacoustic emissions,'' Proc. Natl. Acad. Sci. U.S.A. **95**, 2709–2713.
Meenderink, S. W. F., and Van Dijk, P. (**2004**). ''Level dependence of distortion product otoacoustic emissions in the leopard frog, *Rana pipiens pipiens*,'' Hear. Res. **192**, 107–118.
Mills, D. M. (**2002**). ''Interpretation of standard distortion product otoacoustic emission measurements in light of the complete parametric response,'' J. Acoust. Soc. Am. **112**, 1545–1560.
Mills, D. M., and Shepherd, R. K. (**2001**). ''Distortion product otoacoustic emission and auditory brainstem responses in the echidna (*Tachyglossus aculeatus*),'' J. Assoc. Res. Otolaryngol. **2**, 130–146.
Mom, T., Bonfils, P., Gilain, L., and Avan, P. (**2001**). ''Origin of cubic difference tones generated by high-intensity stimuli: Effect of ischemia and auditory fatigue on the gerbil cochlea,'' J. Acoust. Soc. Am. **110**, 1477–1488.
Narins, P. M., and Capranica, R. R. (**1976**). ''Sexual differences in the auditory system of the tree frog *Eleutherodactylus coqui*,'' Science **192**, 378–380.
Narins, P. M., and Hillery, C. (**1983**). ''Frequency coding in the inner ear of anuran amphibians,'' in *Hearing–Physiological Bases and Psychophysics*, edited by R. Klinke and R. Hartmann (Springer-Verlag, New York), pp. 70–76.
Newmark, M., Merlob, P., Bresloff, I., Olsha, M., and Attias, J. (**1997**). ''Click evoked otoacoustic emissions: inter-aural and gender differences in newborns,'' J. Basic Clin. Physiol. Pharmacol. **8**, 133–139.
Owens, J. J., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1992**). ''Influence of otitis media on evoked otoacoustic emission in children,'' Semin. Hear. **13**, 53–65.
Owens, J. J., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1993**). ''Otoacoustic emissions in children with normal ears, middle ear dysfunction, and ventilating tubes,'' Am. J. Otol. **14**, 34–40.
Palmer, A. R., and Wilson, J. P. (**1982**). ''Spontaneous and evoked acoustic emissions in the frog *Rana esculenta*,'' J. Phonetics **324**, 66.
Pitchford, S., and Ashmore, J. F. (**1987**). ''An electrical resonance in hair cells of the amphibian papilla of the frog *Rana temporaria*,'' Hear. Res. **27**, 75–83.
Probst, R., Lonsbury-Martin, B. L., and Martin, G. K. (**1991**). ''A review of otoacoustic emissions,'' J. Acoust. Soc. Am. **89**, 2027–2067.
Puria, S. (**2003**). ''Measurements of human middle ear forward and reverse acoustics: Implications for otoacoustic emissions,'' J. Acoust. Soc. Am. **113**, 2773–2789.
Robbins, R. G., Bauknight, B. S., and Honrubia, M. D. (**1967**). ''Anatomical distribution of efferent fibers in the VIIIth cranial nerve of the bullfrog (*Rana catesbeiana*),'' Acta Oto-Laryngol. **64**, 436–448.
Robles, L., and Ruggero, M. A. (**2001**). ''Mechanics of the mammalian cochlea,'' Physiol. Rev. **81**, 1305–1352.
Ronken, D. A. (**1991**). ''Spike discharge properties that are related to the characteristic frequency of single units in the frog auditory nerve,'' J. Acoust. Soc. Am. **90**, 2428–2440.
Rosowski, J. J., Peake, W. T., and White, J. R. (**1984**). ''Cochlear nonlinearities inferred from two-tone distortion products in the ear canal of the alligator lizard,'' Hear. Res. **13**, 141–158.
Schneider, S., Prijs, V., and Schoonhoven, R. (**2003**). ''Amplitude and phase of distortion product otoacoustic emissions in the guinea pig in an (f_1, f_2) area study,'' J. Acoust. Soc. Am. **113**, 3285–3296.
Shera, C. A., and Guina, Jr., J. J. (**1999**). ''Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs,'' J. Acoust. Soc. Am. **105**, 782–798.
Shera, C. A., and Guinan, Jr., J. J. (**2003**). ''Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning,'' J. Acoust. Soc. Am. **113**, 2762–2772.
Shofner, W. P., and Feng, A. S. (**1984**). ''Quantitative light and scanning electron microscopic study of the developing auditory organs of the bullfrog: Implications on their functional characteristics,'' J. Comp. Neurol. **224**, 141–154.
Simmons, D. D., Bertolotto, C., and Narins, P. M. (**1992**). ''Innervation of the amphibian and basilar papillae in the leopard frog: Reconstructions of single labeled fibers,'' J. Comp. Neurol. **322**, 191–200.
Stover, L. J., Gorga, M. P., Neely, S. T., and Montoya, D. (**1996**). ''Toward optimizing the clinical utility of distortion product otoacoustic emission measurements,'' J. Acoust. Soc. Am. **100**, 956–967.
Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (**1999**). ''Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **105**, 275–292.
Talmadge, C. L., Tubis, A., Long, G. R., and Piskorski, P. (**1998**). ''Modeling otoacoustic emission and hearing threshold fine structures,'' J. Acoust. Soc. Am. **104**, 1517–1541.
Tubis, A., Talmadge, C. L., Long, G. R., and Dhar, S. (**2000**). ''On the relationships between the fixed-f_1, fixed-f_2, and fixed-ratio phase derivatives of the 2 f_1-f_2 distortion product otoacoustic emission,'' J. Acoust. Soc. Am. **108**, 1772–1784.
Van Dijk, P., and Manley, G. A. (**2001**). ''Distortion product otoacoustic emissions in the tree frog *Hyla cinerea*,'' Hear. Res. **153**, 14–22.
Van Dijk, P., and Wit, H. P. (**1987**). ''Temperature dependence of frog spontaneous otoacoustic emissions,'' J. Acoust. Soc. Am. **82**, 2147–2150.
Van Dijk, P., Mason, M. J., and Narins, P. M. (**2002**). ''Distortion product otoacoustic emissions in frogs: correlation with middle and inner ear properties,'' Hear. Res. **173**, 100–108.
Van Dijk, P., Narins, P. M., and Mason, M. J. (**2003**). ''Physiological vulnerability of distortion product otoacoustic emissions from the amphibian ear,'' J. Acoust. Soc. Am. **114**, 2044–2048.
Van Dijk, P., Narins, P. M., and Wang, J., (**1996**). ''Spontaneous otoacoustic emissions in seven frog species,'' Hear. Res. **101**, 102–112.
Van Dijk, P., Wit, H. P., and Segenhout, J. M. (**1989**). ''Spontaneous otoacoustic emissions in the European edible frog (*Rana esculenta*): Spectral details and temperature dependence,'' Hear. Res. **42**, 273–282.
Wagner, W., and Plinkert, P. K. (**1999**). ''The relationship between auditory threshold and evoked otoacoustic emissions,'' Eur. Arch. Otorhinolaryngol. **256**, 177–188.
Whitehead, M. L. (**1998**). ''Species differences of distortion-product otoacoustic emissions: Comment on Interpretation of distortion-product otoacoustic emission measurements. I. Two stimulus tones [J. Acoust. Soc. Am. **102**, 413–429 (1997)],'' J. Acoust. Soc. Am. **103**, 2740–2742.
Whitehead, M. L., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995a**). ''Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears: I. Effects of decreasing L_2 below L_1,'' J. Acoust. Soc. Am. **97**, 2346–2358.
Whitehead, M. L., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995b**). ''Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears: II. Asymmetry in L_1, L_2 space,'' J. Acoust. Soc. Am. **97**, 2359–2377.
Wit, H. P., Van Dijk, P., and Segenhout, J. M. (**1989**). ''DC injection alters otoacoustic emission frequency in the frog,'' Hear. Res. **41**, 199–204.
Zelick, R. D., and Narins, P. M. (**1985**). ''Temporary threshold shifts, adaptation, and characteristics of frog auditory nerve fibers,'' Hear. Res. **17**, 161–176.
Zhang, M., and Abbas, P. J. (**1997**). ''Effects of middle ear pressure on otoacoustic emission measures,'' J. Acoust. Soc. Am. **102**, 1032–1037.
Zheng, J., Shen, W., He, D. Z. Z., Long, K., Madison, L. D., and Dallos, P. (**2000**). ''Prestin is the motor protein of cochlear outer hair cells,'' Nature (London) **405**, 149–155.

Habitat-dependent ambient noise: Consistent spectral profiles in two African forest types

Hans Slabbekoorn[a)]
Center for Tropical Research, Department of Biology, San Francisco State University, San Francisco, California 94132

(Received 30 May 2001; accepted for publication 11 August 2004)

Many animal species use acoustic signals to attract mates, to defend territories, or to convey information that may contribute to their fitness in other ways. However, the natural environment is usually filled with competing sounds. Therefore, if ambient noise conditions are relatively constant, acoustic interference can drive evolutionary changes in animal signals. Furthermore, masking noise may cause acoustic divergence between populations of the same species if noise conditions differ consistently among habitats. In this study, ambient noise was sampled in a replicate set of sites in two habitat types in Cameroon: contiguous rainforest and ecotone forest patches north of the rainforest. The noise characteristics of the two forest types show significant and consistent differences. Multiple samples taken at two rainforest sites in different seasons vary little and remain distinct from those in ecotone forest. The rainforest recordings show many distinctive frequency bands, with a general increase in amplitude from low to high frequencies. Ecotone forest only shows a distinctive high-frequency band at some parts of the day. Habitat-dependent abiotic and biotic sound sources and to some extent habitat-dependent sound transmission are the likely causes of these habitat-dependent noise spectra.
[DOI: 10.1121/1.1811121]

PACS numbers: 43.80.Nd [WWA] Pages: 3727–3733

I. INTRODUCTION

Many animal species use acoustic signals to communicate with each other. Males may defend their territories using songs or calls; females may find and choose mates based on acoustic courtship signals. Individuals may also aggregate guided by contact calls or flee and disperse triggered by alarm calls. Long-distance transmission of the acoustic signal is often a crucial factor in any of these functions, as conveyance of the message encoded in a sound depends on remote detection and recognition (Dusenbery, 1992; Bradbury and Vehrencamp, 1998). Characteristics of the environment used for acoustic communication will affect the transmission of sounds by signal attenuation and degradation (Wiley and Richards, 1982; Slabbekoorn, 2004). Another major factor that influences the perceptibility of an acoustic signal in the natural environment is the amount and nature of ambient noise (Klump, 1996; Slabbekoorn, 2004). Ambient noise affects the signal-to-noise ratio, and therefore may interfere with the efficiency of a signal depending on the frequency matching (Klump, 1996; Langeman *et al.*, 1998; Lohr *et al.*, 2003; Dooling, 2004). Noise interference may have direct fitness consequences, especially through the impact on signal efficiency in the context of sexual selection (Searcy and Andersson, 1986; Catchpole, 1987; Searcy and Yasukawa, 1996), which may result in signal evolution towards an acoustic design more efficient under prevailing noise conditions (Endler, 1992; Forrest, 1994; Slabbekoorn and Smith, 2002a; 2002b; De la Torre and Snowdon, 2002).

Sources of ambient noise are diverse and of all frequencies (Morton, 1975; Brenowitz, 1982a; 1982b; Ryan and Brenowitz, 1985; Waser and Brown, 1986). Rustling leaves and twigs may produce sounds over a wide range of frequencies, while air passing over substrates primarily causes low-frequency sounds. Animal communities may be complex and natural habitats can harbor many vocal species filling the air continuously. The frequency of sounds produced by animals roughly depends on their body size, explaining general differences among taxonomic groups, with insects mainly restricted to relatively high frequencies (e.g., Bennet-Clark, 1970; Michelsen and Nocke, 1974), and frogs and toads (e.g., Gerhardt, 1994), birds (e.g., Wallschläger, 1980; Tubaro and Mahler, 1998), and mammals (e.g., Waser and Waser, 1977) vocalizing at lower frequencies. In addition to all natural abiotic and biotic noise sources, the world becomes increasingly affected by anthropogenic noise, which concerns primarily low-frequency rumbles and roars from cars, aircraft, and all sorts of machinery (De la Torre and Snowdon, 2002; Slabbekoorn and Peet, 2003; Katti and Warren, 2004).

The spectral profile of ambient noise is likely to be habitat specific. The impact of wind will depend on the openness of the vegetation, while the composition of the sound-producing animal community will also be habitat specific. Several studies have found variation in noise spectra associated with habitat type (Morton, 1975; Ryan and Brenowitz, 1985; Waser and Brown, 1986; De la Torre and Snowdon, 2002), but the sampling in these studies was limited, both with respect to replication over the day, and with respect to collecting data at multiple sites. Therefore, empirical evidence for consistent habitat-dependent noise conditions is

[a)]Electronic mail: slabbekoorn@rulsfb.leidenuniv.nl; Current address: Behavioural Biology, Institute of Biology, Leiden University, P.O. Box 9516, 2300 RA Leiden, The Netherlands.

TABLE I. Overview of field sites per habitat type. DJA reserve was sampled three times (DJA1, DJA2, and DJA3) and Kribi twice (KRB1 and KRB2).

	Site	Code	Latitude	Longitude	Visiting date
Ecotone forest	Betare Oya	BTO	N 5° 34′	E 14° 05′	17 February
	Wakwa	WKW	N 7° 16′	E 13° 31′	25 February
	Ngaoundaba	RDN	N 7° 08′	E 13° 42′	27 February
	Meiganga	MGG	N 6° 27′	E 14° 15′	2 March
	Djohong	DJH	N 6° 50′	E 14° 40′	4 March
	Yoko	YKO	N 4° 02′	E 9° 06′	11 March
Rainforest	Dja Reserve	DJA1	N 3° 11′	E 12° 49′	22 January
	Dja Reserve	DJA2	N 3° 11′	E 12° 49′	28 January
	Kribi	KRB1	N 2° 43′	E 9° 52′	6 February
	Dja Reserve	DJA3	N 3° 11′	E 12° 49′	15 April
	Eboumetoum	EBE	N 3° 44′	E 13° 04′	23 April
	Ndibi	NDB	N 3° 47′	E 12° 13′	27 April
	Zoebefame	ZBF	N 2° 39′	E 13° 23′	6 May
	Kribi	KRB2	N 2° 43′	E 9° 52′	12 June
	Akok	AKK	N 2° 45′	E 11° 14′	23 June
	Nyangon	NGN	N 2° 58′	E 10° 45′	26 June
	Lendi	LND	N 2° 55′	E 9° 57′	28 June

still largely lacking. In this paper, I present a replicated set of ambient noise samples, and test whether spectral characteristics differ consistently between two tropical forest types.

II. METHODS

A. Sites and visiting dates

Ambient noise samples were collected at 14 sites in Cameroon, central Africa, from January to June 1999 (Table I): eight sites in rainforest, and six sites in the ecotone region towards dryer savanna in the north (Fig. 1). Rainforest sites concern primary or secondary mature forest with a continuous canopy at about 30-m height. Forest in the ecotone is restricted to relatively small patches along streams, typically referred to as gallery forests. The canopy of gallery forests is generally lower compared to rainforest, and in particular the middle forest strata are relatively open (Letouzey, 1968; Longman and Jenik, 1992). Patterns of rainfall in Cameroon vary geographically, but in general there is a minor wet season in April–May and a major wet season that peaks in September–October (Smith, 1990). Seasonal changes in animal activity may vary between habitats, but little is known. Little greenbul (*Andropadus virens*), a widespread forest passerine, sings primarily in February and March in gallery forests while at that time the species is quiet in the rainforest, where it starts to sing in April (Slabbekoorn and Smith, 2002b).

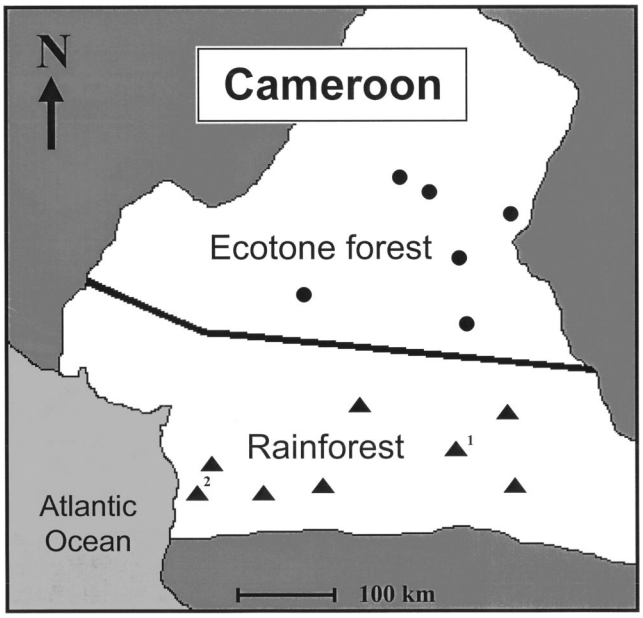


FIG. 1. Map of Cameroon showing the location of eight rainforest sites (triangles below the line indicating the habitat boundary) and six ecotone sites (dots above line). The rainforest site labeled with a 1 is sampled three times (DJA) and the one labeled with a 2 is sampled twice (KRB).

B. Sampling procedure

I used a Sennheiser ME67 directional microphone and a Sony TCM-5000EV tape recorder. Ambient noise was recorded in a standardized way, using the same equipment, the same procedure, and the same settings for all sites. The microphone was held at 1.5 m from the ground during all recordings. All recording sites were within the forest with moderately wet soil and leaf litter on the ground and more or less dense shrub foliage nearby. I took 33 samples per location, taken in three different directions, at 11 hourly sampling times throughout the day, from 7:00 in the morning to 17:00 in the afternoon. At the first recording time of the day, one direction was chosen randomly, the second direction was opposite to the first, and the third direction was a recording with the microphone in upright position, pointing at the canopy. Gallery forest patches are long and narrow (often between 10 to 20 m wide); here, recording directions were chosen parallel to the forest edge, and unavoidably they were always taken relatively near the edge. The directions were then fixed for the rest of the hourly recordings during the day. None of the recording days included rainy days. Rain on the canopy layer, and dripping water, will add noise in a wide frequency range and is likely to conceal habitat-dependent patterns at least to some extent.

C. Analyses

Ambient noise recordings were digitized at a sampling rate of 25.0 kHz using SIGNAL sound analysis software (Engineering Design, Belmont, 1998). Sonograms were made of 1-s time periods for all samples, which were organized by location and time of day for visual inspection. Short sound events, such as bird song or monkey calls, were avoided to sample the continuous sound background. I measured the amplitude level (rms values) within eight 1.0-kHz frequency ranges from 0 to 8.0 kHz. The FFT size used was 1024, which led to a spectral resolution of 24 Hz and a temporal resolution of 41 ms. No consistent differences were found in the spectral distribution for recordings taken in the three different directions, which indicates that they are neither heavily affected by interference patterns from ground reflections, nor that recordings from an omnidirectional microphone would have led to qualitatively very different results. The rms values were averaged over the three recording directions, before testing for habitat differences in the spectral distribution of sound energy, comparing each of the eight frequency ranges. Three frequency ranges with typical dominant noise sources (two in the rainforest, one in ecotone forest) were separately processed to give additional insight into variation over the day. Finally, the eight amplitude measures across the spectrum were transformed by means of a principal component analysis to visualize intra- and interhabitat variation. All statistical analyses were done using SPSS, version 10.0 from June 2000.

III. RESULTS

A. Sonographic inspection

The noise spectra of rainforest sites were easily distinguished from those of gallery forest sites in the ecotone. Sonograms of rainforest sites showed many relatively discrete frequency bands of sounds at different frequencies, which were less apparent in ecotone sites (Fig. 2). In all rainforest sites, a ''double band'' of sound appeared around 5.0 kHz in the early morning, which increased in frequency in a stepwise manner from hour to hour over the morning, to become less recognizably present during the rest of the day. This typical double band was never present in the sonograms of ecotone sites. The increase in frequency of apparently the same noise bands over the morning was only detected in one of the ecotone sites. Another typical band of sound was present around 3.5 kHz throughout the day in all rainforest sites (except for one: NDB), which never appeared in ecotone forest sites. The gallery forests in the ecotone showed fewer distinctive sound bands, and could be characterized by relatively equal noise levels throughout the spectrum. A high and loud frequency band above 6.0 Hz, peaking around 7.0 kHz, was apparent during long periods of the day in YKO and in WKW [see Fig. 2(A), 10.00–15.00], but was only detected in very few samples in DJH, and not at all in the other three ecotone sites.

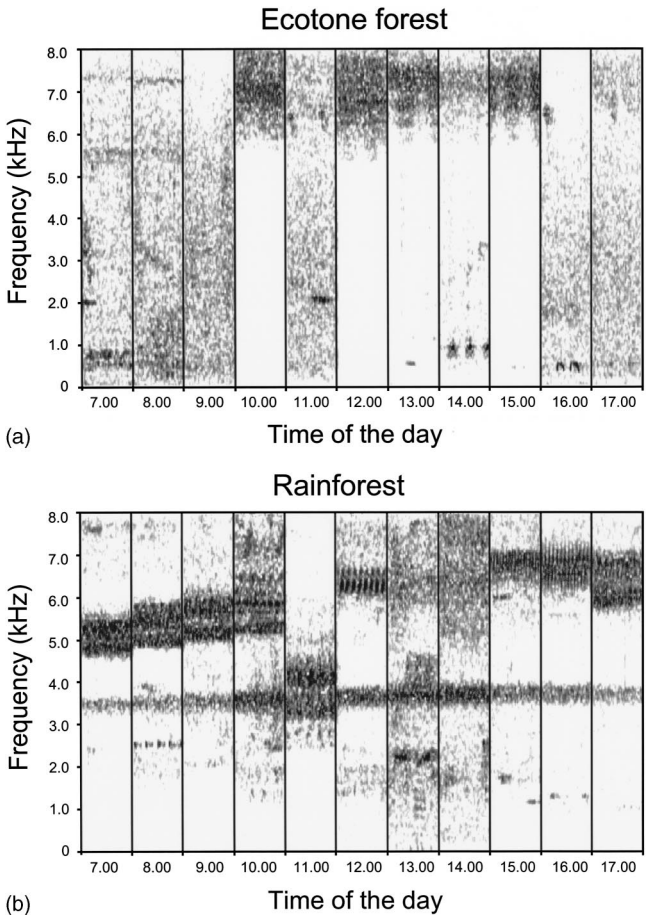


FIG. 2. Compilation of sonograms of 11 1-s sound samples taken from 7.00 to 17.00 at (a) an ecotone forest site (WKW) and (b) a rainforest site (ZBF). Note for the ecotone site the loud high-frequency band during several samples in the middle of the day and for the rainforest site (1) the many different discrete bands in the upper part of the frequency range; (2) the stepwise increase in frequency for the typical double noise band from 7.00 to 10.00; and (3) the noise band at about 3.5 kHz, which is present throughout the day (while completely absent in the ecotone site). The loud double noise band at 11.00 is caused by calling of *Ugada giovanninae* cicadas.

B. Identification of sound sources

One identified source of between-site variation was the presence or absence of a large (>5 cm) species of cicada, *Ugada giovanninae* Boulard (Duffels and Van der Laan, 1985). Individuals of this species were located when calling, typically sitting on bark of large tree trunks at an estimated height of on average 5.6 m (SD$\pm$2.6, $N=20$ individuals, site: DJA). The species was found calling in five of the eight rainforest sites (not in NDB, KRB, and LND), but never in ecotone sites. The apparent absence of the cicada in NDB may be explained by a change in species composition at the site due to human exploitation, whereas KRB and LND are coastal sites, and this particular species was never observed in this or any other coastal site (not included in this study). Calling individuals would chorus mainly during a few hours in the morning (typically between 9:00 and 12:00), and again for a short period just before dusk. The chorus leads to a loud double band of sound between 3.0 kHz and 4.5 kHz [see Fig. 2(B), 11.00].

Only a few bird species sing in a relatively continuous mode, and therefore contributed in some of the sampling

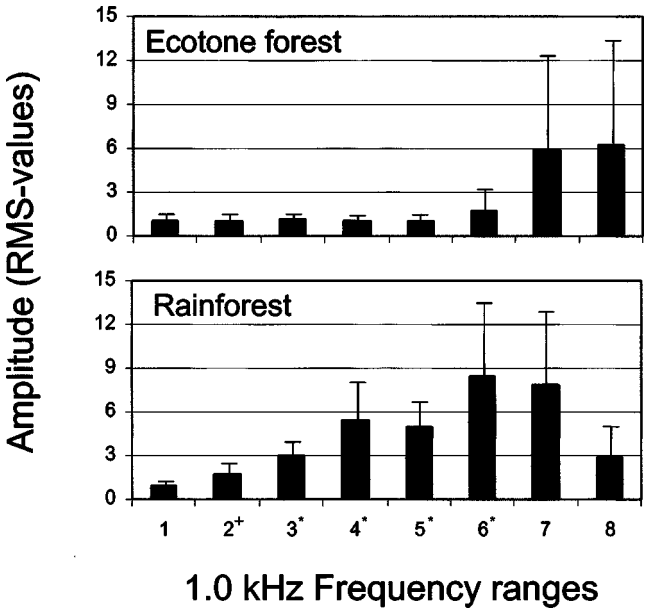


FIG. 3. Relative amplitude levels (rms values) per 1.0-kHz interval (averages and standard deviation) from 0 to 8.0 kHz for eight rainforest sites (lower part) and six ecotone sites (upper part). Each frequency range was tested for significant differences between the two habitats (each site contributing one set of measurements only): + means $P<0.1$;* means $P<0.001$ (Kruskal–Wallis test).

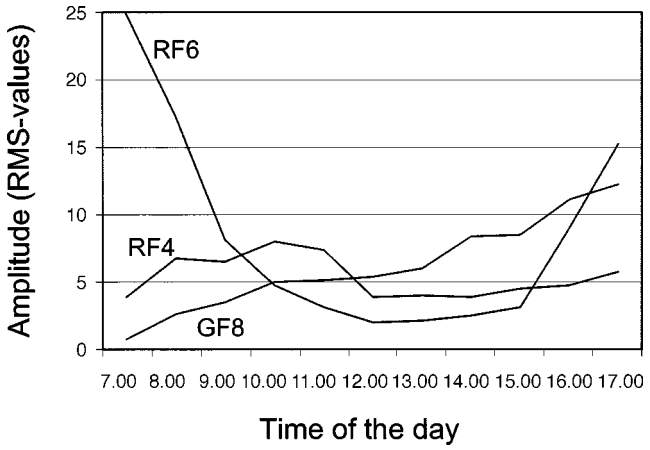


FIG. 4. Variation in relative amplitude over the day for three characteristic frequency bands, two from rainforest and one from ecotone forest. RF4: frequency range 4 from 3.0 to 4.0 kHz, including the typical rainforest noise band at 3.5 kHz; RF6: frequency range 6 from 5.0 to 6.0 kHz, including the double band in the rainforest which is moving upward in frequency over the morning; and GF8: frequency range 8 from 7.0 to 8.0 kHz, including the high-frequency cicadas of the ecotone forests.

periods to the continuous background noise. Typical examples of such species in the rainforest are the yellow-rumped tinkerbird, *Pogoniulus bilineatus* [repeating monotonous series of tones of about 1.0 kHz, see Fig. 2(B), 15.00] and the hairy-breasted barbet, *Tricolaema hirsuta* (repeating monotonous tones of about 500 Hz, at a slower pace than *P. bilineatus*). In the ecotone sites, continuous sounds are produced by the yellow-fronted tinkerbird, *Pogoniulus chrysoconus* (repeating monotonous series of tones of about 1.0 kHz, in a different temporal pattern than *P. bilineatus*), and various species of *Streptopelia* doves, *S. semitorquata*, *S. senegalensis*, and *S. vinacea*, that are abundant in adjacent savanna habitat (different temporal patterns of monotonous cooing of about 500 to 900 Hz (Slabbekoorn *et al.*, 1999), see Fig. 2(A), 14.00 and 16.00).

C. Spectral profiles

In the rainforest sites, the typical spectral profile is characterized by several discrete amplitude peaks in the frequency range from 3.0 to 7.0 kHz. This pattern results in relatively low energy levels in the low-frequency range, with increasing amplitude levels going up in frequency until 7.0 kHz (Fig. 3, lower part). In the ecotone forest sites, the typical spectral profile is characterized by relatively equal noise levels throughout the spectrum and a strong but variable amplitude peak between 6.0 and 8.0 kHz (Fig. 3, upper part). Four of the eight frequency ranges have significantly different amplitude levels for the two habitat types ($P<0.001$, Kruskal Wallis test; see Fig. 3).

I selected three characteristic frequency bands for a closer look at variation over the day in relative amplitude (Fig. 4). RF4: the frequency range in the rainforest in which the characteristic noise band at 3.5 kHz appeared consistently throughout the day was without significant amplitude variation (ANOVA: $F=1.27$, $P=$NS). RF6: the frequency range in the rainforest in which the double band appeared that is moving upward in frequency over the morning from 5.0 to 6.0 kHz. This noise band dropped dramatically in amplitude from 7.00 to 10.00 to rise again before dusk at 16.00 and 17.00 (ANOVA: $F=7.68$, $P<0.001$). GF8: the frequency range in the ecotone recordings in which presumably cicadas produce loud sounds in a wide band around 7.0 kHz. This noise band became louder throughout the day and showed peak levels in the late afternoon, although statistical tests indicated no significant differences among sampling times (ANOVA: $F=0.634$, $P=$NS), nor a significant increase over the day (linear regression: $F=0.49$, $P=$NS).

In addition to the habitat comparisons reported above, I analyzed noise variation among sites without *a priori* categorizing sites into habitat types. I used a principal component analysis (PCA) based on the averaged rms values for the eight frequency ranges. The analysis produced three components with distinctive factor loadings for the eight frequency ranges (Table II). Especially the first component, accounting

TABLE II. Factor loadings, eigenvalues, and % of variance explained for three principal components based on the pattern of noise energy distribution among eight 1.0-kHz frequency ranges. High loadings are indicated in bold for each PC (PC1 and PC2 are used in Fig. 5).

Frequency range (kHz)		PC1	PC2	PC3
(1)	0–1.0	−0.148	−0.171	−0.659
(2)	1.0–2.0	0.576	−0.034	**0.731**
(3)	2.0–3.0	0.571	−0.536	−0.280
(4)	3.0–4.0	**0.728**	−0.320	−0.236
(5)	4.0–5.0	**0.917**	−0.151	−0.218
(6)	5.0–6.0	**0.865**	0.316	0.254
(7)	6.0–7.0	0.469	**0.826**	−0.040
(8)	7.0–8.0	−0.032	**0.874**	−0.250
Eigenvalue		0.021	1.988	1.279
Variance explained		37.8%	24.8%	16.0%

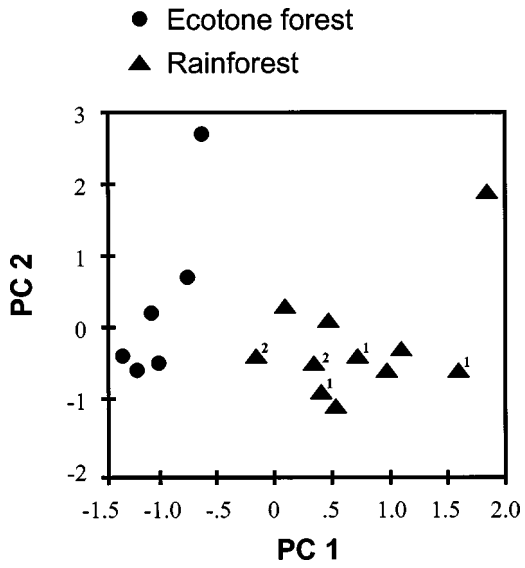


FIG. 5. Acoustic variation in noise characteristics among sites, as reflected by variation in the first two principal components (PC1 and PC2), based on the relative amplitude levels in eight 1.0-kHz intervals. The rainforest site labeled with a 1 is sampled three times (DJA), twice in the dry season and once in the rainy season, and the one labeled with a 2 is sampled twice (KRB), once in both seasons. PC1 accounts for 37.8% of the variation, PC2 for 24.8% (see Table II for factor loadings).

for 38% of the variation and reflecting primarily variation in the frequency range between 3.0 to 6.0 kHz, separates the samples clearly according to habitat. Figure 5 depicts all sites in the multidimensional space of the first two principal components. Different measurements in the same rainforest sites in different seasons show some variation, but fall out within the space typical for ''rainforest-type'' ambient noise.

IV. DISCUSSION

A. Habitat-dependent noise characteristics

The ambient noise characteristics of rainforest and ecotone forest show significant and consistent differences. The recordings from contiguous rainforest are characterized by many distinctive frequency bands not found in the ecotone, which lead to a general increase in amplitude from low to high frequencies. The recordings from ecotone forest are characterized by amplitude levels that are relatively equal throughout the frequency range, while a distinctive loud high-frequency band, presumably produced by some species of cicada, was present at three of the six ecotone sites in some periods of the day. The two most likely explanations for the differences in noise patterns between the two forest types are a different impact of wind-generated noise, and a different sound-producing animal community. The relatively closed rainforest sites are less affected by wind than the ecotone forest sites, which are mostly gallery forests forming elongated forest patches along streams in an open savanna landscape. Although few of the species responsible for the generated sounds were identified, it seems likely that communities of insects, frogs, and birds will differ considerably in two habitat types that differ in many abiotic factors (Letouzy, 1968; Smith, 1990; Basset *et al.*, 1992). The identified large *Ugada* cicadas give a clear example of how insect species, which contribute significantly to the ambient noise profile, can have a distribution closely linked to habitat types. The presence of such a species has a marked effect on making rainforest noise distinct from ecotone forest noise, and additionally, may provide distinct noise profiles for coastal and inland rainforest.

B. Cross-continent comparison

The most extensive study to date on habitat-dependent ambient noise was conducted in east Africa (Waser and Brown, 1986), comparing rainforest, riverine forest, and savanna habitat. The first two habitats are very similar to the central African forest types that I examined in this study, the forest labeled as ''riverine'' in east Africa being comparable to ecotone forest in Cameroon. Generally, the results from both studies reflect very similar patterns. Rainforest recordings in east Africa also showed relatively low levels of low-frequency noise, with increasing amplitude levels in particular between 2.0 and 4.0 kHz. Insect-generated sounds in the riverine forest were in a higher frequency range (between 4.0 and 8.0 kHz) compared to the rainforest sites, concordant with the noise amplitude peak between 6.0 and 8.0 kHz in some of the ecotone sites in this study. Wind and moving vegetation were reported to be important factors leading to increased noise levels in the riverine forest, especially in the early afternoon. The data set from Cameroon shows a diurnal pattern consistent with gradually increasing wind levels over the morning, leading to most wind-generated noise between 11:00 and 14:00. The strikingly similar patterns found in both studies suggest that the consistency in noise characteristics related to habitat type may apply over even broader scales than the 400 km between the two most distant sites within the same habitat in the present study. The spectral profiles at the two rainforest sites that were sampled in different seasons showed some variation, but patterns remained distinct from ecotone noise patterns. A study on two Ecuadorian forest types also revealed larger habitat differences than seasonal variation in spectral profiles despite seasonal variation in overall amplitude (De la Torre and Snowdon, 2002).

C. Temperature-dependent frequency shifts

An intriguing pattern that appeared on the sonagrams, of the rainforest sites in particular, is the stepwise increase in frequency of distinctive noise bands for subsequent sampling times over the morning. This pattern corresponds to an increase in temperature that starts a few hours after sunrise: forest temperatures rise most quickly from 8:00 to 12:00, while temperature starts to decline gradually again after 15:00 or 16:00 (east Africa: Haddow and Corbet, 1960; west Africa: Cachan, 1963, both in Waser and Waser, 1977). Therefore, a likely explanation for this phenomenon seems to be that the distinctive noise bands are produced by particular insect species, and that the frequency of their calls increases with the rise in temperature in the forest. The positive relationship between call frequency and temperature is relatively well documented for insects (e.g., Walker, 1962; Sismondo, 1979; Desutter-Grandcolas, 1998), while this relationship has

not been reported for birds, and is known to be of limited magnitude for frogs (e.g., Gerhardt and Mudry, 1980). The temperature-dependent patterns were prominent in the rainforest sites, but hardly present in any of the ecotone sites. This difference could be due to an absence of insect species in the ecotone producing distinctive noise bands for long periods of time. Consequently, the temperature increase over the morning would not be reflected by the stepwise increase of the continued presence of any particular noise band. Temperature-dependent noise bands could mask acoustic signals of animals that are not temperature-dependent themselves in producing signals. This means that temperature variation could also play a role in signal evolution of species which are traditionally thought to signal independently of for example diurnal or seasonal cycling between warm and cold.

D. Ambient noise and signal evolution

High background noise levels are known to affect signal efficiency in a variety of species, including humans (see, e.g., Pisoni, 1996; Fu *et al.*, 1998). More noise leads to a decrease in the distance at which the signal-to-noise ratio is still sufficient for conspecifics to detect and recognize the signal. The ambient noise level has been reported to affect the ''active space'' in birds (Brenowitz, 1982b) and primates (Brown, 1989), and several studies report on noise affecting the mate attraction function in frog calls (e.g., Gerhardt and Klump, 1988; Schwartz, 1993, Wollerman, 1999). Generally, the level of signal interference depends on the frequency matching between the signal and the background noise (Klump, 1996; Langeman *et al.*, 1998; Lohr *et al.*, 2003; Dooling, 2004). As a consequence, different noise patterns will have a different impact on the perception of a particular signal (Brown and Sinnott, 1990; Lohr *et al.*, 2003).

Acoustic signals that are predominantly used in a particular habitat are likely to be shaped by the ambient noise characteristics of that habitat. Absolute amplitude levels may affect the strength of ambient noise as a selection pressure for acoustic signals. However, the sound energy distribution throughout the frequency spectrum determines whether ambient noise leads to directional selection. Differences in amplitude levels for different frequency ranges between habitat types may therefore indicate divergent selection pressures on acoustic signals (Slabbekoorn and Smith, 2002a; 2002b).

Brenowitz (1982b) already emphasized the potential relationship between evolutionary signal changes and noise characteristics, and he showed how the frequency use of the red-winged blackbird (*Agelaius phoeniceus*) matches a ''noise window,'' a restricted frequency range within which amplitude levels of ambient noise are relatively low. Similarly, a study on vocalizations of pygmy marmosets (*Cebuella pygmaea*) and their relationship with ambient noise revealed that spectral call characteristics matched relatively quiet regions in the ambient noise in two different tropical forest types in South America (De la Torre and Snowdon, 2002). The matching of vocal frequencies of individual species to ambient noise profiles may have contributed to the typical habitat-dependent acoustic features found in comparative studies across species (Chappuis, 1971; Morton, 1975; Ryan and Brenowitz, 1985; Slabbekoorn, 2004).

It is not only at the species level that masking noise may play a role in general signal convergence within habitat and signal divergence between habitats. For example, Slabbekoorn and Smith (2002a) found evidence for habitat-associated song divergence among populations within a passerine species, the little greenbul (*Andropadus virens*), and part of the song divergence could be explained by habitat-dependent noise characteristics. Greenbul songs in the rainforest are distinct from songs in the ecotone by having a lower minimal frequency, a frequency range coinciding with relatively low noise levels, as outlined in detail in the current paper. A study on urban great tits (*Parus major*) even showed that song characteristics of individual birds may be correlated to local noise conditions in their individual territories (Slabbekoorn and Peet, 2003). Male tits with noisy territories had higher minimum song frequencies leading to less masking by the predominantly low-frequency traffic noise. Habitat-dependent signal divergence within species driven by variation in ambient noise profiles could potentially affect reproductive divergence among populations and thereby influence speciation (Slabbekoorn and Smith 2002b).

E. Conclusion

The present study clearly shows that noise energy present may vary considerably throughout the spectrum, and that spectral profiles can be significantly and consistently different depending on the habitat. Furthermore, the findings in the literature on a match between animal signals and relatively silent frequency ranges in the ambient noise of their typical habitat support the hypothesis that ambient noise may drive signal evolution. It appears that masking noise may affect song characteristics of individual animals, change signal features of whole populations, and shape species-specific songs and calls over evolutionary time. Given that natural habitats are often very noisy and that anthropogenic noise is becoming more and more dominant in both urban and natural areas, ambient noise should certainly not be treated as ''just noise'' in the context of ecology and evolution of animal signals.

ACKNOWLEDGMENTS

I thank the government of the republic of Cameroon for permission to conduct the field work, ECOFAC for the collaboration at the Réserve de Dja, and the Wildlife Conservation Society for hospitality and logistic support during my stay in Cameroon. I am grateful to Luis Baptista and the California Academy of Sciences for the use of their bioacoustic laboratory. TDK Recording Media Europe S.A. donated the supply of audio tapes. Eliot Brenowitz, Sonya Clegg, Kristen Ruegg, and Tom Smith made helpful comments on earlier versions of the manuscript. The work was funded by the Netherlands Organization for Scientific Research (NWO, Grant S 84-467), the Society for the advancement of research in the tropics (TREUB), the National Science Foundation (Grants DEB-9726425 and IRCEB-0077072 to T. B. Smith), and the National Geographic Society.

Basset, Y., Aberlenc, H.-P., and Delvore, G. (**1992**). ''Abundance and stratification of foliage arthropods in a lowland rain forest of Cameroon,'' Ecol. Entomol. **17**, 310–318.
Bennet-Clark, H. C. (**1970**). ''The mechanism and efficiency of sound production in mole crickets,'' J. Exp. Biol. **52**, 619–652.
Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer, Sunderland).
Brenowitz, E. A. (**1982a**). ''The active space of red-winged blackbird song,'' J. Comp. Physiol. [A] **147**, 11–522.
Brenowitz, E. A. (**1982b**). ''Long-range communication of species identity by song in the red-winged blackbird,'' Behav. Ecol. Sociobiol. **10**, 29–38.
Brown, C. H. (**1989**). ''The active space of the blue monkey and grey-cheeked mangabey vocalizations,'' Anim. Behav. **37**, 1023–1034.
Brown, C. H., and Sinnott, J. M. (**1990**). ''The perception of complex acoustic patterns in noise by blue monkey (*Cercopithecus mitis*) and human listeners,'' Int. J. Comp. Psychol. **4**, 79–90.
Cachan, P. (**1963**). ''Signification ecologique des variations microclimatiques dans la forêt sempervirente de basse Côte d'Ivoire,'' Ann. Fac. Sciences Dakar **8**, 89–155.
Catchpole, C. K. (**1987**). ''Bird song, sexual selection and female choice,'' Trends Ecol. Evol. **2**, 94–97.
Chappuis, C. (**1971**). ''Un example de l'influence du milieu sur les emissions vocales des oiseaux: l'evolution des chants en forêt equatoriale,'' Terre et Vie **118**, 183–202.
De la Torre, S., and Snowdon, C. T. (**2002**). ''Environmental correlates of vocal communication of wild pygmy marmosets, *Cebuella pygmaea*,'' Anim. Behav. **63**, 847–856.
Desutter-Grandcolas, L. (**1998**). ''Broad-frequency modulation in cricket (*Orthoptera, Grylloidea*) calling songs: Two convergent cases and a functional hypothesis,'' Can. J. Zool. **76**, 2148–2163.
Dooling, R. J. (**2004**). ''Audition: Can birds hear everything they sing?'' in *Nature's Music: The Science of Birdsong*, edited by P. Marler and H. Slabbekoorn (Elsevier-Academic, San Diego), pp. 206–225.
Duffels, J. P., and Van der Laan, P. A. (**1985**). *Catalogue of the Cicadoidea (Homoptera, Auchenorhyncha) 1956–1980* (Dr. W. Junk, Dordrecht).
Dusenbery, D. B. (**1992**). *Sensory Ecology. How Organisms Acquire and Respond to Information* (Freeman, New York).
Endler, J. (**1992**). ''Signals, signal conditions, and the direction of evolution,'' Am. Nat. **139**, 125–153.
Forrest, T. G. (**1994**). ''From sender to receiver: propagation and environmental effects on acoustic signals,'' Am. Zool. **34**, 644–654.
Fu, Q.-J., Shannon, R. V., and Wang, X. (**1998**). ''Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,'' J. Acoust. Soc. Am. **104**, 3586–3596.
Gerhardt, H. C. (**1994**). ''The evolution of vocalizations in frogs and toads,'' Annu. Rev. Ecol. Syst. **25**, 293–324.
Gerhardt, H. C., and Klump, G. M. (**1988**). ''Masking of acoustic signals by the chorus background noise in the green treefrog: A limitation on mate choice,'' Anim. Behav. **36**, 1247–1249.
Gerhardt, H. C., and Mudry, K. M. (**1980**). ''Temperature effects on frequency preferences and mating call frequencies in the green treefrog, *Hyla cinerea* (Anura: Hylidae),'' J. Comp. Physiol. **137**, 1–6.
Haddow, A. J., and Corbet, P. S. (**1960**). ''Entomological studies from a high tower in Mpanga forest, Uganda. II. Observations on certain environmental factors at different levels,'' Trans. R. Ent. Soc. Lond. **113**, 249–269.
Katti, M., and Warren, P. S. (**2004**). ''Tits, noise and urban bioacoustics,'' Trends Ecol. Evol. **19**, 109–110.
Klump, G. M. (**1996**). ''Bird communication in the noisy world,'' in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University, New York), pp. 321–338.
Klump, G. M., and Langemann, U. (**1995**). ''Comodulation masking release in a song bird,'' Hear. Res. **87**, 157–164.
Langemann, U., Gauger, B., and Klump, G. M. (**1998**). ''Auditory sensitivity in the great tit: Perception of signals in the presence and absence of noise,'' Anim. Behav. **56**, 763–769.
Letouzy, R. (**1968**). *Etude Phytogéographique du Cameroun* (Edition Paul Lechevalier, Paris).
Lohr, B., Wright, T. F., and Dooling, R. J. (**2003**). ''Detection and discrimination of natural calls in masking noise by birds: Estimating the active space of a signal,'' Anim. Behav. **65**, 763–777.
Longman, K. A., and Jenik, J. (**1992**). ''Forest-savanna boundaries: General considerations,'' in *Nature and Dynamics of Forest-savanna Boundaries*, edited by P. A. Furley, J. Proctor, and J. A. Ratter (Chapman and Hall, New York), pp. 3–20.
Michelsen, A., and Nocke, H. (**1974**). ''Biophysical aspects of sound communication in insects,'' Adv. Insect Physiol. **10**, 247–296.
Morton, E. S. (**1975**). ''Ecological sources of selection on avian sounds,'' Am. Nat. **109**, 17–34.
Nieder, A., and Klump, G. M. (**1999**). ''Adjustable frequency selectivity of auditory forebrain neurons recorded in a freely moving songbird via radiotelemetry,'' Hear. Res. **127**, 41–54.
Pisoni, D. B. (**1996**). ''Word identification in noise,'' Lang. Cogn. Proc. **11**, 681–687.
Ryan, M. J., and Brenowitz, E. A. (**1985**). ''The role of body size, phylogeny, and ambient noise in the evolution of bird song,'' Am. Nat. **126**, 87–100.
Schwartz, J. J. (**1993**). ''Male calling behavior, female discrimination and acoustic interference in the neotropical treefrog *Hyla microcephala* under realistic acoustic conditions,'' Behav. Ecol. Sociobiol. **32**, 401–414.
Searcy, W. A., and Anderson, M. (**1986**). ''Sexual selection and the evolution of song,'' Annu. Rev. Ecol. Syst. **17**, 507–533.
Searcy, W. A., and Yasukawa, K. (**1996**). ''Song and female choice,'' in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell University Press, New York), pp. 454–473.
Sismondo, E. (**1979**). ''Stridulation and tegminal resonance in the tree cricket *Oecanthus nigricornis* (*Orthoptera: Gryllidae: Oecanthinae*),'' J. Comp. Physiol. [A] **129**, 269–279.
Slabbekoorn, H. (**2004**). ''Singing in the wild: The ecology of birdsong,'' in *Nature's Music: The Science of Birdsong*, edited by P. Marler and H. Slabbekoorn (Elsevier-Academic, San Diego), pp. 178–205.
Slabbekoorn, H., and Peet, M. (**2003**). ''Birds sing at a higher pitch in urban noise,'' Nature (London) **424**, 267.
Slabbekoorn, H., and Smith, T. B. (**2002a**). ''Habitat-dependent song divergence in the little greenbul: An analysis of environmental selection pressures on acoustic signals,'' Evolution (Lawrence, Kans.) **56**, 1849–1858.
Slabbekoorn, H., and Smith, T. B. (**2002b**). ''Bird song, ecology, and speciation,'' Philos. Trans. R. Soc. London, Ser. B **357**, 493–503.
Slabbekoorn, H., de Kort, S., and ten Cate, C. (**1999**). ''Comparative analysis of perch coo vocalizations of the *Streptopelia* doves,'' Auk **116**, 737–748.
Smith, T. B. (**1990**). ''Patterns of morphological and geographic variation in trophic bill morphs of the African finch *Pyrenestes*,'' Biol. J. Linn. Soc. **41**, 381–414.
Tubaro, P. L., and Mahler, B. (**1998**). ''Acoustic frequencies and body mass in New World Doves,'' Condor **100**, 54–61.
Walker, T. J. (**1962**). ''Factors responsible for intraspecific variation in calling songs of crickets,'' Evolution (Lawrence, Kans.) **16**, 407–428.
Wallschläger, D. (**1980**). ''Correlation of song frequency and body weight in passerine birds,'' Experientia **36**, 412.
Waser, P. M., and Brown, C. H. (**1986**). ''Habitat acoustics and primate communication,'' Am. J. Primatol **10**, 135–154.
Waser, P. M., and Waser, M. S. (**1977**). ''Experimental studies of primate vocalization: Specializations for long-distance propagation,'' Z. Tierpsychol **43**, 239–263.
Wiley, R. H., and Richards, D. G. (**1982**). ''Adaptations for acoustic communication in birds: Sound transmission and signal detection,'' in *Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Academic, New York), pp. 131–181.
Wollerman, L. (**1999**). ''Acoustic interference limits call detection in a Neotropical frog *Hyla ebraccata*,'' Anim. Behav. **57**, 529–536.

Measurement of viscoelastic properties of homogeneous soft solid using transient elastography: An inverse problem approach

S. Catheline,[a] J.-L. Gennisson, G. Delon, M. Fink, R. Sinkus,[b] S. Abouelkaram,[c] and J. Culioli[c]

Laboratoire Ondes et Acoustique, ESPCI, Université Paris VII, U.M.R. C.N.R.S. 7587, 10 rue Vauquelin, 75231 Paris cedex 05, France

(Received 2 February 2004; revised 1 September 2004; accepted 20 September 2004)

Two main questions are at the center of this paper. The first one concerns the choice of a rheological model in the frequency range of transient elastography, sonoelasticity or NMR elastography for soft solids (20–1000 Hz). Transient elastography experiments based on plane shear waves that propagate in an Agar-gelatin phantom or in bovine muscles enable one to quantify their viscoelastic properties. The comparison of these experimental results to the prediction of the two simplest rheological models indicate clearly that Voigt's model is the better. The second question studied in the paper deals with the feasibility of quantitative viscosity mapping using inverse problem algorithm. In the ideal situation where plane shear waves propagate in a sample, a simple inverse problem based on the Helmholtz equation correctly retrieves both elasticity and viscosity. In a more realistic situation with nonplane shear waves, this simple approach fails. Nevertheless, it is shown that quantitative viscosity mapping is still possible if one uses an appropriate inverse problem that fully takes into account diffraction in solids. [DOI: 10.1121/1.1815075]

PACS numbers: 43.80.Vj, 43.80.Ev, 43.20.Qf [Fd] Pages: 3734–3741

I. INTRODUCTION

Elasticity mapping is an effective method for detecting tumors in soft tissues. Images of elasticity can be achieved with static elastography[1] or dynamic elastography (sonoelasticity,[2,3] NMR elastography,[4,5] or transient elastography[6,7]). The dynamic methods present the advantage of potentially revealing the medium dynamic properties[8,9] such as viscosity. In the simple case of a plane monochromatic shear wave for example, the speed and attenuation are deduced from the phase and amplitude measurements, which can be converted into elasticity and viscosity. This last computational step requires choosing an appropriate rheological model. The simplest viscoelastic models (utilizing two parameters) are Voigt's model and Maxwell's model. Although the first one is more often used for solids and the second one for liquids, the choice is not obvious in liquid-like solids such as soft tissues since both models are used in the literature.[10–12] Moreover, in the dynamic elastography frequency range (20–1000 Hz), no commercially available rheometer exists (typically 0.01–10 Hz). Homemade rheometers moreover tend to present resonance and propagation difficulties.[13,14] This paper may contribute to fill this gap in experimental data.

II. EXPERIMENT

In our experimental study, we used an ultrafast ultrasonic scanner.[15,16] It provides images of the echogenecity of tissues similar to a standard echographic device but with a rate of 5000 frames/s (about 100 times higher than conventional scanners). It includes a medical ultrasonic array with 128 channels working at 5 MHz. Each channel is connected to a 2 Mbytes memory, the echoes are sampled at 50 MHz and digitized with 9 bit resolution. In a typical experiment, 250 echographic images are recorded in memory with a cadence of 3000 frames/s. A movie of the displacement is obtained using correlation algorithms between successive speckle images. Thus, it allows one to detect fast tissue motion induced by low frequency shear waves inside the medium. The apparatus can measure displacements as small[17] as 1 μm. As shown in Fig. 1, the low frequency (50–500 Hz) shear wave is generated by transversally shaking a rigid aluminum plate ($110\times170\ \mathrm{mm}^2$) applied on one side of the phantom with a vibrator (Bruel&Kjaer 4810 type). The phantom is a mixture of water with 3% gelatin and 2% agar. The transverse displacement field of the shear wave is measured in a plane perpendicular to the rigid plate ($40\times40\ \mathrm{mm}^2$). The displacement field, displayed in Fig. 2, confirms the plane nature of the shear wave. The 100 Hz shear wave propagates from bottom to top in the four sequential images selected from the displacement field movie. The use of such plane waves simplifies the analysis of the viscoelastic properties of soft solids since viscoelastic theories often utilize this assumption. From the displacement field along the axis of the rigid plate ($z=20$ mm), Fig. 3(a), the phase and the amplitude as a function of depth are computed, Figs. 3(b) and 3(c). The speed, $C_T=2.26\pm0.003\ \mathrm{m\,s}^{-1}$, and the attenuation, $\alpha_T=23\pm1\ \mathrm{m}^{-1}$, are then deduced. The experimental errors are estimated from the linear fit deviations. We point out that the estimate of the phase slope as a function of depth is less subjected to noise than the amplitude slope. Consequently, attenuation measurements may be difficult in noisy situations. The experiment is repeated for frequencies rang-

[a]Electronic mail: stefan.catheline@espci.fr
[b]Also at: Philips Research Laboratories, Division Technical System, Röntgenstrasse 24-26, D-22335, Hamburg, Germany.
[c]Also at: Station de Recherches sur la Viande, INRA, Theix, 63122 St Genès Champanelle, France.

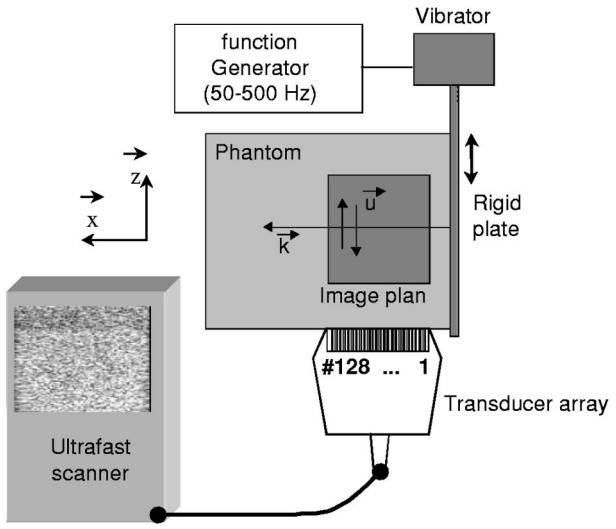


FIG. 1. Experimental setup. A low frequency plane shear wave is generated in an Agar-gelatin phantom by shaking a rigid plate with a vibrator. The wave propagates along the direction **k**. In the meantime, a medical transducer array, connected to an ultrafast scanner, insonifies the medium on a 40×40 mm^2 area (gray area). Then a correlation algorithm, on speckle images stored in memory, computes the displacement field **u** within the phantom.

ing from 50 to 500 Hz with a 25 Hz step. The shear wave measurements as a function of the central frequency, Fig. 4 (circles), show a rather flat speed distribution and a strongly varying attenuation distribution (from 5 to 80 m^{-1}). The maximum error occurs at 500 Hz, because the amplitude of the shear wave drops with the increase in frequency: $\Delta C_T = 0.007$ m s^{-1} and $\Delta \alpha_T = 4$ m^{-1}. This reliable experimental data will be the reference measurements in the following.

III. THE INVERSE PROBLEM APPROACH

The elasticity mapping in transient elastography is performed with some inverse problem algorithm. We propose

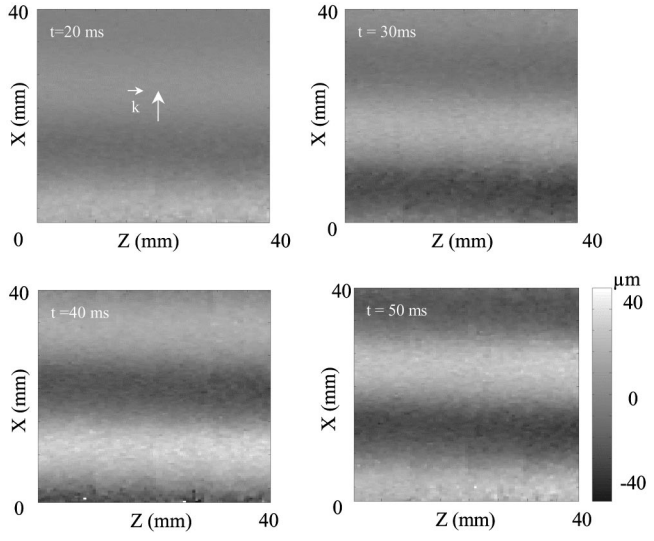


FIG. 2. Typical experimental transverse displacement field (z component) of a 100 Hz shear wave. These images are extracted from a movie at different acquisition times 20, 30, 40, 50 ms after a few cycles of sinusoid are generated from the bottom of the images. The arrow indicates the direction of propagation **k**. The plane property of the shear wave is clearly visible: it is a transverse wave.

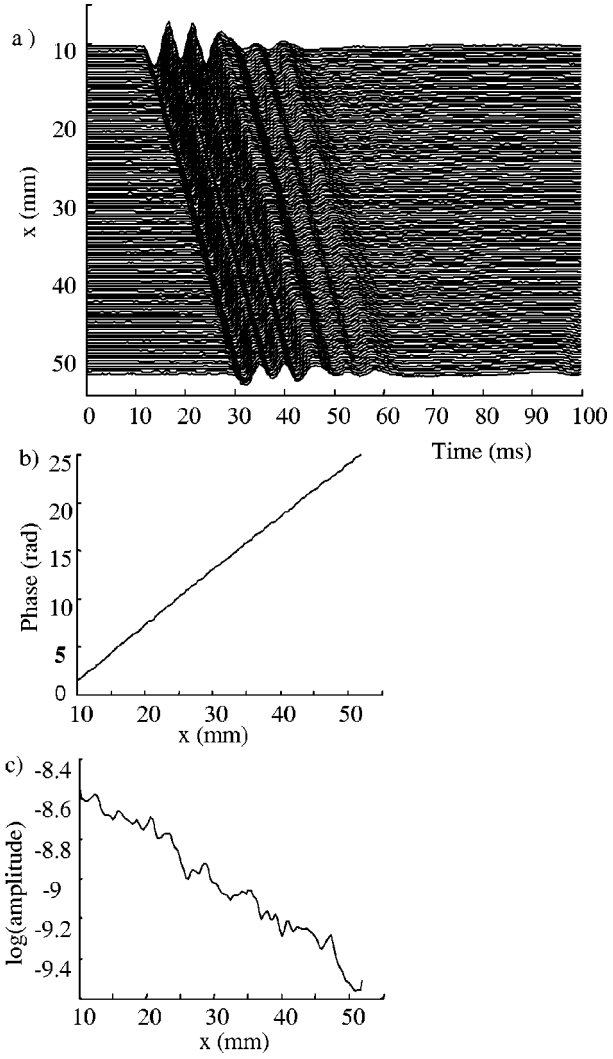


FIG. 3. (a) Seismic representation of the transverse displacement field. (b) The phase and (c) the amplitude are computed from a spectral analysis at the central frequency 100 Hz. Linear fits give the speed ($C_T = 2.265 \pm 0.003$ m s^{-1}) and the attenuation ($\alpha_T = 23 \pm 1$ m^{-1}).

here a very simple one to measure both the local elasticity and viscosity. In the simple case of a plane monochromatic shear wave, the starting point is the one-dimensional (1D) Helmholtz equation:

$$\frac{\partial^2 \mathrm{FT}_t(u_z(\mathrm{x}))}{\partial x^2} + k^2 \mathrm{FT}_t(u_z(\mathrm{x})) = 0, \tag{1}$$

where FT_t, u_z, and x stand for the time Fourier transform, the transverse displacement, and the longitudinal coordinate, respectively. Thus the local complex wave vector is given by

$$k = \sqrt{\frac{\partial^2 \mathrm{FT}_t(u_z(\mathrm{x}))}{\partial x^2} \Big/ \mathrm{FT}_t(u_z(\mathrm{x}))}, \tag{2}$$

and the local speed C_T and attenuation α_T are expressed as

$$C_T = \frac{\omega}{\mathrm{Re}\{k\}}, \quad \alpha_T = \mathrm{Im}\{k\}. \tag{3}$$

Thus, the latter set of equations constitute the basis of an algorithm that permits one to compute the local speed and

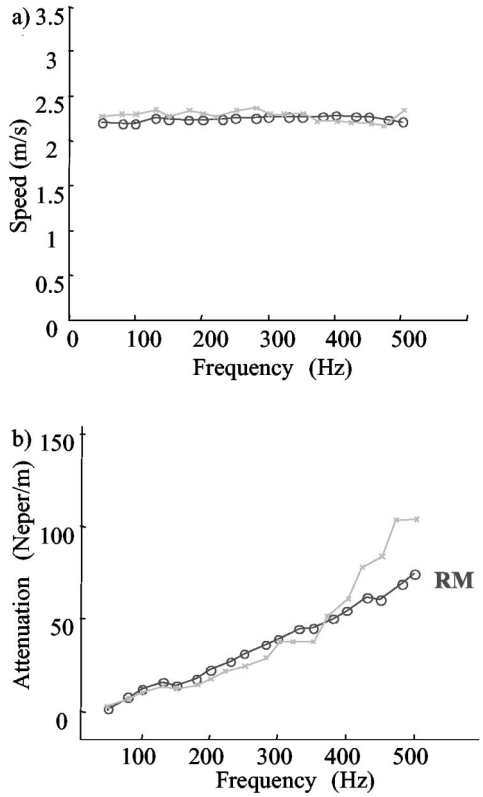


FIG. 4. Speed and attenuation measurements of the transverse wave as a function of the frequency. (a) The inverse problem algorithm gives a good estimation of speed (crosses) compared to the reference measurement obtained with a spectral analysis of the displacement field method (more reliable but nonlocal). (b) However the inverse problem approach for attenuation measurements seems to be more sensitive to noise since a deviation from the reference measurements (circles) is visible beyond 400 Hz.

the local attenuation from the local displacement field $u_z(x)$. Its application to a homogeneous Agar-gelatin based phantom serves to test the feasibility of the technique in a very simple configuration. The results obtained from the inverse problem algorithm and those obtained from the phase and amplitude, Fig. 4 (cross), are very close: within 5% for the speed and within 11% for the attenuation below 400 Hz. The deviation in the reconstructed values of attenuation above 400 Hz is due to a poor signal-to-noise ratio in displacement. Thus, not only is the speed correctly recovered using the inverse problem algorithm, which has already been demonstrated,[18] but also attenuation even though it is more sensitive to noise. As a consequence, viscosity mapping should be possible if a suitable rheological model is chosen.

IV. VOIGT'S VERSUS MAXWELL'S MODEL

At a given frequency, two experimental measurements of phase and amplitude allow one to compute the local speed and attenuation, from which the elasticity and viscosity estimates are deduced. This last step requires a rheological model. These viscoelastic models, which require only two parameters, are Voigt (V) and Maxwell's (M) models, Fig. 5. Both models are composed of a dashpot and a spring, but they are placed in parallel (V) or in series (M) depending on the model. The stress–strain relationships are, respectively,

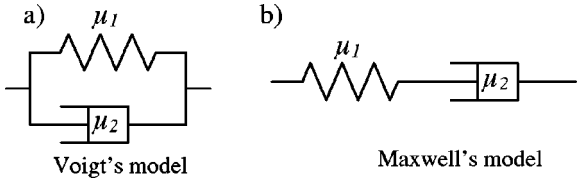


FIG. 5. (a) The Voigt's model is composed of a spring, with spring constant μ_1, and a dashpot, with coefficient of viscosity μ_2, in parallel. (b) The Maxwell's model is a spring and a dashpot in series.

$$\sigma=\left(\mu_1-\mu_2\frac{\partial}{\partial t}\right)\varepsilon, \quad \left(\mu_1+\mu_2\frac{\partial}{\partial t}\right)\sigma=\mu_1\mu_2\frac{\partial\varepsilon}{\partial t}. \tag{4}$$

The stress σ is linked to the strain ε via the shear elasticity μ_1, the shear viscosity μ_2, and the time-derivative operator $\partial/\partial t$. Inserting these relations for monochromatic excitation into the equation of motion, one obtains the 1D Helmholtz equation for Voigt's model and Maxwell's model, respectively,

$$\begin{aligned}&\frac{\partial^2U}{\partial x^2}+\frac{\rho\omega^2}{(\mu_1+i\omega\mu_2)}U=0,\\&\frac{\partial^2U}{\partial x^2}+\frac{\rho\omega^2(\mu_1+i\omega\mu_2)}{i\omega\mu_1\mu_2}U=0,\end{aligned} \tag{5}$$

where ρ is the density and U the Fourier transform of the displacement. Then, from the complex wave vector, the transverse wave speed (C_T^V, C_T^M) as well as the attenuation (α_T^V, α_T^M) can be computed for each model:

$$\begin{aligned}C_T^V&=\sqrt{\frac{2(\mu_1^2+\omega^2\mu_2^2)}{\rho(\mu_1+\sqrt{\mu_1^2+\omega^2\mu_2^2})}},\\C_T^M&=\sqrt{\frac{2\mu_1}{\rho\left(1+\sqrt{1+\dfrac{\mu_1^2}{\omega^2\mu_2^2}}\right)}},\end{aligned} \tag{6}$$

$$\begin{aligned}\alpha_T^V&=\sqrt{\frac{\rho\omega^2(\sqrt{\mu_1^2+\omega^2\mu_2^2}-\mu_1)}{2(\mu_1^2+\omega^2\mu_2^2)}},\\\alpha_T^M&=\sqrt{\frac{\rho\omega^2\left(\sqrt{1+\dfrac{\mu_1^2}{\omega^2\mu_2^2}}-1\right)}{2\mu_1}}.\end{aligned} \tag{7}$$

In order to compare the theoretical predictions of Voigt and Maxwell's models to the experimental data, one frequency is chosen arbitrarily, for example, 400 Hz. With the measured values of Fig. 6, $C_T(400)=2.282\pm0.004\ \mathrm{m\,s^{-1}}$ and $\alpha_T(400)=54\pm1\ \mathrm{m^{-1}}$, it is possible to evaluate μ_1 and μ_2 using Eqs. (6) and (7). Voigt's model gives $\mu_1=5.67\pm0.02$ kPa and $\mu_2=0.22\pm0.01$ Pa s and Maxwell's model $\mu_1=5.73\pm0.02$ kPa and $\mu_2=23\pm1$ Pa s. Both models provide similar values for the elasticity. However, there is a difference of two orders of magnitude for the viscosity. In the literature, the viscosity for gelatin rarely exceeds 1 Pa s. For example, with Zener's model in gelatin at 1.5 Hz, Djabourov *et al.*[19] found 0.21 Pa s, very close to the value ob-

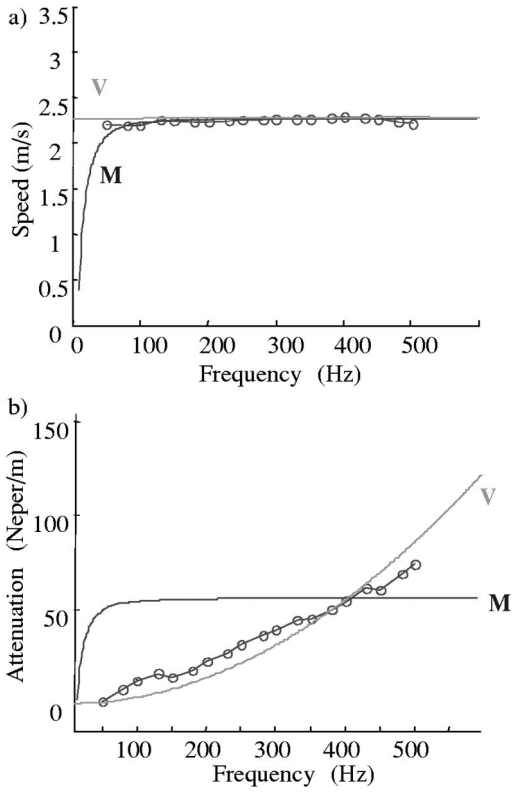


FIG. 6. Transverse wave speed and attenuation measurements vs frequency. (a) The flat speed distribution (circles) is correctly predicted by the Voigt's (*V*) and the Maxwell's model (*M*). (b) On the other hand, the Maxwell's model attenuation predictions cannot properly account for the experimental distribution (circles). Voigt's model is better.

tained with the Voigt's model. This result indicates that Voigt's model seems to be more appropriate for the description of viscous effects in case of the Agar-gelatin phantom. Now, a study of the prediction of each model for all the other frequencies reveals that both models are correct as far as the speed is concerned, Fig. 6(a). The dispersion curve of the speed, therefore, does not allow for differentiating between the two models. On the other hand, the curve of attenuation versus frequency indicates a dramatic difference between Voigt and Maxwell's models and thus allows one to distinguish the good model: Voigt's model. Although this model is not perfect, it gives much better predictions for attenuation than Maxwell's model. The same general behavior is observed independent of the frequency chosen to fit the experimental data: an increasing law as ω^2 for Voigt's model and a plateau for Maxwell's model. Indeed, for frequencies beyond 100 Hz, the dashpot becomes ''harder'' than the spring. When these two elements are placed in series, as in Maxwell's model Fig. 5(b), the viscoelastic properties are entirely driven by the spring. Thus the speed and attenuation no longer depend on frequency: they present a plateau. Now, no plateau is reached in the measured attenuation curve, Fig. 6(b). As a result, the conclusion so far is that for Agar-gelatin based phantom, the best of the simplest rheological models is Voigt's model.

V. EXCISED POSTRIGOR BEEF MUSCLE

The next step is to conduct the plane shear wave experiment in real soft tissues. We have chosen an *excised post rigor* beef muscle: the biceps *femoris*. Figure 7 shows, from right to left, the vibrator, the duralumin rigid plate applied on the back face of the muscle sample, and the transducer array applied on the left side of the sample. When shaken, the plate transmits low frequency vibrations as transverse waves with a polarization perpendicular to the fibers in this situation. The displacements are detected using the transducer array connected to the ultrafast scanner. The speed and the attenuation are represented in Fig. 8 as functions of frequency which ranges from 50 to 350 Hz. The reason why the fre-

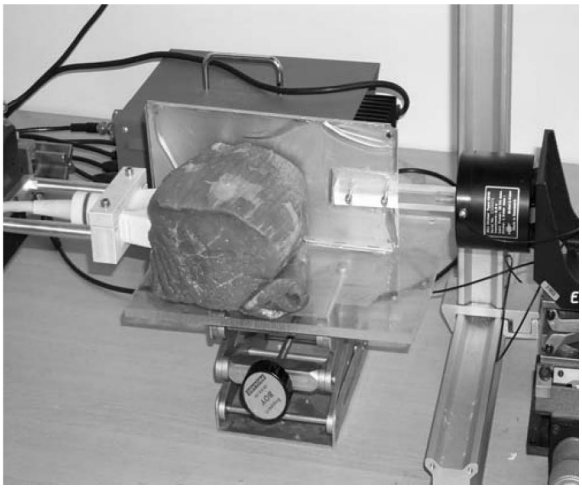


FIG. 7. Picture of the experimental setup in a beef muscle (biceps *femoris*). The vibrator transmits longitudinal oscillations to a rigid plate applied on the back side of the muscle sample. The transverse wave propagates in the direction of the fibers. The displacements are detected with a medical transducer array placed on the left-hand side of the muscle sample and connected to an ultrafast scanner.

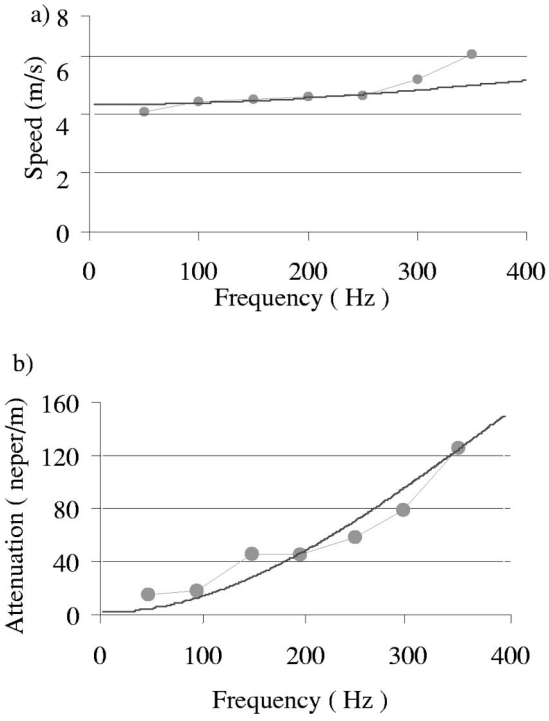


FIG. 8. Speed and attenuation vs frequency in a beef muscle. The transverse wave has a polarization perpendicular to the fibers. Theoretical predictions (solid lines) are based on a Voigt's model.

quency range ends at 350 Hz is that the attenuation ($\alpha \approx 80\ \mathrm{m}^{-1}$ at 300 Hz) is approximately twice as large as in the Agar-gelatin based phantom ($\alpha \approx 40\ \mathrm{m}^{-1}$ at 300 Hz) and thus the displacements are too small to be exploited until 500 Hz. Nevertheless, it is clear that these data are once again well described by Voigt's model since no plateau is present in the attenuation curve, Fig. 8(b). In addition, the biceps femoris muscle presents a transverse isotropy. General theoretical considerations on anisotropic viscoelastic media[20,21] show that two shear elasticities and two shear viscosities are needed to properly describe the transverse isotropy. It should be noted that these theories generalize Hooke's law by adding a Newtonian viscous term in the stress tensor T_{ij}. Thus, the implicit viscoelastic model is Voigt's solid,

$$T_{ij} = c_{ijkl} S_{kl} + \eta_{ijkl} \frac{\partial S_{kl}}{\partial t}, \tag{8}$$

with c_{ijkl} denoting Christophel's tensor, S_{kl} the strain tensor, and η_{ijkl} the viscosity tensor.

A slow and a fast wave are found depending on the polarization with respect to the fibers. Two elasticities and two viscosities are then deduced: $\mu_1^{\perp} = 25 \pm 1$ kPa, $\mu_2^{\perp} = 3.3 \pm 0.4$ Pa s, $\mu_1^{\parallel} = 49 \pm 2$ kPa, $\mu_2^{\parallel} = 15 \pm 2$ Pa s.

Thus the experimental measurements obtained with the plane shear wave experiment are in good agreement with previous work on muscles.[22,23] In other soft bovine tissues such as calfskin, liver, or cardiac muscle, Voigt's model was shown to give fair predictions of ultrasonic shear wave speed and attenuation.[24]

VI. THE 1D SHEAR ELASTICITY PROBE

The plane shear wave experiments presented in the first part of this paper can hardly be applied in real *in vivo* soft tissues. So what results can one expect from the inverse problem algorithm in a more realistic situation with nonplane shear waves?

We used the 1D shear elasticity probe.[25,26] As shown in Fig. 9(a), the shear elasticity probe is composed of a single ultrasonic transducer mounted on a vibrator. It has the advantage of using simple electronics while giving a quantitative global estimate of elasticity. It can potentially be applied to the diagnosis of liver diseases.[27] In the following experiments, the same phantom as in the first part was tested with the shear elasticity probe. The shear wave is generated by the front face of the transducer meanwhile working in a pulse echo mode. The central frequency of the 7-mm-diam transducer is 5 MHz. Its focal is 35 mm. A scans are recorded in a 9 bit digitizer at a sampling frequency of 50 MHz. The repetition frequency between successive A-scans is fixed in the experiments at 3000 Hz. The longitudinal component of the displacement along the ultrasonic beam is computed using correlation algorithms. The complex diffraction pattern of a low frequency vibrator,[28] assures that we are not working with plane shear waves. A typical experimental displacement field is shown in Fig. 9(b). When a 100 Hz central frequency pulse is applied, a fast compression wave (P), a slow shear wave (S), as well as a reflected shear wave from the bottom of the sample are visible. If the same algorithm used for plane shear waves is directly applied to the latter measurements, a very good speed estimate is obtained, Fig. 11(a). Nevertheless, the attenuation is overestimated, Fig. 11(b). The reasons for this apparent overattenuation must be found in the inverse problem algorithm. In order to carefully study the underlying approximations of the algorithm, one can start with the most general wave equation in an isotropic and linear solid, Eq. (8). For the sake of simplicity, the medium is considered as purely elastic. Adding viscosity does not change the essential argument:

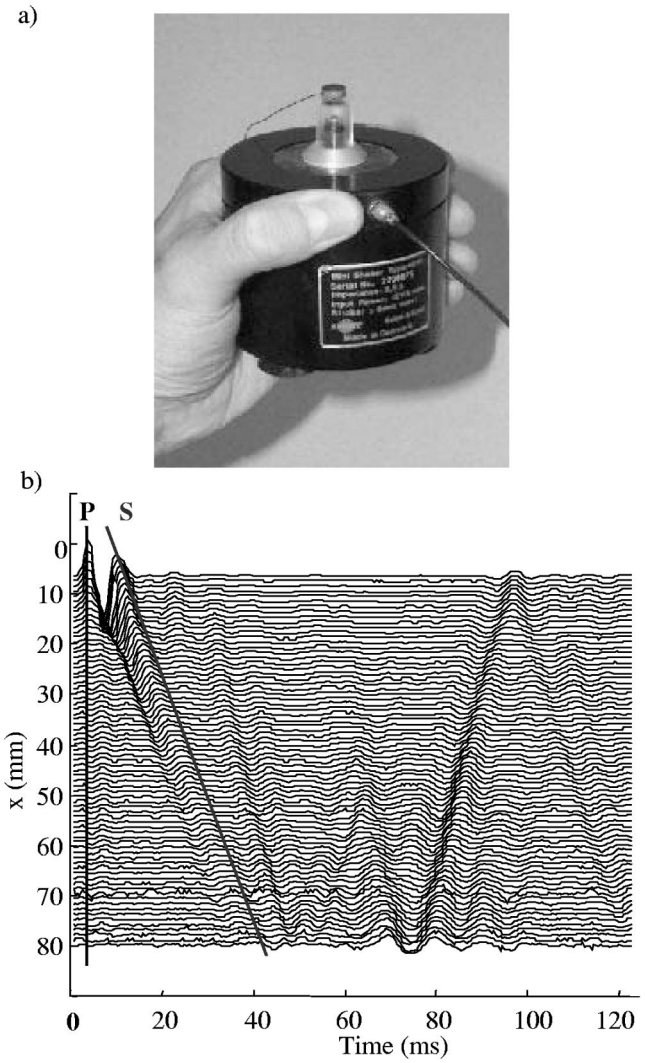


FIG. 9. (a) Picture of the 1D shear elasticity probe. A single ultrasonic transducer is mounted on a vibrator. (b) Sismic representation of the longitudinal component (along the x axis) of the displacement field. A compression (P), a shear (S), and a reflected shear waves are visible when a 100 Hz central frequency pulse is applied from the top ($x=0$) of an Agar-gelatin based phantom.

$$(\lambda + 2\mu)\nabla\nabla\mathbf{u} + \mu\nabla^2\mathbf{u} - \rho\frac{\partial^2\mathbf{u}}{\partial t^2} = \mathbf{0}. \tag{9}$$

The first assumption one needs in order to obtain the Helmholtz equation used in the inverse problem algorithm is that the compression term, the first one in Eq. (9), is negligible. It is referred to as the compression-free approximation (also called the Helmholtz approximation). In order to justify this assumption, one notes that it is possible to clearly separate

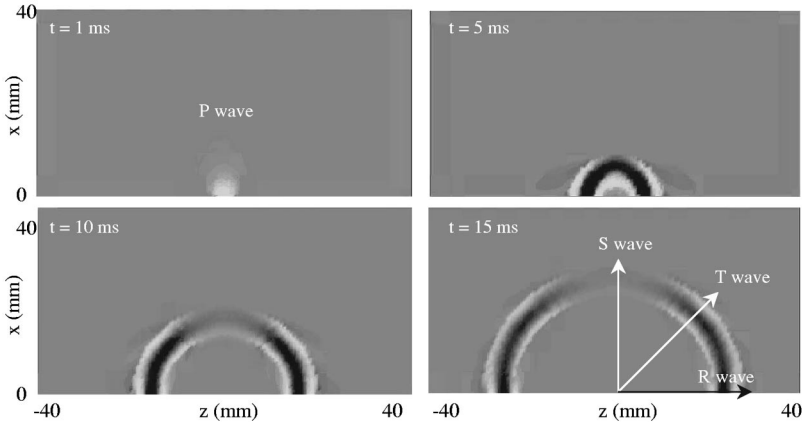


FIG. 10. Simulation of the displacement field using the exact Green's functions in a semi-infinite elastic solid. A pulse is applied from the bottom by a 7 mm circular piston source. At $t=1$ ms, the compression wave is visible. Then the shear wave propagates as a Rayleigh wave (R) at the surface, as a transverse wave (T) in the oblique direction and as a near field shear wave (S) in the axis of the piston.

the displacement induced by the compression wave and that induced by the shear wave in homogeneous media with transient elastography, Fig. 9(b). Nevertheless, getting rid of the compression wave is not so obvious as we shall see, especially in the usual experimental situation where waves are in the near field. The problem has thus been greatly simplified. The components of the displacement field u_i are decoupled and obey the same wave equation,

$$\mu \Delta u_i - \rho \frac{\partial^2 u_i}{\partial t^2} = 0, \quad i = x,y,z. \tag{10}$$

Since it is the only one measured experimentally, the x component is selected. Then the Laplacian operator reads:

$$\Delta u_x = \frac{\partial^2 u_x}{\partial x^2} + \frac{\partial^2 u_x}{\partial y^2} + \frac{\partial^2 u_x}{\partial z^2}. \tag{11}$$

One last simplification is needed. Indeed, the experimental configuration using the 1D shear elasticity probe allows measurement of only one component of the displacement in one dimension along the ultrasonic beam. The experiment requires us to neglect two terms in the Laplacian operator:

$$\frac{\partial^2 u_x}{\partial x^2} \gg \frac{\partial^2 u_x}{\partial y^2} + \frac{\partial^2 u_x}{\partial z^2}. \tag{12}$$

This second assumption is referred as the quasiplane approximation. Then, from the 1D wave equation,

$$\frac{\partial^2 u_x}{\partial x^2} - \frac{\rho}{\mu} \frac{\partial^2 u_x}{\partial t^2} = 0, \tag{13}$$

the Helmholtz equation is finally deduced for the case of monochromatic excitation:

$$\frac{\partial^2 u_x}{\partial x^2} + k^2 u_x = 0. \tag{14}$$

Consequently, what is a perfect inverse problem for (plane) transverse waves is only an approximate inverse problem for general diffracted shear waves (although well founded as far as speed measurements are concerned). So how can one compensate for the lack of information resulting from the compression-free and the quasiplane approximations? We used the analytical solution of the displacement field in a semi-infinite elastic solid induced by a point source applied perpendicularly to the surface.[29,30] The theoretical longitudinal component (x component) of the displacement field is represented on gray scale images, Fig. 10, after a 100 Hz central frequency pulse is applied from the bottom. One can distinguish a Rayleigh wave on the surface (R), a transverse wave (T) at 45°, and a shear wave along the axis of the source. The latter wave is precisely the one detected in the experiments using the 1D shear elasticity probe. If the inverse problem algorithm is applied to the theoretical longitudinal displacement in one dimension along the source axis just as in the experiment, one obtains a good estimation of the speed and a surprising (at first sight) nonzero attenuation even though the medium is ideally lossless, Fig. 11(b)

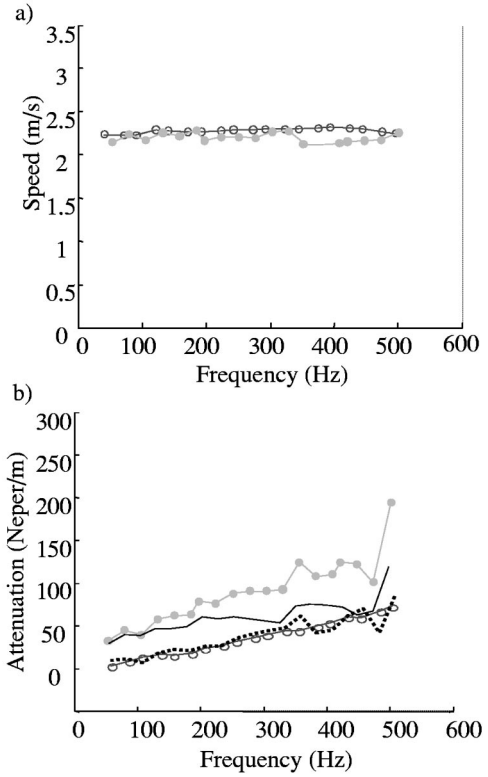


FIG. 11. Speed and attenuation measurements vs frequency. (a) The inverse problem speed estimation (dots) is very close (within 8%) to the reference measurements (circle). (b) But an overattenuation is clearly visible with the inverse problem method (dots) compared to the reference measurements (circle). Applied to the simulated data in a perfectly elastic solid, the inverse problem gives an overattenuation as well (solid line). Subtracting these latter data from the experimental measurements (dots) compensates for diffraction effects and allows one to compute the medium attenuation (dotted line).

(solid line). This paradox is a consequence of the approximations in the inverse problem. Indeed, diffraction, is seen by the inverse problem algorithm as a dissipative effect. It therefore computes an apparent attenuation due in reality to diffraction. Thus if one subtracts this apparent attenuation from the experimental measurements, the diffraction effects are eliminated and what remains is the attenuation due to the medium, Fig. 11(b) (dotted line). Consequently, speed and attenuation (or elasticity and viscosity) can be easily measured with the shear elasticity probe.

VII. DISCUSSION

Thus the Green's function simulation allows one to avoid bias due to diffraction. Another benefit of simulation is that the compression-free and quasiplane approximations, used in the inverse problem algorithm, can be tested. Surprisingly, it appears that the quasiplane approximation is very well founded since the left term of Eq. (12) is at least ten times bigger than the right-hand side terms in the experimental configuration.[31] This ratio increases for the displacements measured farther and farther away from the source. Thus the Laplacian operator is correctly evaluated and the conclusion is that the main flaw in the approximated inverse problem algorithm comes from the first approximation, namely the compression-free approximation. Should not it be eliminated? The divergence of the displacement is almost zero and yet it would be a mistake to ignore the whole compression term. It would be ignoring the huge coefficient $(\lambda+2\mu)$ in front that makes the whole compression term non-negligible: the longitudinal wave is clearly visible, Fig. 9. The problem in a quasi-incompressible medium such as tissues is that no experimental devices (using ultrasounds or NMR) has the accuracy to experimentally evaluate this compression term. It should be noted that for the same reason, any Poisson's ratio mapping is not feasible with elastography except in the special case of porous soft materials.[32] Nevertheless, as long as one is interested in phase (or speed) measurements, the compression-free approximation is valid. But it fails in amplitude (or attenuation) estimations and introduces important biases.

We have presented an experimental situation in which a single transducer is as good as a transducer array or a NMR device. A compression-free inverse problem algorithm does not lead to the correct value of attenuation along the axis of a point source. From a more general point of view, this experiment reveals that in order to correctly reconstruct the attenuation map (and thus a viscosity map), one has to fully take into account diffraction effects. It means that the inverse problem must follow from the exact wave equation, Eq. (9). So how can one reconcile the experimental limitation (the compression term cannot be evaluated) with the theoretical requirement (the exact wave equation has to be taken into account)? As shown previously in the simple case of a homogeneous medium, a first answer would be to use simulation. For heterogeneous media, the question is still open. A feasible exact inverse problem[33] is still needed in order to create a quantitative viscosity image.

VIII. CONCLUSION

Through a transient elastography experiment using plane shear waves in soft solids, this paper presents a simple inverse problem algorithm to estimate the speed and the attenuation. Moreover, it is demonstrated that the best of the simplest rheological model for an Agar-gelatin based phantom and a beef muscle sample is Voigt's solid. The conclusion of this first part is that quantitative viscosity can be retrieved from an inverse problem algorithm. Some difficulties remain in the more general case where nonplane shear waves are used. This is the subject of the second part. A simulation based on theoretical Green's function is shown to properly compensate for diffraction biases. Thus, a quantitative estimate of shear elasticity and shear viscosity is obtained with the so-called 1D shear elasticity probe in a homogeneous soft solid (Agar-gelatin phantom). For viscosity imaging (in heterogeneous media), diffraction effects require one to solve an exact inverse problem whose solution is yet to be imagined.

[1] B. S. Garra, E. I. Cespedes, J. Ophir, S. R. Spratt, and R. A. Zuurbier, "Elastography of breast lesions: Initial clinical results," Radiology **202**, 79–86 (1997).

[2] S. F. Levinson, M. Shinagawa, and T. Sato, "Sonoelastic determination of human skeletal muscle elasticity," J. Biomech. **28**, 1145–1154 (1995).

[3] F. Lee, J. P. Bronson, R. M. Lerner, K. J. Parker, S. R. Huang, and D. J. Roach, "Sonoelasticity Imaging: Results in In vitro tissue specimens," Radiology **181**, 237–239 (1991).

[4] R. Muthupillari, D. J. Lomas, P. J. Rossman, J. F. Greenleaf, A. Manduca, and R. L. Ehman, "Magnetic resonance elastography by direct visualisation of propagating acoustic strain wave," Science **269**, 1854–1857 (1995).

[5] D. B. Plewes, I. Betty, S. N. Urchuk, and I. Soutar, "Visualizing tissue compliance with MR imaging." J. Magn. Reson Imaging **5**, 733–738 (1995).

[6] S. Catheline, F. Wu, and M. Fink, "A solution to diffraction biases in sonolelasticity: The acoustic impulse technique," J. Acoust. Soc. Am. **105**, 2941–2950 (1999).

[7] L. Sandrin, M. Tanter, S. Catheline, and M. Fink, "Shear modulus imaging with 2-D transient elastography," IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 426–435 (2002).

[8] T. E. Oliphant, A. Manduca, R. L. Ehman, and J. F. Greenleaf, "Complex valued stiffness reconstruction for magnetic resonance elastography by algebraic inversion of the differential equation," Magn. Reson. Med. **45**, 299–310 (2001).

[9] M. Sridhar, H. Du, C. Pellot-Barakat, S. I. Simon, and M. F. Insana, "Ultrasonic mechanical relaxation imaging," Proc. SPIE **5373**, 202 (2004).

[10] E. E. W. Van Houten, J. B. Weaver, M. I. Miga, F. E. Kennedy, and K. D. Paulsen, "Elasticity reconstruction from experimental MR displacement data: Initial experience with an overlapping subzone finite element inversion process," Med. Phys. **27**, 101–107 (2000).

[11] H. L. Oestrecher, "Field and impedance of an oscillating sphere in a viscoelastic medium with an application to biophysics," J. Acoust. Soc. Am. **23**, 707–714 (1951).

[12] T. J. Royston, Y. Yazicioglu, and F. Loth, "Surface response of a viscoelastic medium to subsurface acoustic sources with application to medical diagnosis," J. Acoust. Soc. Am. **113**, 1109–1121 (2003).

[13] A. Bot, R. H. W. Wientjes, and K. H. de Haas, "The transition zone for gelatin gels as measured by high-frequency rheology," Imaging Sci. J. **45**, 191–196 (1997).

[14] I. R. Titze, S. A. Klemuk, and S. Gray, "Methodology for rheological testing of engineered biomaterials at low audio frequencies," J. Acoust. Soc. Am. **154**, 392–401 (2004).

[15] L. Sandrin, S. Catheline, M. Tanter, X. Hennequin, and M. Fink, "Time resolved pulsed elastography with ultrafast ultrasonic imaging," Ultrason. Imaging **21**, 259–272 (1999).

[16] J. Bercoff, S. Chaffai, M. Tanter, L. Sandrin, S. Catheline, M. Fink, J.-L. Gennisson, and M. Meunier, "*In vivo* breast tumor detection using transient elastography," Ultrasound Med. Biol. **29**, 1387–1396 (2003).

[17] W. F. Walker and G. E. Trahey, ''A fundamental limit on the performance of correlation based on phase correction and flow estimation technique,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **41**, 644–654 (1994).
[18] U. Hamhaber, F. A. Grieshaber, J. H. Nagel, and U. Klose, ''Comparison of quantitative shear wave MR-elastography with mechanical compression tests,'' Magn. Reson. Med. **49**, 71–77 (2003).
[19] M. Djabourov, J. Leblond, and P. Papon, ''Gelation of aqueous gelatin solutions. II. Rheology of the sol-gel transition,'' J. Phys. (France) **49**, 333–343 (1988).
[20] L. Landau, E. Lifchitz, *Théorie de l' élasticité* (Mir, Moscou, 1990), Vol. 7, Chap. 5, Sec. 33.
[21] D. Royer and E. Dieulesaint, *Ondes Elastiques dans les Solides* (Masson, Paris, 1996), Vol. 1, Sec. 426.
[22] J. L. Gennisson, S. Catheline, S. Chaffaï, and M. Fink, ''Transient elastography in anisotropic medium: Application to the measurement of slow and fast shear waves velocities in muscles,'' J. Acoust. Soc. Am. **114**, 536–541 (2003).
[23] I. Sack, J. Bernarding, and J. Braun, ''Analysis of wave pattern in MR elastography of skeletal muscle using coupled harmonic oscillator simulations,'' J. Magn. Reson Imaging **20**, 95–104 (2002).
[24] E. L. Madsen, H. J. Sathoff, and J. A. Zagzebski, ''Ultrasonic shear wave properties of soft tissues and tissuelike materials,'' J. Acoust. Soc. Am. **74**, 1346–1354 (1983).
[25] L. Sandrin, M. Tanter, J.-L. Gennisson, S. Catheline, and M. Fink, ''Shear elasticity probe for soft tissues with 1D transient elastography,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 436–446 (2002).
[26] S. Catheline, L. Sandrin, J.-L. Gennisson, M. Tanter, and M. Fink, ''Ultrasound-based noninvasive shear elasticity probe for soft tissues,'' IEEE Ultrasonic Symp. **2**, 1799–1801 (2000).
[27] L. Sandrin, B. Fourquet, J.-M. Hasquenoph, S. Yon, C. Fournier, F. Mal, C. Christidis, M. Ziol, B. Poulet, F. Kazemi, M. Beaugrand, and R. Palau, ''Transient elastography: A new noninvasive method for assessment of hepatic fibrosis,'' Ultrasound Med. Biol. **29**, 1705–1703 (2003).
[28] S. Catheline, J. L. Thomas, F. Wu, and M. Fink, ''Diffraction field of a low frequency vibrator in soft tissues using transient elastography,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **46**, 1013–1019 (1999).
[29] D. C. Gakenheimer and J. Miklowitz, ''Transient excitation of an half space by a point load traveling on the surface,'' J. Appl. Mech. **36**, 505–514 (1969).
[30] H. Lamb, ''On the propagation of tremors over the surface of an elastic solid,'' Philos. Trans. R. Soc. London **203**, 1–42 (1904).
[31] J.-L. Gennisson, ''The 1D shear elasticity probe: A new tool to study biological tissues,'' thesis manuscript (in French), Paris VI University, France, 2003, Chap. 5, Sec. I.b.
[32] R. Righetti, J. Ophir, S. Srinivasan, and T. Kroutskop, ''The feasibility of using elastography for imaging the Poisson's ratio in porous media,'' Ultrasound Med. Biol. **30**, 215–228 (2004).
[33] R. Sinkus, M. Tanter, T. Nisius, and C. Kuhl, ''Anisotropic properties of breast cancer. *In Vivo* results utilizing a cubic model,'' Second International Conference on the Ultrasonic measurement and Imaging of Tissue Elasticity, Corpus Christi, TX, October 2003.

Theoretical gas body pulsation in relation to empirical gas-body destabilization and to cell membrane damage thresholds

Douglas L. Miller[a] and Chunyan Dou
University of Michigan Department of Radiology, Ann Arbor, Michigan 48109-0553

(Received 15 July 2004; revised 21 September 2004; accepted 30 September 2004)

Contrast agent gas bodies attached to phagocytic monolayer cells pulsate in response to ultrasound exposure and damage the cells above thresholds, which increase in proportion to frequency. This study considered the physical basis for the thresholds and their frequency dependence. Theory for the pulsation was evaluated using empirical pulse waveforms acquired at thresholds for 1.0, 2.25, 3.5, 5.0, 7.5, and 10 MHz. For optimum-sized gas bodies, the amplitudes calculated at the thresholds were about 11% of the initial radii. At the cell membrane damage thresholds, theoretical negative shell stresses were approximately constant with frequency at about 50 MPa. This stress appears to be sufficient to induce failure of the shell, and gas body destabilization was confirmed by increases in transmission of ultrasound pulses through the monolayer and by microscopically-observed shrinkage of the gas bodies. A model of acoustic microstreaming was used to calculate the shear stress during the pulses. The maximum shear stress increased from about 1500 to 4500 Pa from 1 to 10 MHz, sufficient for the cell membrane damage. This theoretical analysis shows that both the gas body destabilization and the cell membrane damage could be expected at similar peak rarefactional pressure amplitudes, with thresholds having the observed proportionality to frequency.
 [DOI: 10.1121/1.1823212]

PACS numbers: 43.80.Gx, 43.35.Ei, 43.35.Wa [FD] Pages: 3742–3749

I. INTRODUCTION

A theory has been developed for the interaction of ultrasound with the encapsulated gas bubbles used for diagnostic ultrasound contrast agents such as Optison® (Amersham Health, Princeton, NJ) (de Jong *et al.*, 1994; Church, 1995). Essentially, the shell adds elasticity and viscosity-related damping parameters to the theory for the pulsation of free spherical bubbles. This theory has been utilized to describe the scattering properties of suspensions of the agents, leading to empirical values for the shell parameters (Marsh *et al.*, 1998). The theories appear to be reasonable models of the gas body behavior for low levels of excitation. As the incident ultrasound pressure amplitude increases from zero, the stabilizing shell experiences oscillating stresses. Above rather modest excitation peak rarefactional pressure amplitudes (PRPAs), the stresses may be sufficient to expect shell failure (Stride and Saffari, 2003). However, uncertainty remains as to how the theories can be used to describe the destabilization of the gas bodies, and the complex behavior at higher amplitudes. The fate of destabilized gas bodies can include gradual shrinkage or rapid fragmentation, depending on physical conditions (Chomas *et al.*, 2001a, 2001b). Ultrasound exposure can also result in cavitation nucleation (i.e., production and proliferation of free microbubble cavities) (Miller and Thomas, 1995). In some circumstances the remnants of destabilized gas bodies (no longer capable of producing image contrast) can still serve as cavitation nuclei (Dalecki *et al.*, 1997).

The interaction of ultrasound with contrast agent gas bodies is a form of acoustic cavitation, which can lead to biological effects (Nyborg *et al.*, 2002). The mechanical action of the pulsating gas bodies on nearby cells or tissues can lead to a wide range of bioeffects, such as cell membrane damage (Miller *et al.*, 1996; Miller and Dou, 2004a) and capillary rupture (Skyba *et al.*, 1998; Li *et al.*, 2003). The bioeffects are considered to be cavitational bioeffects and potentially pose risks of adverse consequences of contrast enhanced diagnostic ultrasound (Nyborg *et al.*, 2002). For low pressure amplitudes, the gas bodies may remain stable, with mechanical perturbations produced by the pulsation. There appear to be PRPA thresholds for bioeffects, such as cell membrane damage, which presumably occur when cell perturbation reaches critical levels for damage. In addition, destabilization of the gas bodies appears to occur at comparable PRPAs, particularly for shell-encapsulated designs. For relatively high pressure amplitudes, the gas bodies are destroyed but can serve as cavitation nuclei with subsequent effects resulting from free-bubble cavitation phenomena.

A sensitive system for studying cellular bioeffects of contrast agent pulsation is a cell monolayer grown on the upper window of an exposure chamber. This allows gas bodies in the medium filling the chamber to rise and come into direct contact with the cells before exposure. This method is particularly useful for pulsed ultrasound, and has been used to study ATP release (membrane damage) from CHO cells during 3.3 MHz pulsed ultrasound exposure (Miller and Bao, 1998), sonoporation of epidermoid cells using an actual diagnostic ultrasound scanner (Miller and Quddus, 2000), and membrane damage in phagocytic macrophagelike cells (Miller and Quddus, 2001). The phagocytic cells, but not the epidermoid cells, tended to retain the gas bodies when monolayers incubated with contrast agent were rinsed before ex-

[a]Electronic address: douglm@umich.edu

TABLE I. Empirical thresholds and theoretical results for gas body pulsation under threshold conditions.

	Frequency (MHz)					
	1.0	2.25	3.5	5.0	7.5	10.0
Threshold (MPa)	0.066	0.15	0.24	0.37	0.44	0.62
Standard error (MPa)	0.018	0.038	0.05	0.10	0.09	0.19
Pulse duration (μs)	1.9	0.79	0.60	0.40	0.27	0.20
Optimal R_0 (μm)	7.5	4.4	3.3	2.6	2.1	1.8
Maximum $R-R_0$ (μm)	0.98	0.53	0.45	0.35	0.174	0.144
Minimum $R-R_0$ (μm)	−0.83	−0.47	−0.36	−0.30	−0.156	−0.132
Relative amplitude	0.12	0.11	0.12	0.13	0.079	0.078
Tensile stress (MPa)	55.5	51.2	59.2	59.9	40.4	42.3
Compress. stress (MPa)	65.8	65.2	68.1	76.6	52.6	57.7
Maximum U_1 (m/s)	5.5	6.8	7.6	9.1	6.3	7.1
β (μm)	0.56	0.38	0.30	0.25	0.21	0.18
SS stress (Pa)	1230	2130	3630	5050	2970	3750

posure, thereby assuring direct cell-gas body contact during even the briefest pulses.

The phagocytic cell monolayer therefore provides a model system suitable for the study cavitational bioeffects at low pulse pressure amplitudes. For brief incubation, the gas bodies stick to the cells without complete phagocytosis. A 3.5 MHz laboratory ultrasound system was used to explore the influence of exposure parameters and agent properties on cell membrane damage (Miller and Dou, 2004a). The two different agents Optison® (Amersham Health Inc., Princeton, NJ) and Definity® (Bristol-Myers Squib Medical Imaging, N. Billerica, MA) produced similar thresholds for bioeffects, which indicates a minor role for small differences in gas body properties in cellular bioeffects. Thresholds for damage were only weakly dependent on timing parameters. For example, thresholds for continuous exposure with Optison® were 0.05 MPa for 60 s increasing only to 0.84 MPa for 0.6 μs (a factor of 10^8 in duration). The membrane damage thresholds also depended on the frequency of ultrasound exposure (Miller and Dou, 2004b). Focused ultrasound with two cycle pulses and 1% duty cycle was used, which simulated pulse timing parameters of diagnostic imaging. PRPA thresholds were found to increase approximately in proportion to frequency from 0.066 MPa at 1 MHz to 0.62 MPa at 10 MHz. These thresholds were relatively low for cavitational bioeffects, and the physical basis for the proportionality to frequency was uncertain. For comparison, the inertial cavitation threshold estimated even for free microbubbles of optimal size are greater and have a different frequency dependence for this frequency range. Approximate calculations indicated a frequency dependence of this minimum inertial cavitation threshold of the form $\mathrm{PRPA}^2 f^1 = 0.06\,\mathrm{MPa}^2\,\mathrm{MHz}^{-1}$ (Apfel and Holland, 1991), or $\mathrm{PRPA}\, f^{-1/2} = 0.24\,\mathrm{MPa}\,\mathrm{MHz}^{-1/2}$ (expressed in the form of the Mechanical Index). This contrast between inertial cavitation theory and experimental results does not support a hypothesis that the inertial cavitation thresholds predict the membrane damage thresholds for this model system of contrast agent bioeffects.

Development of theoretical models of pulsation, destabilization, and cell damage could elucidate the relationship between the ultrasound interaction and the bioeffects potential associated with contrast agents. Detailed relationships between the pulsation theory and bioeffects experiments including the influence of agent destabilization and mechanisms for biological effects have not been established. The purpose of this study was to assess the physical basis for the cell membrane damage at relatively low PRPAs and the observed proportionality of the cell damage threshold to frequency. Measurements were made of ultrasound transmission through the monolayers and the destabilization of the gas bodies was microscopically observed. The theory for ultrasonic pulsation of the contrast agent gas bodies was applied to describe gas body responses for short pulse waveforms measured for membrane damage thresholds in the 1–10 MHz range. This behavior was related to stresses in stabilizing shells of the gas bodies and to acoustic microstreaming shear stress on cells. The shell stresses were sufficient for gas body destabilization, the shear stresses were sufficient for cell membrane damage, and both critical stresses predicted threshold PRPAs proportional to ultrasound frequency.

II. METHODS

A. Experimental methods

Experimental methods used to obtain the PRPA thresholds for membrane damage were described previously (Miller and Dou, 2004b). Briefly, monolayers of phagocytic cells growing on thin Mylar sheets were loaded with Option® contrast agent gas bodies. The gas bodies were attached to, but not engulfed by the cells. The cells and gas bodies were exposed to 1 min pulsed ultrasound under free-field conditions using two cycle pulses and 1% duty cycle. Six ultrasound frequencies were tested: 1.0, 2.25, 3.5, 5.0, 7.5, and 10.0 MHz. The pulse timing produced a decreasing pulse duration and increasing pulse repetition frequency with increasing frequency, simulating timing parameters of pulsed ultrasound in diagnostic imaging, but a constant total on-time of 0.6 s. Cells were scored for dye exclusion at the center of the focal zone where the incident PRPA was approximately constant over the area. Thresholds for membrane damage were determined from exposure-response curves and are listed in Table I, together with the statistical uncertainties of

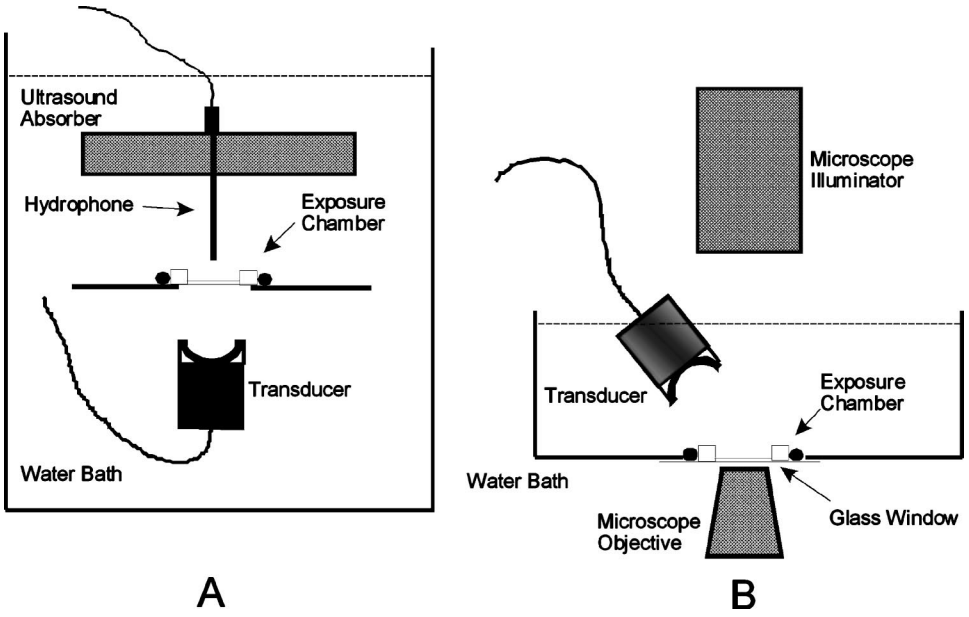


FIG. 1. Illustrations of the experimental apparatus used in this study: (a) the arrangement for transmission measurements and (b) the arrangement for microscopical observation of gas-body behavior. The exposure chamber consisted of a stainless steel ring, which clamped the Mylar sheet with cell monolayer and an O ring, which held a plastic cover sheet in place enclosing the cell culture medium.

the thresholds and the measured pulse durations (Miller and Dou, 2004b).

For this study pulse pressure waveforms were recorded for use as the driving pressure function in the theoretical analysis. The exposure fields were measured using a calibrated hydrophone (model 805 PVDF bilaminar membrane hydrophone, Sonora Medical Systems, Longmont, CO) placed at the focus of the transducers, in the absence of the exposure chamber. Corrections were made for spatial averaging in the focal zone, as described previously (Miller and Dou, 2004b). The system drive voltage was adjusted to provide the threshold pressure amplitude (or a low multiple of that voltage consistent with a good signal to noise ratio, later readjusted to the threshold) and the two cycle pulse waveforms were recorded at 5000 time point resolution with a 500 MHz digitizing oscilloscope (model TDS 520B, Tektronix, Beaverton, OR).

The acoustical behavior of the ensemble of gas bodies on the monolayer under near threshold conditions was explored by measuring the transmission of the ultrasound through exposure chambers. The chambers with cell monolayers were set up as for the bioeffects tests involving incubation with 2% Optison® in the medium followed by rinsing to remove unattached gas bodies. The apparatus for transmission measurements is illustrated in Fig. 1(a). The transmission coefficient was calculated as the ratio of the voltage detected by a hydrophone with the chamber in place to the voltage without the chamber. The transmission loss was primarily due to the gas body population, with transmission through an empty chamber better than 0.95 for all frequencies.

Finally, microscopical observations of gas body behavior were performed using an inverted microscope (model MZ FLIII, Leica Inc. Deerfield, IL) with a still or video camera. The apparatus for microscopical observations is illustrated in Fig. 1(b). A dish of water with a 0.24-mm-thick glass window glued to the bottom served as the exposure water vessel, with the transducer aimed at a 30° angle from vertical toward a chamber placed on the window. This located the monolayer 1 mm above the coverslip, which may have compromised free-field conditions to some extent. However, this apparatus was used only with the 3.5 MHz transducer with 0.6 μs duration pulses, which would have minimal field distortion by reflections.

B. Theoretical methods

For this study, the theoretical model for shelled gas bodies developed by Church (1995) was used to calculate radius time curves. The differential equation for the gas body dynamics was solved using Mathematica® (version 5, Wolfram Research, Champaign, IL). The differential equation for the radial velocity $\mathbf{U}$ for a gas body of initial radius R_0, including the linear radiation damping term suggested by Church (1995), was (dotted symbols represent differentiation with respect to time t):

$$\dot{U}=\left(-U^2\left(\frac{3}{2}+\left(\frac{\rho_L-\rho_S}{\rho_S}\right)\left(\frac{4(R+\Delta)^3-R^3}{2(R+\Delta)^3}\right)\frac{R}{(R+\Delta)}\right)+\frac{1}{\rho_S}\left(P_i\left(\frac{R_0}{R}\right)^{3\kappa}-(P_0+p)-\frac{2\sigma_1}{R}-\frac{2\sigma_2}{R+\Delta}-4\frac{U}{R}\left(\frac{\mu_S V_S+\mu_L R^3}{(R+\Delta)^3}\right)\right)-(\omega R_0)^2\times\left(\frac{R_0+\Delta}{c(1+(\omega(R_0+\Delta)/c)^2)}\right)\frac{U}{R}-4\rho_S\frac{V_S G_S}{(R+\Delta)^3}\times\left(1-\frac{R_e}{R}\right)\Bigg/\ R\left(1+\left(\frac{\rho_L-\rho_S}{\rho_S}\right)\frac{R}{R+\Delta}\right) \quad (1)$$

in which

$$V_S=(R_0+\Delta)^3-R_0^3 \quad (2)$$

and

TABLE II. Definitions and values of the constants used in Eqs. (1)–(3) for the shelled gas body.

Constant	Definition	Assumed value
P_0	Ambient pressure	10^5 Pa
ρ_L	Density of water	10^3 kg m^{-3}
μ_L	Viscosity of water	10^{-3} Pa s
σ_1	Gas-shell surface tension	0.04 N m^{-1}
κ	Polytropic exponent	1.0
P_i	Internal pressure	P_0
ρ_S	Density of shell	1.1×10^3 kg m^{-3}
μ_S	Viscosity of shell	1.11 Pa s
G_S	Modulus of rigidity	83.3 MPa
Δ	Shell thickness	15 nm
R_0	Stable gas body radius	Test variable
ω	2π times frequency	Test variable

$$R_e = R_0\left(1+\left(\frac{2\sigma_1}{R_0}+\frac{2\sigma_2}{R_0+\Delta}\right)\frac{(R_0+\Delta)^3}{4V_S G_S}\right). \tag{3}$$

Here, the radial velocity **U**, radius **R**, and incident pressure **p** are functions of time. The other symbols represent constants: ambient pressure P_0, density of water ρ_L, viscosity of water μ_L, gas-shell surface tension σ_1, polytropic exponent κ, internal pressure P_i, density of shell ρ_S, viscosity of shell μ_S, modulus of rigidity G_S, shell thickness Δ, and 2π times frequency ω. The values of the symbols appropriate for the shelled Optison® gas body are listed in Table II. The values of the constants were from Church (1995), except for values of G_s and μ_S, which were from Marsh *et al.* (1998). The differential equation for the radius **R** was simply

$$\dot{R} = U. \tag{4}$$

This system of equations was solved using the ND solve function. This model neglects thermal damping, which was appropriate for the small gas bodies considered. Measured values of **p** were acquired using a calibrated hydrophone for the experimental pulses as described earlier.

The shell experiences stresses as the gas body pulsates (Church, 1995; Stride and Saffari, 2003). The approximate effective stress within the shell at the inner surface is given by (Stride and Saffari, 2003):

$$T_{\text{eff}} = -6\left(G_S\frac{R-R_e}{R}+\mu_S\frac{U}{R}\right). \tag{5}$$

The modification of the model needed to describe the dynamic behavior after destabilization is not known. The behavior depends on the strength of the interaction, and may include slow dissolution, emergence of free bubbles or break-up of the gas body into quickly vanishing fragments (Chomas *et al.*, 2001a, 2001b).

Detailed theory for estimating the amount of biological perturbation expected from a given exposure situation is not available for use in medical applications of contrast agents. Even for moderate pressure amplitudes, a number of different mechanisms may plausibly have a role in bioeffects (Miller *et al.*, 1996). Many gas bodies were nonspherical or changed in shape during an exposure. Furthermore, asymmetric pulsation is expected near solid boundaries, which can result in asymmetrical collapse and fluid jets directed at the surface for high amplitudes (see for example, Marmottant and Hilgenfeldt, 2003; Krasovitski and Kimmel, 2004). For the *in vitro* experiments, gas bodies were absent from stained (i.e., killed) cells when the effect was scored after suprathreshold exposure (e.g., Miller and Quddus, 2002). This suggests a possible role for gas body destabilization in the cell membrane damage. However, significant steady (nonoscillatory) fluid shear stress occurs for gas body pulsation near solid surfaces even at relatively low amplitudes which are expected for threshold conditions. Thus, a simplified theory for shear-stress induced by acoustic microstreaming was adopted for modeling cell injury.

Fluid shear stress generated in the vicinity of a pulsating gas body can produce mechanical cell membrane damage (Rooney, 1972; Nyborg, 1977). The pulsation near a solid surface produces acoustic microstreaming flow and steady shear (SS) stress. The fluid velocity decreases to approximately zero (nonslip condition) at a solid surface within a thin boundary layer of thickness β (Nyborg, 1977), given by

$$\beta = \left(\frac{2\mu_L}{\rho\overline{\omega}}\right)^{1/2}. \tag{6}$$

This results in a velocity gradient with a maximum value (amplitude) given approximately by U_1/β, in which U_1 is the velocity amplitude, and in an oscillatory shear stress. However, biophysical data suggest that the role of the oscillatory stress may be small compared to the role of steady stresses (Rooney, 1972).

The SS stresses or unidirectional forces result in actual displacement, allowing the accumulation of perturbations over multiple cycles and pulses. The critical stress needed for damage has been determined for several fluid shear situations, and the microstreaming shear stress model has proven useful in describing several bioeffects involving ultrasonic gas body activation (Miller, 1987). Microstreaming flow is generated with a velocity gradient near the boundary which results in a SS stress approximately given by

$$S_{\text{SS}} = \frac{\mu_L\omega R_1^2}{R_0\beta}. \tag{7}$$

Here, R_1 was the linearized amplitude of the radial oscillation, which was calculated as the mean of the maximum and minimum values of $|\mathbf{R}-\mathbf{R}_0|$. When this shear stress exceeds a critical value then cell membrane damage may be expected. For pulse mode exposure, the effects of multiple pulses appear to accumulate to produce critical stress at lower PRPAs than for single pulses. The actual processes occurring near ultrasonically pulsating gas bodies in the megahertz frequency range are only approximated by this model. The boundary layer thickness is typically smaller than the cell size in the megahertz frequency range, so that the cells cannot be subjected to uniform stress, as is the case for large scale shearing flow. Furthermore, the time course and scale of actual fluid flow is uncertain for the pulsed mode exposure. Nevertheless, this simplified theory provided a framework for understanding the cell membrane damage.

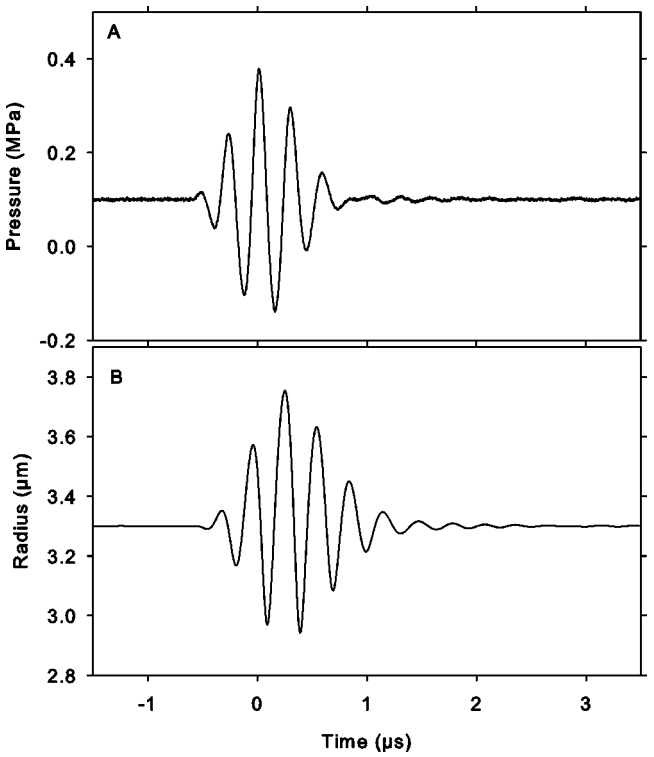


FIG. 2. (a) The measured pulse waveform for 3.5 MHz exposures and (b) the theoretical radial response of the most strongly responding (i.e., optimal radius) gas body. The amplitude is small relative to the radius, an essentially linear response.

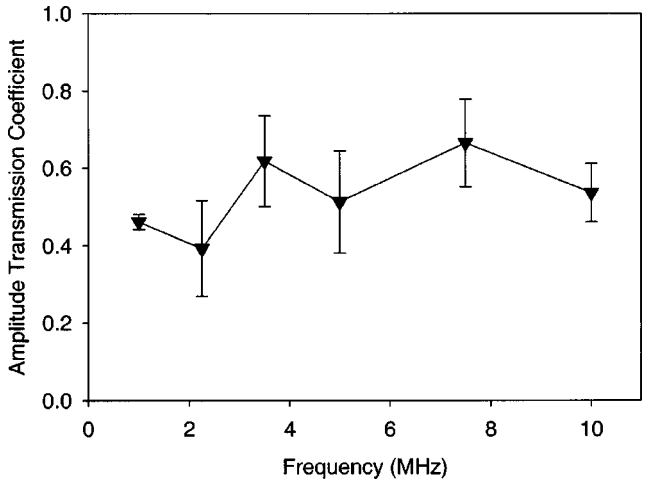


FIG. 3. The amplitude transmission coefficient through phagocytic cell monolayers loaded with contrast agent gas bodies for the six frequencies tested, plotted as the means of four measurements, with standard error bars. There was no apparent trend with frequency.

III. RESULTS

A. Experimental results

Pulse waveforms were recorded using excitation conditions corresponding to membrane damage thresholds at each frequency, in the absence of a chamber. The pulse waveform recorded for 3.5 MHz for the 0.24 MPa threshold PRPA is shown in Fig. 2(a). All the pulse waveforms had a similar appearance, but different durations as listed in Table II. The discrete data stored in a spreadsheet was imported into the Mathematica program, and an interpolation function was created for use with Eq. (1).

The presence of the chambers with cells and gas bodies influenced the ultrasound field. To characterize this interaction, the amplitude transmission coefficients were measured for four chambers at each frequency for two PRPA values used for the threshold determinations (Miller and Dou, 2004b). The values corresponded to the PRPA levels nearest the membrane damage threshold and twice this PRPA, measured both before and after 1 min of exposure. The transmission coefficients for the start of exposure near the threshold are shown in Fig. 3. There were no significant differences between the results at different frequencies. The overall average transmission coefficient for all the frequencies was 0.53 ± 0.21 SD. This attenuation indicates perturbation of the exposure field. For sparsely distributed point scatterers in a plane surface, the transmitted pressure reflects the actual pressure amplitude in the plane (Miller, 1980). Thus, the field actually interacting with the gas bodies might have been about half the calibrated measurements. Unfortunately, this aspect of the exposure conditions is uncertain given the attachment of multiple gas bodies (i.e., not sparse) to the cells. However, the observations indicate that the behavior was about the same for all the frequencies tested and therefore not a bias in the frequency trends.

There were increases in transmission between the beginning and ending of the 1 min exposures, as shown for the overall averages for the six frequencies in Fig. 4. For the near threshold PRPAs the increase was about 7%, while for the higher PRPAs the increase was about 30% (significantly different, $P < 0.01$). This indicates that the ultrasound modified the gas-body population during the exposure. Observation using the inverted microscope apparatus revealed that the gas bodies tended to shrink during exposure as shown in Fig. 5. For even higher PRPAs, the transmission increased essentially to unity (i.e., the transmission coefficient in the absence of gas bodies), and all gas bodies apparently disappeared on effected cells.

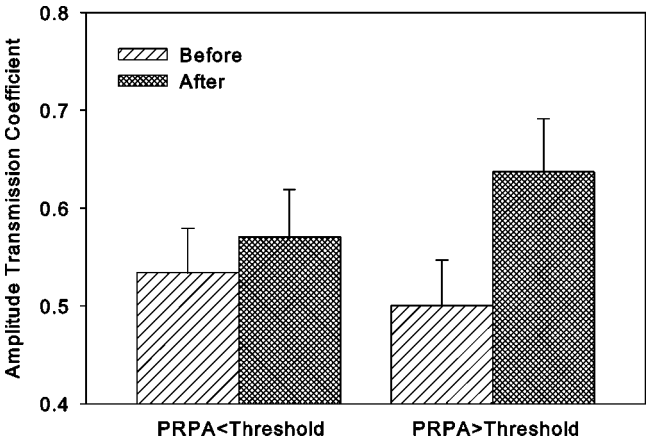


FIG. 4. Amplitude transmission coefficients averaged for all six frequencies before and after 1 min pulsed exposures at (or slightly less) than the threshold PRPAs, and at about twice the threshold PRPAs. Above the thresholds, a significant increase in transmission was induced by the exposures.

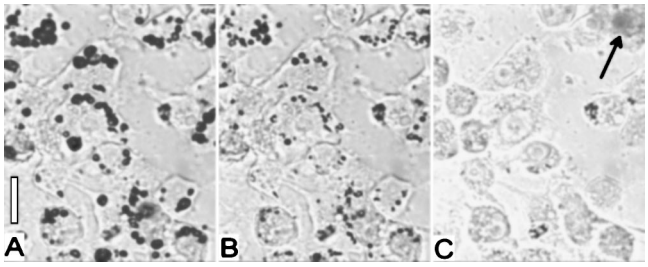


FIG. 5. Frames taken from a video recording of phagocytic cells loaded with gas bodies and exposed to 3.5 MHz pulsed ultrasound: (a) gas bodies and cells immediately before exposure, (b) after 1 s of exposure, and (c) 5 min after the end of a 10 s exposure. The scale bar in (a) represents a length of 20 μm, and applies to all three photomicrographs. The arrow in (c) indicates a cell with membrane damage resulting from the exposure, which was stained by trypan blue dye in the medium (in greyscale, the nucleus is slightly darker). Other cells appeared to be unaffected by the exposure, even though the gas bodies were destabilized.

B. Theoretical results

For each frequency the measured pulse waveform was used with Eqs. (1)–(4) to obtain the radial pulsation as a function of time for 100 values of R_0 varied incrementally over the 1–10 μm range of radii. The pulse-maximum amplitude R_1 was determined during the pulse response for each radius. From this set of data the optimum (essentially the resonance size) radius was found and used in subsequent calculations for the maximum response within the gas body population. The optimum size at 3.5 MHz was 3.3 μm, which had a radius-time response as shown in Fig. 2(b). The response was of small amplitude and very nearly linear. The maximum amplitude response at 3.5 MHz for 1–10 μm radii is shown in Fig. 6.

The optimal radius (initial radius giving the maximum amplitude) decreased with increasing frequency, as expected for the linear resonance frequency. The pulsation amplitudes also decreased by nearly a factor of 10, even as the threshold PRPA values increase by approximately a factor of 10. However, because the optimum radius also decreased, this resulted in an approximately constant relative amplitude R_1/R_0 of about 0.11 ± 0.02. Theoretical results are listed in Table II for each frequency. An approximately constant

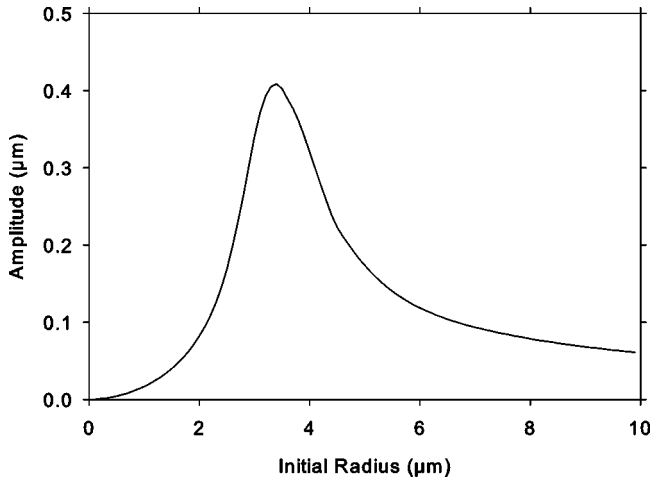


FIG. 6. A plot of the theoretical response of gas bodies of radii 0.1–10 μm for 3.5 MHz pulsed exposure.

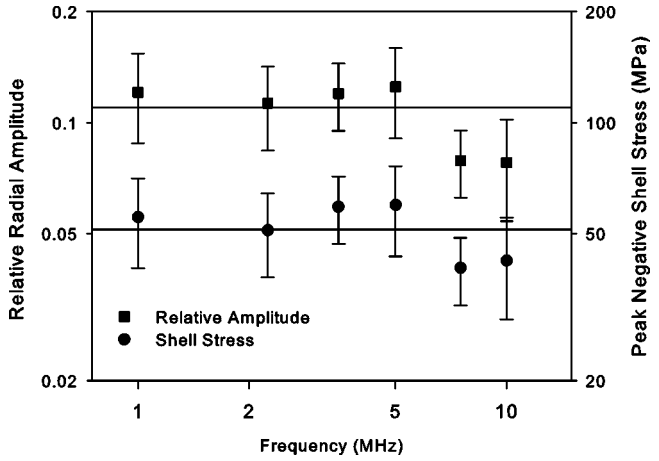


FIG. 7. Plots of the relative amplitude and effective shell tensile stress for the six frequencies tested. The horizontal lines indicated the averages of the six values.

(given the uncertainties in the thresholds, and other factors) shell stress was calculated for the 6 frequencies of 51.4 ± 8.4 MPa. The relative amplitudes and effective shell tensile stresses are plotted in Fig. 7.

The decreasing amplitude needed for constant relative response R_1/R_0 with increasing frequency resulted in an approximately constant maximum radial velocity amplitude U_1 of 7.1 ± 1.2 m s^{-1}. The calculated SS stresses were rather variable over the frequency range, making a trend difficult to discern (it should be noted that the dependence on the square of the amplitude exaggerated the fractional variation compared to first-order quantities). The velocity amplitudes and SS stresses are plotted in Fig. 8. Linear regression ($r^2=0.65$) on the plot of SS stress indicated an increase from about 1500 to 4500 Pa from 1 to 10 MHz (with an average of 3127 ± 1338 Pa).

IV. DISCUSSION

A. Summary

Stabilized gas bodies in Optison® ultrasound contrast agent can attach to phagocytic monolayer cells, which pro-

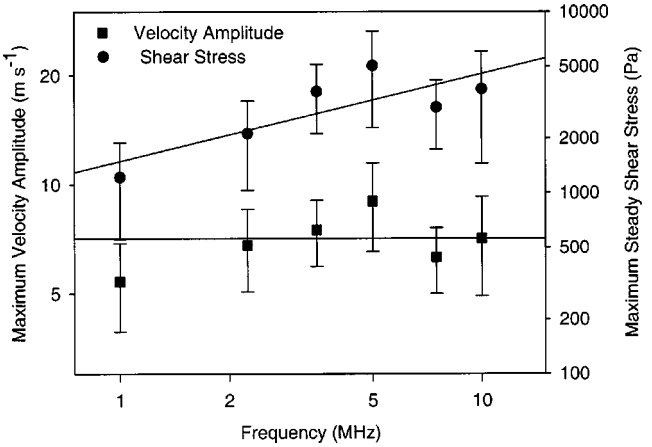


FIG. 8. Plots of the pulse peak velocity amplitudes and steady shear stresses. The line for the velocity amplitudes represents the average of the six values. The line for the shear stresses, which appear to increase somewhat with frequency is a linear regression ($r^2=0.65$).

vides a model *in vitro* system for the study of bioeffects induced by the interaction of ultrasound with the gas bodies. Pulsed ultrasound exposure of the gas bodies has been shown to induce cell membrane damage (cell death) above apparent thresholds, which were approximately proportional to frequency for six frequencies from 1 to 10 MHz (Miller and Dou, 2004b). Empirical pulse waveforms were acquired for threshold PRPAs [e.g., Fig. 2(a)]. These wave forms approximated the pressure waveforms at the cell-gas body locations, but may overestimate the true local values as indicated by reduced transmission through the monolayers loaded with gas bodies (Fig. 3). Measurement of transmission below and above the cell damage thresholds indicated that the gas bodies were modified by the higher PRPA pulses, such that the transmission increased (Fig. 4). Microscopical observation confirmed destabilization of the gas bodies, which shrank during exposure (Fig. 5).

Established theory for the gas body pulsation (Church, 1995) was evaluated using the empirical pulse waveforms for the driving pressure to determine theoretical behavior of the gas bodies [e.g., Fig. 2(b)]. The shell properties restricted the theoretically calculated pulsation to small amplitudes (i.e., compared to free gas bubbles), which may be categorized as noninertial cavitation. The relative radial amplitudes calculated at the thresholds were approximately constant with frequency at only about 11% of the optimal radii giving maximum response. This small amplitude response may nevertheless be important for gas body stability, and resulted in estimated shell stress amplitudes of about 50 MPa. As noted by Stride and Saffari (2003), this level of pulsation and shell stress may be sufficient to induce failure of the shell. Empirical observations confirmed that shell destabilization was apparent under pulse PRPAs above the cell membrane damage thresholds.

A model of near boundary microstreaming with steady fluid flow, which has been established as a mechanism for cell membrane effects (Nyborg *et al.*, 2002), was used to calculate the peak unidirectional shear stress during the pulsed exposure. The critical SS stress for-cell membrane damage has been observed in several different fluid shear situations (Miller, 1987). The critical shear stress was about 190 Pa for 1 s, and increased slowly for decreasing duration T of shear, approximately as $T^{-0.29}$. This expected trend was reflected in the slowly decreasing membrane damage thresholds for increasing exposure duration (Miller and Dou, 2004a). PRPA thresholds at 3.5 MHz were 0.84 MPa for single 0.6 μs pulses and 0.08 MPa for 0.6 s continuous exposure. Extrapolation from the calculations for this study yield SS stresses of 42 kPa and 400 Pa for these PRPA thresholds, compared to expected critical stress values of 12.1 kPa and 220 Pa. For multiple-pulse exposures, damage seemed to accumulate, and the thresholds fell between that for a single pulse and that for continuous exposure of duration equal to the total on-time (0.6 s for the pulsed exposures considered here). The calculated SS stress increased slowly from about 1500 to 4500 Pa (Fig. 8) for the cell damage thresholds for 1–10 MHz pulsed exposure (Miller and Dou, 2004b). Critical stresses of approximately 8.8–16.7 kPa were expected for the single pulse durations (1.8–0.2 μs over the frequency range) and 220 Pa for a duration equal to the total on-time of 0.6 s. These empirical values bracket the theoretically calculated estimates, which supports the hypothesis that the steady shear stress generated by the pulsating gas bodies was the mechanism responsible for the cell membrane damage for the multiple pulse exposures.

B. Discussion of limitations

This result is essentially in agreement with conclusions by Rooney (1972), Miller (1985), Miller and Bao (1998), Wu (2002) and others (see Nyborg *et al.*, 2002). However, it should be noted that details of the phenomena are not entirely clear for this model system. There is uncertainty as to the actual acoustic field existing at individual gas bodies, indicated by transmission loss and the close proximity of gas bodies on different cells or on the same cell. The detailed nature of the pulsation including asymmetrical oscillation for gas bodies near a boundary would influence the cell perturbation to some extent even for linear conditions (Marmottant and Hilgenfeldt, 2003; Krasovitski and Kimmel, 2004). Furthermore, the internal cell viscosity is much higher than the surrounding medium, which can influence the gas-body pulsation and the shear stress or strain applied to the membrane Dayton *et al.* (2001).

Another important uncertainty arises from the apparent coincidence between the PRPA values needed to destabilize the gas bodies and the values needed to damage the cells. Microscopical observation, as described in Sec. II, showed that the gas bodies shrank during the exposure at near threshold conditions. This behavior apparently was similar to the phenomenon observed by Chomas *et al.*, (2001a). Near the threshold, the gas bodies sometimes shrink, often into nonspherical shapes, with little or no cell killing. At twice the near threshold PRPA, the shrinkage was more rapid and complete, and some cell membranes were damaged, as shown in Fig. 5 for 0.5 MPa at 3.5 MHz. Most cells with shrinking gas bodies survived the exposure, which implies that the destabilization *per se* was not the cause of the cell membrane damage. However, destabilization and gas body loss could influence the cell membrane damage threshold in at least four ways. First, the reduced number of gas bodies would reduce the attenuation caused by the array of gas bodies covering the cell monolayer, allowing higher PRPAs at the remaining gas bodies. Second, a large gas body could shrink to the optimal size for the frequency applied, subjecting the cell to higher SS stresses. Third, the strained shell failure could cause concomitant cell membrane failure at the point of gas-body-cell attachment. Finally, at suprathreshold PRPAs destabilization leading to behavior more like that of a free bubble might also occur and lead to much larger cell perturbation than expected from a shelled gas body (Miller and Bao, 1998; Wu, 2002).

C. Conclusion

Cell membrane damage induced by 1–10 MHz pulsed ultrasound interaction with contrast agent gas bodies attached to monolayer cells was examined using physical theory for gas body pulsation. The observed PRPA thresholds

corresponded to approximately constant relative pulsation amplitudes. The resulting shell stresses were expected to be sufficient to induce destabilization, which was confirmed by measurements and observations. The thresholds also corresponded to approximately constant radial velocity amplitudes. The resulting steady microstreaming shear stress slowly increased for decreasing 1.8–0.2 μs pulse durations in the 1–10 MHz range and were sufficient to damage cell membranes, based on previous empirical observations in a variety of fluid flows. These theoretical insights provide rational explanations for the gas body destabilization and cell membrane damage, for the occurrence of the two phenomena at similar PRPAs, and for their similar frequency dependence. For this model system, contrast agent gas body destabilization and pulsation-induced cell membrane damage occur in theory and experiment for diagnostically relevant pulsed ultrasound when the parameter $\mathbf{p\,f^{-1}}$ exceeds about 0.06 MPa MHz^{-1}.

ACKNOWLEDGMENTS

The authors thank Dr. Oliver Kriphgans for help with the Mathematica® software. This research was supported by the US National Institutes of Health via Grant No. EB000338.

Apfel, R. E., and Holland, C. K. (**1991**). ''Gauging the likelihood of cavitation from short-pulse, low-duty cycle diagnostic ultrasound,'' Ultrasound Med. Biol. **17**, 179–185.

Chomas, J. E., Dayton, P., May, D., and Ferrara, K. (**2001a**). ''Threshold of fragmentation for ultrasonic contrast agents,'' J. Biomed. Opt. **6**, 141–150.

Chomas, J. E., Dayton, P., Allen, J., Morgan, K., and Ferrara, K. (**2001b**). ''Mechanisms of contrast agent destruction,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **48**, 232–248.

Church, C. C. (**1995**). ''The effects of an elastic solid surface layer on the radial pulsations of gas bubbles,'' J. Acoust. Soc. Am. **97**, 1510–1521.

Dayton, P. A., Chomas, J. E., Lum, A., Allen, J. S., Lindner, J. R., Simon, S. I., and Ferrara, K. W. (**2001**). ''Optical and acoustical dynamics of microbubble contrast agents inside neutrophils,'' Biophys. J. **80**, 1547–1556.

Dalecki, D., Raeman, C. H., Child, S. Z., Penny, D. P., and Carstensen, E. L. (**1997**). ''Remnants of Albunex nucleate acoustic cavitation,'' Ultrasound Med. Biol. **23**, 1405–1412.

de Jong, N., Cornet, R., and Lancee, C. T. (**1994**). ''Higher harmonics of vibrating gas-filled microspheres. Part one: Simulations,'' Ultrasonics **32**, 447–453.

Krasovitski, B., and Kimmel, E. (**2004**). ''Shear stress induced by a gas bubble pulsating in an ultrasonic field near a wall,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **51**, 973–979.

Li, P., Cao, L., Dou, C., Armstrong, W. F., and Miller, D. L. (**2003**). ''Impact of myocardial contrast echocardiography on vascular permeability: An *in vivo* dose response study of delivery mode, ultrasound power, and contrast dose,'' Ultrasound Med. Biol. **29**, 1341–1349.

Marmottant, P., and Hilgenfeldt, S. (**2003**). ''Controlled vesicle deformation and lysis by single oscillating bubbles,'' Nature (London) **423**, 153–156.

Marsh, J. N., Hughes, M. S., Hall, C. S., Lewis, S. H., Trousil, R. L., Brandenburger, G. H., Levine, H., and Miller, J. G. (**1998**). ''Broadband through-transmission signal loss measurements of Albunex suspensions at concentrations approaching *in vivo* doses,'' J. Acoust. Soc. Am. **104**, 1654–1666.

Miller, D. L. (**1980**). ''Acoustical interaction of spherical and cylindrical bubbles on plane sheets and ribbons,'' Ultrasonics **18**, 277–282.

Miller, D. L. (**1985**). ''Microstreaming shear as a mechanism of cell death in Elodea leaves exposed to ultrasound,'' Ultrasound Med. Biol. **11**, 285–292.

Miller, D. L. (**1987**). ''A review of the ultrasonic bioeffects of microsonation, gas-body activation, and related cavitation-like phenomena,'' Ultrasound Med. Biol. **13**, 443–470.

Miller, D. L., and Bao, S. (**1998**). ''The relationship of scattered subharmonic, 3.3 MHz fundamental and second harmonic signals to damage of monolayer cells by ultrasonically activated Albunex,'' J. Acoust. Soc. Am. **103**, 1183–1189.

Miller, D. L., and Quddu, J. (**2000**). ''Sonoporation of monolayer cells by diagnostic ultrasound activation of contrast-agent gas bodies,'' Ultrasound Med. Biol. **26**, 661–667.

Miller, D. L., and Quddus, J. (**2001**). ''Lysis and sonoporation of epidermoid and phagocytic monolayer cells by diagnostic ultrasound activation of contrast agent gas bodies,'' Ultrasound Med. Biol. **27**, 1107–1113.

Miller, D. L., and Quddus, J. (**2002**). ''Diagnostic ultrasound induced membrane damage in phagocytic cells loaded with contrast agent and its relation to Doppler mode images,'' IEEE Trans. Ultrason. Ferroelectr. Freq. Control **49**, 1094–1102.

Miller, D. L., and Thomas, R. M. (**1995**). ''Ultrasound contrast agents nucleate inertial cavitation *in vitro*,'' Ultrasound Med. Biol. **21**, 1059–1065.

Miller, D. L., and Dou, C. (**2004a**). ''Membrane damage thresholds for pulsed or continuous ultrasound in phagocytic cells loaded with contrast agent gas bodies,'' Ultrasound Med. Biol. **30**, 405–411.

Miller, D. L., and Dou, C. (**2004b**). ''Membrane damage thresholds for 1–10 MHz pulsed ultrasound exposure of phagocytic cells loaded with contrast agent gas bodies *in vitro*,'' Ultrasound Med. Biol. **30**, 973–977.

Miller, M. W., Miller, D. L., and Brayman, A. A. (**1996**). ''A review of *in vitro* bioeffects of inertial ultrasonic cavitation from a mechanistic perspective,'' Ultrasound Med. Biol. **22**, 1131–1154.

Nyborg, W. L. (**1977**). *Physical Mechanisms for Biological Effects of Ultrasound* (Bureau of Radiological Health, Rockville, MD).

Nyborg, W. L., Carson, P. L., Carstensen, E. L., Dunn, F., Miller, M. W., Miller, D. L., Thompson, H. E., and Ziskin, M. C. (**2002**). *Exposure Criteria for Medical Diagnostic Ultrasound: II. Criteria Based on All Known Mechanisms*. Report No. 140 (National Council on Radiation Protection and Measurements, Bethesda, MD).

Rooney, J. A. (**1972**). ''Shear as a mechanism for sonically induced biological effects,'' J. Acoust. Soc. Am. **52**, 1718–1724.

Skyba, D. M., Price, R. J., Linka, A. Z., Skalak, T. C., and Kaul, S. (**1998**). ''Direct *in vivo* visualization of intravascular destruction of microbubbles by ultrasound and its local effects on tissue,'' Circulation **98**, 290–293.

Stride, E., and Saffari, N. (**2003**). ''On the destruction of microbubble ultrasound contrast agents,'' Ultrasound Med. Biol. **29**, 563–573.

Wu, J. (**2002**). ''Theoretical study on shear stress generated by microstreaming surrounding contrast agents attached to living cells,'' Ultrasound Med. Biol. **28**, 125–129.

Vocal tract filtering and the "coo" of doves

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

Tobias Riede
Institute for Theoretical Biology, Humboldt-University Berlin, Invalidenstrasse 43, 10115 Berlin, Germany

Gabriël J. L. Beckers
Behavioural Biology, Institute of Biology, Leiden University, P.O. Box 9516, 2300 RA Leiden, The Netherlands

Roderick A. Suthers
School of Medicine, Jordan Hall, Indiana University, Bloomington, Indiana 47405

(Received 16 March 2004; revised 12 August 2004; accepted 12 September 2004)

Ring doves (*Streptopelia risoria*) produce a ''coo'' vocalization that is essentially a pure-tone sound at a frequency of about 600 Hz and with a duration of about 1.5 s. While making this vocalization, the dove inflates the upper part of its esophagus to form a thin-walled sac structure that radiates sound to the surroundings. It is a reasonable assumption that the combined influence of the trachea, glottis and inflated upper esophagus acts as an effective band-pass filter to eliminate higher harmonics generated by the vibrating syringeal valve. Calculations reported here indicate that this is indeed the case. The tracheal tube, terminated by a glottal constriction, is the initial resonant structure, and subsequent resonant filtering takes place through the action of the inflated esophageal sac. The inflated esophagus proves to be a more efficient sound radiating mechanism than an open beak. The action of this sac is only moderately affected by the degree of inflation, although an uninflated esophagus is inactive as a sound radiator. These conclusions are supported by measurements and observations that have been reported in a companion paper. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1811491]

PACS numbers: 43.80.Ka [JAS] Pages: 3750–3756

I. INTRODUCTION

Birdsong has a wide variety of forms, as described in the classic book by Greenewalt (1968). Some birds produce melodically rich extended songs with individual notes that are almost pure-tone in some species but rich in harmonics in others. Other birds, such as crows, may produce simple vocalizations with harmonically rich spectra that are shaped into formant bands as in human speech. Some cockatoos even produce chaotic nonharmonic cries (Fletcher, 2000). Surveys of the mechanisms involved in some of these cases have been given by Brackenbury (1982) and others, while quantitative models for some of these systems have been developed by Casey and Gaunt (1985), Fletcher (1988), and Fletcher and Tarnopolsky (1999). The case of Ring doves is rather different in that they often produce simple pure-tone coos in a relatively narrow frequency range, and do so with their upper esophagus inflated to a large sac that effectively radiates the sound while the beak remains closed. Apart from this feature, the vocal system of the dove differs little from that of other birds (Ballintijn *et al.*, 1995). Extensive studies of doves have recently been published by Beckers *et al.* (2003a, b) and by Riede *et al.* (2004), and these provide the experimental basis of the present paper. These papers contain numerous references to the relevant literature and provide the experimental data upon which the present paper is based.

The prominence of the inflated esophagus or esophageal sac (the abbreviation IE will be used henceforth—it is distinct from that part of the esophagus or crop, the *ingluvies*, in which food is stored) during song invites the conclusion that it is acoustically important, and this is the hypothesis that is explored here. It is possible, however, that it is also a visual display feature used in courtship, and it is certainly used in this way by some species. The fact that the beak is closed during song leads to the tentative conclusion that the resonances of the mouth and beak that are so important in the song of some other birds are irrelevant in the case of the dove, but that leaves the length of trachea connecting the syrinx to the expanded esophagus as a possible contributor to resonance. The sound passes through the glottis at the entry to the trachea, so that constriction of this passageway may also contribute to overall behavior.

The purpose of the present paper is thus to examine the role of these anatomical structures in providing a very efficient filter for the generally harmonically rich sound that is expected to be generated by the vibrating syringeal valve. Since mere speculation and modeling are inadequate, the model will be tested by comparison with the experimental data on song in the Ring dove (*Streptopelia risoria*) provided by the study of Riede *et al.* (2004).

II. ACOUSTIC MODEL

X-ray photographs of a dove while singing are given by Riede *et al.* (2004), and from these it is possible to derive

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

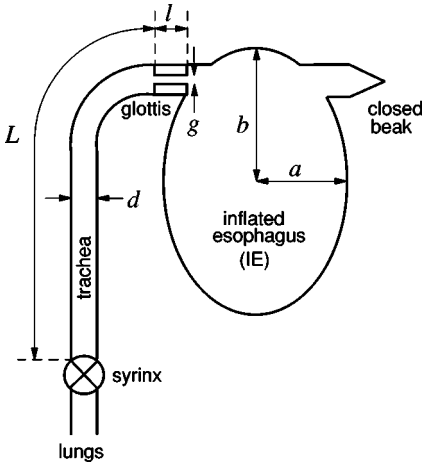


FIG. 1. Simplified geometry of the vocal tract of the Ring dove.

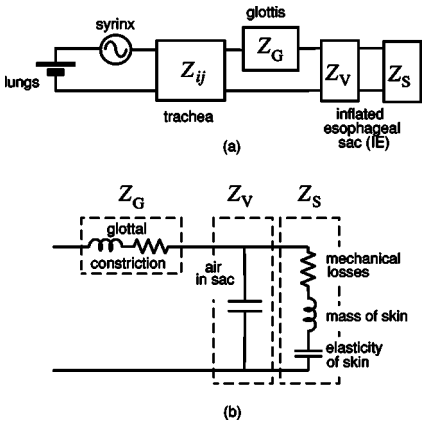


FIG. 2. (a) Electric network simulating the acoustic behavior of the vocal tract of the dove. Voltage is the analog of acoustic pressure, and current is the analog of acoustic volume flow. (b) More detailed representation of the analog impedances of the glottal constriction and the esophageal sac.

anatomical details. For the purpose of acoustic modeling, details such as the precise location and curvature of the trachea are unimportant, and the vocal system can be treated in terms of the simplified anatomy shown in Fig. 1. The anatomical dimensions, given by the data of Riede *et al.*, are shown in Table I.

Figure 2(a) shows an electric analog circuit that represents the acoustic system of Fig. 1. In such a circuit the analog of acoustic volume flow is electric current and the analog of acoustic pressure is electric potential (Fletcher, 1992). The lungs provide a constant pressure (voltage) of about 1.5 kPa (15 cm water gauge) and the syrinx behaves like a valve oscillating at a frequency $f=\omega/2\pi$ of about 600 Hz. Since the lung pressure is high relative to that in the trachea, the impedance of the syringeal source is high, and it can be considered to inject a volume flow of constant amplitude. Because the valve almost closes once in each cycle, the acoustic volume flow into the base of the trachea is rich in harmonic overtones (Fletcher, 1988).

It is interesting to note that, while the trachea and glottis constitute a series impedance, the components of the inflated esophageal sac constitute a parallel impedance. The interaction between these two impedances, themselves in series with one another, is a little complex and generally results in a pair of resonances, even if they are adjusted to the same resonance frequency. This need not be of concern, since it is automatically taken into account in the analysis.

A. Tracheal and glottal impedances

The tracheal tube is represented by a four-terminal element with impedance coefficients Z_{ij} given by (Kinsler *et al.*, 1982; Fletcher, 1992)

TABLE I. Anatomical dimensions.

Length of trachea	$L=75$ mm
Diameter of trachea	$d=3$ mm
Length of glottal constriction	$l=5$ mm
Diameter of glottal constriction	$g=0.8$ mm
Diameter of inflated esophagus	$2a=35$ mm
Length of inflated esophagus	$2b=50$ mm
Fraction of IE surface vibrating	0.5
Total mass of IE walls	$m=2$ g
Quality factor of IE resonance	$Q=10$

$$Z_{11}=Z_{22}=-jZ_0\cot kL,$$
$$Z_{12}=Z_{21}=-jZ_0\,\mathrm{cosec}\,kL. \tag{1}$$

Here L is the length of the trachea, $Z_0=\rho c/S_T$, where $\rho\approx1.2$ kg/m^3 is the density and $c\approx350$ m/s the velocity of sound in air at dove body temperature, $S=\pi d^2/4$ is the area of the trachea, $k=(\omega/c)-j\alpha$ is the complex wave number for sound of angular frequency $\omega=2\pi f$ and attenuation coefficient $\alpha\approx2.4\times10^{-5}\omega^{1/2}/d$ in the tracheal tube of diameter d, and $j=\sqrt{-1}$. The acoustic pressures p_i and acoustic volume flows U_i at the two ends of the tube are then related by

$$p_i=\sum_j Z_{ij}U_j. \tag{2}$$

If the impedance connected to the upper end of the trachea is $Z_L=p_2/U_2$, then it follows from these equations that the input impedance Z_{in} at the syringeal end of the trachea is

$$Z_{in}=\frac{p_1}{U_1}=\frac{Z_LZ_{11}-Z_{11}Z_{22}+Z_{12}Z_{21}}{Z_L-Z_{22}}$$
$$=Z_0\frac{Z_L\cos kL+jZ_0\sin kL}{jZ_L\sin kL+Z_0\cos kL}. \tag{3}$$

Since the syrinx is a high-impedance source, it is tempting to simply select a frequency to maximize the real part of this input impedance, since maximum power will then be transferred to the vocal tract, but this does not give information about radiated acoustic power—the input power may be simply dissipated in viscous and thermal losses within the tract.

The glottal constriction to a narrow tube of length l and diameter g can be represented, since it is so short and narrow, by a simple inductive impedance in series with a resistive viscous loss (Fletcher, 1992), giving an impedance

$$Z_G\approx\frac{4\rho c}{\pi g^2}\left(\frac{2.4\times10^{-5}l}{g}+j\frac{4\rho(l+0.6g)\omega}{\pi g^2}\right), \tag{4}$$

where l is the length and g is the diameter of the constriction. This is a valid approximation since the length of the

glottal constriction is much less than the sound wavelength involved.

B. Impedance of the inflated esophagus

When an acoustic current flows into the inflated esophagus from the trachea it tends to both compress the contained air and also to expand the elastic containing wall. The acoustic impedances of these two elements are effectively in parallel, as shown in Fig. 2, since they both experience the same acoustic pressure but the acoustic flow is divided between them. The acoustic impedance of the air contained within the IE, neglecting the effect of the small overpressure created by wall tension, is

$$Z_{\mathrm{V}} \approx -j\frac{\rho c^2}{V\omega} = -j\frac{\gamma p_0}{V\omega}, \tag{5}$$

where

$$V \approx \tfrac{4}{3}\pi a^2 b \tag{6}$$

is the volume of the IE, p_0 is the normal atmospheric pressure (100 kPa), and $\gamma = 1.4$ is the ratio of specific heats for air.

The walls of the IE, which expand and contract under the influence of the oscillating acoustic pressure within the IE and carry with them the overlying tissues and feathers, present an impedance that is the sum of an inertance due to the wall mass, a springlike term due to the elasticity of the walls, and a resistive term due to viscous and other losses within the material of the wall. The radiation load on the outside of the walls is small enough to be neglected. This leads to an expression for the wall impedance of the form

$$Z_{\mathrm{W}} \approx R + j\frac{m\omega}{S_{\mathrm{W}}^2} + jX, \tag{7}$$

where m is the total mass of the walls,

$$S_{\mathrm{W}} \approx 4\pi ab \tag{8}$$

is the total wall area, R takes account of losses in the walls, and the final term jX takes account of the elastic resilience of the walls, the form and magnitude of which will be discussed below. (Actually jX turns out to be negligible compared with the second term in the equation.) The total impedance presented by the IE to the trachea is then

$$Z_{\mathrm{S}} = \frac{Z_{\mathrm{W}} Z_{\mathrm{V}}}{Z_{\mathrm{W}} + Z_{\mathrm{V}}}. \tag{9}$$

Exact calculation of the contribution of elasticity to the total wall impedance is complicated by the fact that the esophagus has expanded primarily in the radial direction a, while the length $2b$ has only increased a little. The skin tension is thus not isotropic but probably concentrated in the plane normal to the length b of the esophagus, though the structure of the esophagus wall may influence this. If the uninflated esophageal tube is taken to have radius a_0 and length b_0, and the inflated sac to be approximately a prolate spheroid with short radius a and long radius $b \approx b_0$, then the surface area has expanded from about $2\pi a_0 b_0$ to about βab, where the factor β lies somewhere in the range $2\pi \leqslant \beta \leqslant 4\pi$ depending upon the ratio b/a, being smaller if $b > a$. As will be shown below, the exact value is not significant here. If K is the relevant elastic modulus of the wall material, multiplied by the wall thickness and adjusted for stretch, then the tension in the wall is largely in the equatorial plane and has a value close to

$$T \approx K(a - a_0), \tag{10}$$

and the equilibrium internal pressure excess Δp created to balance this tension is approximately

$$\Delta p \approx \frac{T}{a} \approx \frac{K(a - a_0)}{a}. \tag{11}$$

Equivalently, we can write

$$K \approx \frac{a\Delta p}{a - a_0}. \tag{12}$$

The elastic force per unit area of wall for an expansion of amplitude δa at frequency ω is $K\delta a/a$ while the inertial force is about $m\omega^2 \delta a/S$ or $m\omega^2 \delta a/\beta ab$. The ratio of inertial to elastic force at frequency ω is thus about

$$\frac{4\pi b \Delta p}{m\omega^2}\left(\frac{a}{a - a_0}\right) \sim 10^{-5}\Delta p\left(\frac{a}{a - a_0}\right), \tag{13}$$

where the second expression assumes that the frequency is about 600 Hz and the total wall mass about 1 g. Since $a \approx 10a_0$, the final factor is close to unity, and the IE overpressure Δp is only a few hundred pascals (i.e., a few centimeters of water pressure). The elastic force is therefore only about 10^{-2} times the inertial force, and so the term jX in Eq. (7) can be neglected.

Evaluation of the resistive term R in (7) is difficult without experimental measurements. Most biological materials are rather lossy because of the liquid-filled cells of which they are composed, so that the quality factor, or Q value, of the IE resonance, given by

$$Q = \frac{2\pi f^* m}{R}, \tag{14}$$

where f^* is the resonance frequency, is likely to be only about 10 and perhaps even less. This implies that R is probably about one-tenth of the magnitude of the inertial term at 600 Hz. Since the magnitude of the IE-wall damping is unknown, the choice of $Q = 10$ is speculative. A check of the calculated results for lower values of Q shows, however, that the only significant change is a reduction in the radiated power. There is little change in the range of required glottal constriction, since this serves mainly to tune the tracheal resonance.

C. Effective mass of the IE wall

The other major parameter required for the model is an estimate of the mass m of the IE walls. In the absence of measurements, the following argument provides an approximate value.

From the x-ray photographs of the neck of a dove with inflated esophagus (Riede *et al.*, 2004), it appears the diameter of the neck of the bird is about 15 mm, so that its

cross-sectional area is about 2 cm^2. Approximately half of this cross-section is occupied by the spine, the trachea, and the interior of the esophagus tube, so that the area of tissue available to cover the surface of the expanded esophagus is about 1 cm^2, or equivalently about 1 g of tissue per centimeter length of neck. Since the neck is about 5 cm in length, this provides an upper limit of about 5 g for the mass of the membrane surrounding the IE. To adopt this value would, however, be a gross overestimate, since some of the tissue is muscles, sinews and blood vessels. A more conservative estimate of 1 to 2 g does, however, seem reasonable.

The simplified treatment leading to (7), however, is based upon the assumption that all parts of the IE walls vibrate equally, so that the expression for the acoustic impedance involves the simple factor m/S_W^2, where S_W is the total area of the IE walls. Because, however, the IE is about the shape of a prolate spheroid with diameter ratio about 5:3, it is most likely that the walls near the equator are thinner and vibrate with greater amplitude than those near the upper and lower ends of the IE. Suppose, as an extreme case, that vibration was confined to an equatorial band of mass m' and area $S'=S_W/2$. To achieve the same acoustic impedance, since the vibrating area is halved, the vibrating mass must be reduced to one-quarter of the total original value, or to half of the mass originally spread over that area. Thus a total wall mass of 2 g, for example, would have the same acoustic effect as 4 g distributed evenly over a symmetrically vibrating sphere of the same surface area. In the calculations to follow, the mass will therefore be regarded as a parameter that can be adjusted to some degree but that must be kept within these reasonable anatomical limits.

These estimates have been confirmed by examination of a dove body that had been kept in a sealed bag in a refrigerator for about 12 months. Dissection of the neck and discarding of those tissues that are clearly not part of the inflatable esophagus left a residual mass of 1.14 g along a neck length of between 4 and 5 cm. This is probably an underestimate of the real mass, however, partly because of slight drying of the tissues during storage, and partly because of loss of a few feathers during dissection. The value of 2 g adopted in the calculation is therefore supportable.

D. Effect of esophageal inflation

During the ''coo'' vocalization the dove exhales air into the esophageal sac, further inflating it. It is important to know the effect of this inflation on the resonance frequency of the IE. The ''coo'' of a single vocalization lasts for about 1.5 s and, during that time, the dove exhales about 10 cm^3 of air into the IE. The resonance frequency of the IE is given approximately by

$$\omega^* \approx \left(\frac{\gamma p_0 S^2}{Vm} \right)^{1/2}. \tag{15}$$

Since $V \propto a^2 b$ and $S \propto ab$, this indicates that $\omega^* \propto b^{1/2}$. But the length b of the IE is nearly constant and only the radius a increases with inflation, so that the resonance frequency ω^* is nearly independent of further inflation.

Even in the case of a IE expanding uniformly in all directions, the resonance frequency varies only as $b^{1/2}$ or $V^{1/6}$. The initial IE volume is about 30 cm^3 and this could be expanded to about 40 cm^3 during the coo, an increase of about 30%. This would lead to an increase of only about 5% in the resonance frequency, and this, it should be stressed, is an upper limit to the real situation. A realistic interpolation is to assume that the longitudinal axis bends slightly, since it is constrained on one side by the bird's neck, and to take

$$b = b_0 \left(1 + \epsilon \frac{a^2}{b_0^2} \right), \tag{16}$$

where b_0 is the uninflated sac length and the numerical coefficient $\epsilon \leqslant 0.6$ gives an indication of the extent to which sac expansion occurs in the lengthwise b-direction for a given lateral expansion in the a-direction. A choice of $\epsilon = 0.1$ seems a reasonable estimate from the appearance of x-ray images of the singing bird, and will be used in a later evaluation. The frequency shift in this case will be intermediate between the ''no-shift'' result if b is constant and the small shift that would be the result of uniform expansion.

The effect of complete deflation of the IE so that a is reduced to about 1.5 mm is, however, quite a different matter. As will be seen in a later calculation, such deflation increases the effective impedance of the IE by several orders of magnitude and effectively prevents significant sound radiation.

III. PERFORMANCE CALCULATION

Because the syrinx is fed from the relatively high overpressure in the bird's lungs, it injects an oscillating volume flow into the trachea, the magnitude of which depends very little upon the acoustic impedance presented by the trachea and associated structures. At the other end of the system, the acoustic power P radiated in the ''coo'' sound is

$$P = C U_S^2 \omega^2, \tag{17}$$

where C is a constant. This power is thus proportional to the square of the amplitude of the acoustic current passing through the extreme right branch of the analog circuit in Fig. 2(b), multiplied by the square of the frequency. This simplification is appropriate since the IE diameter is small compared with the wavelength of sound at the frequencies considered, so that it can be treated as a ''simple source'' (Morse, 1948; Fletcher, 1992, Chap. 7).

From a standard analysis of the analog circuit in Fig. 2(b), and writing each impedance Z_i as $R_i + jX_i$, it can be readily shown that the acoustic current U_S through the impedance Z_S representing the IE walls when a current U is injected at the syrinx is given by

$$U_S^2 = \left[\frac{(X_{12}X_V)^2 + (R_{12}X_V)^2}{A^2 + B^2} \right] U^2, \tag{18}$$

where

$$\begin{aligned} A &= R_S(R_{22}+R_G) - (X_S+X_V)(X_{22}+X_G) - X_V X_S, \\ B &= R_S(X_{22}+X_G) + (X_S+X_V)(R_{22}+R_G) + X_V R_S. \end{aligned} \tag{19}$$

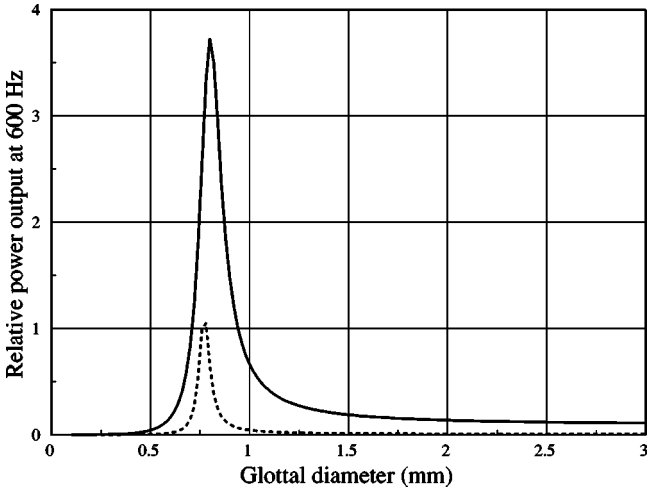


FIG. 3. Calculated relative acoustic power output at 600 Hz as a function of the diameter of the glottal constriction, assuming the parameter values shown in Table I. The broken curve shows the result of reducing the effective mass of the IE walls to 1 g, thus increasing the IE resonance frequency.

Since the most readily adjustable parameter from the viewpoint of the bird is the diameter of the glottal constriction, the full curve in Fig. 3 shows an analysis of the effect that this has on output power, assuming a ''coo'' frequency of 600 Hz, and with all the other parameters having the values given in Table I. It is clear that the power output depends critically upon this vocal adjustment. The full curve in Fig. 4 shows the relative power output as a function of frequency, assuming that the glottal constriction has been optimized in this way. It seems clear that the dove must learn to constrict the glottis in just this manner in order to be able to produce a ''coo'' sound with reasonable power. It is possible that the dove can also make some adjustment to the length of the trachea, as is done by some other birds, and alter the tracheal resonance frequency in this way, but the required extension is by about a factor of 2, unless glottal constriction is also

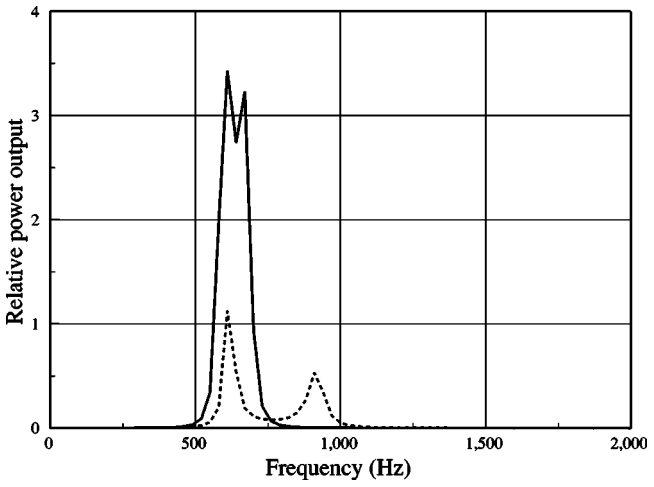


FIG. 4. Calculated relative power output as a function of coo frequency assuming the glottal constriction to be adjusted to the value given in Table I so as to optimize power output at 600 Hz. (The double peak is due partly to slight misalignment of resonance frequencies and partly to the series-parallel resonance arrangement shown in Fig. 2.) The broken curve shows the effect on power output of reducing the effective mass of the IE walls to 1 g, thus increasing the IE resonance frequency to about 900 Hz.

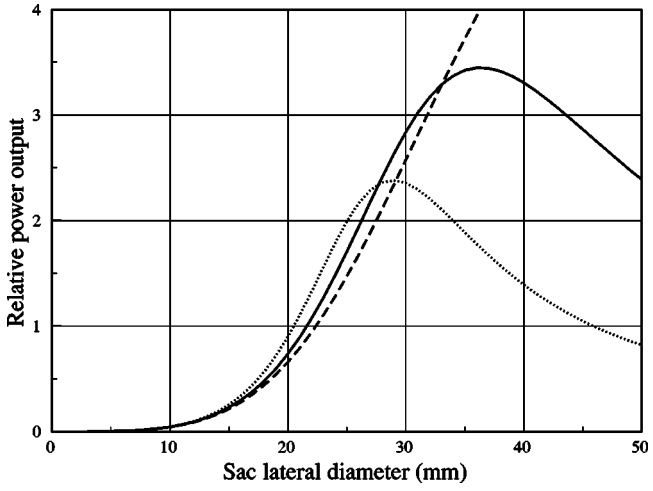


FIG. 5. Calculated relative acoustic power at 600 Hz as a function of the IE lateral diameter $2a$, assuming that the expansion parameter $\epsilon=0.1$ and that other parameters are as in Table I. The broken curve shows the calculated result if $\epsilon=0$, so that the longitudinal dimension $2b$ of the IE remains constant during the expansion, while the dotted curve shows the result if $\epsilon=0.3$.

invoked, and there is no evidence that the dove can do this.

Figure 4 also has another implication. The syrinx is a nonlinear flow regulator and produces not only a flow at its fundamental oscillation frequency, but also harmonic components at frequencies that are precise integer multiples of that frequency. The flow amplitudes associated with these harmonics are less than that of the fundamental, but still not negligible. A pressure sensing microphone placed in the trachea will therefore detect these harmonics, as has been established by Riede *et al.* (2004). The calculation leading to Fig. 4 shows, however, that the efficiency with which these harmonics influence the motion of the skin of the IE is vanishingly small, so that they do not show up in the radiated sound. The IE effectively acts as an acoustic band-pass filter tuned to the song frequency near 600 Hz.

Since the full curves in Figs. 3 and 4 were calculated on the basis of a IE resonance that is close to the coo frequency of 600 Hz, it is interesting to see what happens if the IE resonance is adjusted away from this frequency. The broken curves in Figs. 3 and 4 are calculated with the vibrating mass of the IE wall reduced by a factor of 2 to 1 g, so that its resonance frequency is raised by a factor of $2^{1/2}\approx 1.4$. The output power at 600 Hz is greatly reduced even with an optimized glottal constriction, as shown in Fig. 3, though there is a minor peak at about 900 Hz in Fig. 4. The coo frequency could, of course, be adjusted to 900 Hz and a related adjustment made to the glottal constriction to take advantage of this resonance, but it would then be the call of a different bird.

Figure 5 examines the effect of IE inflation upon radiated sound. The full curve is calculated on the assumption that the longitudinal inflation coefficient ϵ of (16) has the value 0.1, which appears to be a reasonable approximation to the real situation. There is a broad peak in the output near the actual IE diameter value $2a=35$ mm. More importantly, however, the curve shows that the output declines almost to zero when the IE is deflated, agreeing with experimental ob-

servations. Since the exact value of ϵ influences the position of the peak and thus the optimal level of IE inflation, it is useful to examine also the limiting case $\epsilon=0$, corresponding to a fixed longitudinal dimension $2b$. This is shown as a broken curve in the figure. The output now increases steadily with IE inflation, but is still reduced to near zero for a deflated IE. Finally, the dotted curve in the figure shows the result if we assume a value $\epsilon=0.3$, which is probably rather larger than is appropriate in reality.

A. Tracheal vibration

In the dove, the trachea is a ringed tube with rather thin walls between the rings, which raises a question about ways in which this structure might contribute directly to the radiated sound, since the acoustic pressure within it is quite high. There are two ways in which this might happen, first by direct radiation through the surrounding tissue, and second by mechanical transmission of vibrations to excite the esophageal sac.

Quantitative consideration shows, however, that these two contributions are probably negligibly small. In regard to direct radiation, the radiated sound power for a given wall motion is proportional to the square of the area of the vibrating wall, and the sac wall area is an order of magnitude larger than that of the tracheal walls. The tracheal wall vibrational amplitude is also unlikely to be much larger than that of the sac walls because of the restricting effect of the cartilage rings along its length and the mass of the surrounding tissue.

Transmission of vibration directly from the tracheal walls to the inflated sac could certainly occur but, because of the nature of the quarter-wave tracheal resonance, it would actually be 180° out of phase with the excitation produced by air flow through the glottis into the sac, and would thus reduce rather than increase the radiated sound. Once again, however, the actual magnitude of this effect is likely to be negligible compared with the major mechanism discussed above, as can be established by examining resonance conditions in a slightly expansible tube.

B. Comparison with open-beak calls

It is interesting to examine the acoustic reasons underlying the strategy of esophageal inflation in doves, as contrasted with simple vocalization through an open beak. Certainly the pure-tone nature of the dove coo provides clear species identification, but much of the reason for vocalization lies with the communication range achievable. How does the acoustic output from an inflated esophagus compare with that which the bird could achieve in an open-beak call?

The simplest way to examine this is to compare the acoustic power in each type of call with that which could be produced from the trachea and glottis tuned to the same frequency and simply vented to the environment. In all cases, glottis, beak or esophageal sac, the maximum dimensions of the radiating structure are small compared with the acoustic wavelength at the coo frequency of 600 Hz. An analysis treating the radiation as that from a ''simple source'' (Morse, 1948; Fletcher, 1992, Chap. 7) therefore suffices, and the expression in (17) is an appropriate relative output measure.

A problem arises, however, in deciding what to keep constant in such a comparison. The actual acoustic power produced in any biological, or indeed musical, system is usually only a small fraction of the input power, which is in this case the product of the lung pressure (about 1.5 kPa) and the mean volume flow (about 10 $\mathrm{cm}^3\,\mathrm{s}^{-1}$), giving about 15 mW. The radiated acoustic power, on the other hand, is typically less than 1 mW or about 80 dB at a distance of 1 m. It is therefore reasonable to assume that, in terms of vocal effort, it is the total input power that should be kept constant in any comparison, rather than the acoustic input power, and this amounts to keeping the volume flow amplitude U_1 through the syrinx constant.

The analysis developed earlier in Sec. III can now be used to compare the acoustic outputs from the IE and from a simple vented glottis. The first case has already been calculated and displayed in the figures. To simulate radiation from the glottis in the absence of the IE, it is adequate to simply set the effective sac mass m equal to zero in the calculation. The result, when output power is plotted against frequency, is a peak similar to that in Fig. 4, but reduced in amplitude by a factor of about 12, corresponding to a decrease in radiated sound level of about 11 dB, and with a smaller peak at around 2300 Hz representing the mistuned second resonance of the trachea and glottal constriction.

A similar calculation is now, in principle, required for radiation from the beak. The beak is acoustically complex, however, even at low frequencies where beak resonances are not involved. The acoustic behavior of a beak model has, however, been investigated in detail by Fletcher and Tarnopolsky (1999), and their results show that, to a reasonable approximation, a partly opened beak imparts an acoustic power gain of about 6 dB, or a factor of 4, compared with the power that would be radiated from the open glottis, assuming that the acoustic input power is the same in each case. If the total input power, rather than the acoustic power, is kept constant, then the beak gain will be rather less than this, say 4 or 5 dB.

From this analysis it is thus reasonable to conclude that the inflated-esophagus strategy gains the dove a significant advantage of about 6 dB in terms of radiated power at the fundamental coo frequency, compared with a normal call through the beak. This numerical value will depend upon the value of the quality factor Q of the IE resonance, increasing with increasing Q. The beak-radiated call will, however, also contain some power at higher harmonic frequencies that is filtered out by the IE. The IE coo possesses the probable advantage of being very different in acoustic spectrum from a typical beak-radiated call, which may aid in conspecific communication.

IV. CONCLUSIONS

This detailed analysis of the acoustics of the vocal system of the dove leads to several related conclusions, as follows.

(i) The vocal tract as a whole is tuned to the fundamental frequency of the dove coo. This involves both passive anatomical tuning of the inflated esophageal sac and

also active tuning by the bird of the glottal constriction coupling the trachea to the esophagus.

(ii) To produce maximum acoustic output, the resonance frequency of the IE must be approximately matched to the song frequency. This matching depends upon the total vibrating mass of the IE walls and the degree to which the IE is inflated. The resonance frequency is, however, not critically dependent upon IE volume, so that the inflation caused during a single coo does not greatly affect this matching.

(iii) To produce maximum acoustic power, the glottis must also be constricted so that the resonance of the trachea with glottal termination approximately matches the song frequency.

(iv) It is reasonable to suppose that young birds have to learn to constrict the glottis appropriately before they can produce satisfactory coos.

(v) As well as eliminating upper harmonics to produce a characteristic pure-tone coo, the evolved strategy of radiating the coo of the dove by means of an inflated esophagus rather than an open beak gives a significant increase in radiated acoustic power, and thus in the range of audibility of the call.

The analysis reported here assumes a simple source/filter mechanism and neglects the feedback influence of the vocal tract upon the syringeal oscillation. This coupling, the mechanics of which has been discussed elsewhere (Fletcher, 1988), can be expected to magnify the resonance effects discussed here and to lead to an even closer dependence of power output upon glottal tuning.

While this analysis does not claim to resolve all the questions associated with dove calls, it does appear to provide a reasonable basis for understanding the origin of the almost pure-tone sound and the function of the glottal constriction. By presenting a clear and quantitative model for the way in which the vocal filtering system might operate, it paves the way for detailed experimental exploration of the hypotheses underlying the theory, in particular,

(i) the assumption that the glottis is actively constricted during the coo song to produce a tracheal resonance at the coo frequency and

(ii) the values assumed for the physical variables such as IE wall mass and vibration distribution.

In passing, it could be noted that there is a surprising resemblance between the dove, with its inflated esophagus and pure-tone coo, and the bladder cicada *Cystosoma Saundersii* (Westwood), whose abdomen consists simply of a large hollow ellipsoid, the resonance of which selectively reinforces the fundamental of the 800-Hz sound generated by repeatedly buckling tymbals (Fletcher and Hill, 1978).

ACKNOWLEDGMENTS

TR was supported by a fellowship within the Postdoctoral Program of the German Academic Exchange Service (DAAD). General support of the program from NIH-NINDS Grant No. NS029467 to RAS is also acknowledged.

Ballintijn, M. R., ten Cate, C., Nuijens, F. W., and Berkhoudt, H. (**1995**). ''The syrinx of the collared dove (*Streptopelia decaocto*): structure interindividual variation and development,'' Netherlands J. Zool. **45**, 455–479.

Beckers, G. J. L., Suthers, R. A., and ten Cate, C. (**2003a**). ''Pure-tone birdsong by resonance filtering of harmonic overtones,'' Proc. Natl. Acad. Sci. U.S.A. **100**, 7372–7376.

Beckers, G. J. L., Suthers, R. A., and ten Cate, C. (**2003b**). ''Mechanisms of frequency and amplitude modulation in ring dove song,'' J. Exp. Biol. **206**, 1833–1843.

Brackenbury J. H. (**1982**). ''The structural basis of voice production and its relationship to sound characteristics,'' in D. E. Kroodsma and E. H. Miller (editors) *Acoustics Communication in Birds* Vol. **1** p. 53–73 (Academic Press, New York).

Casey, R. M., and Gaunt, A. S. (**1985**). ''Theoretical models of the avian syrinx,'' J. Theor. Biol. **116**, 45–64.

Fletcher, N. H. (**1988**). ''Bird song—a quantitative acoustic model,'' J. Theor. Biol. **135**, 455–481.

Fletcher, N. H. (**1992**). *Acoustic Systems in Biology*, (Oxford U.P., New York), Chaps. 8–12.

Fletcher, N. H. (**2000**). ''A class of chaotic bird calls?'' J. Acoust. Soc. Am. **108**, 821–826.

Fletcher, N. H., and Hill, K. G. (**1978**). ''Acoustics of sound production and of hearing in the bladder cicada *Cystosoma Saundersii* (Westwood),'' J. Exp. Biol. **72**, 43–55.

Fletcher, N. H., and Tarnopolsky, A. (**1999**). ''Acoustics of the avian vocal tract,'' J. Acoust. Soc. Am. **105**, 35–49.

Greenwalt, C. H. (**1968**). *Bird Song: Acoustics and Physiology* (Smithsonian Institution, Washington, D.C.).

Kinsler, L. E., Frey, A. R., Coppens, A. B., and Sanders, J. V. (**1982**). *Fundamentals of Acoustics* (Wiley, New York).

Morse, P. M. (**1948**). *Vibration and Sound* (McGraw–Hill, New York; reprinted by Acoustical Society of America, 1981), pp. 312–314.

Riede, T., Beckers, G., Blevins, W., and Suthers, R. (**2004**). ''Inflation of the esophagus and vocal tract filtering in Ring doves,'' J. Exp. Biol. **207**, 4025–4036.

Geographic variations in the whistles of spinner dolphins (*Stenella longirostris*) of the Main Hawai'ian Islands

Carmen Bazúa-Durán[a)]
Department of Oceanography, SOEST, University of Hawai'i, 1000 Pope Road MSB, Honolulu, Hawai'i 96822

Whitlow W. L. Au
Hawai'i Institute of Marine Biology, University of Hawai'i, P.O. Box 1106, Kailua, Hawai'i 96734

(Received 20 August 2003; revised 25 June 2004; accepted 3 July 2004)

Geographic variations in the whistles of Hawai'ian spinner dolphins are discussed by comparing 27 spinner dolphin pods recorded in waters off the Islands of Kaua'i, O'ahu, Lāna'i, and Hawai'i. Three different behavioral states, the number of dolphins observed in each pod, and ten parameters extracted from each whistle contour were considered by using clustering and discriminant function analyses. The results suggest that spinner dolphin pods in the Main Hawai'ian Islands share characteristics in approximately 48% of their whistles. Spinner dolphin pods had similar whistle parameters regardless of the island, location, and date when they were sampled and the dolphins' behavioral state and pod size. The term ''whistle-specific subgroup'' (WSS) was used to designate whistle groups with similar whistles parameters (which could have been produced in part by the same dolphins). The emission rate of whistles was higher when spinner dolphins were socializing than when they were traveling or resting, suggesting that whistles are mainly used during close-range interactions. Spinner dolphins also seem to vary whistle duration according to their general behavioral state. Whistle duration and the number of turns and steps of a whistle may be more important in delivering information at the individual level than whistle frequency parameters.
 [DOI: 10.1121/1.1785672]

PACS numbers: 43.80.Ka [FD] Pages: 3757–3769

I. INTRODUCTION

In some species, geographically separated groups can be identified by differences in their acoustic emissions or phonations. Differences between neighboring groups are usually named microgeographic variations, while differences occurring between groups over long distances which do not normally mix are called macrogeographic variations (Krebs and Kroodsma, 1980; Conner, 1982; Mundinger, 1982; McGregor *et al.*, 2000). Geographic variations can provide valuable information because they may reflect dispersal capabilities of a species (Mundinger, 1982; McGregor *et al.*, 2000), adaptation to local ecological conditions (Marler, 1960; Thielcke, 1965), and the extent of isolation and genetic divergence between groups or populations (Lemon, 1966; McGregor *et al.*, 2000; Ford, 2002). The study of differences in the phonations of cetaceans may be useful in providing insights into the social organization and association patterns among individuals or pods (Whitehead *et al.*, 1998; Deecke *et al.*, 1999; Miller and Bain, 2000; Yurk *et al.*, 2002).

It is extremely difficult, if not impossible, to design passive acoustic experiments with wild cetaceans. Encounter rates, location, pod size, behavioral context, and number of dolphins phonating cannot be controlled. Therefore, in this study sampling was performed across the fullest possible range of potential situations so that comparisons incorporated components that are known to affect the production of phonations, such as behavioral state, location, environmental conditions, and feeding strategy. The existence of microgeographic variations in the whistles of Hawai'ian spinner dolphins is discussed in this paper by comparing the whistles of spinner dolphins recorded off four of the Main Hawai'ian Islands (MHI), the islands of Kaua'i, O'ahu, Lāna'i, and Hawai'i. The whistles of the different spinner dolphin pods recorded in this study were compared between and within each island. In addition, behavioral state and pod size were considered in order to investigate if these variables influenced the differences found between whistles. To our knowledge, no other study has examined the existence of geographic variations in the whistles of spinner dolphins.

Along the Kona coast of the Island of Hawai'i (Norris *et al.*, 1994; Östman, 1994) and in the Lāna'i–Maui area (Pacific Whale Foundation, 1998, 2000, personal communication), spinner dolphins live in a fission-fusion society. Individual dolphins in a pod change from day to day, and even within a day. A pod may split into several subpods and pods that come in contact may exchange individuals or merge completely. Within each island, spinner dolphins are known to travel along the coast (Bazúa Durán, 2001). In O'ahu, spinner dolphins travel between the western (Wai'anae) and south shores, tend to spend time at the Wai'anae Shore, and just traverse along the South Shore (Fig. 1). In Lāna'i, spinner dolphins frequent the South Shore quite regularly and spend time at Mānele Bay, while in the Island of Hawai'i, spinner dolphins tend to frequent the western (Kona) coast and spend time in Kealakekua Bay (Fig. 1). Spinner dolphins

[a)]Current address (corresponding author): Dep. Física, Fac. Ciencias, UNAM, Ciudad Universitaria, 04510 México, D.F., México. Electronic mail: bazua@servidor.unam.mx

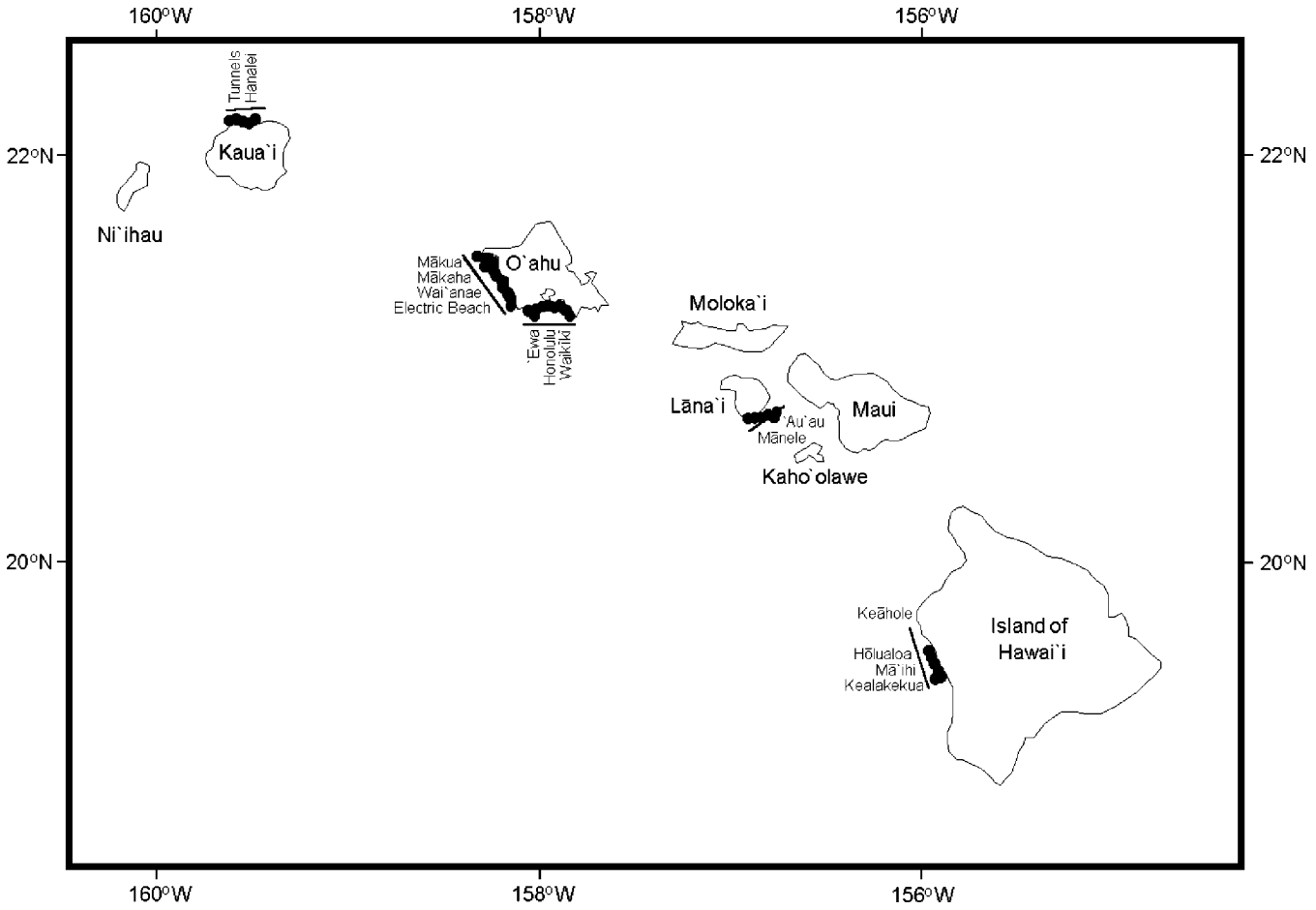


FIG. 1. Locations of sightings of spinner dolphin pods along the coastal waters of four Hawai'ian Islands during the field trips from 1998 to 2000. Lines represent the regions along the coast inspected for spinner dolphins.

travel between the islands of O'ahu and Moloka'i (Maldini, 2003), however, it is not known how much mixing exists between pods found in the different islands. Spinner dolphins in the Main Hawai'ian Islands are in such close geographic proximity that direct interbreeding and social interaction can occur.

II. METHODOLOGY

Field trips along the coastal waters of the islands of Kaua'i, O'ahu, Lāna'i, and Hawai'i in the tropical north Pacific Ocean (21°N, 157°W) were conducted from 1998 to 2000 as shown in Fig. 1. Data were collected on the whistles emitted, general behavioral state, and pod size of spinner dolphins using small, outboard-driven motorboats that ranged from 4 to 10 m in length. In this study the term pod was used to describe the assembly of dolphins encountered in a specific place and time, without considering the composition of the dolphin assemblage. The use of the word group was only used to refer to whistle groups. Spinner dolphins were recorded three times in 2000 for Kaua'i; once in 1998, twice in 1999, and seven times in 2000 for O'ahu; three times in 1998 and five times in 1999 for Lāna'i; and twice in 1998 and four times in 1999 for the Island of Hawai'i (Table I). The acoustic recordings of each spinner dolphin pod were considered independent samples. However, as identification of individual dolphins within a pod was not obtained, pods sighted on different days could be composed of the same animals. We could only be certain that the composition of the pod was different for encounters had on the same day at widely separated (more than 5 km) locations.

Behavior parameters were simultaneously annotated by voice on the second channel of the tape recorder and manually on data sheets. The general behavioral state of a spinner dolphin pod could be unambiguously classified as resting, socializing, or traveling [spinner dolphins are nocturnal feeders (Norris *et al.*, 1994) and recordings were only made during daylight hours]. A pod was considered resting if the interval between successive surfacings was greater than 2 min and animals were coordinated in their movements. It was considered socializing when dolphins were physically interacting among themselves. Spinner dolphins were traveling if they were not resting or socializing and were moving towards a specific direction without detours or stops. The maximum size of each spinner dolphin pod was estimated to within ± 5 if individuals numbered below 40, by ± 10 if individuals numbered between 40 and 80, by ± 20 if individuals numbered between 80 and 150, and by ± 50 if individuals numbered more than 150.

Whistles were recorded using an International Transducer Corporation (ITC) spherical hydrophone 1032 [receiving response is flat to about 35 kHz (± 3 dB)] and a portable two-channel Sony® Digital Audio Tape (DAT) TCD-D8 re-

TABLE I. Pods of spinner dolphins that compose the whistle-specific subgroups in four Hawai'ian Islands (Kaua'i, O'ahu, Lāna'i, and the Island of Hawai'i) with the dates and locations where each pod was recorded, the number of spinner dolphins in each pod, and the number of whistles analyzed for each spinner dolphin pod.

Whistle-specific subgroup		Island	Whistle group or pod recorded	Date recorded	Location	Pod size	No. of whistles
WSS1	Oa	O'ahu	OG1	13 Aug 1998	Wai'anae	80	315
			OG6	29 Jun 2000	Electric Beach	80	369
			OG7	30 Jun 2000	Electric Beach	50	139
			OG8	01 Jul 2000	Electric Beach	60	251
			OG9	01 Jul 2000	Wai'anae	50	94
			OG10	02 Jul 2000	Mākaha Beach	50	224
	Oc	O'ahu	OG4	15 Jun 2000	Honolulu-Waikīkī	60	199
	Lb	Lāna'i	LG3a,b	17 Aug 1998	W Mānele-Mānele	150	291
			LG4	15 Jun 1999	'Auau Channel	100	269
			LG8	19 Jun 1999	Mānele Bay	130	275
	Ha	Hawai'i	HG1	19 Jun 1998	Kealakekua Bay	50	126
			HG2	23 Jun 1998	Kealakekua Bay	70	215
			HG4	23 Mar 1999	Kealakekua Bay	35	10
			HG6	25 Mar 1999	Kealakekua Bay	50	151
		Kaua'i	KG1	12 Aug 2000	Hanalei Bay	60	499
			KG3	15 Aug 2000	Hanalei mouth	60	172
WSS2	Lc	Lāna'i	LG5	17 Jun 1999	E Mānele Bay	60	142
			LG6	17 Jun 1999	E Mānele Bay	60	25
			LG7	17 Jun 1999	E Mānele Bay	60	114
	Hb	Hawai'i	HG3	23 Mar 1999	Keauhou Bay	50	277
WSS3	Ob	O'ahu	OG2	31 Mar 1999	Mākua Beach	110	403
			OG3	02 Apr 1999	Mākua Beach	80	297
			OG5	22 Jun 2000	Honolulu	100	322
	La	Lāna'i	LG1	15 Aug 1998	'Auau Channel	100	241
			LG2a,b	16 Aug 1998	E Mānele-Mānele	60	548
		Kaua'i	KG2	14 Aug 2000	Tunnels	60	312
none	Hc	Hawai'i	HG5	24 Mar 1999	Keauhou Bay	80	182

corder with an upper frequency limit of 24 kHz. The boat engines were turned off while recording whistles. The acoustic recording stopped when the animals were too far away (approximately 400 m depending on environmental conditions) to be clearly seen and the boat was then repositioned. The likelihood of recording whistles from different dolphins was enhanced by recording for several hours, repositioning the boat to different locations with respect to the dolphin pod (spinner dolphins tend to retain spatial structure within the pod regardless of the position of the boat), and by the fact that pods were large (between 35 and 150 individuals, Table I). However, the number of dolphins whistling in a pod was never ascertained and the recordings probably contained some whistles produced by the same individuals. The behavioral state (and possibly pod size) contributed the most in the amount of whistles recorded. Socializing spinner dolphins emit more whistles (7.1 ± 4.0 whistles/min, $N=16$) than traveling (5.8 ± 4.3 whistles/min, $N=7$) and resting ones (3.7 ± 2.6 whistles/min, $N=11$). A total of 6462 whistles were selected for analysis from the recordings made (22.75 h of acoustic recording). All selected whistles for each spinner dolphin pod sampled were used. No adjustments were made in accordance with sampling effort and pod size. From 3.22 h of acoustic recording off the island of Kaua'i a total of 983 whistles were used, 2613 whistles were used from 10.35 h recorded off O'ahu, 1905 whistles were used from 5.13 h recorded off Lāna'i, and 961 whistles were used from 4.05 h recorded off the Island of Hawai'i (Table I).

CoolEdit® was used to redigitize the DAT recordings at a sampling rate of 48 kHz and 16 bits, and the recordings were stored as sound files (i.e., wav format files). The sound files were inspected using spectrograms computed with CoolEdit® to select the whistles suitable for detailed analysis. Whistles selected had a good signal-to-noise ratio (greater than 6 dB) and did not overlap with more than two whistles. To consider successive whistle production as individual whistles, a gap between the whistles had to be larger than 200 ms as described in Bazúa-Durán and Au (2002). About 80% of the whistles recorded were selected for analysis. The whistles selected for analysis simultaneously describe the different number of whistle types (whistle repertoire) and the context in which they were emitted (whistle usage).

Canary® was used to generate the spectrogram of each whistle selected for analysis (FFT size=1024 points, Hamming window, overlap=50%). A frame length of 512 points (or 10.67 ms) was used for the analysis resolution, thus a filter bandwidth of 380.62 Hz and a time resolution of 5.33 ms were obtained. Ten parameters shown in Fig. 2 were extracted from the fundamental frequency of each whistle contour. Begin (BF), end (EF), peak (PF: frequency at which the maximum intensity is found), maximum (MaF), and mini-

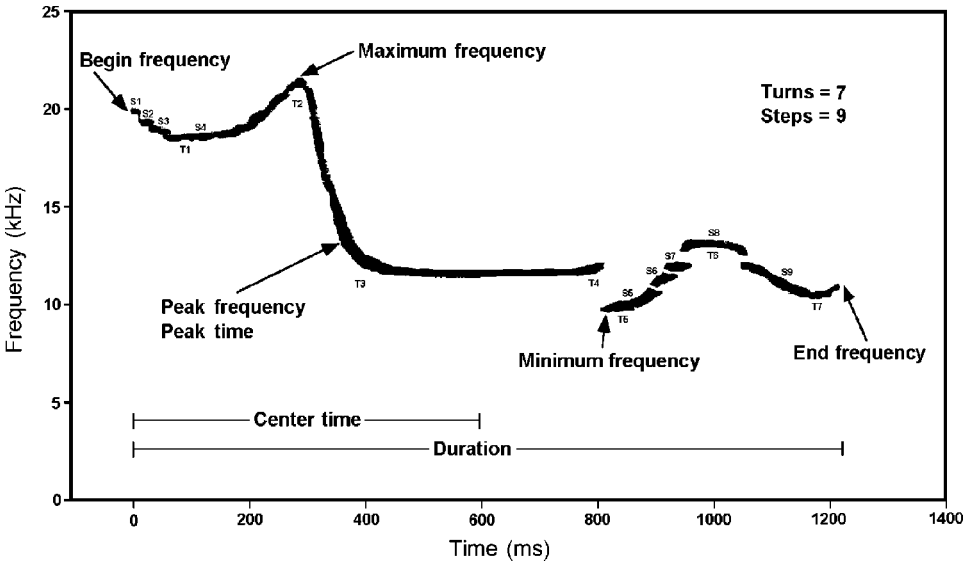


FIG. 2. The ten parameters measured with Canary® for the fundamental frequency of each whistle or whistle contour. Each turn and step measured are depicted [modified from Bazúa-Durán and Au (2002)].

mum frequencies (MiF), peak (PT: time at which the maximum intensity is found) and center times (CT: time at which the collective energy is half the total), and duration (Dur) were manually extracted using the measurement tool of Canary®. Number of turns (NT) and steps (NS) were determined visually (Fig. 2) as described in Bazúa-Durán and Au (2002). These parameters were chosen to be consistent with previous studies (e.g., Steiner, 1981; Wang *et al.*, 1995; Rendell *et al.*, 1999).

The whistles selected from the recordings of each dolphin pod comprised the different groups (or whistle groups) to be compared. The ten parameters extracted from each whistle contour were used with Homoscedastic or linear-Canonical (L-DFA) and Heteroscedastic or quadratic (Q-DFA) Discriminant Function Analyses (Huberty, 1994; Krzanowski and Marriott, 1995). Q-DFA was computed in order to test if both types of DFA would give the same results. DFA with a tolerance level of 0.01 and *a priori* classification probabilities proportional to the amount of whistles analyzed for each pod were used to investigate differences in spinner dolphin whistles. Differences in the whistles were examined using the following outputs of the DFA: (1) within-whistle group means of the first two canonical functions (CV1 and CV2), (2) squared Mahalanobis distances (D^2; for L-DFA) and generalized distance (G^2; for Q-DFA) of the between-whistle group pairwise comparisons, (3) *p*-values from the *F* test of the between-whistle group pairwise comparisons, and (4) plug-in correct classification scores (CCscores) of individual whistles to the whistle group of the spinner dolphin pod they were recorded from. For the L-DFA all four outputs were calculated, while for the Q-DFA only outputs two to four could be computed.

Between-whistle group pairwise *F* test *p*-values and CCscores were used to determine if a whistle group could be discriminated from others. The pairwise *F* test *p*-value was used to determine if two whistle groups were significantly different at the $p<0.05$ level. CCscores indicate how many individual whistles are distinctive enough in their parameters to be correctly assigned by functions to their original whistle group and, when compared to CCscores expected by chance alone (*a priori* classification probabilities proportional to the amount of whistles per whistle group), they indicate the performance of the DFA in finding differences between whistle groups. If the CCscore for a whistle group was not greater than the classification expected by chance, then that whistle group was not discriminated from others.

The graph of within-whistle group means of the first two CV was used to visualize relative differences between whistle groups by looking at distances between canonical means using each CV separately and using the combination of both. The first two CV explained more than 75% of the variance in the data in all comparisons. The D^2 and G^2 from between-group pairwise comparisons of whistle groups were used to determine the magnitude of the differences found between whistle groups. The larger D^2 and G^2 the more different the whistle groups were.

III. RESULTS

A. All whistle groups' comparison I

The relative difference in the whistles of the 27 spinner dolphin pods recorded is depicted in Fig. 3 using the within-whistle group means of the first two CV. The relative difference between some whistle groups recorded in the same island (e.g., OG1 and OG2) could be as large as the relative difference between whistles recorded off different islands (e.g., OG2 and LG8; Fig. 3). D^2 and G^2 were of about the same magnitude between whistle groups of spinner dolphins recorded off the same (e.g., $D^2_{OG1\text{-}OG2}=2.69$, $G^2_{OG1\text{-}OG2}=2.36/3.58$) or different island (e.g., $D^2_{OG2\text{-}LG8}=2.77$, $G^2_{OG2\text{-}LG8}=2.51/4.06$).

To investigate possible reasons for finding whistle differences within islands being as large as between islands, whistles were compared using the whistle groups from each Hawai'ian Island separately (i.e., ten for O'ahu, eight for Lāna'i, and six for the Island of Hawai'i). Kaua'i whistle groups were not analyzed separately because only three different whistle groups were involved for Kaua'i (see Bazúa Durán, 2001).

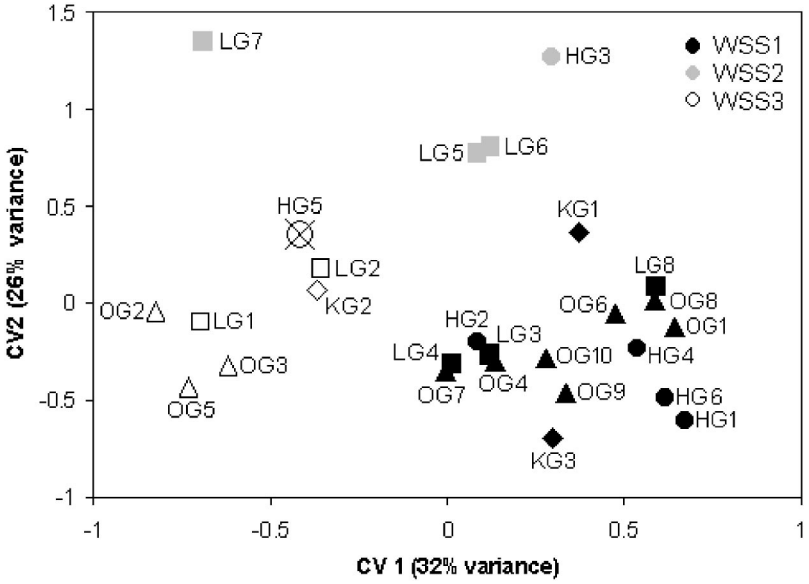


FIG. 3. Values of the within-whistle group means for the first two canonical functions (CV) of the Homoscedastic-Canonical Discriminant Function Analysis (L-DFA) for the comparison of the whistles of all 27 spinner dolphin pods recorded off Kaua'i, O'ahu, Lāna'i, and the Island of Hawai'i. Between-whistle group squared Mahalanobis distances and generalized squared distances, correct classification scores, and between-whistle group *p*-values were used to identify whistle groups that clustered in each whistle-specific subgroup (WSS). Shadings of points represent the WSS they belong to (WSS1, WSS2, or WSS3).

B. Within islands' comparisons

In the graph of all values for the first two CVs from the within-island comparison, there is no obvious difference in whistle groups (Fig. 4); canonical values for all whistle groups clustered together. Although statistically significant differences were obtained using L-DFA, differences are very small (O'ahu: Wilks' $\lambda=0.57$, approximate $F[90,17\,603]=16.90$, $p<0.000\,01$; Lāna'i: Wilks' $\lambda=0.54$, approximate $F[90,12\,801]=13.43$, $p<0.000\,01$; Island of Hawai'i: Wilks' $\lambda=0.53$, approximate $F[50,4317]=13.10$, $p<0.000\,01$). Whistles of the different whistle groups seem to have common characteristics. The number of whistles from whistle groups that had common characteristics was calculated by dividing each whistle group into subsamples. Each subsample contained whistles recorded consecutively in 15- to 60-min intervals and was compared to the other subsamples from the same island using L-DFA. If a subsample was not significantly different ($p>0.05$) and/or their CCscore was lower than the classification expected by chance alone, then the whistles of that subsample were considered as common with those of other whistle groups.

The graphs of the within-island whistle group means for the first two CVs (Fig. 5) indicate that whistles of three islands clustered into several groups. A new term, ''whistle-specific subgroup'' (WSS), was used to name these whistle clusters regardless of the location, day, and year of recording. These whistle clusters were named subgroups since they are part of all the whistles produced by spinner dolphins in the Hawai'ian Islands. Joining Clustering Analysis (JCA) was resorted to form clusters because the massive amount of data was beyond the capabilities of our computational resources. However, to substantiate the clustering of whistle groups into WSSs using DFA, JCA was computed using the means and medians of the absolute measurements of the ten whistle parameters of each whistle group.

Several criteria were used to identify whistle groups that clustered in each WSS (Fig. 6). CCscores and pairwise *F* test *p*-values were first examined. When the CCscore of a whistle

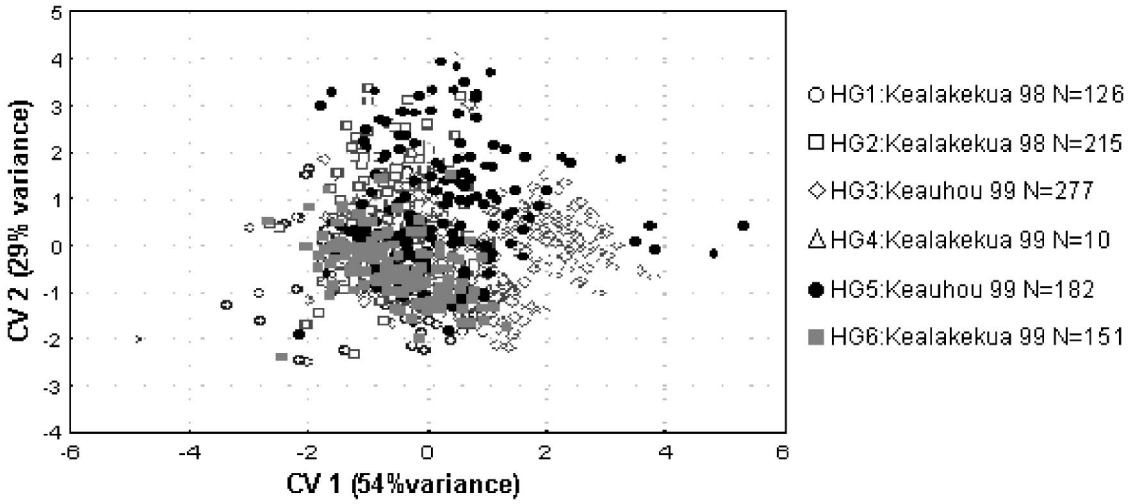


FIG. 4. Canonical values of the within-whistle group comparison for the first two functions (CV) of the Homoscedastic-Canonical Discriminant Function Analysis (L-DFA) for the comparison of the whistles of the six spinner dolphin pods recorded off the Island of Hawai'i. HG2, HG4, and HG6 canonical values are completely overlapped with those of other whistle groups.

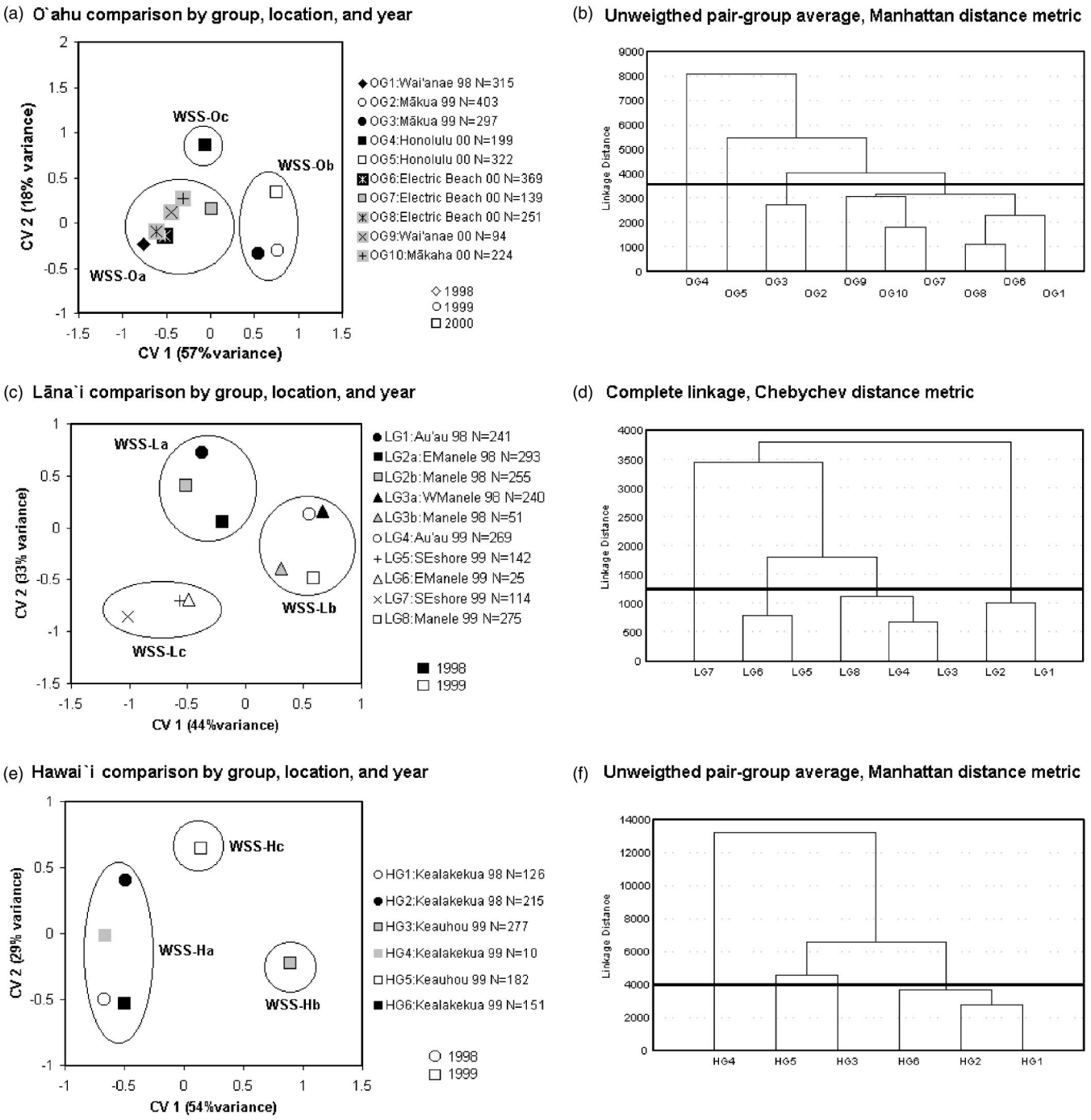


FIG. 5. Values of the within-whistle group means for the first two canonical functions (CV) of the Homoscedastic-Canonical Discriminant Function Analysis (L-DFA) (a,c,e) and Joining Clustering Analysis (JCA) (b,d,f) for the comparison of the whistles by location for (a,b) the ten spinner dolphin pods recorded off O`ahu, (c,d) the eight pods recorded off Lāna`i, and (e,f) the six pods recorded off the Island of Hawai`i. For the L-DFA, between-whistle group squared Mahalanobis distances and generalized squared distances, correct classification scores, and between-whistle group *p*-values were used to identify groups that clustered in each whistle-specific subgroup (WSS). Shape (a,e) or shading (c) of points represents the year whistles were recorded. JCA was computed using the means (d,f) and medians (b) of the ten whistle parameters measured for the fundamental frequency of whistles of each whistle group. The solid thick line represents the linkage distance value to cut different WSS.

group was smaller than expected by chance and/or the pairwise *F* test *p*-value was greater than 0.05, this whistle group was clustered with the whistle group having the smallest pair-wise D^2 and G^2. Comparisons including whistle groups with sample sizes much larger than for other whistle groups produced CCscores not always greater than chance, biasing CCscores towards the larger sample [CCscores depend on sample size (Beecher, 1989)].

After examining CCscores and pairwise *F* test *p*-values, the remaining whistle groups were clustered according to their pairwise D^2 and G^2 values. If pairwise D^2 and G^2 were less than 1.0, those two whistle groups were clustered into the same WSS. However, a whistle group occasionally had pairwise D^2 and G^2 less than 1.0 with two or more other whistle groups, yet these other whistle groups did not have a pairwise D^2 and G^2 less than 1.0 between them. This whistle

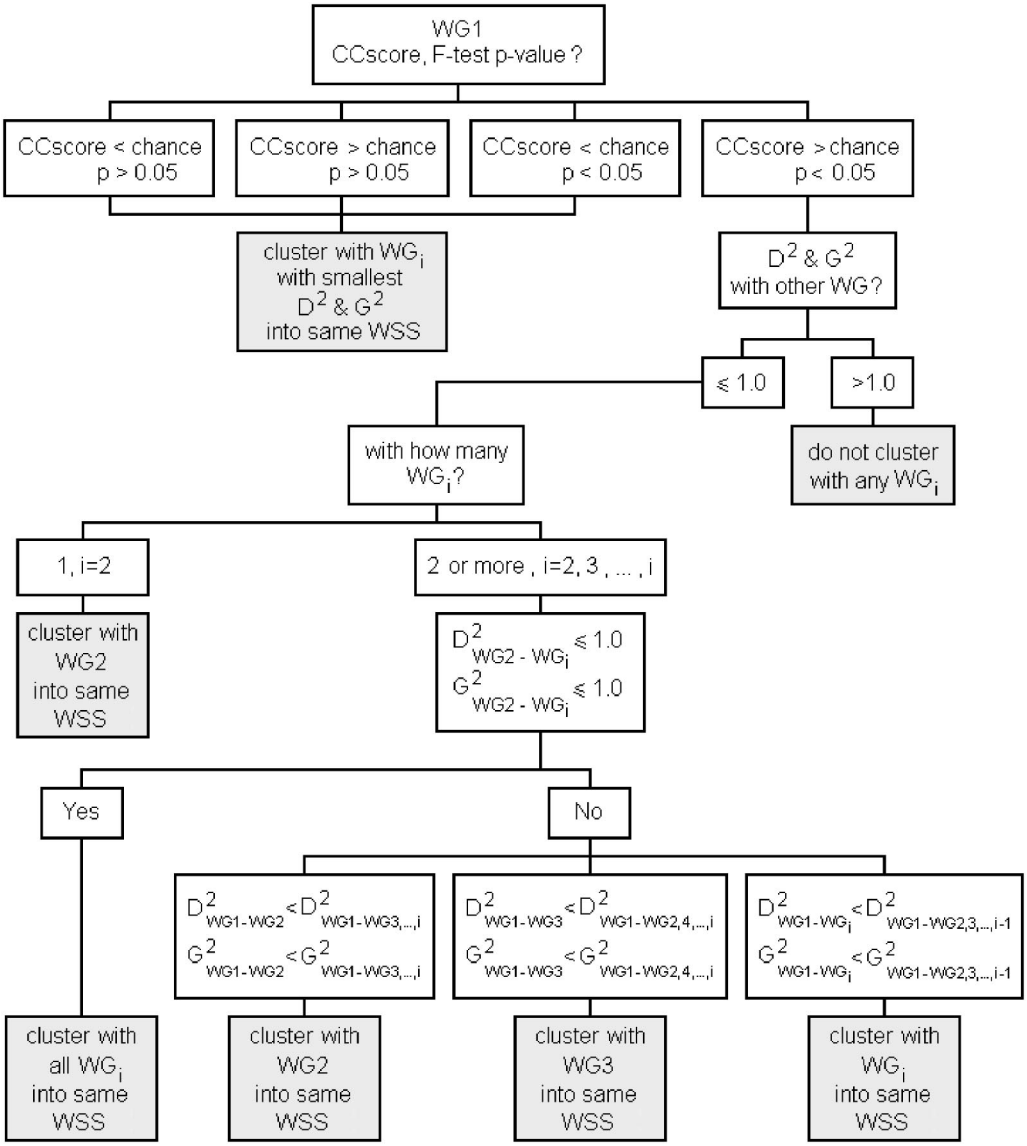


FIG. 6. Criteria used to identify whistle groups (WG) that clustered together into a whistle-specific subgroup (WSS). CCscores are the plug-in correct classification scores and F-test *p*-values, D^2, and G^2 are the *p*-values, the squared Mahalanobis distances, and the generalized distance of the between-whistle group pairwise comparisons, respectively.

group was clustered with the whistle group with which it had the smallest pairwise D^2 and G^2 only when two other whistle groups that did not have a pairwise D^2 and G^2 less than 1.0 between them were involved.

Three possible WSSs exist for Oʻahu: WSS-Oa consisting of OG1, OG6, OG7, OG8, OG9, and OG10; WSS-Ob consisting of OG2, OG3, and OG5; and WSS-Oc consisting of OG4 [Table I and Fig. 5(a)]. JCA with the medians of each whistle group using unweighted or weighted pair-group averages with a Manhattan distance metric revealed almost the same clusters when the linkage distance was cut at 3500 [Fig. 5(b)].

When considering subsamples, all whistles from WSS-Oa ($N=1392$) seem to be common with those of other whistle groups. All whistle groups could not be discriminated; they had pairwise *p*-values greater than 0.05 and CCscores were lower than chance for OG7, OG8, and OG9. For WSS-Ob, only some whistles from OG2, OG3, and OG5 ($N=411$) could be discriminated from the whistles of WSSs Oa and Oc ($p<0.001$, CCscores>chance). Whistles recorded off Waikīkī Beach ($N=35$) belonging to WSS-Oc were the only ones from OG4 that could be discriminated from whistle of other WSSs ($p<0.001$, CCscores>chance). Approximately 75% of the whistles in Oʻahu were common amongst whistle groups (Table II).

Three possible WSSs exist for Lānaʻi: WSS-La consisting of LG1 and LG2; WSS-Lb consisting of LG3, LG4, and LG8; and WSS-Lc consisting of LG5, LG6, and LG7 [Table I and Fig. 5(c)]. Whistles from whistle groups LG5 and LG6 could not be discriminated from each other ($p=0.37$; CCscore=0 for LG6). JCA with the means of each whistle group using either the Ward method, complete linkage, or unweighted or weighted pair-group averages with a Chebychev distance metric revealed almost the same clusters when the linkage distance was cut at 1200 [Fig. 5(d)].

When considering subsamples with LG2 (corresponding to WSS-La; $N=31$) and LG3, LG4, and LG8 (corresponding

TABLE II. Classification of cases (CCscore) and whistles not in common between whistle groups for the comparison of the whistles for the three whistle-specific subgroups (WSS) of O'ahu, the three WSSs of Lāna'i, the three WSSs of the Island of Hawai'i, and for the three WSSs for all islands combined.

Whistle-specific subgroup	L-DFA CCscore[a]	Q-DFA CCscore[a]	Chance[b]	No. of whistles not in common	Percentage of whistles not in common
Oa	0.7162	0.8642	0.5329	0	0
Ob	0.6184	0.5333	0.3909	411	40
Oc	0.3166	0.1306	0.0762	35	18
O'ahu	0.6475	0.6789	···	···	25
La	0.5805	0.5779	0.4142	137	18
Lb	0.7198	0.7245	0.4384	82	10
Lc	0.4164	0.4219	0.1474	66	23
Lāna'i	0.6173	0.6218	···	···	17
Ha	0.7151	0.8293	0.5223	75	22
Hb	0.6823	0.6462	0.2882	120	43
Hc	0.3791	0.3956	0.1894	108	59
Island of Hawai'i	0.6420	0.6930	···	···	38
WSS1	0.8528[c]	0.9583[c]	0.5528	945	32
WSS2	0.2849[c]	0.3927[c]	0.1053	328	59
WSS3	0.4899[c]	0.9494[c]	0.3419	1197	66
ALL	···	···	···	···	52

[a]These values are approximations since the correct classification scores (CCscores) for each WSS where calculated from the results of the comparison between all whistle groups by island, not between whistle-specific subgroups.
[b]Chance=*a priori* classification score expected by chance alone proportional to the amount of whistles per whistle group.
[c]These values are the correct classification scores (CCscores) for the comparison between each whistle-specific subgroup (WSS1, 2, and 3).

to WSS-Lb; $N=239$), some pair-wise *p*-values were greater than 0.05. These whistle subsamples could not be discriminated and seem to be common with those of other subsamples. Approximately 83% of the whistles in Lāna'i were common between whistle groups (Table II).

Three possible WSSs exist for the Island of Hawai'i: WSS-Ha consisting of HG1, HG2, HG4, and HG6, WSS-Hb consisting of HG3, and WSS-Hc consisting of HG5 [Table I and Fig. 5(e)]. Whistles of HG4 and HG6 were poorly discriminated from each other ($p=0.03$; CCscore=0 for HG4, $N=10$), probably because of the small sample size of HG4. JCA using the means of each whistle group and either (1) the single linkage with Euclidean or power distance metrics or (2) the Ward method or unweighted or weighted pair-group averages with a Manhattan distance metric revealed almost the same clusters when the linkage distance was cut at 1800 and 4000, respectively [Fig. 5(f)].

When considering subsamples, some whistles of HG1, HG2, HG4, and HG6 ($N=289$; corresponding to WSS-Ha), of HG3 ($N=231$; corresponding to WSS-Hb), and of HG5 ($N=74$; corresponding to WSS-Hc) had pairwise *p*-values greater than 0.05 and could not be discriminated. These whistles of WSSs-Ha, -Hb, and -Hc seem to be common with those of whistle groups from other WSSs. Approximately 62% of the whistles in the Island of Hawai'i were common amongst whistle groups (Table II).

C. All whistle groups' comparison II

From the analysis within the islands of O'ahu, Lāna'i, and Hawai'i it is clear that finding whistle differences within islands as large as between islands (Fig. 3) is not accounted for the effects of adding or subtracting whistle groups. In order to describe how the WSSs of the islands of O'ahu, Lāna'i, and Hawai'i compared to each other, whistles were compared using the WSSs for each island. The use of WSSs was supported by the fact that discriminant functions correctly classified whistles to each WSS almost twice as much as expected by chance alone (Table II). The WSSs by island were grouped together according to the same criteria used to form them (Fig. 6) and these new WSSs were designated with a number. WSS1 consists of Oa, Oc, Lb, and Ha, WSS2 consists of Lc and Hb, and WSS3 consists of Ob and La [Table I and Fig. 7(a)]. Hc (or HG5) was not ascribed to any numbered WSS because it could not be discriminated (CCscore=0, $N=182$) and had pairwise D^2 and G^2 less than 1.0 with Ob, Lb, and Lc. The same numbered WSSs containing the same whistle groups were formed when using the DFA output of the comparison between the original 27 whistle groups (including Kaua'i; Fig. 3).

The numbered WSSs also had some common characteristics, making discrimination impossible in some situations, i.e., Oc ($N=199$) and Ha ($N=502$) (from WSS1) could not be discriminated (CCscore<chance). Approximately 48% of whistles of numbered WSSs had common characteristics (Table II). When all whistles belonging to a numbered WSS were combined, WSS1, WSS2, and WSS3 could be discriminated from each other (Wilks' $\lambda=0.71$, approximate $F[20,10\,570]=99.36$, $p<0.000\,01$), indicating that there were sufficient number of whistles with distinct characteristics (Table II). The distance between WSS2 and both WSS1

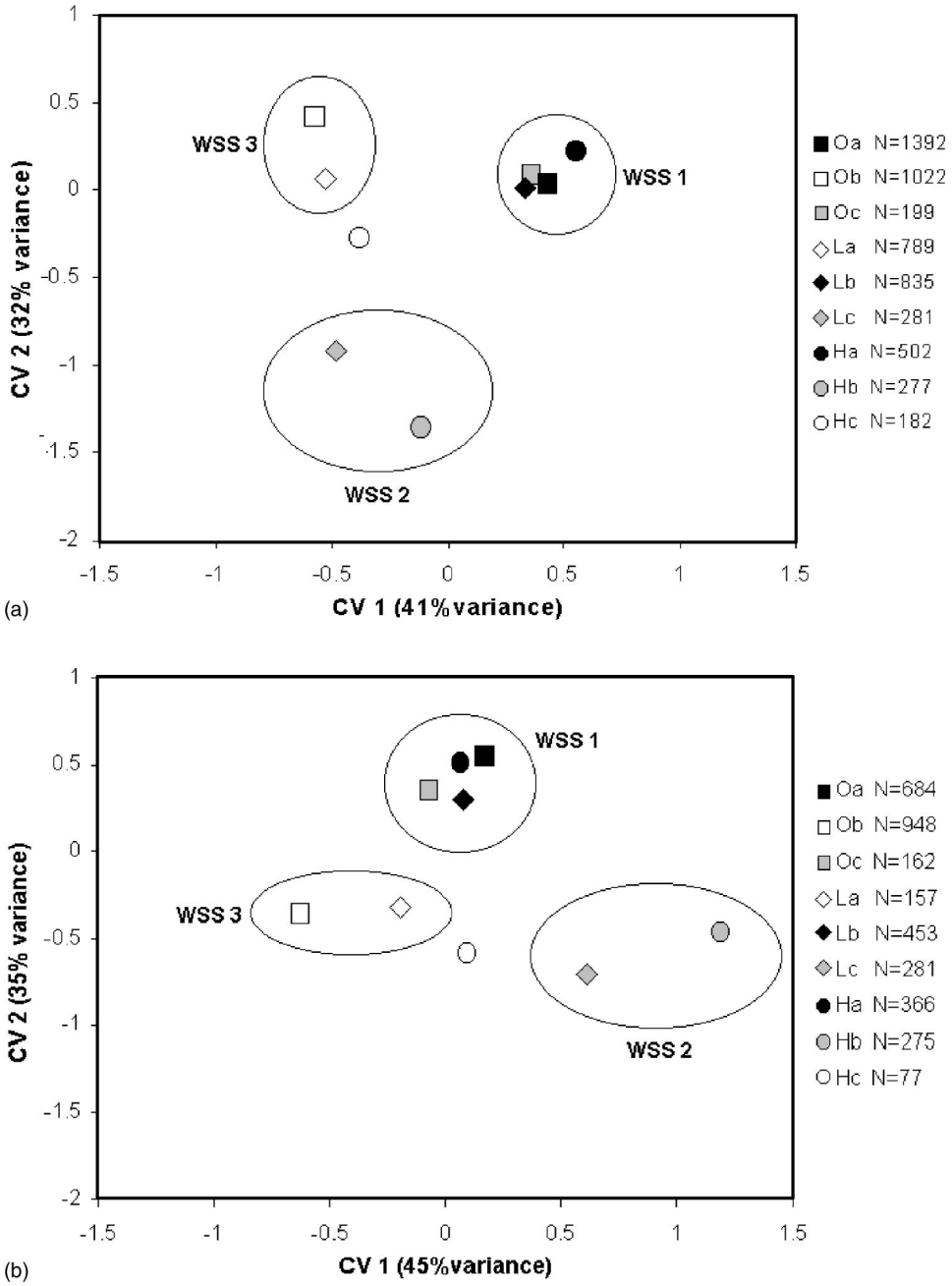


FIG. 7. Values of the within-whistle group means for the first two canonical functions (CV) of the Homoscedastic-Canonical Discriminant Function Analysis (L-DFA) for the comparison of whistles by whistle-specific subgroup (WSS) for the three WSSs of O'ahu, the three WSSs of Lāna'i, and the three WSSs of the Island of Hawai'i for (a) spinner dolphins engaged in all behavioral states and (b) only for whistles recorded when spinner dolphins were socializing. Between-whistle group squared Mahalanobis distances and generalized squared distances, correct classification scores, and between-whistle group *p*-values were used to identify whistle groups that clustered in each whistle-specific subgroup (WSS).

and WSS3 was larger ($D^2=1.95$ and 1.96; $G^2=2.69/1.41$ and 1.77/1.78, respectively) than between WSS1 and WSS3 ($D^2=1.02$ and $G^2=1.49/0.77$).

D. Whistle parameters important in the discrimination

Differences found in the whistles of spinner dolphins could be explained by six of the ten whistle parameters (Wilks' $\lambda=0.71$, approximate $F[12,10\,578]=162.14$, $p<0.000\,01$; *F*-remove $[2,5285]>20$, $p<0.000\,01$). All descriptive parameters (NT and NS), almost all frequency parameters (BF, EF, and MaF), and one time parameter (Dur) were responsible for the differences found (Table III). WSS1 was characterized by whistles of small MaF and large BF, WSS2 by whistles of small NT and large NS and EF, and WSS3 by long whistles (Table IV).

E. Effect of general behavioral state and pod size

In order to determine if general behavioral state influenced the differences found, whistles were compared between islands using the WSSs for O'ahu, Lāna'i, and the Island of Hawai'i as performed previously. Whistles were also compared earlier using WSSs to avoid having whistle groups with less than 50 whistles when testing for the effects of general behavioral state. Only whistles recorded when spinner dolphins were socializing were considered because in other behavioral states few whistles were produced.

When comparing whistles of socializing dolphins grouped by the different WSS found for each island ($N=3403$ or 62% of all whistles), significant differences between WSSs were also found (Wilks' $\lambda=0.57$, approximate $F[80,21\,477]=24.59$, $p<0.000\,01$). WSSs by island tended to cluster into the same numbered WSS as when all general

TABLE III. Whistle parameters that contributed to the discrimination between the three whistle-specific subgroups (WSSs) found for three Hawai'ian Islands (O'ahu, Lāna'i, and the Island of Hawai'i) and the WSSs they discriminated for whistles recorded during all behavioral states and only for whistles recorded when spinner dolphins were socializing. Parameters are listed in increasing degree of contribution to the discrimination.

WSSs separated	All whistles parameters	Social whistles parameters
WSS1-WSS2	No. of turns	No. of turns
	No. of steps	No. of steps[a]
	Maximum frequency[a]	Maximum frequency[a]
	Duration[a]	Duration[a]
	Begin frequency[a]	Begin frequency[a]
	End frequency[a]	End frequency[a]
WSS1-WSS3	Duration	Duration
	No. of turns	No. of turns
	Begin frequency[a]	
WSS2-WSS3	No. of turns	No. of turns[a]
	No. of steps	Duration[a]
	Duration[a]	No. of steps[a]
	Maximum frequency[a]	Maximum frequency[a]
	End frequency[a]	End frequency[a]

[a]These parameters have only a small contribution to the discrimination when compared to the other listed parameters.

behavioral states were taken into account [Fig. 7(b)] and the same six whistle parameters explained the differences between numbered WSSs (Wilks' $\lambda = 0.67$, approximate $F[12,6636] = 123.11$, $p < 0.000\,01$; F-remove $[2,3318] > 20$, $p < 0.000\,01$; Table III). Therefore, general behavioral state did not affect the conclusions drawn. Nonetheless, differences between whistle groups were smaller when only using whistles recorded from socializing animals. The relative distance between the means of the first two CV [Fig. 7(a) vs. 7(b)] and the pairwise D^2 and G^2 between groups were smaller, and NS, NT and BF changed their relative contribution (Table III). Some of the differences found between whistle groups could be due to differences in general behavioral state. Whistle characteristics seem to change with the general behavioral state of the dolphin pod they were recorded from, but retaining the properties of their WSS.

The size of spinner dolphin pods whose whistles clustered into the same WSS varied (Table I). Therefore, spinner dolphin pod size does not seem to affect the conclusions drawn either.

IV. DISCUSSION

For three Hawai'ian Islands, whistle characteristics did not seem to change according to location and date of recording. Whistles recorded in the same location on different days were clustered either to the same (OG1 and OG9, OG2 and OG3, and HG1, HG2, HG4 and HG6) or different WSS (OG4 and OG5, LG1 and LG4, LG2 and LG8, and HG3 and HG5). Whistles recorded on consecutive days were also clustered either into the same (OG6–OG10, OG2 and OG3, LG1 and LG2, and HG4 and HG6) or different WSS (OG4 and OG5, OG5 and OG6–OG10, LG1–LG2 and LG3, LG4 and LG5–LG7, LG5–LG7 and LG8, and HG5 and HG4, HG6). Whistles recorded from pods in the same location on the same day were determined to be part of the same WSS (LG5, LG6, and LG7) and the whistle groups from dolphin pods which were known to contain different dolphins were sometimes determined to be in the same (OG8 and OG9) and at other times in different WSS (HG3 and HG4). The composition and relatedness of spinner dolphin pods are probably important factors in determining differences in the characteristics of their whistles.

A. Spinner dolphin group composition and differences in whistle characteristics

In the comparisons within and between Hawai'ian Islands using whistle contour parameters and Discriminant Function Analyses remarkable similarities in the whistle characteristics were found; spinner dolphin pods had com-

TABLE IV. Descriptive statistics (mean, standard deviation [SD], and highest and lowest values) of the ten parameters extracted from each whistle contour for the three whistle-specific subgroups (WSSs) found in all four Hawai'ian Islands sampled (Kaua'i, O'ahu, Lāna'i, and the Island of Hawai'i). Span frequency is the maximum minus the minimum frequency.

WSS	Descriptive statistic	Frequency (kHz)						Time (s)			No. of		N
		Begin	End	Minimum	Maximum	Span	Peak	Peak	Center	Duration	Turns	Steps	
WSS1	Mean	11.70	13.54	10.40	15.57	5.17	12.76	0.184	0.195	0.408	3.28	2.37	3599
	SD	3.66	3.45	2.77	3.34	3.31	2.72	0.203	0.166	0.339	2.58	2.54	
	High	23.31	24.23	22.35	25.25	17.26	22.50	2.177	1.318	2.321	32	19	
	Low	1.60	1.07	0.85	1.71	0.22	2.22	0.000	0.004	0.016	0	0	
WSS2	Mean	11.56	16.15	10.72	17.56	6.84	13.57	0.194	0.209	0.446	1.59	3.86	558
	SD	3.29	3.73	2.69	3.30	3.89	2.51	0.197	0.181	0.379	2.13	3.38	
	High	23.78	23.42	18.33	23.98	16.93	20.95	1.547	1.574	3.350	17	19	
	Low	4.87	3.62	2.57	5.40	0.39	4.97	0.001	0.011	0.029	0	0	
WSS3	Mean	11.26	13.33	9.66	16.61	6.95	12.72	0.301	0.336	0.716	3.73	2.85	2123
	SD	3.85	3.69	2.42	3.47	3.28	2.70	0.243	0.191	0.402	2.64	3.07	
	High	23.52	23.98	21.03	23.98	17.71	21.37	1.721	1.403	3.042	24	31	
	Low	2.06	1.99	1.62	2.14	0.29	1.78	0.000	0.005	0.017	0	0	
Hc	Mean	12.21	15.17	10.29	17.21	6.92	12.78	0.265	0.301	0.639	3.02	3.24	182
	SD	4.13	3.92	2.45	3.66	3.82	2.39	0.258	0.200	0.418	2.07	4.01	
	High	22.90	24.08	16.23	24.18	16.91	19.17	1.658	0.981	2.093	14	24	
	Low	3.76	7.16	3.66	8.41	0.49	6.84	0.006	0.016	0.033	0	0	

mon characteristics in approximately 48% of their whistles. HG5 could not be ascribed to any numbered WSS since it had common whistle characteristics with all other whistle groups. The fluidity in the composition of spinner dolphin pods may be reflected in the large proportion of common whistle characteristics. Spinner dolphins that at some point in time occur in the same pod may be copying whistle characteristics from each other. The more pods interact, the greater fluidity in the composition of pods because of animals exchanging pods, which may result in a higher percentage of common whistles amongst spinner dolphins.

Despite the fluidity of pod composition, some individuals tend to associate or belong to the same pod (Marten and Psarakos, 1999; Pacific Whale Foundation, 1998, 2000, personal communication; Norris *et al.*, 1994; Östman, 1994). Along the Kona coast of the Island of Hawai'i, Östman (1994) found that, even though there is a tendency to underestimate stability because of the difficulty in identifying individuals within large pods, association patterns among spinner dolphins varied with location and with sex. He compared the coefficient of association among 11 females, 32 males, and 84 unsexed individuals and concluded that associations were nonrandom; males on average had tighter bonds with more individuals than did females, although there were extensive individual variations among sexes. Female–female associations were less extensive and weaker than male–male associations, with female–male associations being intermediate (Östman, 1994).

The similarities found in whistle characteristics and the presence of whistle-specific subgroups could indicate associations between individual dolphins. According to Rendell *et al.* (1999) and Tyack and Clark (2000), for species where social bonds center on individual relationships phonations will vary between individuals, while in species living in stable pods phonations will vary markedly between pods. Phonations in spinner dolphins do not seem to follow any of these patterns and could vary between individuals that do not spend much time together. The dolphin pods that were associated with the same WSS could have been composed in part of (1) individuals having similar whistles that spend time together, (2) family related individuals that have similar whistles, and/or (3) the same individual dolphins. WSSs in the Main Hawai'ian Islands may represent dialects. Spinner dolphins in the MHI have the potential for intermixing and interbreeding. The isolation in this spinner dolphin population could be due to behavioral differences rather than through geographical barriers. If dialects exist in Hawai'ian spinner dolphin whistles, they would not be as distinct as in killer (*Orcinus orca*) (vocal clans: Ford 1991; Strager, 1995; Yurk *et al.*, 2002) and sperm whales (*Physeter macrocephalus*) (Moore *et al.*, 1993; Weilgart and Whitehead, 1997; Whitehead *et al.*, 1998) because they live in a fission-fusion society whereas killer and sperm whales live in very stable pods. In order to determine what whistle-specific subgroups represent, concurrent studies that investigate the composition of spinner dolphin groups need to be carried out; genetic sampling is recommended since photo-identification methods give less than 30% identification success because these animals do not bear many scars and live within large pods (Östman, 1994).

The percentages of common whistles calculated from the DFA are not comparable to the percentages obtained from analyzing whistle contour shapes (or whistle types), such as the discrete calls of killer whales (e.g., Ford, 1991), the codas of sperm whales (e.g., Weilgart and Whitehead, 1997), and some calls of birds (e.g., Lemon, 1966). No attempt was made in quantifying the percentage of common whistles by whistle types since it would require categorizing whistles according to contour shape and spinner dolphin whistles seem to be more graded than discrete (Taruski, 1979; Herman and Tavolga, 1980; Bazúa-Durán and Au, 2002).

B. Spinner dolphin movements and differences in whistle characteristics

Spinner dolphins cover at least 20 km in 16 h in the Eastern Tropical Pacific (Perrin and Gilpatrick, 1994), therefore, they are capable of traveling between the MHI. It has been documented that spinner dolphins travel between O'ahu and Moloka'i using the Kaiwi Channel (Maldini, 2003), which is about 40 km wide. To investigate how spinner dolphins move within and between the MHI, it was assumed that spinner dolphin pods whose whistles were ascribed to the same WSS were composed of individuals that had spent or are spending time together. It was also assumed that greater similarities in whistle characteristics result from movements within and between islands. Along the northern Gulf of Mexico coast, the degree of movement of bottlenose dolphins between adjacent areas corresponds to the degree of similarities in whistle characteristics (Bazúa Durán, 1997).

Along the Wai'anae coast of O'ahu (Fig. 1), it seems that dolphin pods split and merge so that different animals that spend time together are sighted along this coast on different days (Bazúa Durán, 2001). OG6 through OG10 were whistle groups of spinner dolphin pods recorded on consecutive days in the same general area and their whistles were clustered into the same WSS. It was known that the composition of spinner dolphin pods from whistle groups OG8 and OG9 was different and their whistles clustered into the same WSS, suggesting that individual dolphins that usually spend time together have more whistle characteristics in common. Similarly, WSS-Ob from O'ahu may indicate that there is some movement from animals between the Wai'anae and South Shores. Larger similarities in the whistles of spinner dolphin pods recorded along the same coast, such as Wai'anae Shore, O'ahu (WSS-Oa) and Kealakekua Bay, Hawai'i (WSS-Ha), may suggest some degree of residency.

Off the South Shore of Lāna'i, whistle groups LG5, LG6, and LG7 were the only ones from WSS-Lc and could not be discriminated from each other. These three whistle groups were from spinner dolphin pods recorded on the same day in almost the same location after losing contact with the original pod and encountering another pod some time later. These three pods were probably the same assembly of spinner dolphins that was lost and encountered some time later. The same spinner dolphins also seem to have more whistle characteristics in common.

We were certain that the composition of dolphin pods from whistle groups HG3 and HG4 from the Kona coast was different and, contrary to O'ahu, their whistles were clustered into different WSSs. Spinner dolphin pods that are not composed of the same dolphins and may not spend much time together may not have many whistle characteristics in common.

Using the similarities found between WSS from each island, the roaming of Hawai'ian spinner dolphins does not seem to be limited to a single island. Spinner dolphins may travel more between Kaua'i, O'ahu, and Lāna'i (WSS3) and between Lāna'i and the Island of Hawai'i (WSS2) than between O'ahu and the Island of Hawai'i (WSS1).

C. Possible function of whistles and importance of whistle parameters' variation

In whistle perceptual studies with bottlenose dolphins, Ralston and Herman (1995) suggested that frequency contour is imperative in the judgment, whereas absolute frequency is ultimately ignored. The number of turns and steps in a whistle are parameters that partially describe the frequency contour of dolphin whistles. NT and NS contributed significantly to the discrimination between WSSs but also varied according to behavioral state (Table III). NT and NS in a whistle, therefore, are parameters that have high intraspecific variations. These whistle parameters may play an important role in dolphin communication (Herman and Tavolga, 1980; Caldwell *et al.*, 1990) and they may be important in individual identification (Norris *et al.*, 1994).

Whistle duration was an important parameter in the discrimination between WSSs. Rendell *et al.* (1999) also found in five odontocete species that whistle duration had a high intraspecific and a low interspecific variation. Whitten and Thomas (2001) found that in Pacific white-sided dolphins (*Lagenorhynchus obliquidens*) whistle duration varied between individuals. Besides varying the duration of whistles according to behavioral state or context [Table III and Janik *et al.* (1994)], dolphins also seem to modify whistle duration to convey individual or ''family'' identity information (Steiner, 1981; Wang *et al.*, 1995). Whistle duration may be used to communicate information at the individual level, either to communicate identity or to communicate what an individual is experiencing.

Whistle frequency parameters, such as begin, end, maximum, and minimum, did not contribute significantly in the discrimination between WSSs (Table III); they were relatively constant across spinner dolphin groups [also suggested by Wang *et al.* (1995) for bottlenose dolphins in the Gulf of Mexico]. Rendell *et al.* (1999) also found that whistle frequency parameters had both a high intraspecific and interspecific variation. Frequency characteristics appear to be less important at the individual level.

Besides varying whistle characteristics with behavioral state, spinner dolphins also modified the whistling rate [see also Norris *et al.* (1994)]. When animals were in close contact and actively interacting (i.e., during socializing) whistles were produced frequently and less often when resting and traveling without interactions. The whistling rate also varied with behavioral state in other delphinids such as the pilot whale [*Globicephala melaena* (Taruski, 1979)], the bottlenose dolphin (dos Santos and Almada, 1998), the Atlantic spotted dolphin [*Stenella frontalis* (Herzing, 2000)], the false killer whale [*Pseudorca crassidens* (Nester *et al.*, 2001)], and the killer whale (Thomsen *et al.*, 2002). Whistles seem to play an important part in the acoustic communication of most delphinids (Herman and Tavolga, 1980), especially at close ranges. They may be important in defining the limits and arrangement of a pod, and they may convey information on emotional state and individual identity as Norris *et al.* (1994) suggested.

V. CONCLUSIONS

We were able to compare whistle characteristics for different spinner dolphin pods in different locations, behavioral states, and of different sizes, showing that differences in whistle characteristics exist. Although it was not possible to make strong statements on why differences occur, we were able to discard location, date of recording, number of dolphins in a pod, and, to some degree, general behavioral state. We are suggesting that it may be possible to investigate some aspects of the population structure of a coastal delphinid species acoustically, however, they need to be verified with studies that identify individual spinner dolphins and investigate the stability of association patterns and the relatedness of associates. Aspects of the population structure of spinner dolphins that may be explained by investigating whistle variations could include (1) degree of residency, (2) movements along the shore, and (3) movements between islands or regions.

ACKNOWLEDGMENTS

We thank Richard Gill, Marc Lammers, Joe Mobley, Pacific Whale Foundation, Marsha Green, Samone Yust, Jeff Bearman, Terry Weaver, and all the volunteers for their help and support during fieldwork. We also thank Daniel Blaine, Andrew Taylor, and Leticia Gracia-Medrano for their statistical advise. This project was funded by a Leonida Memorial Scholarship and a University of Hawai'i Seed Money Grant, and partially funded by Grant No. N00014-98-1-0687 from the Office of Naval Research, Dr. Robert Gisiner. CBD was a Fulbright-García Robles-CONACyT and DGAPA, UNAM fellow. Fieldwork was carried out under a letter of authorization from the National Marine Fisheries Service. We also thank Renata Sousa-Lima Mobley and four anonymous reviewers for helpful comments to earlier versions of the manuscript. The work reported here is in part reported in a dissertation submitted as partial fulfillment of the requirements for the Ph.D. degree in the Department of Oceanography, University of Hawai'i at Mānoa. This is SOEST Contribution No. 6438 and Hawai'i Institute of Marine Biology Contribution No. 1187.

Bazúa Durán, M. C. (**1997**). ''Comparisons of whistles among different populations of dolphins (*Tursiops truncatus*) from the Gulf of Mexico,'' master's thesis, Universidad Nacional Autónoma de México, Mexico, DF, Mexico.
Bazúa Durán, M. C. (**2001**). ''The whistles of Hawai'ian spinner dolphins (*Stenella longirostris*): Description and geographic variations,'' doctoral dissertation, University of Hawai'i at Manoa, Honolulu.
Bazúa-Durán, C., and Au, W. W. L. (**2002**). ''The whistles of Hawai'ian spinner dolphins,'' J. Acoust. Soc. Am. **112**, 3064–3072 [DOI: 10.1121/1.1508785].
Beecher, M. D. (**1989**). ''Signaling systems for individual recognition: An information theory approach,'' Anim. Behav. **38**(2), 248–261.
Caldwell, M. C., Caldwell, D. K., and Tyack, P. L. (**1990**). ''Review of the signature-whistle hypothesis for the Atlantic bottlenose dolphin,'' in *The Bottlenose Dolphin*, edited by S. Leatherwood and R. R. Reeves (Academic, San Diego, CA, USA), pp. 199–234.
Conner, D. S. (**1982**). ''Dialects versus geographic variation in mammalian vocalizations,'' Anim. Behav. **30**(1), 297–298.
Deecke, V. B., Ford, J. K. B., and Spong, P. (**1999**). ''Quantifying complex patterns of bioacoustic variation: Use of a neural network to compare killer whale (*Orcinus orca*) dialects,'' J. Acoust. Soc. Am. **105**, 2499–2507.
dos Santos, M. E., and Almada, V. C. (**1998**). ''Acoustic emissions and activity patterns in bottlenose dolphins,'' in *Abstracts of The World Marine Mammal Science Conference*, 20–24 January, 1998, Monaco, p. 36.
Ford, J. K. B. (**1991**). ''Vocal traditions among resident killer whales (*Orcinus orca*) in coastal waters of British Columbia,'' Can. J. Zool. **69**(6), 1454–1483.
Ford, J. K. B. (**2002**). ''Dialects,'' in *The Encyclopedia of Marine Mammals*, edited by W. F. Perrin, B. Würsig, and J. G. M. Thewissen (Academic, New York), pp. 322–323.
Herman, L. M., and Tavolga, W. N. (**1980**). ''The communication system of cetaceans,'' in *Cetacean Behavior: Mechanisms and Functions*, edited by L. M. Herman (Wiley, New York), pp. 149–209.
Herzing, D. L. (**2000**). ''Acoustics and social behavior of wild dolphins: implications for a sound society,'' in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. F. Fay (Springer Verlag, New York), pp. 225–272.
Huberty, C. J. (**1994**). *Applied Discriminant Analysis* (Wiley, New York).
Janik, V. M., Dehnhardt, G., and Todt, D. (**1994**). ''Signature whistle variations in the bottlenosed dolphin, *Tursiops truncates*,'' Behav. Ecol. Sociobiol. **35**(4), 243–248.
Krebs, J. R., and Kroodsma, D. E. (**1980**). ''Repertoires and geographical variation in bird songs,'' Adv. Stud. Behav. **11**, 143–177.
Krzanowski, W. J., and Marriott, F. H. C. (**1995**). *Multivariate Analysis. Part 2: Classification, Covariance Structures and Repeated Measurements* (Arnold, London, Great Britain).
Lemon, R. E. (**1966**). ''Geographic variation in the song of cardinals,'' Can. J. Zool. **44**(3), 413–428.
Maldini, D. (**2003**). ''Abundance and distribution patterns of Hawai'ian odontocetes: focus on O'ahu,'' doctoral dissertation, University of Hawai'i at Manoa, Honolulu.
Marler, P. (**1960**). ''Bird songs and mate selection,'' in *Animal sounds and communication, Publ. No. 7*, edited by W. E. Lanyon and W. N. Tavolga (American Institute of Biological Sciences, Washington, DC), pp. 348–367.
Marten, K., and Psarakos, S. (**1999**). ''Long-term site fidelity and possible long-term associations of wild spinner dolphins (*Stenella longirostris*) seen off O'ahu, Hawai'i,'' Marine Mammal Sci. **15**(4), 1329–1336.
McGregor, P. K., Peake, T. M., and Gilbert, G. (**2000**). ''Communication, behaviour, and conservation,'' in *Behaviour and Conservation*, edited by L. M. Gosling and W. J. Sutherland (Cambridge U.P., Cambridge, Great Britain), pp. 261–285.
Miller, P. J. O., and Bain, D. E. (**2000**). ''Within-pod variation in killer whale calls,'' Anim. Behav. **60**(5), 617–628.
Moore, K. E., Watkins, W. A., and Tyack, P. L. (**1993**). ''Pattern similarity in shared codas from sperm whales (*Physeter catodon*),'' Marine Mammal Sci. **9**(1), 1–9.
Mundinger, P. C. (**1982**). ''Microgeographic and macrogeographic variation in the acquired vocalizations of birds,'' in *Acoustic Communication in Birds: Song Learning and its Consequences. Vol. 2*, edited by D. E. Kroodsma, E. H. Miller, and H. Ouellet (Academic, San Diego), pp. 147–208.
Nester, A. E., Thomas, J. A., and Acevedo-Gutiérrez, A. (**2001**). ''The underwater whistle repertoire of wild false killer whales (*Pseudorca crassidens*),'' J. Acoust. Soc. Am. **109**, 2390.
Norris, K. S., Würsig, B., Wells, R. S., and Würsig, M. (**1994**). *The Hawai'ian Spinner Dolphin* (Univ. of California, Berkeley).
Östman, J. S. O. (**1994**). ''Social organization and social behavior of Hawai'ian spinner dolphins (*Stenella longirostris*),'' doctoral dissertation, University of California at Santa Cruz.
Pacific Whale Foundation (**1998**).
Pacific Whale Foundation (**2000**). Pacific Whale Foundation, 300 Maalaea Road, Suite 211, Wailuku, HI 96793, USA, Tel +1-808-249-8811, Fax +1-808-243-9021, info@pacificwhale.org
Perrin, W. F., and Gilpatrick, Jr., J. W. (**1994**). ''Spinner dolphin *Stenella longirostris* (Gray, 1828),'' in *Handbook of Marine Mammals. Vol. 5*, edited by S. H. Ridgway and R. Harrison (Academic, London, Great Britain), pp. 99–128.
Ralston, J. V., and Herman, L. M. (**1995**). ''Perception and generalization of frequency contours by a bottlenose dolphin (*Tursiops truncatus*),'' J. Comp. Psychol. **109**, 268–277.
Rendell, L. E., Matthews, J. N., Gill, A., Gordon, J. C. D., and Macdonald, D. W. (**1999**). ''Quantitative analysis of tonal calls from five odontocete species, examining interspecific and intraspecific variation,'' J. Zool. Lond. **249**(4), 403–410.
Steiner, W. W. (**1981**). ''Species-specific differences in pure tonal whistle vocalizations of five western north Atlantic dolphin species,'' Behav. Ecol. Sociobiol. **9**(4), 241–246.
Strager, H. (**1995**). ''Pod-specific call repertoires and compound calls of killer whales (*Orcinus orca* Linnaeus 1758), in the waters of Norway,'' Can. J. Zool. **73**(6), 1037–1047.
Taruski, A. G. (**1979**). ''The whistle repertoire of the north Atlantic pilot whale (*Globicephala melaena*) and its relationship to behavior and environment,'' in *Behavior of Marine Animals. Current Perspectives in Research. Vol. 3: Cetaceans*, edited by H. E. Winn and B. L. Olla (Plenum, New York), pp. 345–368.
Thielcke, G. (**1965**). ''Gesangsgeographische variation des gartenbaumlaufers (*Certhia brachydactyla*) im hinblick auf das artbildungsproblem,'' Z. Tierpsychol **22**, 542–566.
Thomsen, F., Franck, D., and Ford, J. K. B. (**2002**). ''On the communicative significance of whistles in wild killer whales (*Orcinus orca*),'' Naturwissenschaften **89**(9), 404–407 [DOI:10.1007/s00114-002-0351-x].
Tyack, P. L., and Clark, C. W. (**2000**). ''Acoustic communication in whales and dolphins,'' in *Hearing by Whales and Dolphins*, edited by W. W. L. Au, A. N. Popper, and R. F. Fay (Springer Verlag, New York), pp. 156–224.
Wang, D., Würsig, B., and Evans, W. (**1995**). ''Whistles of bottlenose dolphins: Comparisons among populations,'' Aquat. Mamm. **21**(1), 65–77.
Weilgart, L. S., and Whitehead, H. (**1997**). ''Group-specific dialects and geographical variation in coda repertoire in South Pacific sperm whales,'' Behav. Ecol. Sociobiol. **40**(5), 277–285.
Whitehead, H., Dillon, M., Dufault, S., Weilgart, L., and Wright, J. (**1998**). ''Non-geographically based population structure of south Pacific sperm whales: Dialects, fluke-markings and genetics,'' J. Anim. Ecol. **67**(2), 253–262.
Whitten, J. L., and Thomas, J. A. (**2001**). ''Whistle repertoire of Pacific white-sided dolphins (*Lagenorhynchus obliquidens*) at the John G. Shedd Aquarium,'' J. Acoust. Soc. Am. **109**(5 Pt. 2), 2391.
Yurk, H., Barrett-Lennard, L., Ford, J. K. B., and Matkin, C. O. (**2002**). ''Cultural transmission within maternal lineages: vocal clans in resident killer whales in southern Alaska,'' Anim. Behav. **63**(6), 1103–1119.

Within-individual variation in bullfrog vocalizations: Implications for a vocally mediated social recognition system

Mark A. Bee[a)]
AG Zoophysiologie und Verhalten, Institut für Biologie und Umweltwissenschaften, Carl von Ossietzky Universität Oldenburg, Postfach 2503, D-26111 Oldenburg, Germany

(Received 8 March 2004; revised 1 May 2004; accepted 30 June 2004)

Acoustic signals provide a basis for social recognition in a wide range of animals. Few studies, however, have attempted to relate the patterns of individual variation in signals to behavioral discrimination thresholds used by receivers to discriminate among individuals. North American bullfrogs (*Rana catesbeiana*) discriminate among familiar and unfamiliar individuals based on individual variation in advertisement calls. The sources, patterns, and magnitudes of variation in eight acoustic properties of multiple-note advertisement calls were examined to understand how patterns of within-individual variation might either constrain, or provide additional cues for, vocal recognition. Six of eight acoustic properties exhibited significant note-to-note variation within multiple-note calls. Despite this source of within-individual variation, all call properties varied significantly among individuals, and multivariate analyses indicated that call notes were individually distinct. Fine-temporal and spectral call properties exhibited less within-individual variation compared to gross-temporal properties and contributed most toward statistically distinguishing among individuals. Among-individual differences in the patterns of within-individual variation in some properties suggest that within-individual variation could also function as a recognition cue. The distributions of among-individual and within-individual differences were used to generate hypotheses about the expected behavioral discrimination thresholds of receivers. © *2004 Acoustical Society of America.* [DOI: 10.1121/1.1784445]

PACS numbers: 43.80.Ka [JMS] Pages: 3770–3781

I. INTRODUCTION

Vocal communication plays important roles in social recognition by allowing animals to direct appropriate behaviors toward other individuals after learning to recognize individually distinct properties of their acoustic signals. Patterns of individual variation in acoustic signals contribute toward identifying individuals in a number of taxa, including fish (Myrberg *et al.*, 1993; Crawford *et al.*, 1997), frogs (Bee and Gerhardt, 2001a; Bee *et al.*, 2001), reptiles (Polakow, 1997); birds (Weary *et al.*, 1990; Robertson, 1996; Lengagne *et al.*, 1997), and mammals (Gelfand and McCracken, 1986; Hauser, 1991; Insley, 1992; Fischer *et al.*, 2002). Field and laboratory playback studies in a wide range of taxa have confirmed that animals behaviorally discriminate among individuals based on the learned recognition of individually distinct acoustic signals in a number of different behavioral contexts, such as kin recognition (Beecher, 1991; Balcombe, 1990; Aubin and Jouventin, 2002), mate recognition (Robertson, 1996; Lengagne *et al.*, 2000), territorial neighbor recognition (Brooks and Falls, 1975a; Myrberg and Riggio, 1985; Bee and Gerhardt, 2001b, c, 2002), and individual recognition (Beecher *et al.*, 1996; Cheney and Seyfarth, 1999; Gentner and Hulse, 1998, 2000).

Relatively few studies, however, have sought to understand the mechanisms and evolution of vocally mediated social recognition by systematically investigating the acoustic and perceptual bases of recognition either within or across taxa (e.g., Beecher, 1991; Gentner and Hulse, 1998, 2000; Bee and Gerhardt, 2001a, b, 2002; Bee, 2001, 2003). Clearly an important aspect of acoustically mediated recognition systems involves the patterns of variation in acoustic signals that identify individuals (Beecher, 1989, 1991; Sherman *et al.*, 1997). Only after characterizing the potential for specific acoustic properties to identify individuals can researchers investigate in an informed way the perceptual basis of recognition using playback experiments.

Here I report results from a study of a vocally mediated social recognition system in a frog. Davis (1987) demonstrated that territorial males of the North American bullfrog (*Rana catesbeiana*) exhibit relatively lower levels of aggression in response to playbacks of their adjacent neighbor's advertisement calls from the neighbor's territory in comparison to responses to the calls of strangers. Hence, as in territorial songbirds (reviewed in Stoddard, 1996), territorial male bullfrogs learn to recognize their neighbor's vocalizations and rely on individual variation in vocalizations to discriminate among neighbors and strangers. Bee and Gerhardt (2001a) investigated the acoustic basis of neighbor-stranger discrimination in bullfrogs by analyzing the patterns of individual variation present in the first note of advertisement calls, which are more commonly produced as multiple-note calls (Fig. 1). The fundamental frequency (and correlated spectral properties) exhibited the greatest among-individual variation relative to within-individual variation, and contributed most toward statistically discriminating among individuals in multivariate statistical analyses.

Using the habituation-discrimination paradigm, Bee and

a)Electronic mail: mark.bee@uni-oldenburg.de

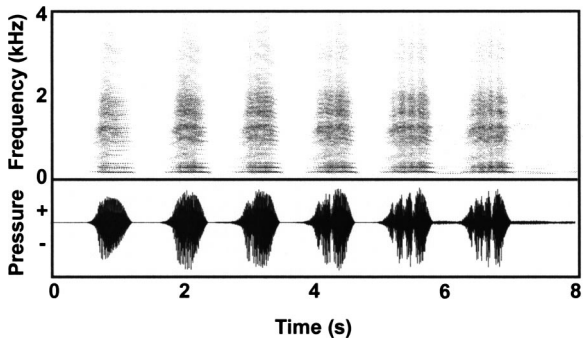


FIG. 1. Sonogram (top) and oscillogram (bottom) of a six-note advertisement call.

Gerhardt (2001b, c, 2002; Bee, 2003) tested the hypothesis that repeated exposure to a territorial neighbor's advertisement calls would allow males to learn to recognize acoustic properties related to the pitch of a neighbor's advertisement calls. In a discrimination phase following habituation trials, changes in the fundamental frequency of the synthetic stimulus elicited a recovery of habituated territorial aggressive responses, as predicted if territory residents use features related to call pitch to identify neighbors as familiar.

Based on the results of their acoustic analyses and field playback experiments, Bee and Gerhardt (2001a, b, c) estimated that the ''just-meaningful difference'' (JMD) (Nelson and Marler, 1990; Gerhardt, 1992) in fundamental frequency required to elicit renewed territorial aggression following habituation training was between 5% and 10%. Bee and Gerhardt (2001b, c, 2002) stressed, however, that this conclusion was provisional for two important reasons. First, Bee and Gerhardt (2001a) reported, but did not describe, variation in several acoustic properties, including fundamental frequency, that occurred *among* the separate notes *within* the multiple note advertisement calls produced by an individual (Fig. 1). This note-to-note variation in advertisement calls could make discriminating among individuals more difficult due to increased overlap of the signal properties of different individuals' calls. Second, the synthetic stimuli used in the habituation-discrimination tasks were invariant, meaning that the note-to-note within-individual variation present in multiple-note advertisement calls was not incorporated in the perceptual learning and discrimination task. Thus, an important and unanswered question concerns the implications of the patterns and magnitudes of within-individual variation in multiple-note advertisement calls for the behavioral discrimination thresholds that receivers should use to discriminate among familiar and unfamiliar signals. As a first step toward addressing this issue, I describe here the sources, magnitudes, and patterns of variation in multiple-note advertisement calls. The main objective of this study was to determine the ways in which note-to-note within-individual variation in multiple-note calls can either constrain or provide additional cues for acoustically mediated neighbor recognition.

II. METHOD

A. Study organism

During their breeding season, male bullfrogs establish and defend territories in permanent bodies of water from which they emit advertisement calls to attract gravid females and to repel rival males (Howard, 1978). During periods of active calling, which typically occur between 2300 and 0400 hours, most advertisement calls consist of four to seven notes (Fig. 1; Bee unpublished data). The individual notes comprising multiple-note advertisement calls are about 700 ms in duration, are separated by intervals of about 500 ms, and have a quasiperiodic fine-temporal waveform with a periodicity that varies inversely with body size and corresponds to the fundamental frequency of the call (Capranica, 1965; Bee and Gerhardt, 2001a). The frequency spectrum of a call note consists of a series of harmonics with a bimodal distribution of sound energy. The lower frequency peak is centered between 200 and 400 Hz; a broader, higher frequency band occurs between 1000 and 2000 Hz (Fig. 2). The fundamental frequency (90–130 Hz) is usually absent from the frequency spectrum or contains relatively little sound energy compared to other spectral components (Capranica, 1965).

B. Recording of vocalizations

Between May and August 1998, the advertisement calls of 25 territorial male bullfrogs were recorded in ponds located in the Little Dixie Lake Conservation Area (Callaway Co., Missouri, USA). Recordings were made during active choruses under ambient light conditions, usually between 2300 and 0400 hours, central daylight time. At least five multiple-note advertisement calls were recorded from each of the 25 males using an HHb PDR-1000 DAT recorder (sampling rate=32 kHz) and a Sennheiser MKH 70 shotgun microphone (with a Windtech SG-3 windscreen) that was mounted on a tripod placed 1–2 m from a subject and angled downward between 25° and 45° from parallel with the surface of the pond. Immediately after each recording, the air and water temperatures at the frog's position were determined to the nearest 0.1 °C. Each subject had been captured and individually marked, and had its snout-to-vent length (SVL) measured, on a night prior to that of recording [following Bee and Gerhardt (2001a)].

C. Acoustical analysis of vocalizations

Most of the calls analyzed in this study consisted of between three and seven notes. To examine the patterns of within-individual variation among separate notes in multiple-note advertisement calls, I analyzed eight acoustic properties for the first, middle, and last notes of five multiple-note calls from each individual (5 calls×3 notes×25 individuals=375 total call notes) using a Kay DSP Sona-Graph Model 5500.[1] In calls having an even number of notes, one of the two middle notes was chosen randomly. I analyzed call notes in the relative positions of the first, middle, and last notes of calls, instead of notes in fixed positions within a call (e.g., the second, third, and fourth notes), because preliminary analyses of sonograms indicated that this would be the most effective method for capturing the range of within-individual variation in multiple-note calls, as adjacent call notes are often more similar to one another than are call notes spaced further apart (Fig. 1). The person performing the acoustical analyses was different from the person who made the record-

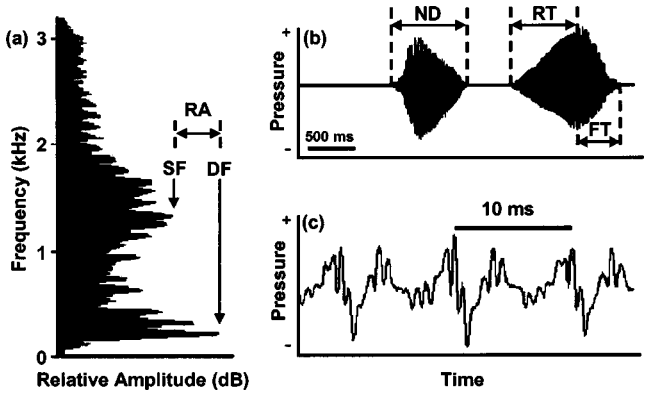


FIG. 2. (a) Power spectrum of an advertisement call note averaged over a 100-ms section from the middle of the note, beginning at the peak of the amplitude envelope. (b) Oscillogram depicting two call notes. (c) Oscillogram of the middle of the first call note in (b) illustrating the fine-temporal waveform. Dominant frequency (DF), secondary frequency (SF), and the relative amplitude (RA) were measured from power spectra as in (a). Dominant frequency is the harmonic of single greatest amplitude and corresponds to the second harmonic. Secondary frequency is defined as the harmonic of greatest relative amplitude in the higher frequency band of the bimodal spectrum. The amplitude (in dB) of the dominant frequency, relative to that of the secondary frequency, was expressed as the ratio of the sound pressure of the dominant frequency relative to that of the secondary frequency [dB $=20\times\log_{10}$(pressure ratio); 0 dB=1, +6 dB=2, +12 dB=4, etc.]. Note duration (ND), rise time (RT), and fall time (FT) were measured from oscillograms as in (b). Note duration was defined as the time between the onset and offset of an advertisement call note. Rise time was defined as the time from note onset to the time of maximum amplitude; fall time was defined as the time from maximum amplitude to note offset. Fundamental frequency (FF) was determined as the reciprocal of the average period of five repetitions of the fine-temporal waveform as shown in (c). Horizontal bars indicate time scales in (b) and (c).

ings and, although aware of each bullfrog's individual number (e.g., BF9-98), had no *a priori* expectations about the values to be determined in the analyses. The calls analyzed in this study are the same as those analyzed by Bee and Gerhardt (2001a) for the purpose of comparing among-individual variability in analyses of one call note versus multiple call notes; however, that study did not describe the patterns and magnitudes of individual variation in multiple-note calls.

Three spectral properties of advertisement calls were measured from power spectra generated by fast Fourier transformation (FFT) over a 100-ms interval from the middle of a call note, beginning at the point of peak signal amplitude (transform size=1024 points, filter bandwidth=14.5 Hz, 0–4 kHz setting). The spectral properties that were measured included dominant frequency and secondary frequency (to the nearest 10 Hz), and the amplitude of the dominant frequency relative to the secondary frequency (to the nearest dB) [Fig. 2]. I also determined the ratio of secondary frequency to dominant frequency (hereafter ''frequency ratio'') as a measure of the distribution of sound energy in the bimodal frequency spectrum. The fundamental frequency is equal to the reciprocal of the waveform periodicity and is related to the ''periodicity pitch'' of the signal (Simmons and Ferragamo, 1993). Using an oscillogram with an expanded time base, fundamental frequency was determined (to the nearest 1 Hz) as the reciprocal of the average period of five repetitions of the fine-temporal waveform measured from the middle of a call note [Fig. 2]. Three gross-temporal properties were measured (to the nearest 1 ms) from oscillograms, including note duration, rise time, and fall time [Fig. 2]. The recordings had sufficiently high signal-to-noise ratios to readily distinguish the onsets and offsets of individual notes from the background noise.

D. Statistical analysis of vocalizations

The statistical analyses described below were performed on call properties that were corrected for variation in water temperature at the times different individuals were recorded using the procedures outlined in Bee and Gerhardt (2001a).

1. Sources and patterns of variation

Two goals of the present study were to test for the presence of systematic *within-individual* variation in multiple-note advertisement calls that occurs among the separate notes within a call, and to compare this variation to the magnitude of among-individual variation present in the population. I used model II ANOVAs (Sokal and Rohlf, 1994) to examine the variation in call properties that was due to among-individual differences (*individual*) and among-note differences (*note*) by treating *individual* and *note* as main effects in a factorial design.[2] These analyses were designed to address three questions. First, the effect of *individual* tested whether the among-individual variation in call properties was significantly greater than the within-individual variation, which included variation due to among-note differences and among-call differences in the 15 call notes recorded from each individual. Second, the effect of *note* tested whether any systematic variation among separate notes within multiple-note calls was significantly greater than the within-note variation that was due to among-individual and among-call differences in each of the 125 exemplars (25 individuals×5 calls) of notes in the first, middle, and last positions of a call. Bonferroni *posthoc* comparisons were used to examine differences among first, middle, and last notes if the main effect of *note* was significant. Finally, the *individual*×*note* interaction term tested for any among-individual differences in the patterns of systematic variation among notes within an individual's calls. Because calls with different numbers of notes were included in these analyses, any variation due to the *individual*×*note* interaction must be interpreted with caution, as some of this variation could be due to differences in the number of call notes that different bullfrogs produced in their multiple note calls. For these and all subsequent statistical analyses, a criterion of $\alpha=0.05$ was used to determine significance.

2. Relative variation among acoustic properties

A third goal of this study was to compare the magnitude of within-individual variation among different call properties. Within-individual variability in each call property was estimated in two standardized ways. First, within-individual coefficients of variation (=SD/mean·100%) for each call property were calculated separately for each individual using the mean and standard deviation from the sample of 15 call notes (5 calls×3 notes) recorded from the individual. Second, the within-individual range of variation for each call property was calculated separately for each individual as a

percentage of the within-individual mean from the sample of 15 call notes recorded from the individual. These estimates of within-individual variation were log-transformed to improve normality and compared among call properties using a MANOVA approach to test for within-subjects effects. Subsequent Bonferroni *posthoc* tests were used to compare individual call properties.

3. Individual distinctiveness

The fourth goal of this study was to assess the extent to which advertisement calls can be considered individually distinct given the within-individual variation that occurs among separate notes within a call. Discriminant function analysis (DFA) is a commonly used multivariate statistical approach for examining the individual distinctiveness of acoustic signals (Nelson and Marler, 1990; Weary *et al.*, 1990; Hauser, 1991; Insley, 1992; Bee *et al.*, 2001; Fischer *et al.*, 2002). Instead of performing a DFA on the values for all eight call properties, which were highly intercorrelated (see the Appendix), I subjected the data to a principal components analysis (PCA) to extract orthogonal principal components and performed a DFA using the factor scores from the PCA as input variables. The discriminant functions were used to classify each of the 375 call notes as belonging to a particular individual based on a cross validation procedure.

TABLE I. Results of model II ANOVAs for each call property comparing the variation due to among-individual differences (*individual*), among-note differences (*note*), and among-note differences that varied among individuals (*individual*×*note*).

Call property	Source[a]	*F*	*P*
Fundamental frequency	individual	145.0	<0.01
	note	67.6	<0.01
	individual×note	1.9	<0.01
Dominant frequency	individual	113.2	<0.01
	note	51.3	<0.01
	individual×note	2.6	<0.01
Secondary frequency	individual	8.2	<0.01
	note	5.2	0.01
	individual×note	5.9	<0.01
Frequency ratio	individual	11.3	<0.01
	note	1.5	0.24
	individual×note	5.4	<0.01
Relative amplitude	individual	9.8	<0.01
	note	5.9	<0.01
	individual×note	3.0	<0.01
Note duration	individual	8.4	<0.01
	note	0.0	0.97
	individual×note	5.5	<0.01
Rise time	individual	10.7	<0.01
	note	41.1	<0.01
	individual×note	2.3	<0.01
Fall time	individual	3.8	<0.01
	note	32.5	<0.01
	individual×note	3.0	<0.01

[a]Degrees of freedom: *individual* (24,48), *note* (2,48), *individual*×*note* (48,300).

4. Generating hypotheses about JMDs

The final goal of the present study was to generate testable hypotheses about a receiver's behavioral discrimination thresholds, or JMDs (Nelson and Marler, 1990; Gerhardt, 1992), based on the patterns of among-individual and within-individual variation present in multiple-note advertisement calls. Signal detection theory provides an appropriate framework for understanding the relationships between individual variation in signals and behavioral discrimination thresholds (Wiley, 1994; Sherman *et al.*, 1997; Bradbury and Vehrencamp, 1998; Gerhardt and Huber, 2002). In a social recognition system, receivers are faced with the general problem of discriminating between familiar and unfamiliar signals, or between multiple familiar signals. From the receiver's perspective, there are four basic associations between the perceived signal and the behavioral response: correct detection, correct rejection, false alarm, and missed detection. As Wiley (1994) and others (Sherman *et al.*, 1997; Bradbury and Vehrencamp, 1998; Gerhardt and Huber, 2002) have pointed out, natural selection should favor behavioral discrimination thresholds that optimize the sum of the net advantages of each of these four behavioral outcomes weighted by its probability. The probability of each outcome depends on the overlap in the distributions of the signal properties of the individuals to be discriminated, which in turn depends on the magnitudes of differences among individuals in the population and on the variance in the signal distributions within individuals. Comparisons of the distribution of within-individual differences to that of among-individual differences provide a way to generate hypotheses about a receiver's behavioral threshold differences in the acoustic properties of familiar and unfamiliar signals that should lead to a discriminative behavioral response.

I used the following procedure to estimate JMDs. First, a distribution of among-individual differences was created by computing the absolute value of the difference between each of an individual's 15 call notes and each of the other 360 call notes from the remaining 24 individuals. This was repeated for each of the 25 individuals to create a distribution of among-individual differences based on a total of 67 500 pairwise among-individual comparisons. Second, a distribution of within-individual differences was created by computing the absolute value of the difference between each call note from an individual and the other 14 call notes recorded from that individual. This was repeated for each of the 25 individuals to create a distribution of within-individual differences based on 2625 pairwise within-individual comparisons. The among-individual and within-individual differences were standardized as percentages of the grand mean determined from all 375 call notes. Finally, as an estimate of a discrimination threshold, I computed the percentage difference in each signal property at which the cumulative proportion of among-individual differences falling below the threshold first exceeded the cumulative proportion of within-individual differences falling above this threshold. The JMD estimated in this way corresponds to the level of difference between two signals at which the difference is more likely attributable to an among-individual difference than to a within-individual difference.

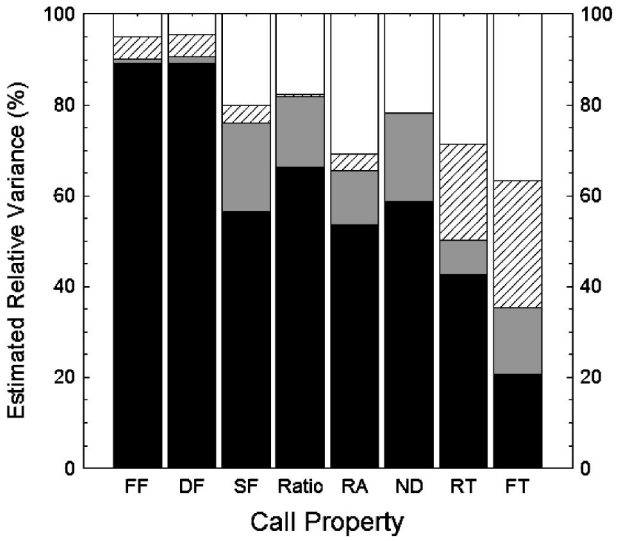


FIG. 3. Estimated relative variance accounted for by the main effects of *individual* (solid black) and *note* (diagonal lines), and the *individual*×*note* interaction (gray) in model II ANOVAs in which *individual* and *note* were tested as main effects in a factorial design. The white bars indicate the relative unexplained error variance. (FF=fundamental frequency; DF =dominant frequency; SF=secondary frequency; Ratio=ratio of SF:DF; RA=sound pressure of DF relative to that of SF; ND=note duration; RT =rise time; FT=fall time).

III. RESULTS

A. Sources and patterns of variation

Overall, the effects included in the model II ANOVAs explained substantial amounts of variation in each call property, as indicated by R^2 values that ranged from 0.59 for fall time to 0.95 for fundamental frequency and dominant frequency. Variation among individuals was significantly greater than that within individuals, even after including variation that occurred among three separate notes within each of five exemplars of an individual's multiple-note calls (Table 1). The proportions of variation that were due to among-individual differences (Fig. 3, black bars) were highest for fundamental frequency and dominant frequency (89%), intermediate for secondary frequency, frequency ratio, relative amplitude, and note duration (54%–66%), and lowest for rise time (43%) and fall time (21%).

There was also significant systematic variation that depended on the position of a note within a multiple-note call (*note*) for several call properties (Table I). The proportions of variation explained by systematic among-note differences (Fig. 3, diagonally patterned bars) were highest for rise time (21%) and fall time (28%). Fundamental frequency, dominant frequency, secondary frequency, and relative amplitude exhibited relatively less variation among notes (4%–5%). The variance components attributable to systematic among-note differences were negligible for frequency ratio and note duration (<0.5%).

For fine-temporal and spectral properties, *posthoc* comparisons revealed that fundamental frequency [Fig. 4(a)] and dominant frequency [Fig. 4(b)] were significantly lower in the first note of a call compared to the middle and last notes, which did not differ significantly. Secondary frequency [Fig. 4(c)] was significantly lower at the beginning and end of a call compared to notes from the middle of the call, and notes at the beginning of a call had significantly lower secondary frequencies than notes at the end of a call. The amplitude of the dominant frequency relative to that of the secondary frequency [Fig. 4(e)] was significantly lower in the first note of a call compared to the middle and last notes, which did not differ significantly.

The general similarities among the patterns of variation in fundamental frequency, dominant frequency, and secondary frequency reflect the fact that dominant frequency and secondary frequency are harmonically related to the fundamental. Slight differences among these patterns of variation can arise, however, because the secondary frequency can correspond to different harmonic components among different notes within a call, while the dominant frequency of each call note was always the second harmonic. Although there were no significant differences in the ratio of secondary frequency to dominant frequency [Fig. 4(d)], there was a slight trend for frequency ratio to be higher in notes from the middle of a call. Together, the results for secondary frequency and frequency ratio suggest that the peak of the spectral energy in the upper mode of the frequency spectrum shifted to slightly higher harmonics in the middle of the calls given by some individuals (see below).

Average note duration [Fig. 4(f)] was constant within a call, while there were significant differences in rise times [Fig. 4(g)] and fall times [Fig. 4(h)] that depended on the position of notes within a call. *Posthoc* comparisons indicated that rise times increased, and fall times decreased, between notes within a call [Figs. 4(g) and (h); see also Fig. 1].

Patterns of within-individual variation among first, middle, and last notes similar to those depicted in Fig. 4 were also observed after separately standardizing the values for each call note from each individual as a percentage deviation from the within-individual mean for that individual (Fig. 5). On average, fine-temporal and spectral call properties tended to vary little around the individual mean: fundamental frequency (3%), dominant frequency (3%), secondary frequency (7%), and frequency ratio (4%). The exception was relative amplitude, which varied about 18% around the individual mean. Among gross temporal properties of the amplitude envelope, the average note duration was constant across multiple notes, varying less than 1% around the individual mean, although the variability in note duration, as depicted by the size of the error bars in Fig. 5, was greater than that for fundamental frequency and dominant frequency. Compared to note duration, considerably more variation was observed for rise time (25%) and fall time (24%).

The *individual*×*note* interactions were significant for all call properties (Table I). The variance that could be attributed to the *individual*×*note* interaction (Fig. 3, gray bars) was greatest for secondary frequency, frequency ratio, relative amplitude, note duration, and fall time (15%–20%). The variance components for fundamental frequency and dominant frequency were the smallest (1%–2%), and that for rise time was intermediate (8%). Notice, in Fig. 3, that the proportion of variation in several call properties that could be attributed to the *individual*×*note* interaction was greater than that explained by systematic differences due to note position

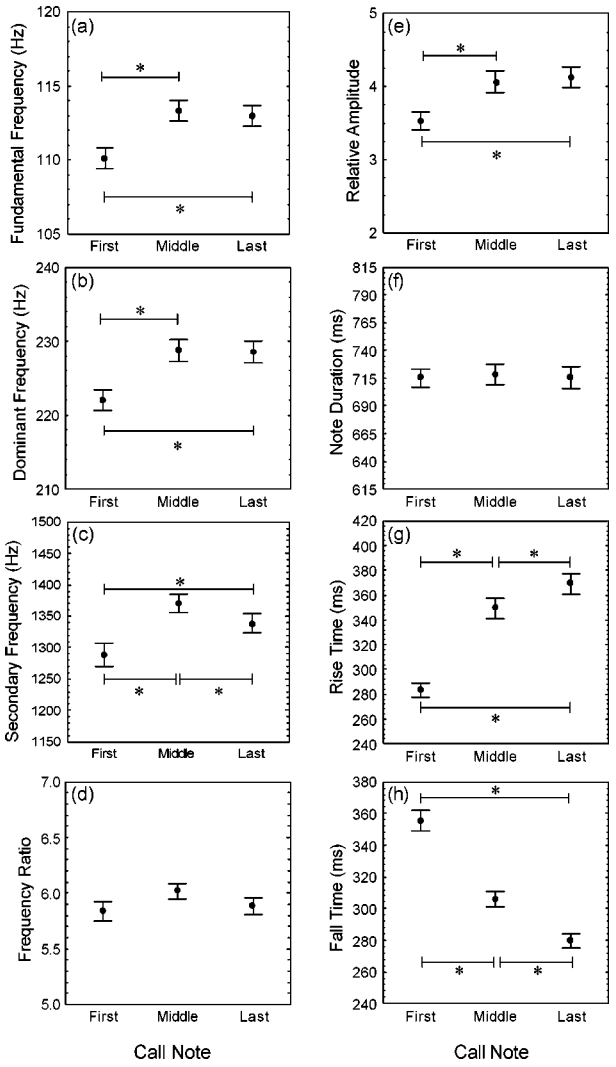


FIG. 4. Mean (±SE) values of the first, middle, and last notes for (a) fundamental frequency, (b) dominant frequency, (c) secondary frequency, (d) frequency ratio, (e) relative amplitude, (f) note duration, (g) rise time, and (h) fall time. These mean values are based on 125 exemplars of each call note (5 calls×25 individuals). The range of values depicted on each *y* axis approximately extends from one standard deviation below to one standard deviation above the sample means for each property, averaged over 3 notes, 5 calls, and 25 individuals. Asterisks indicate significant differences ($P<0.05$) in Bonferroni *posthoc* tests for model II ANOVAs in which the main effect of *note* was significant ($P<0.05$).

alone. This point is illustrated in Fig. 6 for the call properties of secondary frequency and note duration. As Fig. 6(a) illustrates, most individuals had secondary frequencies that were either relatively constant or increased slightly during multiple-note calls. However, at least a few individuals had average secondary frequencies that were considerably more variable among notes within calls [Fig. 6(b)]. Because male bullfrogs generally do not change position during the production of a single call, it is unlikely that differences in secondary frequency among notes within calls could arise due to changes in the positions of some individuals relative to the microphone, which could affect the relative amplitudes of harmonic components, and thus determination of secondary frequency. Hence, the patterns of changes in the frequency of peak sound energy in the upper mode of the bimodal frequency spectrum that occurred among notes within multiple note calls differed reliably (i.e., statistically) among individuals. The reliable individual differences in note-to-note variation were even more pronounced for the gross temporal property of note duration [Figs. 6(c)–(f)], for which less than 0.5% of the variation was explained by note position alone. The duration of notes within multiple-note calls was nearly constant within the calls of some individuals while for other individuals note duration changed by 100–200 ms during a call and exhibited either systematic decreases or increases. It seems unlikely that these distinct individual differences in the patterns of among-note variation exhibited in the calls of some individuals could result only from the fact that calls with different numbers of notes were included in the analysis, although additional work may be needed to clarify this point.

B. Relative variation among acoustic properties

There were significant differences in the variability among call properties both for the within-individual coefficients of variation [Wilks' $\lambda<0.01$, $F_{7,18}=423.1$, $P<0.0001$ and for within-individual ranges [Wilks' $\lambda=0.01$, $F_{7,18}=223.7$, $P<0.0001$. *Posthoc* tests revealed the same pattern of significant differences for both estimates of within-individual variation (Fig. 7). Fundamental frequency and

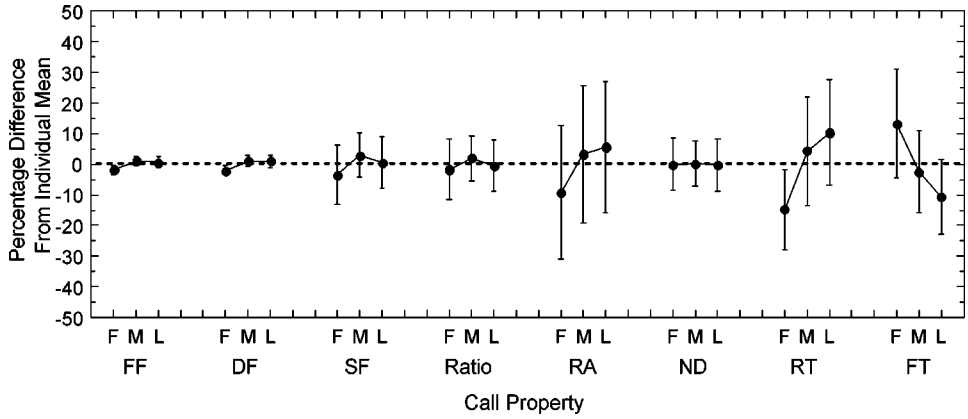


FIG. 5. Magnitudes of variation in the eight call properties among the first (F), middle (M), and last (L) notes within a call standardized as deviations from the within-individual means. Points and error bars depict the mean and ±1 SD, respectively. Percentage deviation scores were calculated separately for each individual ($N=25$) based on the deviation of values for each single note from the within-individual mean averaged over all 15 call notes (5 calls×3 notes) recorded from the individual. (FF=fundamental frequency; DF=dominant frequency; SF=secondary frequency; Ratio=ratio of SF:DF; RA=sound pressure of DF relative to that of SF; ND=note duration; RT=rise time; FT=fall time).

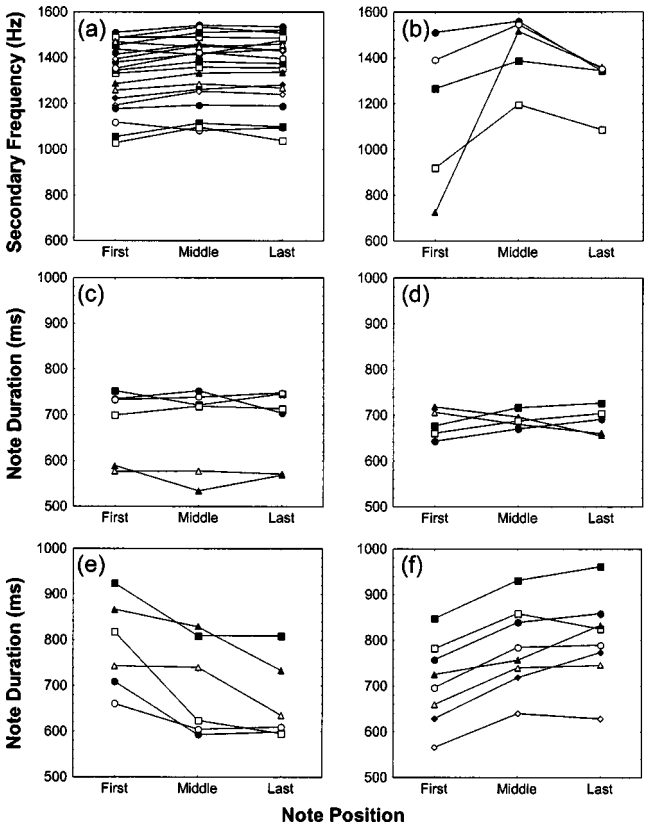


FIG. 6. *Individual*×*note* interactions from model II ANOVAs showing individual differences in the patterns of among-note variation in multiple-note advertisement calls for the call properties of (a,b) secondary frequency and (c–f) note duration. The value for each call note from each individual represents the average value of five calls.

dominant frequency exhibited significantly lower variability within individuals compared to the other six call properties. Secondary frequency, frequency ratio, and note duration exhibited significantly greater variability within individuals compared to fundamental frequency and dominant frequency. Note duration was also significantly more variable than secondary frequency and frequency ratio. The relative amplitude of the dominant and secondary frequencies, and the gross temporal properties of rise time and fall time, exhibited significantly more variability within individuals than the other call properties.

C. Individual distinctiveness

A PCA of the values for each of the eight acoustic properties generated four factors with eigenvalues greater than 1.0. However, because the fifth factor had an eigenvalue of 0.75 and explained an additional 10% of the variance, I included the factor scores from the first five factors as input variables to the DFA (Table II). Together, the first five factors from the PCA explained 98.4% of the variance in the acoustic properties. The DFA generated five significant discriminant functions (P's<0.01). Three discriminant functions had eigenvalues greater than 1.0 and, together, they accounted for 96% of the variation (Table III). The first discriminant function accounted for 80% of the variation and was highly correlated with scores from the first PCA factor, which loaded heavily on fundamental frequency and dominant frequency (Table III). The second discriminant function, which accounted for an additional 10% of the variance, was correlated with PCA factor 3, which loaded most heavily on secondary frequency and frequency ratio (Table III). The third discriminant function explained an additional 6% of the variance and was highly correlated with PCA factor 2, which loaded heavily on note duration and rise time (Table III). The DFA successfully assigned 283 of the 375 recorded call notes (75.5%) to the correct individual, which is significantly higher than the classification success of 4% expected by randomly assigning notes to individuals (Fisher's exact test, $P<0.01$). Therefore, individuals can be statistically distinguished by their multiple-note advertisement calls.

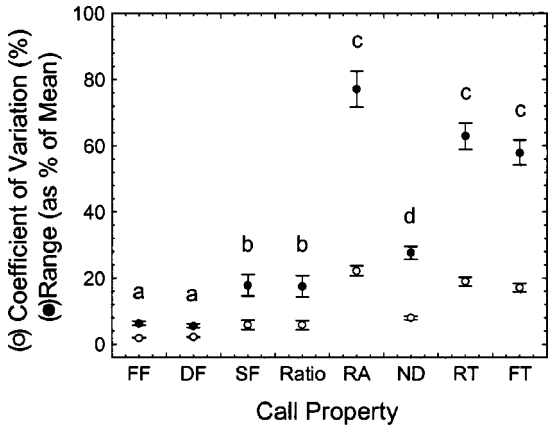


FIG. 7. Comparisons of the relative variability in different call properties. Points and error bars depict the mean and ±1 SE, respectively. Different letters above the error bars indicate significant differences ($P<0.05$) between call properties in *posthoc* tests following MANOVAs of the log-transformed data. (FF=fundamental frequency; DF=dominant frequency; SF=secondary frequency; Ratio=ratio of SF:DF; RA=sound pressure of DF relative to that of SF; ND=note duration; RT=rise time; FT=fall time).

D. Estimated JMDs

Figure 8 depicts histograms showing the distributions of among-individual differences (black bars) and within-individual differences (white bars) in acoustic properties of bullfrog advertisement calls. Table IV shows the estimated JMDs for each call property computed from similar histograms having 1%-wide bins. These estimated JMDs are also compared to the within-individual coefficients of variation and the mean-standardized within-individual ranges of variation that were depicted in Fig. 7. Table IV also shows the proportions of among-individual differences and within-individual differences that fall above and below the estimated JMDs. The estimated JMD of most call properties generally falls between the within-individual coefficient of variation and the mean-standardized range, with the exception of fall time, which has an estimated JMD that is slightly less than the coefficient of variation (Table IV). With the exception of relative amplitude, the estimated JMDs for fine-temporal and spectral call properties are generally lower (4%–6%) than those for gross temporal call properties (10%–21%).

In the context of territorial neighbor recognition, in which individuals respond more aggressively toward unfamiliar individuals, a correct detection is the equivalent of

TABLE II. Normalized factor loadings (varimax rotation) from a principal components analysis of the eight acoustic properties. Bold type indicates the heaviest factor loadings.

	Factor				
Call property	1	2	3	4	5
Fundamental frequency	**0.98**	0.02	0.11	0.02	0.10
Dominant frequency	**0.98**	0.03	0.05	0.06	0.12
Secondary frequency	0.12	−0.01	**−0.99**	0.00	0.08
Frequency ratio	−0.40	−0.02	**−0.92**	−0.02	0.00
Relative amplitude	−0.16	0.10	0.07	−0.06	**−0.98**
Note duration	0.02	**0.92**	−0.01	0.33	−0.04
Rise time	0.04	**0.93**	0.04	−0.29	−0.09
Fall time	0.06	0.01	0.01	**0.99**	0.06

responding aggressively to the signals of stranger, while a false alarm is the equivalent of aggressively responding to a familiar neighbor that already shares an established territorial boundary. A correct rejection involves withholding an aggressive response to the signals of a familiar neighbor, while a missed detection is the equivalent of failing to respond aggressively toward an unfamiliar individual. Because a territory holder is much more likely to hear its nearby neighbors than the occasional stranger, at least after stable boundaries have been established, and because aggressive interactions are probably costly, we should assume that there would be a selective advantage to having a behavioral discrimination threshold that leads to high correct rejection rates and low false alarm rates. As shown in Table IV, the likelihood that individuals would respond appropriately to the calls of a familiar neighbor (correct rejection) are considerably higher than the likelihood of mistakenly responding to the neighbor's calls (false alarm). Notice, also, that *differences* in the estimated rates of correct rejections and false alarms are highest for fundamental frequency and dominant frequency. However, as the results in Table IV also make clear, if discrimination were based on the estimated JMD for a single call property, there could be a relatively high rate of missed detections (26%–47%), in which the individual could fail to respond aggressively toward an unfamiliar individual. This result, of course, stems from the overlap in the distributions of among-individual and within-individual differences, and the fact that it is impossible to simultaneously maximize both correct rejections and missed detections.

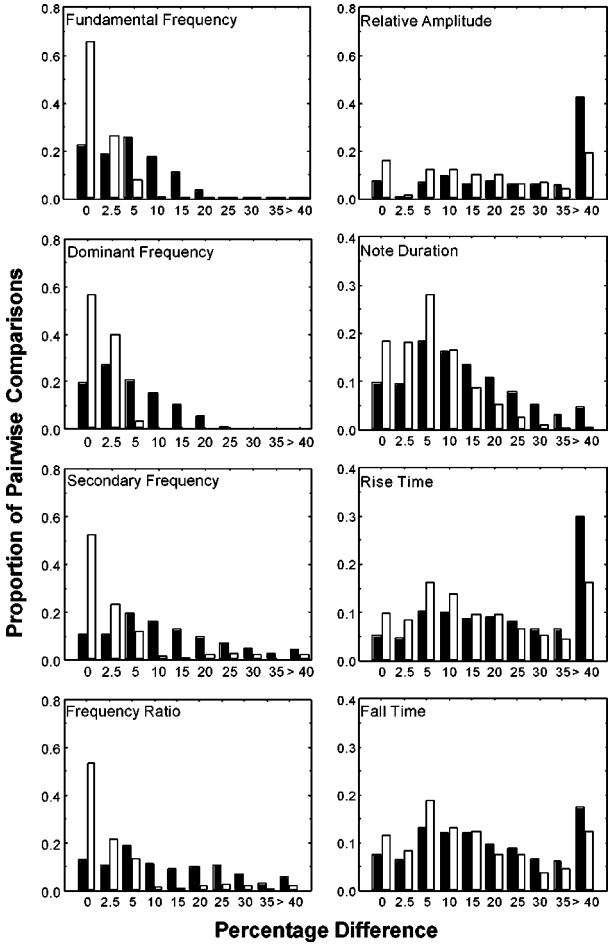


FIG. 8. Histograms showing the distributions of percentage differences in each acoustic property that occurred among individuals (black bars) and within-individuals (white bars). Numbers along the *x* axis depict the lower bounds of the histogram bins.

IV. DISCUSSION

A. Implications of within-individual variation for vocally mediated recognition in bullfrogs

Bee and Gerhardt (2001a) reported that several acoustic properties of bullfrog advertisement calls exhibited significant variation both among individuals and among separate notes within individuals. The results reported above extend this previous work in several important ways. First, the re-

TABLE III. Factor structure from a discriminant function analysis showing the correlations between principal component scores and the canonical roots. For each factor from the principal components analysis (PCA), the acoustic properties with the heaviest factor loadings, as depicted in Table II, are shown in parentheses. Bold type indicates correlations greater than $r = 0.50$.

	Canonical Root				
PCA factor	1	2	3	4	5
Factor 1 (fundamental and dominant frequencies)	**−0.61**	−0.07	−0.11	−0.25	**0.74**
Factor 2 (note duration and rise time)	0.05	−0.39	**−0.79**	0.19	0.43
Factor 3 (secondary frequency and frequency ratio)	−0.04	**−0.61**	0.45	**0.59**	−0.28
Factor 4 (fall time)	−0.04	−0.12	−0.22	−0.31	**−0.92**
Factor 5 (relative amplitude)	−0.06	0.36	−0.24	**0.81**	−0.40
Eigenvalue	22.0	2.8	1.6	1.0	0.2
Cumulative proportion of variance explained	0.80	0.90	0.96	0.99	1.00

TABLE IV. Estimated ''just-meaningful differences'' (JMDs) for each call property in relation to within-individual coefficients of variation and mean-standardized ranges, and the proportion of among-individual and within-individual differences that fall above and below the estimated JMD.

				Among-individual differences (%)		Within-individual differences (%)	
Call property	Within-individual coefficient of variation (%)	Within-individual range (%)	Estimated JMD (%)[a]	Above estimated JMD (correct detections)	Below estimated JMD (missed detections)	Below estimated JMD (correct rejections)	Above estimated JMD (false alarms)
Fundamental frequency	1.9	6.3	4	66	34	84	16
Dominant frequency	2.1	5.6	5	53	47	96	4
Secondary frequency	5.9	17.8	6	74	26	82	18
Frequency ratio	5.8	17.5	6	72	28	80	20
Relative amplitude	22.3	77.2	25	61	39	63	37
Note duration	8.0	27.6	10	62	38	65	35
Rise time	18.9	62.9	21	58	42	60	40
Fall time	17.8	58.0	17	56	44	57	43

[a]JMDs were estimated from histograms of the among-individual and within-individual differences in bins of 1%. See Sec. II D 4. for additional details.

sults describe the systematic patterns of variation in call properties present among notes within an individual's calls. These findings can be used to improve the design of future playback experiments that investigate the perceptual basis of discrimination by providing the data necessary to incorporate natural levels of within-individual variation into perceptual learning and discrimination tasks. Second, the results from the present study suggest that there may be significant among-individual differences in the patterns of within-individual variation in some call properties, as indicated by the significant *individual*×*note* interactions in the model II ANOVAs (Table I). This finding suggests that within-individual variation need not function solely as a constraint on recognition, and suggests the hypothesis that among-individual differences in the patterns of within-individual (i.e., among-note) variation in the calls of some individuals might function as an additional source of acoustic recognition cues. Third, the results show that, with the exception of relative amplitude, fine-temporal and spectral call properties exhibited significantly lower magnitudes of variation within individuals than did gross temporal properties. Finally, the present study demonstrates that fundamental frequency and the correlated property of dominant frequency exhibited the most reliable among-individual differences, contributed most toward statistically discriminating among individuals based on variation in multiple-note calls, and had the lowest estimated JMDs, even though these two correlated call properties also vary significantly among the notes within multiple-note calls. One general conclusion that may be drawn from this study and that by Bee and Gerhardt (2001a) is that fine-temporal and spectral call properties are potentially better vocal recognition cues than gross temporal call properties, despite the fact that fine-temporal and spectral call properties also exhibit significant within-individual variation among the notes in an individual's multiple-note calls.

B. Individual discrimination based on "pitch"

As in some songbirds (Brooks and Falls, 1975b; Nelson, 1989), male bullfrogs appear to use acoustic properties related to pitch to discriminate among familiar and unfamiliar vocalizations. In habituation-discrimination experiments, Bee and Gerhardt (2001b,c) demonstrated that a 5–10% change in fundamental frequency was sufficient to elicit significant levels of recovery of the habituated aggressive response. Hence, the JMDs for fundamental frequency and correlated spectral properties of 4%–6% that were estimated above (Table IV) appear to be in line with the presently available data from field playback tests. Psychophysical estimates of frequency discrimination are available for only one frog, the South African clawed frog, *Xenopus leavis* (Elepfandt *et al.*, 2000). Frequency difference limens for pure tones measured using a GO/NO-GO conditioning paradigm were approximately 5% in the hearing range of the amphibian papilla, and 2.4%–6% in the hearing range of the basilar papilla. The current estimates of JMDs for fine-temporal or spectral properties of bullfrog advertisement calls may, therefore, approach the absolute limits of the anuran auditory system.

C. Individual discrimination based on other acoustic properties

At present, there have been only limited efforts to investigate the use of other call properties as recognition cues in bullfrogs. As pointed out by Bee and Gerhardt (2001a), and as illustrated in Table IV, discrimination based only on call properties related to the pitch of the call would likely be far from perfect because of the non-negligible proportion of relatively small among-male differences in these properties present in the population (Fig. 8, Table IV). Other properties of advertisement calls also have some potential to statistically distinguish among male bullfrogs (Table I, Bee and Gerhardt, 2001a; Simmons, 2004), and therefore we might expect behavioral discrimination to be based on multiple dimensions of signal variability. Although the sample sizes were small, Bee and Gerhardt (2001c) found no evidence in their habituation-discrimination experiments that 10% decreases in note duration and internote interval or 10% increases in rise and fall times elicit recovery of habituated aggressive responses. Novel stimuli having differences in the harmonic fine-structure (presence/absence and relative amplitudes of harmonic components), but a constant fundamen-

tal frequency, also failed to elicit recovery of habituated aggressive responses (Bee and Gerhardt, 2001c). On the one hand, these previous findings are consistent with the relatively higher estimated JMDs for rise time, fall time, and relative amplitude (Table IV) in comparison to those for fine-temporal and spectral call properties. On the other hand, however, if note duration is used as a recognition cue, then these results suggest that the JMD for note duration used by receivers is higher than that estimated here. Estimating JMDs for note duration is complicated by the relatively large among-individual differences in the patterns of note-to-note variation. Clearly, more work is needed to determine whether receivers use gross temporal call properties as recognition cues.

Three properties of bullfrog advertisement calls, for which the patterns of among-individual and within-individual variation have not been investigated thoroughly, are internote interval, note duty cycle, and phase spectrum. At a constant note duty cycle, a 10% decrease in internote interval (and note duration) failed to elicit response recovery (Bee and Gerhardt, 2001c). When note duration was held constant and the duty cycle was increased 20% (from 0.5 to 0.7) by decreasing internote interval by 57% (700 to 300 ms), half of the subjects responded aggressively to the novel stimulus [only 25% responded to the control stimulus (Bee and Gerhardt, 2001c)]. None of the subjects responded to a novel stimulus having a 10% increase in duty cycle (from 0.5 to 0.6) effected by a 33% decrease in internote interval (700 to 467 ms). In a field playback test examining habituation and recovery of evoked advertisement calling, Hainfeld *et al.* (1996) found some evidence to suggest that territorial male bullfrogs might behaviorally discriminate among advertisement calls differing in phase spectrum based on differences in waveform periodicity. If there were reliable individual differences in the phase spectra of advertisement calls that were reliably transmitted through the environment across distances that separate individuals, then the phase spectrum of a neighbor's calls might also function as an additional acoustic cue for recognition. This seems unlikely, however, as Hainfeld *et al.* (1996) argued that phase spectrum was not a particularly salient cue for discriminating among advertisement calls in the natural environment.

D. Relevance of estimating JMDs from patterns of individual variation in signals

Acoustical and statistical analyses of signals cannot determine a receiver's discrimination thresholds, which instead requires experimentation. Therefore, it is important for readers to keep in mind that estimating JMDs from the degree of overlap in the distributions of among-individual and within-individual differences in multiple note calls represents one way to generate hypotheses about a receiver's behavioral discrimination threshold that can be tested in future studies. The value of deriving estimated JMDs in this way is that it provides a reasonable starting point from which to assess actual JMDs in field playback experiments.

The actual discrimination thresholds of receivers could differ from those estimated above for numerous reasons. For example, the estimated JMDs may fall below the resolution of the bullfrog auditory system. This possibility is difficult to address because psychophysical estimates of spectral and temporal resolution of the anuran auditory system are still generally lacking. Alternatively, the actual and estimated JMDs might differ because natural selection favors behavioral discrimination thresholds that are flexible within individuals and can be adjusted according to spatial and temporal variation in the level of competition for territories. If suitable territories were limited, or stable territory boundaries were in flux, this could increase the benefits of correct detections (i.e., responding aggressively toward strangers) and increase the costs of missed detections (i.e., failing to respond aggressively toward strangers), which would shift the discrimination threshold toward lower values for threshold differences. One result of such a threshold shift would be an increase in the false alarm rate, so that individuals would also more often respond aggressively toward neighbors. Indeed, Stoddard (1996) describes a situation in song sparrows (*Melospiza melodia*) in which strong neighbor-stranger discrimination occurred in two seasons in which territories were relatively stable, but not in an intervening season in which territory boundaries were in constant flux due to an unusually high predation rate. As argued by Stoddard (1996), an ultimate level explanation for his finding is that neighbors and strangers were equally threatening during the period of high territory turn over, and thus deserved similar responses. A shift in the birds' behavioral discrimination threshold might explain this finding at a more proximate level.

An important limitation of the above discussion of estimated JMDs is that these estimates assume that within-individual variation functions as a constraint on recognition by increasing the overlap in the distributions of signal properties of different individuals. However, this need not be the case. If particular call properties vary substantially within individuals, but there are also reliable among-individual differences in the *patterns* of within-individual variation, then the pattern of within-individual variation itself may be used as an additional cue to identify the individual. This point is best illustrated by the gross temporal property of note duration [Figs. 6(c)–(f)], for which there was a significant *individual*×*note* interaction that accounted for 20% of the variation in the sample of recorded call notes (Table I, Fig. 3). Determining whether receivers take advantage of this additional source of among-individual variation will require testing in future playback studies.

E. Conclusions

Surprisingly little effort has been made to integrate, either within or across species, an understanding of the underlying acoustic and perceptual bases of vocally mediated social recognition. More commonly, studies of vocal recognition examine either (*i*) the ability of signals to statistically identify individuals without regard for whether or not the animals actually do, or possibly even could, discriminate among individuals based on the patterns of observed signal variation, or (*ii*) the ability of animals to discriminate between the signals of individuals with little or no regard for understanding the underlying acoustic basis of recognition.

The present study represents the latest effort to better understand both the acoustic and perceptual bases of vocally mediated neighbor recognition in North American bullfrogs (Bee and Gerhardt, 2001a,b,c, 2002, Bee, 2001, 2003). A better understanding of the bioacoustic basis of vocally mediated social recognition systems requires a description of the within-signal variation that individuals experience both when learning to recognize, and when discriminating among, the signals of other individuals. The ultimate goal of acoustical and statistical analyses like those reported here should be to relate patterns of signal variation both to the JMD required to elicit evolutionarily adaptive behavioral responses from receivers, as well as to the ''just-noticeable difference'' (JND) of the receiver's auditory system. The habituation-discrimination paradigm represents a powerful experimental tool for providing behavioral estimates of receivers' JMDs under natural conditions. Psychophysical estimates of the JNDs of acoustic signal properties are generally still lacking for frogs. Future psychophysical experiments with bullfrogs that employ techniques based on conditioning (Elepfandt *et al.*, 2000) or reflex modification (Megela-Simmons *et al.*, 1985; Megela-Simmons, 1988) could provide much needed insights into the mechanisms of vocal recognition in bullfrogs by illuminating the relationships between the spectral and temporal resolution of the auditory system, on the one hand, and patterns of signal variation and the behavioral threshold for discriminating between familiar and unfamiliar signals on the other.

ACKNOWLEDGMENTS

Jeff Koppleman and Don Martin provided access to the study sites, Chris Bowling assisted with field recordings, Michael Seaborn assisted with acoustical analyses, and H. Carl Gerhardt and two anonymous referees provided helpful comments on the manuscript. Funding was provided by an NSF Graduate Research Fellowship, an NSF Dissertation Improvement Grant, and an NSF International Research Fellowship Award. Recordings were made while the author was a doctoral student at the University of Missouri—Columbia and were approved by the University of Missouri Animal Care and Use Committee (IACUC # 2944).

APPENDIX: CORRELATIONS AMONG CALL PROPERTIES

Pearson product-moment correlation coefficients (r, below the diagonal) and the associated P-values (above the diagonal) for correlations between call properties calculated over 375 call notes.

	Call Property							
Call property	Fundamental frequency	Dominant frequency	Secondary frequency	Frequency ratio	Relative amplitude	Note duration	Rise time	Fall time
Fundamental frequency		<0.01	0.78	<0.01	<0.01	0.47	0.36	0.08
Dominant frequency	0.95		0.08	<0.01	<0.01	0.25	0.36	0.02
Secondary frequency	0.01	0.09		<0.01	<0.01	0.95	0.36	0.97
Frequency ratio	−0.48	−0.44	0.85		0.93	0.67	0.20	0.30
Relative amplitude	−0.25	−0.27	−0.16	0.00		0.03	<0.01	0.02
Note duration	0.04	0.06	0.00	−0.02	0.11		<0.01	<0.01
Rise time	0.05	0.05	−0.05	−0.07	0.20	0.74		<0.01
Fall time	0.09	0.12	0.00	−0.05	−0.13	0.32	−0.26	

[1]The five exemplars of multiple-note calls analyzed from each individual were selected from a longer sequence of 20 consecutive advertisement calls recorded during a single night. Bee and Gerhardt (2001a) have shown that the acoustic properties of the first note of multiple-note calls are highly repeatable across separate nights. Model II ANOVAs that tested the effect of among-call variation in the five exemplars recorded from each individual in an individual×call factorial design confirmed that little of the variation in the sample of recordings was explained by systematic variation among calls (all F's<1.2, all P's>0.17).

[2]Separate univariate analyses were performed for each acoustic property because the question of interest addressed how each call property varied among individuals and among notes within calls, and not whether there was overall significant variation. This is not meant to imply that all eight acoustic properties uniquely contribute to the total amount of information in the signal (Beecher, 1989). This is certainly not the case, as there are significant correlations among several call properties (see the Appendix).

Aubin, T., and Jouventin, P. (**2002**). ''How to vocally identify kin in a crowd: The penguin model,'' Adv. Study Behav. **31**, 243–277.

Balcombe, J. P. (**1990**). ''Vocal recognition of pups by mother Mexican free-tailed bats, *Tadarida brasiliensis mexicana*,'' Anim. Behav. **39**, 960–966.

Bee, M. A. (**2001**). ''Habituation and sensitization of aggression in bullfrogs (*Rana catesbeiana*): Testing the dual-process theory of habituation,'' J. Comp. Psychol. **115**, 307–316.

Bee, M. A. (**2003**). ''Experience-based plasticity in acoustically evoked aggression in a territorial frog,'' J. Comp. Physiol., A **189**, 485–496.

Bee, M. A., and Gerhardt, H. C. (**2001a**). ''Neighbour-stranger discrimination by territorial male bullfrogs (*Rana catesbeiana*): I. Acoustic basis,'' Anim. Behav. **62**, 1129–1140.

Bee, M. A., and Gerhardt, H. C. (**2001b**). ''Neighbour-stranger discrimination by territorial male bullfrogs (*Rana catesbeiana*): II. Perceptual basis,'' Anim. Behav. **62**, 1141–1150.

Bee, M. A., and Gerhardt, H. C. (**2001c**). ''Habituation as a mechanism of reduced aggression between neighboring territorial male bullfrogs, *Rana catesbeiana*,'' J. Comp. Psychol. **115**, 68–82.

Bee, M. A., and Gerhardt, H. C. (**2002**). ''Individual voice recognition in a territorial frog (*Rana catesbeiana*),'' Proc. R. Soc. London, Ser. B **269**, 1443–1448.

Bee, M. A., Kozich, C. E., Blackwell, K. J., and Gerhardt, H. C. (**2001**). ''Individually distinct advertisement calls of territorial male green frogs, *Rana clamitans*: Implications for individual discrimination,'' Ethology **107**, 65–84.

Beecher, M. D. (**1989**). ''Signalling systems for individual recognition: an information theory approach,'' Anim. Behav. **38**, 248–261.

Beecher, M. D. (**1991**). ''Successes and failures of parent-offspring recognition in animals,'' in *Kin Recognition*, edited by P. G. Hepper (Cambridge U.P., Cambridge), pp. 94–124.

Beecher, M. D., Stoddard, P. K., Campbell, S. E., and Horning, C. L. (**1996**).

''Repertoire matching between neighbouring song sparrows,'' Anim. Behav. **51**, 917–923.
Bradbury, J. W., and Vehrencamp, S. L. (**1998**). *Principles of Animal Communication* (Sinauer, Sunderland, MA).
Brooks, R. J., and Falls, J. B. (**1975a**). ''Individual recognition by song in white-throated sparrows. I. Discrimination of songs of neighbors and strangers,'' Can. J. Zool. **53**, 879–888.
Brooks, R. J., and Falls, J. B. (**1975b**). ''Individual recognition by song in white-throated sparrows. III. Song features used in individual recognition,'' Can. J. Zool. **53**, 1749–1761.
Capranica, R. A. (**1965**). *The Evoked Vocal Response of the Bullfrog* (MIT, Cambridge).
Cheney, D. L., and Seyfarth, R. M. (**1999**). ''Recognition of other individuals' social relationships by female baboons,'' Anim. Behav. **58**, 67–75.
Crawford, J. D., Cook, A. P., and Heberlein, A. S. (**1997**). ''Bioacoustic behavior of African fishes (Mormyridae): Potential cues for species and individual recognition in *Pollimyrus*,'' J. Acoust. Soc. Am. **102**, 1200–1212.
Davis, M. S. (**1987**). ''Acoustically mediated neighbor recognition in the North American bullfrog, *Rana catesbeiana*,'' Behav. Ecol. Sociobiol. **21**, 185–190.
Elepfandt, A., Eistetter, I., Fleig, A., Günther, E., Hainich, M., Hepperle, S., and Traub, B. (**2000**). ''Hearing threshold and frequency discrimination in the purely aquatic frog *Xenopus laevis* (Pipidae): Measurement by means of conditioning,'' J. Exp. Biol. **203**, 3621–3629.
Fischer, J., Hammerschmidt, K., Cheney, D. L., and Seyfarth, R. M. (**2002**). ''Acoustic features of male baboon loud calls: Influences of context, age, and individuality,'' J. Acoust. Soc. Am. **1115**, 1465–1474.
Gelfand, D. L., and McCracken, G. F. (**1986**). ''Individual variation in the isolation calls of Mexican free-tailed bat pups (*Tadarida brasiliensis mexicana*),'' Anim. Behav. **34**, 1078–1086.
Gentner, T. Q., and Hulse, S. H. (**1998**). ''Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*,'' Anim. Behav. **56**, 579–594.
Gentner, T. Q., and Hulse, S. H. (**2000**). ''Perceptual classification based on the component structure of song in European starlings,'' J. Acoust. Soc. Am. **107**, 3369–3381.
Gerhardt, H. C. (**1992**). ''Multiple messages in acoustic signals,'' Semin. Neurosci. **4**, 391–400.
Gerhardt, H. C., and Huber, F. (**2002**). *Acoustic Communication in Insects and Anurans* (Univ. of Chicago, Chicago).
Hainfeld, C. A., Boatright-Horowitz, S. L., Boatright-Horowitz, S. S., and Megela Simmons, A. (**1996**). ''Discrimination of phase spectra in complex sounds by the bullfrog (*Rana catesbeiana*),'' J. Comp. Physiol., A **179**, 75–87.
Hauser, M. D. (**1991**). ''Sources of acoustic variation in rhesus macaque (*Macaca mulatta*) vocalizations,'' Ethology **89**, 29–46.
Howard, R. D. (**1978**). ''The evolution of mating strategies in bullfrogs, *Rana catesbeiana*,'' Evolution (Lawrence, Kans.) **32**, 850–871.
Insley, S. J. (**1992**). ''Mother-offspring separation and acoustic stereotypy: A comparison of call morphology in two species of pinnipeds,'' Behaviour **120**, 103–122.
Lengagne, T., Lauga, J., and Jouventin, P. (**1997**). ''A method of independent time and frequency decomposition of bioacoustic signals: inter-individual recognition in four species of penguins,'' C. R. Acad. Sci. Paris, Sci. Vie **320**, 885–891.
Megela-Simmons, A. (**1988**). ''Masking patterns in the bullfrog (*Rana catesbeiana*). I. Behavioral effects,'' J. Acoust. Soc. Am. **83**, 1087–1092.
Megela-Simmons, A., Moss, C. F., and Daniel, K. M. (**1985**). ''Behavioral audiograms of the bullfrog (*Rana catesbeiana*) and the green tree frog (*Hyla cinerea*),'' J. Acoust. Soc. Am. **78**, 1236–1244.
Myrberg, A. A., and Riggio, R. J. (**1985**). ''Acoustically mediated individual recognition by a coral reef fish (*Pomacentrus partitus*),'' Anim. Behav. **33**, 411–416.
Myrberg, A. A., Ha, S. J., and Shamblott, M. J. (**1993**). ''The sounds of bicolor damselfish (*Pomacentrus partitus*): Predictors of body size and a spectral basis for individual recognition and assessment,'' J. Acoust. Soc. Am. **94**, 3067–3070.
Nelson, D. A. (**1989**). ''Song frequency as a cue for recognition of species and individuals in the field sparrow (*Spizella pusilla*),'' J. Comp. Psychol. **103**, 171–176.
Nelson, D. A., and Marler, P. (**1990**). ''The perception of birdsong and an ecological concept of signal space,'' in *Comparative Perception Vol. 2*, edited by W. C. Stebbins and M. A. Berkley (Wiley, New York), pp. 443–478.
Polakow, D. A. (**1997**). ''Communication and sexual selection in the barking gecko (*Ptenopus kochi*),'' unpublished M.Sc. thesis, University of Capetown, Capetown.
Robertson, B. C. (**1996**). ''Vocal mate recognition in a monogamous, flock-forming bird, the silvereye, *Zosterops lateralis*,'' Anim. Behav. **51**, 303–311.
Sherman, P. W., Reeve, H. K., and Pfennig, D. W. (**1997**). ''Recognition systems,'' in *Behavioral Ecology: An Evolutionary Approach*, edited by J. R. Krebs and N. B. Davies (Blackwell, Malden, MA), pp. 69–96.
Simmons, A. M. (**2004**). ''Call recognition in the bullfrog, *Rana catesbeiana*: Generalization along the duration continuum,'' J. Acoust. Soc. Am. **115**, 1345–1355.
Simmons, A. M., and Ferragamo, M. (**1993**). ''Periodicity extraction in the anuran auditory nerve I. Pitch-shift effects,'' J. Comp. Physiol., A **172**, 57–69.
Sokal, R. R., and Rohlf, F. J. (**1994**). *Biometry* (Freeman, New York).
Stoddard, P. K. (**1996**). ''Vocal recognition of neighbors by territorial passerines,'' in *Ecology and Evolution of Acoustic Communication in Birds*, edited by D. E. Kroodsma and E. H. Miller (Cornell U.P., Ithaca, NY), pp. 356–374.
Weary, D. M., Norris, K. J., and Falls, J. B. (**1990**). ''Song features birds use to identify individuals,'' Auk **107**, 623–625.
Wiley, R. H. (**1994**). ''Errors, exaggeration, and deception in animal communication,'' in *Behavioral Mechanisms in Evolutionary Ecology*, edited by L. A. Real (Univ. of Chicago, Chicago), pp. 157–189.

An echolocation model for the restoration of an acoustic image from a single-emission echo

Ikuo Matsuo[a] and Masafumi Yano
Research Institute of Electrical Communication, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai, 980-8577, Japan

(Received 18 June 2004; revised 18 August 2004; accepted 12 September 2004)

Bats can form a fine acoustic image of an object using frequency-modulated echolocation sound. The acoustic image is an impulse response, known as a reflected-intensity distribution, which is composed of amplitude and phase spectra over a range of frequencies. However, bats detect only the amplitude spectrum due to the low-time resolution of their peripheral auditory system, and the frequency range of emission is restricted. It is therefore necessary to restore the acoustic image from limited information. The amplitude spectrum varies with the changes in the configuration of the reflected-intensity distribution, while the phase spectrum varies with the changes in its configuration and location. Here, by introducing some reasonable constraints, a method is proposed for restoring an acoustic image from the echo. The configuration is extrapolated from the amplitude spectrum of the restricted frequency range by using the continuity condition of the amplitude spectrum at the minimum frequency of the emission and the minimum phase condition. The determination of the location requires extracting the amplitude spectra, which vary with its location. For this purpose, the Gaussian chirplets with a carrier frequency compatible with bat emission sweep rates were used. The location is estimated from the temporal changes of the amplitude spectra. [DOI: 10.1121/1.1811411]

PACS numbers: 43.80.Lb, 43.64.Bt, 43.80.Jz [JAS] Pages: 3782–3788

I. INTRODUCTION

Bats use frequency-modulated (FM) echolocation sounds to locate and discriminate individual objects in three-dimensional (3D) space, even when they are moving (Griffin, 1958; Griffin *et al.*, 1965; Simmons *et al.*, 1995; Webster and Griffin, 1962). The process by which they accomplish this task is not fully understood. If an emission is considered as an acoustic impulse, the acoustic image of an object corresponds to the impulse response, which is referred to as the reflected-intensity distribution (Weißenbacher and Wiegrebe, 2003). The reflected-intensity distribution of an object located at a particular range includes information on its position and shape. The returning echo is produced by convolving an emitted sound with the reflected-intensity distribution. It is therefore an inverse problem to compute the reflected-intensity distribution from the emitted sound and the echo. However, in the case of bat echolocation, it is difficult to directly determine the reflected-intensity distribution from the perceived echo. A single emission from a bat consists of a principal frequency that sweeps through a range of values from ~20 to 100 kHz (Fenton and Bell, 1981), and the time resolution of the bat peripheral auditory system is not sufficient to detect the phase of the perceived echo (Brugge *et al.*, 1969; Rose *et al.*, 1967). This indicates that the output of the peripheral auditory system comprises only the amplitude spectrum of a restricted frequency range, whereas the reflected-intensity distribution is composed of both the amplitude and phase spectra over a range of frequencies. Despite these limitations, bats have been reported to perceive changes in the phase spectra of an echo during range-discrimination tasks (Moss and Simmons, 1993; Simmons, 1979, 1989; Simmons *et al.*, 1990). This raises the question of how the reflected-intensity distribution can be restored from the amplitude spectrum of the restricted frequency range alone, which we describe as an ill-posed problem. In the determination of delay times of multiple glints, Saillant *et al.* (1993) first proposed the model to transform the echo spectrogram into the acoustic image. However, models to restore the reflected intensity distribution of the object from the echo produced by a single emission have not yet been proposed.

The amplitude spectrum of the reflected-intensity distribution varies with changes in its configuration. However, the phase spectrum varies not only with changes in the configuration of the distribution but also with the onset-delay time (T_o), which indicates the location of the object; the former is known as the configuration-dependent phase spectrum, and the latter as the delay-dependent phase spectrum. All three types of spectra are needed to determine the reflected-intensity distribution. We adopted the following steps in order to obtain these measurements. First, we introduced the continuity condition of the amplitude spectrum at the minimum frequency of the emission and the minimum phase condition to determine the configuration-dependent phase spectrum; the configuration of the reflected-intensity distribution over the whole frequency range was restored from the amplitude spectrum of the restricted frequency range. Second, we used Gaussian chirplets for which the carrier frequency agrees with the sweep rate of the emission, since the determination of the location of the object requires extracting the amplitude spectra, which vary not only with the changes in

[a]Electronic mail: matsuo@riec.tohoku.ac.jp

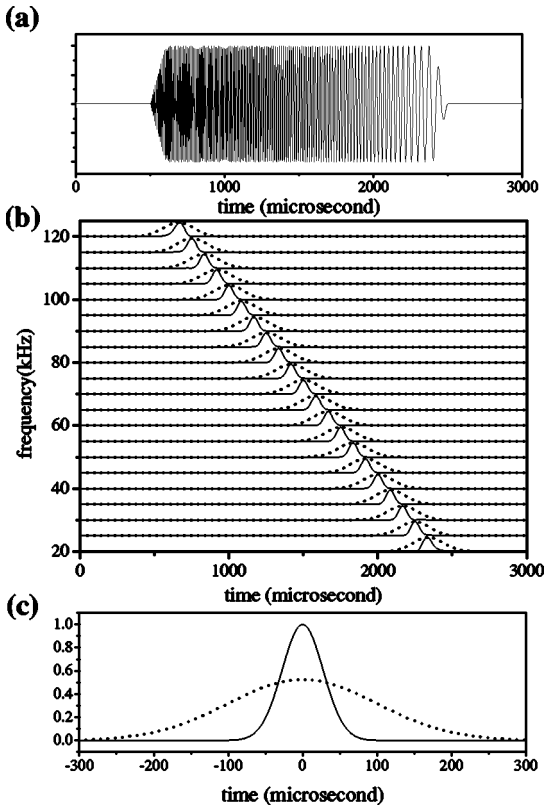


FIG. 1. (a) Emission waveform. (b) Emission spectrogram. The solid and dotted curves represent the spectrograms computed by the convolution of Gaussian chirplet (GC) and Gaussian tone (GT) filters, respectively. (c) Temporal emission patterns, W_c (solid) and W_t (dotted), corresponding to the spectrograms produced with one GC filter and one GT filter, respectively (center frequency=70 kHz).

the configuration of the reflected-intensity distribution but also with its location (Matsuo *et al.*, 2004). The location was estimated from the temporal changes of the amplitude spectra.

II. MODEL

A. Characteristic of emission and echo

Our model uses FM emission, because the purpose of the model is to restore the acoustic image of the object. The emission had a single harmonic linear sweep, the spectrum of which had constant amplitude from 20 to 120 kHz. The 2-ms FM emission begins at 130 kHz and sweeps down to 10 kHz, with 100-μs rise and fall times. Figure 1(a) shows the emitted waveform.

For simplicity, it was assumed that reflection did not modify the phase of the sound. The emitting mouth was midway between two ears, and the distance between the mouth and each ear was 5 cm.

B. Transformation of the waveform into the spectrogram

Waveforms for the emission and the echo were put into this model. These waveforms were transformed into spectrograms in a manner that simulates this process in the mammalian cochlea (Saillant *et al.*, 1993). Two types of bandpass-filter banks were used to detect the configuration of the reflected-intensity distribution and its delay time, Gaussian tone (GT) and Gaussian chirplet (GC) filters (Mann and Haykin, 1992, 1995; Matsuo *et al.*, 2001, 2004), the latter with a carrier frequency in agreement with the sweep rate of emission. The temporal characteristics of these filters can be described by

$$F_t(f_j,t)=\exp\left(-\frac{t^2}{\alpha}\right)\exp(2\pi j f_j t),$$

$$F_c(f_j,t)=\exp\left(-\frac{t^2}{\alpha}\right)\exp\left(2\pi j\left(f_j t+\frac{1}{2}st^2\right)\right),$$

where f_j (kHz) is the center frequency for the jth bandpass filter, t is time (s), s is the sweep rate of the emitted wave, and α is a parameter for the width of the window function. Each bandpass-filter bank was composed of 101 filters, with a center frequency ranging from 20 to 120 kHz, positioned at regular frequency intervals. At 70% of the peak height, the width of the window function for all filters was 160 μs. The waveforms for both the emitted sound and the echo were transformed into a spectrogram by convolution with the filters. The emission spectrogram shown in Fig. 1(b) reveals the smooth-envelope (amplitude) array produced by a linear frequency-modulated input signal. Examples of the temporal patterns of emission convolved with one GT filter and one GC filter are shown in Fig. 1(c). As a result of the linear FM emission, spectrograms produced with the same type of bandpass filter showed similar temporal patterns designated W_t (GT filter) and W_c (GC filter).

In order to demonstrate the output from the cochlear filters for the echo from the object, we considered the situation of a small planar circle (radius=28 mm), which was located at $(x,y,z)=(0,50,700\,\text{mm})$. The values of x, y, and z represent positions on the horizontal, vertical, and depth axes, respectively. Figure 2(a) shows the ideal reflected-intensity distribution for the circle with 1-μs interval. This distribution was obtained by the calculation according to the finite-element method, that is, by the summation of the echoes from the elements when the object is represented as the assembly of mesh elements with the quite small region and an emission is considered as an acoustic impulse. The waveform and spectrograms of the returning echo from the circle are shown in Figs. 2(b) and (c), respectively. The two types of spectrogram produced with the GT and GC bandpass-filter banks were transformed into the range-frequency patterns $S_t(f,\tau)$ and $S_c(f,\tau)$, respectively, with 10-μs intervals by compensating for the sweep rate of the emission-frequency modulation [Figs. 2(d), (e)]. Setting the emission start time to 0, the compensating time (τ) represented the range corresponding to the delay time. The range-frequency pattern of $S_c(f,\tau)$ was more sensitive to temporal changes in the interference patterns of the returning echo than that of $S_t(f,\tau)$, as shown in Fig. 2(f). This indicates that the range-frequency pattern of $S_c(f,\tau)$ varies not only with the changes in the configuration of the reflected-intensity distribution but also with its location.

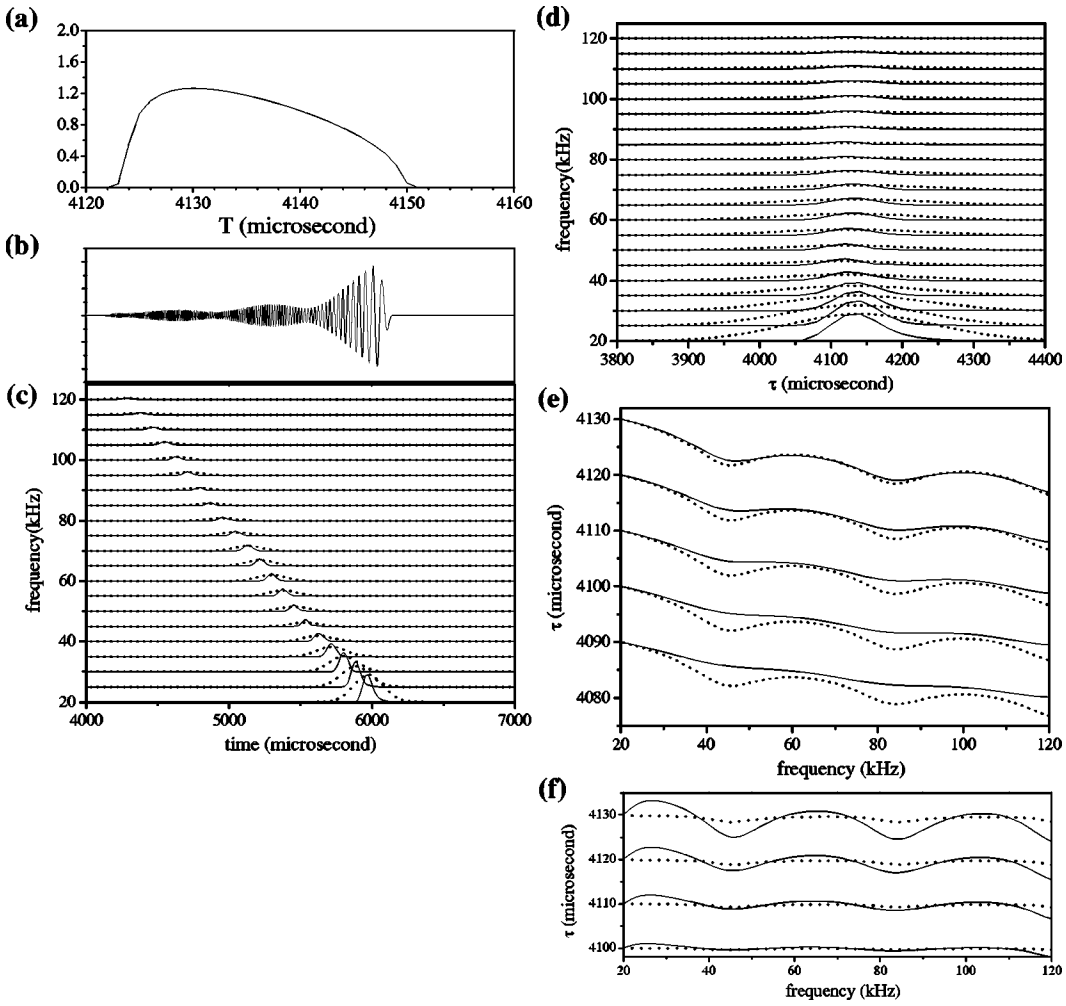


FIG. 2. Reflected-intensity distribution and range-frequency pattern. (a) Ideal reflected-intensity distribution along the range axis equivalent to the delay time T, according to the finite-element calculation. (b) Echo waveform. (c) Echo spectrograms with GC (solid) and GT (dotted) bandpass-filter banks. (d) Range-frequency patterns, $S_c(f,\tau)$ and $S_t(f,\tau)$, computed from the echo spectrograms (S_c, solid; S_t, dotted) by compensating for the sweep rate of the emission-frequency modulation. (e) Range-frequency patterns (S_c, solid; S_t, dotted). The axes are exchanged to represent the amplitude spectrum at each delay τ. (f) Spectra calculated by subtracting the spectrum $S(f,\tau_{\mathrm{on}})$ at the onset delay ($\tau_{\mathrm{on}}=4090\ \mu$s) from the spectra $S(f,\tau)$ at each delay τ (GC, solid; GT, dotted). $S_c(f,\tau)$ is more sensitive to the temporal changes in the interference patterns than $S_t(f,\tau)$.

C. Correspondence of the output of cochlear filters with the reflected-intensity distribution

To determine the reflected-intensity distribution, we examined the correlation between the outputs of the cochlear filters and the reflected-intensity distribution along the range axis (T). As shown in Fig. 3(b), the two range-frequency patterns, $E_t(f,\tau)$ and $E_c(f,\tau)$, which correspond to $S_t(f,\tau)$ and $S_c(f,\tau)$, respectively, can be theoretically calculated from the reflected-intensity distribution. W_c and W_t can be regarded as window functions for the reflected-intensity distribution. Multiplying the reflected-intensity distribution by values of W_c or W_t, the peak time of which corresponds to τ, produces the windowed reflected-intensity distributions $R_c(T,\tau)$ and $R_t(T,\tau)$, respectively. These are represented along the range axis at each delay τ [Fig. 3(a)]. Fourier transforming the windowed reflected-intensity distributions $R_c(T,\tau)$ and $R_t(T,\tau)$ gives the range-frequency patterns $E_t(f,\tau)$ and $E_c(f,\tau)$. The spectral shapes of $S_t(f,\tau)$ and $S_c(f,\tau)$ closely correspond to those of $E_t(f,\tau)$ and $E_c(f,\tau)$, respectively, at each τ [Figs. 2(e) and 3(b)]. This shows that the reflected-intensity distribution can be determined from the windowed reflected-intensity distributions, which can be obtained from the range-frequency patterns if the amplitude and phase spectra over a range of frequencies are estimated from the returning echo.

D. Restoration of the reflected-intensity distribution

Restoration of the reflected-intensity distribution requires estimation of the configuration of the windowed reflected-intensity distribution and the onset delay time T_o, indicating the location of the object. The configuration of the windowed reflected-intensity distribution is determined using the amplitude spectrum $S_t(f,\tau)$, which is suitable for extracting its configuration from 20 to 120 kHz at each delay τ. The amplitude spectrum $S_t(f,\tau)$ at $\tau=4120\ \mu$s is shown in Fig. 4(a), lacking values from 0 to 19 kHz; we compensated for the missing range by setting $S_t(f,\tau)$ to 0. The configuration-dependent phase spectrum is independent of the location of the object equal to the delay time T_o. The

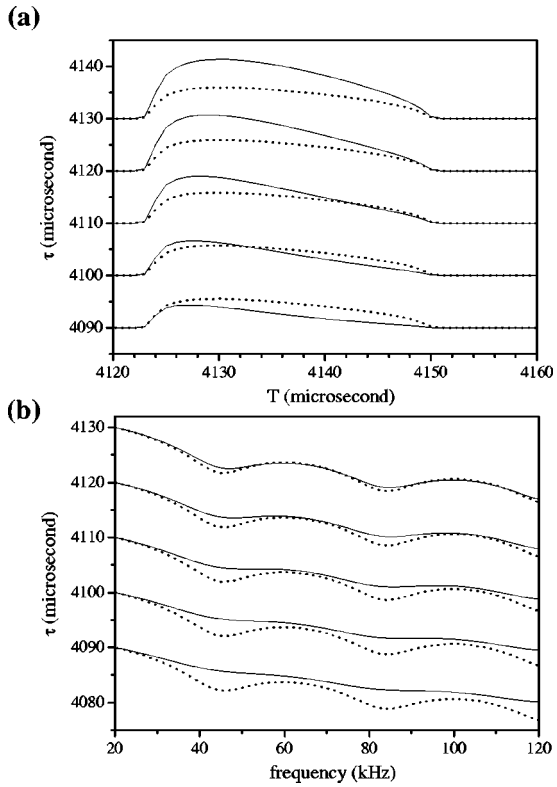


FIG. 3. Correspondence of the output from cochlear filters with the reflected-intensity distribution. (a) Windowed reflected-intensity distributions, $R_c(T,\tau)$ and $R_t(T,\tau)$ (R_c, solid; R_t, dotted). $R_c(T,\tau)$ and $R_t(T,\tau)$ are computed by multiplying the reflected-intensity distribution by values of W_c and W_t, the peak time of which corresponds to τ. (b) Range-frequency patterns, $E_c(f,\tau)$ and $E_t(f,\tau)$, computed by Fourier transforming $R_c(T,\tau)$ and $R_t(T,\tau)$, respectively (E_c, solid; E_t, dotted).

phase spectrum can therefore be estimated from the tentative amplitude spectrum using the minimum phase condition (Oppenheim and Schafer, 1975), which represents the relationship between the amplitude and phase spectra [Fig. 4(b)]. The tentative amplitude spectrum and estimated phase spectrum ϕ [Fig. 4(c)] are converted into the real and imaginary parts. The impulse response, which corresponds to the windowed reflected-intensity distribution, is computed with 4-μs interval by the inverse Fourier transform of these real and imaginary parts from 0 to 120 kHz. It is assumed that the reflected-intensity distribution of one object should include only one region with positive values; therefore, only the first positive component of the impulse response along the time axis is included. The amplitude spectrum $S_t'(f,\tau)$ is obtained using the Fourier transform of this extracted impulse response [Fig. 4(d)]. Here, the amplitude spectrum $S_t(f,\tau)$ from 0 to 19 kHz is estimated from the spectrum $S_t'(f,\tau)$, under the continuity condition that the amplitude of $S_t(f,\tau)$ and its frequency derivative at 20 kHz should agree with those of $S_t'(f,\tau)$ [Fig. 4(e)] (see the Appendix). The phase spectrum ϕ corresponding to the amplitude spectrum $S_t(f,\tau)$ is also calculated using the minimum phase condition [Fig. 4(f)]. The inverse Fourier transform of these amplitude and phase spectra, $S_t(f,\tau)$ and ϕ, produces a revised impulse response [Fig. 4(g)]. Again, only the positive components of the revised impulse response are extracted [Fig. 4(g)]. The Fourier transform of this impulse response leads to the amplitude spectrum $S_t'(f,\tau)$ [Fig. 4(h)]. These steps are repeated until the impulse response contains only one region with positive values; that is, $S_t(f,\tau)$ agrees with $S_t'(f,\tau)$. The procedure is summarized in Fig. 5. Applying this series of steps once to the returning echo is sufficient to restore the amplitude spectrum over the range of frequencies in this example.

The onset-delay time T_o of the reflected-intensity distribution is determined from the configuration of the windowed reflected-intensity distribution with 4-μs interval using the echo spectra S_c, which corresponds to the outputs of the Gaussian chirplet and is suitable to detect the amplitude changes. Candidates for the reflected-intensity distribution

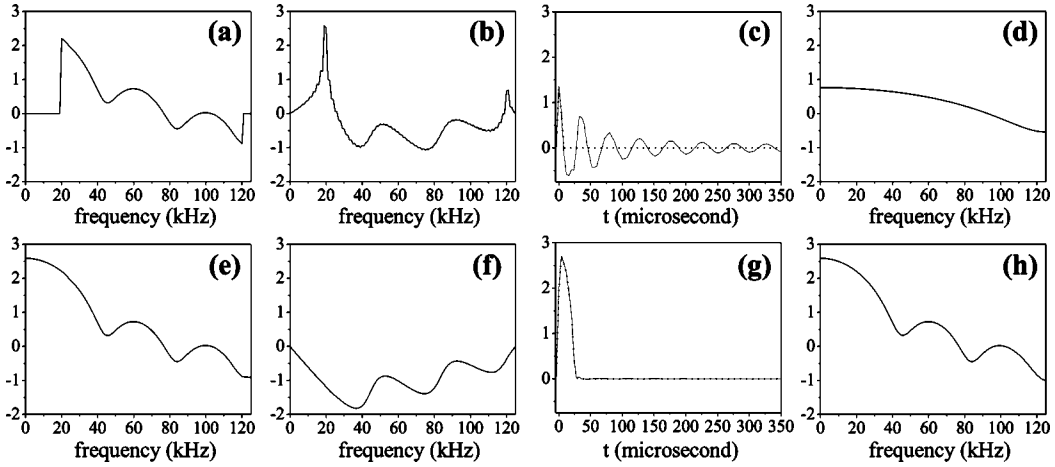


FIG. 4. Determination of the windowed reflected-intensity distribution. (a) Tentative amplitude spectrum $S_t(f,\tau)$ at the delay (τ=4120 μs). Values of the tentative amplitude spectrum $S_t(f,\tau)$ from 0 to 19 kHz were set at 0. (b) Phase spectrum ϕ computed using the minimum phase condition from the tentative amplitude spectrum $S_t(f,\tau)$. (c) Impulse response s (solid) computed by inverse Fourier transform from the tentative amplitude $S_t(f,\tau)$ and phase spectra ϕ. The dotted curve represents the extracted impulse response, which includes only positive components of the impulse response s near $t=0$. (d) Spectrum $S_t'(f,\tau)$ computed by Fourier transform of the extracted impulse response. (e) Extrapolated amplitude spectrum $S_t(f,\tau)$ from 0 to 19 kHz estimated from the spectrum $S_t'(f,\tau)$. (f) Phase spectrum ϕ computed using the minimum phase condition from the extrapolated amplitude spectrum $S_t(f,\tau)$. (g) Revised impulse response s (solid curve) computed by inverse Fourier transform from the new amplitude $S_t(f,\tau)$ and phase spectra ϕ, and the extracted impulse response (dotted curve). (h) Spectrum $S_t'(f,\tau)$ computed by Fourier transform of the extracted impulse response.

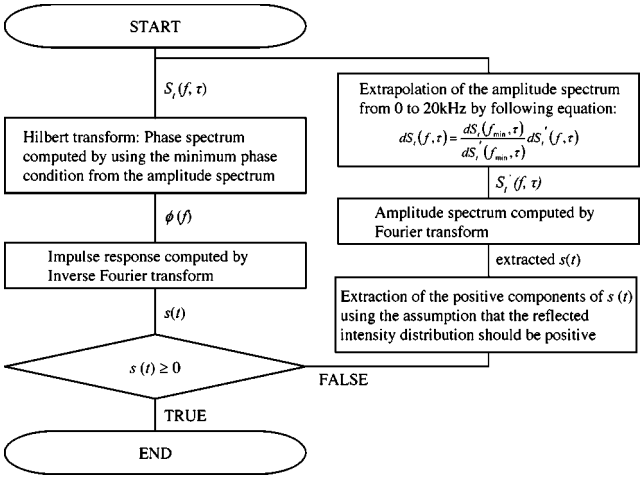


FIG. 5. Iterative algorithm used to reconstruct the windowed reflected-intensity distribution.

with 4-μs interval can be obtained from the configuration of the windowed reflected-intensity distribution by setting the onset-delay time T_o every 1 μs [Fig. 6(a)]. The windowed reflected-intensity distribution is computed by multiplying the candidate reflected-intensity distribution by values of W_c, the peak time of which corresponds to τ. The amplitude spectrum $S_c'(f,\tau)$ is computed by Fourier transforming the windowed reflected-intensity distribution $R_c(T,\tau)$. When the estimated reflected-intensity distribution corresponds to the actual reflected-intensity distribution, the $S_c'(f,\tau)$ corresponds to the $S_c(f,\tau)$ at each delay τ. The spectra S_c' at the present delay τ and 10 μs before are computed from each candidate reflected-intensity distribution [Figs. 6(b) and (c)]. T_o is determined as 4126 μs by comparing the echo spectrum S_c with the estimated spectrum S_c' at the present delay (τ=4120 μs) and 10 μs before (τ=4110 μs). The location of the object, which corresponds to the delay-dependent phase spectrum, can then be determined with 1-μs range accuracy from the temporal change of the amplitude spectrum.

III. RESULTS

We also examined different-shaped objects (rectangle, square, and triangle) located at the same position in the circle but with different rotational angles (0°, 30°, and 45°). The gravity position of each object is $(x,y,z)=(0,50,700\,\text{mm})$. The area of each object is 25 cm^2. Each side of the square is 5 cm. The longer side of the rectangle is ~8.3 cm and the shorter side is 3 cm. Each side of the regular triangle is 7.6 cm. Figures 7(a), 8(a), and 9(a) show the object shape (solid, rotation angle=0°; dotted, rotation angle=30°; broken, rotation angle=45°). The solid curves in Figs. 7(b)–(f), 8(b)–(e), and 9(b)–(f) represent the reflected-intensity distribution restored from the echo with 4-μs interval and with 1-μs range accuracy, and the dotted curves represent the corresponding ideal reflected-intensity distribution calculated using the finite-element method with 1-μs interval. Each calculated reflected-intensity distribution showed good agreement with the corresponding ideal reflected-intensity distribution. This confirms that our method is suitable for extracting the acoustic images of a range of objects.

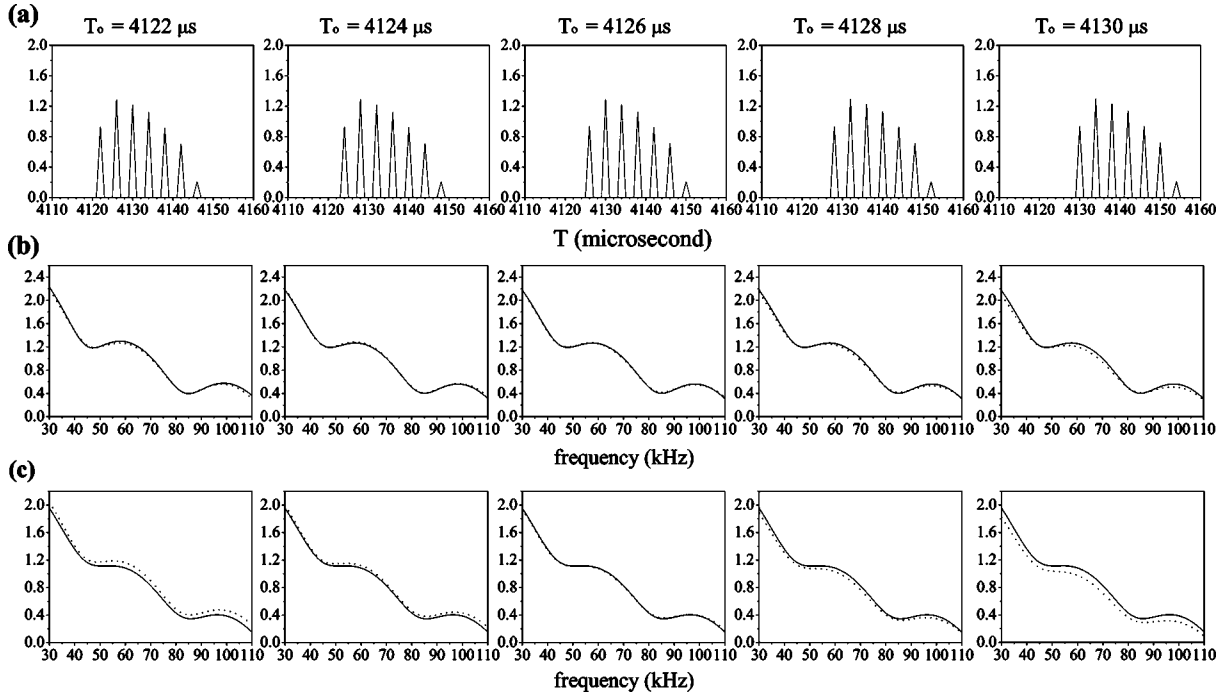


FIG. 6. Determination of the reflected-intensity distribution from the windowed reflected-intensity distribution. (a) Candidates for the reflected-intensity distribution computed from the windowed reflected-intensity distribution by setting the onset delay T_o. (b) Amplitude spectrum $S_c(f,\tau)$ (solid curve) and the spectrum $S_c'(f,\tau)$ (dotted curve) computed from each candidate of the reflected-intensity distribution at the present delay (τ=4120 μs). (c) Amplitude spectrum $S_c(f,\tau)$ (solid curve) and the spectrum $S_c'(f,\tau)$ (dotted curve) computed from each candidate of the reflected-intensity distribution at 10 μs before (τ=4110 μs). T_o is determined as 4126 μs by comparing the echo spectrum $S_c(f,\tau)$ with the estimated spectrum $S_c'(f,\tau)$ at the present delay (τ=4120 μs) and 10 μs before (τ=4110 μs).

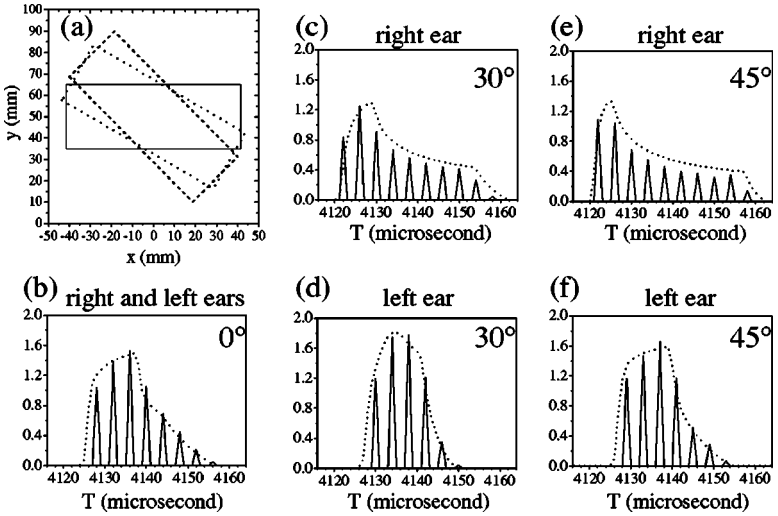


FIG. 7. Outputs produced with rectangle-shaped object. (a) Object shape (solid, rotation angle=0°; dotted, rotation angle=30°; broken, rotation angle=45°). (b)–(f) The solid and dotted curves represent the reflected-intensity distribution computed from the echo and the ideal acoustic image calculated by the finite-element calculation. When the rotation angle=0°, the echo waveforms received by both ears are the same. Therefore, the reflected-intensity distributions estimated from echoes received by both ears are the same. Each calculated reflected-intensity distribution shows good agreement with the corresponding ideal reflected-intensity distribution.

IV. DISCUSSION AND CONCLUSION

Bats can locate and discriminate individual objects even when they are moving (Griffin, 1958; Griffin *et al.*, 1965; Simmons *et al.*, 1995; Webster and Griffin, 1962). Experimental evidence indicates that FM bats are capable of discriminating both the shape and size (Simmons and Vernon, 1971). To identify the object's shape requires estimating the acoustic image, that is, the reflected-intensity distribution from echoes produced by a single emission. This reflected-intensity distribution is composed of amplitude and phase spectra over a range of frequencies, and this distribution is represented by its configuration and the onset-delay time T_o. As shown in Fig. 4, the configuration of the reflected intensity distribution can be restored by using the continuity condition of the amplitude spectrum at the minimum frequency of the emission and the minimum phase condition from the restricted amplitude spectrum (from 20 to 120 kHz). As shown in Fig. 6, the onset-delay time T_o can be determined from the temporal changes of amplitude spectra. Figures 7, 8, and 9 show that the calculated reflected-intensity distributions agree with the ideal reflected-intensity distributions. Therefore, the model performances suggest this proposed method might explain the bat's mechanism for the reconstruction of the acoustic image.

For simplicity, we have only considered objects located in the median plane, for which determining the distance alone is sufficient to locate the object. However, in order to locate objects in 3D space, bats need to process directional information from the echoes as well as the distance data. Recent studies measuring the transfer functions at the external ears have revealed the existence of spectral notches, which are local minima in the amplitude of incident sounds (Wotton *et al.*, 1995). The frequency of these spectral notches, which is known as the external ear-dependent notch frequency (EEDNF), ranged from ~30 to 50 kHz. The amplitude spectrum in the EEDNF region contains information on the shape of objects. We could not separate information on direction and shape in the amplitude spectrum in advance; therefore, the available data for restoring the reflected-intensity distribution of an object were restricted. However, our method was effective for extrapolating the amplitude and phase spectra over a range of frequencies and the iteration times were increased.

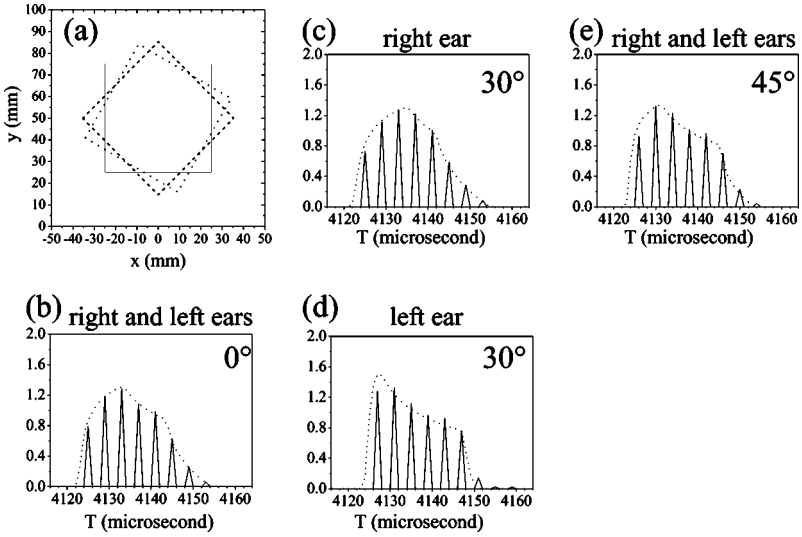


FIG. 8. Outputs produced with square-shaped object. (a) Object shape (solid, rotation angle=0°; dotted, rotation angle=30°; broken, rotation angle=45°). (b)–(e) Computed reflected-intensity distribution. The solid and dotted curves represent the reflected-intensity distribution computed from the echo and the ideal acoustic image calculated by the finite-element calculation. When the rotation angle=0° or 45°, the echo waveforms received by both ears are the same. Therefore, the reflected-intensity distributions estimated from echoes received by both ears are the same. Each calculated reflected-intensity distribution shows good agreement with the corresponding ideal reflected-intensity distribution.

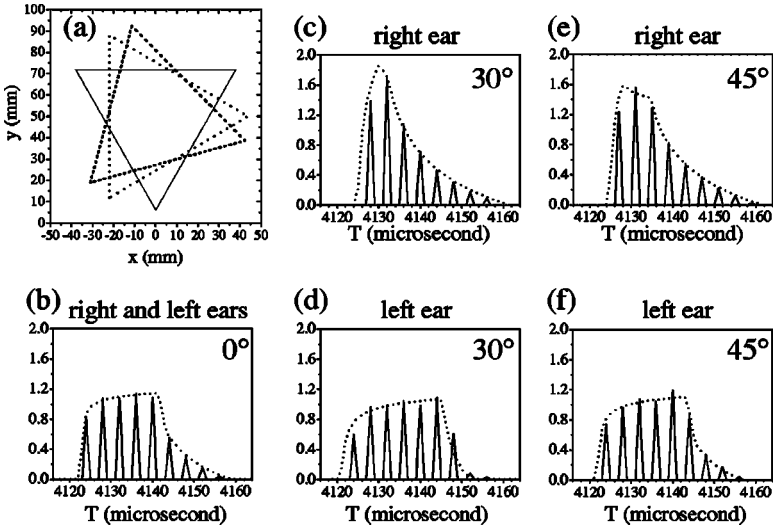


FIG. 9. Outputs produced with triangle-shaped object. (a) Object shape (solid, rotation angle=0°; dotted, rotation angle=30°; broken, rotation angle=45°). (b)–(f) The solid and dotted curves represent the reflected-intensity distribution computed from the echo and the ideal acoustic image calculated by the finite-element calculation. When the rotation angle=0°, the echo waveforms received by both ears are the same. Therefore, the reflected-intensity distributions estimated from echoes received by both ears are the same. Each calculated reflected-intensity distribution shows good agreement with the corresponding ideal reflected-intensity distribution.

Cross correlation between waveforms of the emission and echo is generally used to transform the time domain into the delay time domain. However, we have shown that this method cannot accurately determine the reflected-intensity distribution and delay time of objects in 3D space (Matsuo *et al.*, 2004), in order to explain the observed behavior of bats. Furthermore, the cross-correlation-function method cannot solve the ill-posed problem of extrapolating the amplitude and phase spectra in the whole frequency region from limited information.

APPENDIX: EXTRAPOLATION OF THE AMPLITUDE SPECTRUM FROM 0 TO 19 kHz

The amplitude spectrum $S_t(f,\tau)$ from 0 to 19 kHz is extrapolated from the spectrum $S_t'(f,\tau)$ in accordance with the assumption that the change of the spectrum $S_t(f,\tau)$ corresponds to that of the spectrum $S_t'(f,\tau)$ of 20 kHz. The spectral change $dS_t(f,\tau)$ from 0 to 19 kHz is computed by the following equation:

$$dS_t(f,\tau)=\frac{dS_t(f_{\min},\tau)}{dS_t'(f_{\min},\tau)}dS_t'(f,\tau),$$

where $f_{\min}=20$ kHz implies the minimum frequency of emission. The spectrum $S_t(f,\tau)$ from 0 to 19 kHz is computed from $dS_t(f,\tau)$.

Brugge, J. F., Anderson, D. J., Hind, J. E., and Rose, J. E. (**1969**). "Time structure of discharges in single auditory-nerve fibers of the squirrel monkey in response to complex periodic sounds," J. Neurophysiol. **32**, 386–401.

Fenton, M. B., and Bell, G. P. (**1981**). "Recognition of species of insectivorous bats by their echolocation calls," J. Mammal. **62**, 233–243.

Griffin, D. R. (**1958**). *Listening in the Dark* (Yale University Press, New Haven CT; reprinted by Cornell University Press, Ithaca, NY, 1986).

Griffin, D. R., Friend, J. H., and Webster, F. A. (**1965**). "Target discrimination by the echolocation of bats," J. Exp. Zool. **158**, 155–168.

Mann, S., and Haykin, S. (**1992**). "'Chirplets' and 'wavelets': Novel time-frequency methods," Electron. Lett. **28**, 114–116.

Mann, S., and Haykin, S. (**1995**). "The chirplet transform: physical considerations," IEEE Trans. Signal Process. **43**, 2745–2761.

Matsuo, I., Kunugiyama, K., and Yano, M. (**2004**). "An echolocation model for range discrimination of multiple closely spaced objects: Transformation of spectrogram into the reflected intensity distribution," J. Acoust. Soc. Am. **115**, 920–928.

Matsuo, I., Tani, J., and Yano, M. (**2001**). "A model of echolocation of multiple targets in 3D space from a single emission," J. Acoust. Soc. Am. **110**, 607–624.

Moss, C. F., and Simmons, J. A. (**1993**). "Acoustic image representation of a point target in the bat *Eptesicus fuscus*: Evidence for sensitivity to echo phase in bat sonar," J. Acoust. Soc. Am. **93**, 1553–1562.

Oppenheim, A. V., and Schafer, R. W. (**1975**). *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).

Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). "Phase-locked responses to low-frequency tones in single auditory nerve fibers of the squirrel monkey," J. Neurophysiol. **30**, 769–793.

Saillant, P. A., Simmons, J. A., Dear, S. P., and McMullen, T. A. (**1993**). "A computational model of echo processing and acoustic imaging in frequency-modulated echolocating bats: The spectrogram correlation and transformation receiver," J. Acoust. Soc. Am. **94**, 2691–2712.

Simmons, J. A. (**1979**). "Perception of echo phase information in bat sonar," Science **204**, 1336–1338.

Simmons, J. A. (**1989**). "A view of the world through the bat's ear: The formation of acoustic images in echolocation," Cognition **33**, 155–199.

Simmons, J. A., Ferragamo, M., Moss, C. F., Stevenson, S. B., and Altes, R. A. (**1990**). "Discrimination of jittered sonar echoes by the echolocating bat, *Eptesicus fuscus*: The shape of target images in echolocation," J. Comp. Physiol., A **167**, 589–616.

Simmons, J. A., Ferragamo, M. J., Saillant, P. A., Haresign, T., Wotton, J. M., Dear, S. P., and Lee, D. N. (**1995**). "Auditory dimensions of acoustic images in echolocation," in *Hearing by Bats*, edited by A. N. Popper and R. R. Fay (Springer, New York), pp. 146–190.

Simmons, J. A., and Vernon, J. A. (**1971**). "Echolocation: Discrimination of targets by the bat, *Eptesicus fuscus*," J. Exp. Zool. **176**, 315–328.

Webster, F. A., and Griffin, D. R. (**1962**). "The role of the flight membranes in insect capture by bats," Anim. Behav. **10**, 332–340.

Weißenbacher, P., and Wiegrebe, L. (**2003**). "Classification of virtual objects in the echolocating bat, *Megaderma lyra*," Behav. Neurosci. **117**, 833–839.

Wotton, J. M., Haresign, T., and Simmons, J. A. (**1995**). "Spatially dependent acoustic cues generated by the external ear of the big brown bat, *Eptesicus fuscus*," J. Acoust. Soc. Am. **98**, 1423–1445.

Noise emission during the first powerboat race in an Alpine lake and potential impact on fish communities

Sonja Amoser,[a] Lidia Eva Wysocki,[b] and Friedrich Ladich[c]
University of Vienna, Institute of Zoology, Althanstrasse 14, 1090 Vienna, Austria

(Received 23 April 2004; revised 21 August 2004; accepted 25 August 2004)

In order to assess the effects of high-speed boating on fish communities, noise levels were measured during the first Class 1 powerboat race on the Austrian Lake Traunsee. The noise spectra were compared to natural ambient noise and hearing abilities of four native fish species. Sound pressure levels (SPLs) were significantly elevated during the training heats and the race compared with natural levels, reaching up to 128 dB re 1 μPa (instantaneous SPL) at a distance of 300 m to the powerboats. Continuous equivalent SPLs were significantly lower during training and the pole position race compared to the race itself because fewer boats were simultaneously on the lake. The hearing abilities of the native hearing specialists and generalists were investigated. While carp and roach (two cyprinids) showed enhanced auditory sensitivity typical for hearing specialists, perch and whitefish were much less sensitive to sounds. Comparisons between power boat noise spectra and audiograms showed that the cyprinids can detect the boats up to several hundred meters distance because the main noise energy is well within the most sensitive hearing range. The hearing generalists, however, probably only perceive the first harmonic of the boat noise at close distances.
 [DOI: 10.1121/1.1808219]

PACS numbers: 43.80.Nd, 43.30.Nb, 43.50.Cb [WWA] Pages: 3789–3797

I. INTRODUCTION

Noise is an ever-increasing environmental factor in the aquatic environment due to growing anthropogenic activities such as shipping, drilling, seismic explorations or energy production (hydroelectric power plants, offshore windmills) (Myrberg, 1990; Popper, 2003). The effects of noise on aquatic animals may be manifold. Increased noise levels may impair the detection of sounds relevant for acoustic orientation, acoustic communication, prey capture, or predator avoidance by simply masking an animal's hearing. High-intensity sound such as that arising during seismic explorations can even severely damage the sensory epithelia of the inner ear in cod and pink snapper (Enger, 1981; McCauley *et al.*, 2003), injure other inner organs, and induce endocrinological stress responses in several fishes like European sea bass, Atlantic salmon and European freshwater fishes (Sverdrup *et al.*, 1994; Santulli *et al.*, 1999; Smith *et al.*, 2004; Wysocki *et al.*, 2004).

Beyond direct deleterious impacts on the organisms, noise may also have indirect consequences related to ecology, behavior and fitness over the long term. To this date, the focus has been on the impacts of noise on the marine environment, especially on marine mammals (for a review see Richardson *et al.*, 1995). Certain whale species react to approaching vessels by changing their resting and vocalizing behavior and migration routes (Richardson *et al.*, 1995; Lesage *et al.*, 1999). In freshwater habitats, recreational activities, besides traffic or hydroelectric power plants, are largely responsible for increased noise levels. A study on the effects of disturbances on migrating water birds (Schummer and Eddleman, 2003) on a lake in a national wildlife sanctuary has demonstrated that recreational activities account for nearly 87% of all disturbances. Especially boat fishing was found to induce increased alertedness, escape activities and energy expenditure in various bird species.

The effects of boat noise on fishes have mainly been investigated within the framework of population assessment and better management of catch rates in fishery. Such studies were nearly exclusively conducted in the marine environment.

Behavioral responses (e.g., avoidance reactions) of herrings and cods have been observed in the presence of different types of vessel noise in the lab (Boussard, 1981; Schwarz and Greer, 1984), but also in the field (Vabø *et al.*, 2002; Mitson and Knudsen, 2003; Handegard *et al.*, 2003). In the latter, fishes actively avoided different kinds of vessels by diving reactions and horizontal displacement. Groups of spawning rudds and roaches in the Meuse River (Belgium) were observed to actively avoid high-speed boating (Boussard, 1981). Wysocki *et al.* (2004) are showing that boat noise results in an increase in cortisol secretion in European freshwater fishes.

The first Class 1 Powerboat Race in a freshwater lake took place at the Austrian Lake Traunsee in early September 2003. Powerboat racing (top speed 270 km/h) usually takes place offshore in the marine environment (e.g., in Dubai, Quatar, Portugal, Germany, to name just a few). In Travemünde (Germany), environmental impact assessment on the powerboat races are furnished annually in order to elaborate the factors potentially affecting the adjoining marine and terrestrial protected areas, and to minimize conflicts with the local population and authorities (Morgenroth 2002, 2003). With regard to the aquatic environment, most concern has

[a]Electronic mail: sonja.amoser@univie.ac.at
[b]Electronic mail: lidia.wysocki@univie.ac.at
[c]Electronic mail: friedrich.ladich@univie.ac.at

been expressed about hydrological forces and turbidity in view of the blue mussels and about potential collisions between powerboats and whales. However, the lack of underwater noise measurements is regarded as a main shortcoming for the prognosis of noise effects, and the necessity for such measurements in future surveys has been stressed.

In order to provide a basis for assessing the potential effects of powerboat races on animal communities in and around an alpine lake, the local district authorities in Upper Austria required noise measurements in air and under water, echo-sounder studies on fish distribution before and after the race, and the monitoring of birds during the Class 1 Powerboat Race on Lake Traunsee in early September 2003. The decisive factor for the order to conduct underwater noise measurements and echo-sounder studies was the concern of local fishermen that whitefish stocks might be disturbed. The European whitefish (*Coregonus lavaretus* L.) is not only the dominant but also the most important commercially used species of this and neighboring lakes (Wanzenböck *et al.*, 2002a). During the past years, the population in Lake Traunsee has collapsed due to declining productivity paralleling a pronounced oligotrophication of the ecosystem and continuing high fishing pressure (Wanzenböck *et al.*, 2002b).

Our study is the first to describe underwater noise emissions and noise levels of powerboats in a lake and to determine the hearing abilities of several native fish species (with different hearing abilities). This should enable us to calculate the distance over which powerboats can be detected by fish and to estimate the potential impacts on fish communities. Our paper should enhance public awareness on the effects of recreational boating on fish communities.

II. MATERIAL AND METHODS

A. Study site

Lake Traunsee is situated in the Traun River drainage basin in the lake district ''Salzkammergut'' east of Salzburg, Austria, 422 m above sea level (Fig. 1). It is a typical oligotrophic, deep (mean depth: 90 m, maximum depth: 191 m) alpine lake with rather low retention times due to flushing by the large Traun River (Wanzenböck *et al.*, 2002b). The fish community is dominated by European whitefish (*Coregonus lavaretus* L.) as well as by salmonids such as the arctic charr (*Salvelinus alpinus*), by perches such as *Perca fluviatilis*, and eight cyprinid species, e.g., the roach *Rutilus rutilus* and the carp *Cyprinus carpio* (Wanzenböck *et al.*, 2002a).

Data collection was performed from a 10 m motor sailboat in the afternoon of 13 September and in the morning of 14 September 2003. The boat was anchored at the embankment of the quarry of Karbach (Fig. 1, Position 2), which is on the east bank of the lake. This site was surrounded by mountains and an underwater rock face plunges down to about 140 m water depth. The hydrophone was suspended at a depth of about 1.5 m. The powerboats passed by here at high speed (~270 km/h) because this position was at the longest straights of the course (Fig. 1). The shortest distance to the powerboats was ~300 m (an official security zone prohibited closer access to the powerboats).

The weather was rainy on 13 September (air temperature about 16 °C, water temperature about 15 °C) and became rather dry and partly sunny towards the end of the measurements on 14 September (air temperature 16 °C in the morning, 20 °C around noon; water temperature 16 °C).

In addition, on 12 September (afternoon) and 13 September (morning), measurements were performed at a site near the confluence of the Traun River (hydrophone depth 1–1.5 m) in the southern part of the lake off Ebensee (Fig. 1, Position 1). Because the river was spating due to heavy rains on previous days, strong currents caused the boat to drift and produced high ambient noise levels. Therefore, data from this recording site were excluded from further analysis.

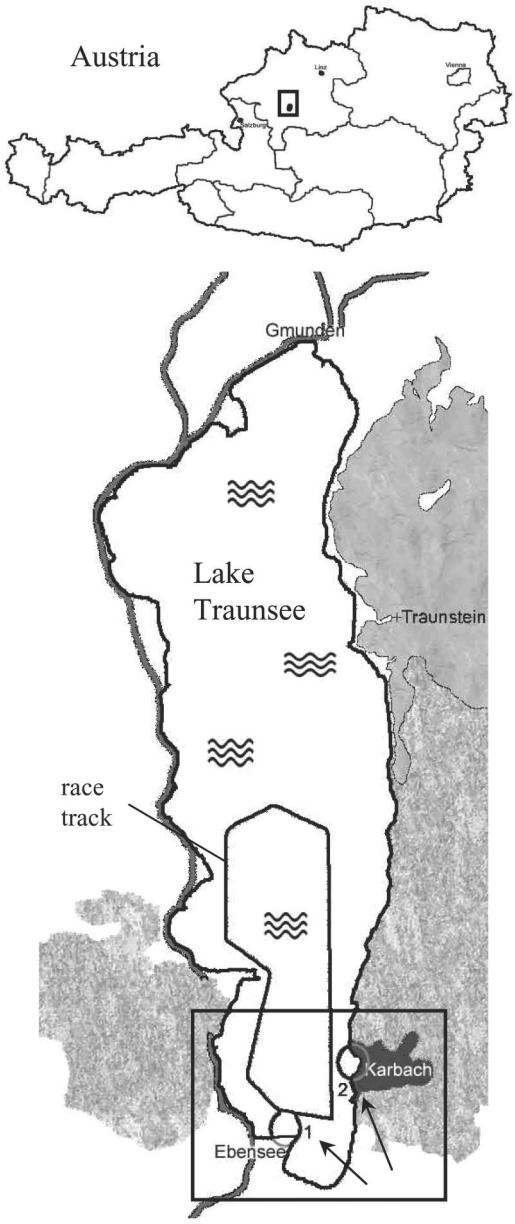


FIG. 1. Map of Lake Traunsee and the race track of the Class 1 powerboat race. (1) and (2) indicate the two measuring positions (Position 1, Position 2; marked by arrows); for more details see text. Gray textures: mountain area.

B. Noise measurements and recordings

Sound pressure levels of ambient and boat noise were measured using a sound level meter (Brüel and Kjaer 2238 Mediator), a Brüel and Kjaer 2804 power supply and a Brüel and Kjaer 8101 hydrophone. Two SPL measures were obtained: (1) The instantaneous SPL (L_{LSP}, *L*-weighted, 5 Hz to 20 kHz, RMS fast), to assess the variability of both the ambient and the powerboat noise, and (2) the equivalent continuous SPL (L_{Leq}), averaged over 60 s. The L_{Leq} is a measure of the averaged energy in a varying sound level and is commonly used to assess environmental noise (ISO 1996). SPL measurements were made for one to two minutes. During each period the L_{LSP} was noted every 5 s, and at the end we noted the L_{Leq} as well (the B & K Mediator allows parallel reading of four different noise measures).

Underwater noise was recorded on a DAT recorder (Sony TCD 100) connected to the power supply and hydrophone. Recordings were made during the heats (training, pole position race, and the race itself) and just before and after the heats, when neither powerboats nor other motor boats were cruising on the lake, to get recordings under ''quiet'' conditions. The minimum recording period was 1 min, but for some recordings—especially during the race itself—this period was elongated up to 5 min and more. Immediately before and/or after the sound recordings, L_{LSP}- and L_{Leq} were measured. Thus, it was possible to attribute SPL values to each of the recordings.

C. Sound analysis

All sound recordings were analyzed using STX 2.17 (sampling frequency 44.1 kHz), the sound processing software developed by the Research Laboratory of Acoustics at the Austrian Academy of Sciences. Two ambient noise recordings were made during a period of 10 h (4 h during the pole position race, 5 h during training, 1 h during the race). A total of 107 s of ambient noise and 3579 s of boat noise (1235 s training, 1018 s pole position race, and 1326 s race) were analyzed.

1. Noise spectra calculations

The relative noise spectra of mean ambient noise and powerboat noise were calculated. Ambient noise (AN) refers to ''quiet'' conditions when neither powerboats nor other motor boats were cruising on the lake. Powerboat noise (PN) is the noise recorded when the powerboats were closest to the recording site (minimum distance of 300 m).

For the absolute AN spectra calculation, FFTs for each recording under quiet conditions were averaged and absolute spectra calculated using the L_{Leq} measured immediately before and after the recordings. For the PN spectra calculations, 59 noise segments of 5 s duration were selected from 9 recordings at those parts with highest sound amplitude (corresponding to minimum distance of the boats) and equated to the corresponding L_{LSP} for spectrum level and distance calculations (Fig. 3).

Averaged sound spectra of the recordings were calculated by a fast Fourier transform (FFT) analysis using a filter bandwidth of 1 Hz. These spectra were then exported as ASCII Files and imported into EXCEL, and the relative spectral values were transformed to linear values using Eq. (1):

$$A_i = 10\hat{}(a_i/10), \tag{1}$$

where A_i are the linear spectral amplitude values, a_i the logarithmic spectral amplitude values. From these values, the relative root-mean square (rms) was calculated by Eq. (2):

$$e = 10*\log \sum Ai, \tag{2}$$

where *e* is the relative rms value calculated from the spectral amplitudes. The relative rms was then equaled to the absolute L_{LSP} measured with the sound level meter immediately before and/or after the recording, and the relative spectral levels were recalculated into absolute spectral levels.

2. Statistical analysis

The L_{Leq} levels measured at four conditions (ambient noise, training, pole position race, and race) were compared by a one-way ANOVA, followed by Scheffé post hoc tests. Instantaneous SPL measures (L_{LSP}) were not normally distributed, nor were variances homogeneous. Therefore, nonparametric statistic tests were used. Values obtained at the four different measurement conditions were compared by a Kruskal–Wallis test. The Mann-Whitney-U test was used for further pair wise comparisons. All statistical tests were run using SPSS 11.5.

D. Auditory sensitivity measurements

1. Animals

Test subjects were six Common carps *Cyprinus carpio* [99.2–155 mm standard length (SL), 29–61.4 g body mass (BM)] from a pond near Vienna, seven European perches *Perca fluviatilis* (85.7–97.8 mm SL; 10.6–16.9 g BM) and six European whitefish *Coregonus lavaretus* (93.7–111.8 mm SL; 8.2–16 g BM). The latter two species were obtained from local hatcheries near Traunsee. In addition, two specimens of roach *Rutilus rutilus* (135 and 136 mm SL; 45.7 and 43.7 g BM) obtained from a fisheries pond in Lower Austria were measured.

All animals were kept in planted aquaria whose bottoms were covered with sand, equipped with half flower pots as hiding places, filtered by external filters, and maintained at a 12L:12D cycle. The temperature in the holding tanks was around 21 °C, for the whitefish 16 °C. The fish were fed live *Tubifex* sp., chironomid larvae or commercially prepared flake food (Tetramin, Tetrapond®) daily. No submerged filters or air stones were used in order to reduce noise in the holding tanks. All experiments were performed with the permission of the Austrian Commission on Experiments in Animals (GZ 68.210/50-Pr/4/2002).

2. Auditory evoked potential (AEP) recordings

The AEP recording protocol used in this study followed that recently described in Wysocki and Ladich (2002, 2003) and Amoser and Ladich (2003). Therefore, only a brief summary of the basic technique is given here. During the experiments, the fish were mildly immobilized with Flaxedil (gallamine triethiodide, Sigma). The dosage used was 0.5–1.3 $\mu g\, g^{-1}$ for *C. carpio*, 1.3–3.3 $\mu g\, g^{-1}$ for *P. fluviatilis*, 2.4–

5.8 $\mu g\,g^{-1}$ for *C. lavaretus*, and 0.8 and 1.5 $\mu g\,g^{-1}$ for *R. rutilus*. This dosage allowed the fish to retain slight opercular movements during the experiments but without significant myogenic noise to interfere with the recording. Test subjects were secured in a bowl-shaped plastic tub (37 cm diameter, 8 cm water depth, 2 cm layer of fine sand) lined on the inside with acoustically absorbent material (air-filled packing wrap) in order to reduce resonances and reflections (see Fig. 1 in Wysocki and Ladich, 2002). Fish were positioned below the water surface (except for the contacting points of the electrodes, which were maximally 1 mm above the surface) in the center of the plastic tub.

A respiration pipette was inserted into the subject's mouth. Respiration was achieved through a simple gravity-fed water circulation system. Temperature during the experiments was 17.6±0.6 °C (SE) for *C. lavaretus* and 21.5 ±0.4 °C (SE) for the other 3 species. The AEPs were recorded using silver wire electrodes (0.25 mm diameter) pressed firmly against the skin. The portion of the head above the water surface was covered by a small piece of Kimwipes tissue paper to keep it moist and to ensure proper contact during experiments. The recording electrode was placed in the midline of the skull over the region of the medulla and the reference electrode cranially between the nares. Shielded electrode leads were attached to the differential input of an a. c. preamplifier (Grass P-55, gain 100×, high-pass at 30 Hz, low-pass at 1 kHz). The plastic tub was positioned on an air table (TMC Micro-g 63-540) which rested on a vibration-isolated concrete plate. The entire setup was enclosed in a walk-in sound-proof room, which was constructed as a Faraday cage (interior dimensions: 3.2 m×3.2 m×2.4 m).

Both sound stimuli presentation and AEP waveform recording were accomplished using a Tucker-Davis Technologies (Gainesville, FL, USA) modular rack-mount system (TDT System 3) controlled by a Pentium 4 PC containing a TDT digital processing board and running TDT BioSig RP Software.

3. Sound stimuli

Sound stimuli waveforms were created using TDT SigGen RP software and fed through a power amplifier (Alesis RA 300). A dual-cone speaker (Tannoy System 600, frequency response 50 Hz–15 kHz±3 dB), mounted 1 m above test subjects in the air, was used to present the stimuli during testing.

Sound stimuli consisted of tone bursts which were presented at a repetition rate of 21 per second. Hearing thresholds were determined at frequencies of 100, 300, 500, 800, 1000, 2000, 3000, and 4000 Hz for *C. carpio* and *R. rutilus*, and of 100, 200, 300, 500, 800, and 1000 Hz for *P. fluviatilis* and *C. lavaretus*, presented in random order. The duration of sound stimuli increased from two cycles at 100–300 Hz up to eight cycles at 4000 Hz. Rise and fall times were one cycle at 100 and 200 Hz and two cycles at all other frequencies. All bursts were gated using a Blackman window. For each test condition, stimuli were presented at opposite polarities (180° phase shifted), and the corresponding AEPs averaged by the Bio-Sig RP software in order to eliminate stimulus artifacts. Sound pressure levels of tone-burst stimuli were reduced in 4 dB steps until the AEP waveform was no longer apparent. The lowest SPL level for which a repeatable AEP trace could be obtained, as determined by overlaying replicate traces, was considered the threshold (Kenyon *et al.*, 1998).

A hydrophone (Brüel and Kjaer 8101, frequency range: 1 Hz–80 kHz±2 dB; voltage sensitivity: −184 re 1 V/μPa) was placed close to the right side of the animals (2 cm apart) in order to determine absolute SPLs underwater in close vicinity of the subjects. Control measurements showed that, in accordance with theoretical expectations (due to increasing distance from the loudspeaker), SPLs decreased with increasing distance from the center of the tub as well as with increasing depth. Our sound-pressure-sensitive hydrophone responded exactly to any attenuation in SPL generated by the BioSig software and played back via the air loudspeaker.

Only measurements of sound pressure were performed; in any acoustic field this is the adequate measure of the degree of auditory stimulation in pressure-sensitive fish such as otophysines (Fay and Popper, 1974). Although hearing generalists such as the perch and whitefish perceive only the kinetic component of sound, hearing thresholds were given as SPL levels in order to enable their auditory sensitivity to be compared with ambient noise spectra (which are always given in pressure units) and to calculate possible effects.

III. RESULTS

A. Radiated noise levels

The L_{Leq} measured at the different times (ambient, pole position race, training, race) differed significantly (one-way ANOVA: $F_{3,35}=76.03$; $p<0.001$). A Scheffé post-hoc test revealed that the noise level during the race differed from all other noise levels ($p<0.001$). Similarly, the ambient noise (AN) was different from all the other conditions ($p<0.001$), whereas the noise levels at the pole position race and during the training were not significantly different from each other ($p=0.156$). On the average, the L_{Leq} was elevated by 14.76±1.01 dB during the race relative to quiet conditions (AN) and was 3.56±0.63 to 5.34±0.74 dB louder than during the pole position race and the training sessions (Fig. 2).

The instantaneous noise levels (L_{LSP}) were not normally distributed, nor were variances homogeneous (Table I). Therefore, nonparametric statistical tests were used. L_{LSP}s were compared between 4 measurement conditions: AN, pole position race, training, race. A Kruskal–Wallis-test revealed significant differences between the different measurements: $\chi^2_{3,216}=153.34$; $p<0.001$. Mann–Whitney-U tests revealed that only L_{LSP} between the pole position race and the race did not differ significantly ($U_{31,36}=547$; $p=0.884$). All other pairings were significantly different.

B. Noise spectra

Figures 3 and 4 illustrate that the main sound energies were concentrated below 5 kHz. PN emissions were harmonic with a mean fundamental frequency of 417±2.7 Hz (range: 400–445 Hz, $n=67$). Four to 11 harmonics were

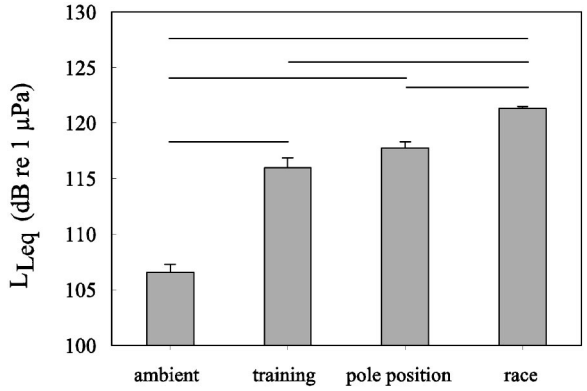


FIG. 2. Mean (±SE) equivalent continuous SPLs (L_{Leq}) of the ambient noise and the noise levels measured during training, the pole position race and the race itself. All possible pairwise tests were statistically different at the level of $p<0.001$ except for training vs pole position race.

detectable in the amplitude spectrum [Figs. 3 and 4(b)]. The drop off at higher frequencies differed between noise conditions. AN levels were relatively high at low frequencies, with a peak of 93 dB at 110 Hz, and decreased by about 1.8 dB per kHz between 0.5 and 12 kHz; they then showed a slight increase again before falling off more steeply (2.9 dB per 1 kHz) to 20 kHz (Fig. 4). The powerboat noise decreased by 0.8 dB per kHz above 3 kHz which is much less than in the AN, indicating that powerboat noise levels remain much higher in the upper frequency range than AN.

C. Hearing sensitivity of the fish species

Both otophysines, *C. carpio* and *R. rutilus*, were sensitive to high-frequency sounds (up to at least 4000 Hz, the highest frequency tested) and had sensitivity maxima between 500 and 1000 Hz (Fig. 5). AEPs of *C. lavaretus* could only be obtained up to 800 Hz, those of *P. fluviatilis* up to 1000 Hz. Sensitivity maxima were found at 300 Hz. The latter two species were up to 50 dB less sensitive to tone bursts than the otophysines (Fig. 5).

IV. DISCUSSION

A. Powerboat noise emissions

Very little data is available on SPL measurements and noise spectra of recreational boats or vessels other than commercial or military ships (Arveson and Vendittis, 2000; Malinowski and Gloza, 2002), and most of these data were assessed in the marine environment. In an attempt to compare the noise characteristics of six research vessels, Mitson (1993) stressed that SPLs and frequency spectrum are extremely variable in relation to speed, load, pitch angle of

TABLE I. Range of instantaneous noise levels (L_{LSP}) measured during the 4 noise conditions. All L_{LSP}-values are given in dB re 1 μPa.

Condition	Min. L_{LSP}	Max. L_{LSP}
Ambient noise	103	114
Training	122	126
Pole Position Race	124	127
Race	124	128

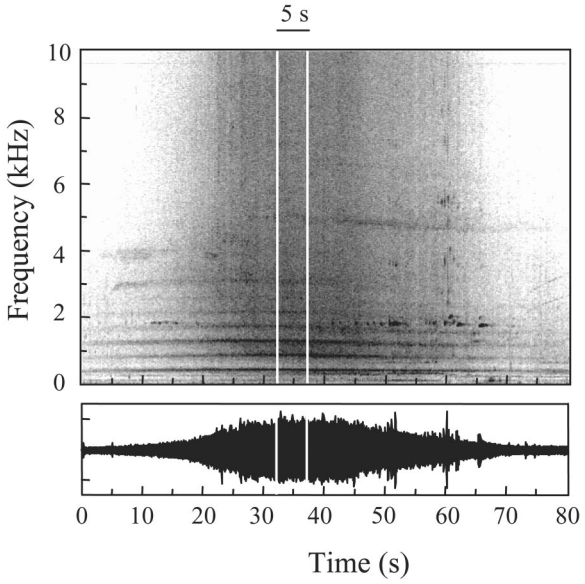


FIG. 3. Sonagram (above) and oscillogram (below) of the powerboat noise recorded during the race. The white bars indicate the 5 s time period chosen for calculating the PN noise spectra. Filter bandwidth 10 Hz, sampling frequency 44.1 kHz, 50% overlap, window: Blackmann Harris.

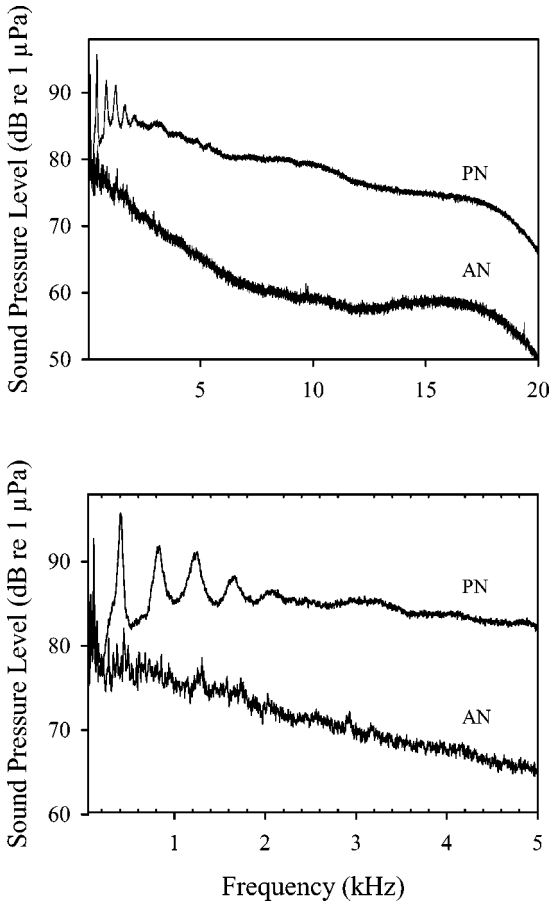


FIG. 4. Noise spectra recorded during the different conditions in Lake Traunsee during the powerboat race in 2003. Shown are the mean spectra computed from 107 s AN and 350 s PN. The lower graph gives a detail of the noise spectra within the hearing range of fishes (50–5000 Hz).

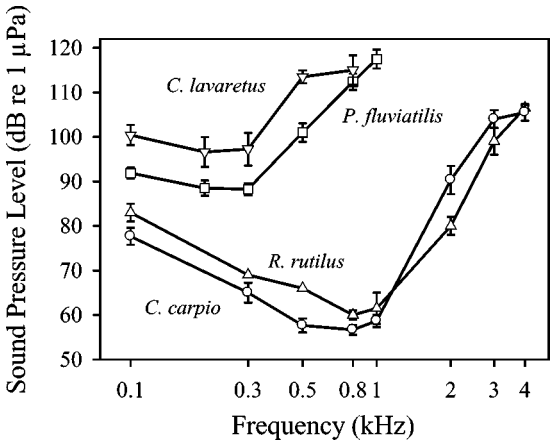


FIG. 5. Audiograms (mean±SE) of the four selected fish species, determined using the AEP-recording technique.

propeller and age of the vessel, which makes it difficult to compare different studies. Noise from merchant ships with keel aspect source levels ranging between 178 and 192 dB re 1 μPa (at 1 m) can elevate the natural ambient noise in the marine environment by 20–30 dB in many areas (Arveson and Vendittis, 2000); smaller ships (<60 m of length) produce keel aspect wideband source levels of 140–167 dB and were found to elevate natural ambient noise by 10–40 dB. The latter observation agrees with our study where, during the race, the equivalent continuous SPL exceeded the natural ambient noise by about 15 dB on the average, whereby the minimum distance to the boats was 300 m and the source level was calculated to be 180 dB at 1.2 m. Boussard (1981) described 1/3 octave band levels of up to 140 and 160 dB near cruising barges and high speed boats, respectively, in the Meuse River.

A characteristic feature for the radiated noise of surface ships is peaks in the frequency range below 100 Hz (Malinowski and Gloza, 2002), sometimes also at higher frequencies, depending on propeller, engine and load. For example, the position of peaks within the frequency spectrum of the FRV ''Thalassa'' changed when trawling conditions were simulated by towing a loaded barge on the surface. Mitson (1993) attributes the increase in high-frequency noise level during the trawling simulation to the different propeller pitch angles. Another research vessel described in that report, the FRV ''Explorer,'' which is a steam-powered vessel, showed a peak at 830 Hz when in free-ranging mode, this ''singing'' being caused by the propeller. This tone is lost during trawling because of the increased cavitation level.

The radiated noise of the powerboats also showed a peak at about 417 Hz, which fits well to the above-mentioned vessels. However, the powerboat noise shows harmonic characteristics as well, a feature not described for the research vessel and clearly audible when listening to the recordings. One feature that powerboat noise has in common with other boat noise is the relatively high levels in the low frequency range and the slow fall off at higher frequencies (in the case of the powerboats above 3 kHz).

Noise levels (L_{Leq}) were significantly higher during the race compared to the training heat and the pole position race. This surely was caused by the larger number of boats racing simultaneously on the lake during the race (maximal 1 and 2 boats during training heat and pole position race as compared to 5–9 boats during the race itself). Noise levels during all heats (training, pole position race, and race) were significantly above the natural ambient noise, even measured at a distance of 300 m. Note that the natural ambient noise measured in Lake Traunsee is quite high for a lake (about 106 dB), which may be attributed to the fact that the Traun River flows rather rapidly through the Lake from its south end in Ebensee to the north end, causing a strong current throughout the lake. The mean noise level in the neighboring Lake Mondsee is ~98,5 dB (pers. obs.).

B. Hearing abilities of lake fishes

The investigated species are typical representatives of hearing specialists and generalists in the lake investigated. In general, fishes without accessory hearing structures (''hearing generalists'') are sensitive to particle motion and are only able to detect low-frequency sounds. Typical hearing generalists in European freshwaters are perch (*Perca fluviatilis*), pike (*Esox lucius*), salmonids (trouts and charrs), and whitefish (coregonids). The ability to transmit oscillations of air-filled cavities within the body enables several groups (''hearing specialists;'' Hawkins and Myrberg, 1983) to perceive not only the kinetic but also the pressure component of sound; this improves their hearing sensitivity by allowing an extension of the hearing range and by lowering the hearing thresholds (Popper and Fay, 1993). The predominant group of hearing specialists in European freshwaters are cyprinids, which belong to the otophysines. This group is characterized by a chain of bony ossicles, known as the Weberian apparatus, connecting the swim bladder to the inner ear (for a recent review, see Ladich and Popper, 2004).

Knowledge of the hearing abilities of native European freshwater fish species is sparse. A sound pressure audiogram up to 300 Hz with a maximum sensitivity of 86.5 dB re 1 μPa at 100 Hz has been established for European perch and pike-perch (up to 800 Hz) (Wolff, 1967, 1968). No data are available on coregonids. Within the salmoniforms, only the Atlantic salmon, *Salmo salar*, was investigated; this species shows particularly low hearing sensitivity with the most sensitive hearing at about 160 Hz (hearing thresholds 95 dB re 1 μPa; Hawkins und Johnstone, 1978). By contrast, cyprinids such as carp and roach possess excellent hearing abilities up to several kHz, with best sensitivities of about 60 dB re 1 μPa between 0.5 and 1.0 kHz. Data on the carp and roach agree well with previous audiograms in cyprinids in respect of hearing range and sensitivity (koi carp: Popper, 1972; goldfish: Ladich 1999, Ladich and Wysocki, 2003). Most cyprinids apparently possess quite similar hearing capabilities, enabling conclusions to be drawn about other cyprinids which have not yet been investigated. The audiograms of perch and whitefish should be interpreted with caution because hearing generalists are only sensitive to particle motion and not to sound pressure. However, it was neither possible to measure boat noise levels in terms of particle motion nor to calculate particle motion sound spectra. The present sound pressure audiograms of the hearing generalists are de-

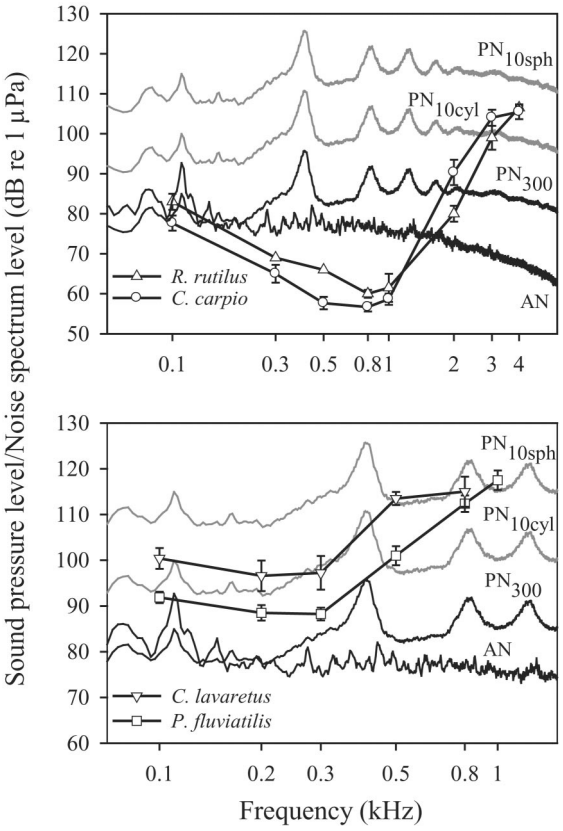


FIG. 6. Audiograms of the cyprinids (above) and the hearing generalists (below) compared with different noise spectra. AN=ambient noise (without boats), PN_{300}=powerboat noise at a distance of 300 m (measured), PN_{10cyl}, PN_{10sph}=noise 10 m away from the boats, considering cylindrical (attenuation of 3 dB/dd) and spherical spreading (attenuation of 6 dB/dd), respectively (calculated). Omitted are potential filter properties of the medium during sound propagation and distortions of the frequency content. Note the difference in *x*-axes ranges in Fig. 4 (linear) and this figure (logarithmic).

termined for comparative purposes and should not be regarded as absolute hearing thresholds.

C. Possible detection distance of boat noise

A comparison of hearing sensitivities measured with absolute powerboat (+ambient) noise spectra shows that the main energies of the boat noise lie within the most sensitive hearing range of the cyprinids [Fig. 6(a)]. Depending on the distance of the boats, their noise emitted is more than 45 dB (distance 10 m) above the hearing thresholds of the fish (at 800 Hz). Even at distances of 300 m, the spectral noise levels are ~30 dB above the hearing threshold measured under relative quiet laboratory conditions. By contrast, hearing generalists such as perch and whitefish most likely only detect the first harmonic (containing the peak energy of the boat noise) at closer distances to the boats [below 10 m, Fig. 6(b)].

When estimating the detection distance of a sound source, however, the prevailing ambient noise must be taken into account. Figure 6(a) illustrates that the ambient noise spectrum levels in Lake Traunsee are higher than the hearing thresholds of cyprinids. Thus, the hearing of these species in the natural habitat is masked, which means that signal detection is impaired. It is therefore insufficient to compare audiograms established under quiet laboratory conditions with field sound spectra in order to calculate the detectable distance of boat noise. Another key factor to consider is the signal-to-noise ratio at threshold level (S/N at threshold) of the fish. It is a measure of how much the signal energy must lie above the background noise for the signal to be detected by the animal and is defined as the difference (in dB) between the masked hearing threshold and the spectrum level of the masking noise (Chapman and Hawkins 1973). In general the S/N ratios at threshold are independent of the actual noise level over wide noise ranges, a phenomenon common to fishes as well as mammals (Fay 1974, 1988, Yost 2000).

Taking those two parameters into account, we calculate—according to Southall *et al.* (2000)—the approximate maximum distance at which the different investigated species can potentially detect the noise emitted by the powerboats. This is only a crude estimate because several other details on the actual sound propagation characteristics at different sites in the lake, i.e., bottom morphology, absorption, shadow zones due to refraction, salinity, temperature clines, etc., will also influence the detection limit. We apply rather strict criteria in our calculation, so this distance is likely to be underestimated. We concentrated on the first harmonic (around 400 Hz) of the boat noise, where the energy maximum is concentrated. S/N ratios at threshold at several frequencies are available for the goldfish, which has an audiogram very similar to carp and roach. Therefore, S/Ns at threshold for other cyprinids are likely to be similar to those of goldfish. In the following calculation, we assumed a $10 \log R$ spreading for shallow waters (according to Southall *et al.*, 2000). Based on a S/N at threshold of 15 dB at 400 Hz, the main energy of noise, (by comparing and interpolating data from the literature; Fay, 1974; Wysocki and Ladich, in press), and on an ambient noise spectrum level of 77 dB re 1 μPa, the hearing threshold would be 92 dB re 1 μPa in this lake. Assuming that the powerboat source spectrum level is 120 dB at 1.2 m under the assumption of cylindrical sound spreading (3 dB/dd), the powerboats should be detectable by otophysine fish at distances of almost 400 m even in the shallow banks of the Lake Traunsee. When we add the S/N at threshold level to the AN spectrum level, the result (92 dB) is 4 dB lower than the powerboat spectrum level at a distance of 300 m (96 dB); this supports the above calculation of the audible powerboat distance [Fig. 6(a)].

For the hearing generalists [Fig. 6(b)], we estimated a S/N ratio at threshold of 20 dB based on interpolations available for other hearing generalists like cod and sunfish (Hawkins and Chapman, 1975; Wysocki and Ladich, in press) at 400 Hz. At this frequency, the European perch's hearing threshold could just be masked (AN spectrum level plus S/N at threshold=97 dB vs interpolated hearing threshold of 95 dB). We therefore included the 97 dB originating from masked hearing into the calculation. Under these assumptions, perch probably perceive the boats up to a distance of 200 m. Whitefish are not likely to be masked in the lake (interpolated hearing threshold of 105 dB at 400 Hz), leading us to operate with absolute hearing threshold values. The maximum distance at which they can detect the boats is roughly 30 m. Note again the restrictions related to interpret-

ing the sound pressure hearing thresholds of perch and whitefish and thus the estimated distance over which they can detect the boat noise. Nonetheless, the general statement that powerboat noise affects hearing-specialized cyprinids over much wider distances than percids or coregonids remains valid. At the audible distances, the boat noise will interact with other sound sources and additionally mask acoustic signals relevant for the fish, such as sounds from prey or predators.

Beyond affecting hearing (Scholik and Yan, 2002), boat noise may also induce disturbances. There are several indications that fishes are disturbed by shipping. Certain investigations, mainly performed in the marine environment, show that fishes react to different extents to ship noise. Cods (*Gadus morhua*) significantly altered their behavior during and after the passage of a bottom-trawling vessel. They initially reacted by diving, then with horizontal movements away from the ship (Handegard *et al.*, 2003). No data on noise levels emitted by the vessel were given. Herring reacted differently to loud and quiet boats (Vabø *et al.*, 2002). Their reactions were already significant at distances of 220–270 m to a 3 t vessel cruising at 10.9 knots. These observations are interesting when regarding the fact that herrings are less sensitive than perches and coregonids measured in our study (best sensitivity according to Mann *et al.* (2001) 100 dB re 1 μPa). Based on these data and vessel speeds, Mitson and Knudsen (2003) calculated that this vessel type radiates 144–164 dB at 8 knots, which has the potential to induce fish reactions between 79 and 790 m. Fernandes *et al.* (2000) reported no reaction in herring near a noise-reduced vessel complying with ICES CRR 209 recommendations. The ICES CRR No. 209 (Mitson 1995) stressed the importance of noise reduction at vessel motors, especially for research vessels surveying fish resources. Based on available data on fish hearing capacities and behavioral reactions to noise, guidelines for noise specifications of vessels that would not affect the fishes were established. In that report, spectral levels (in 1 Hz bands) of 130–134 dB re 1 μPa at 1 m for frequencies up to 2 kHz (the main hearing bandwidth of most fishes) were recommended in order not to alter fish behavior.

Although our knowledge about noise control is increasing in the marine environment, data on freshwaters remain sparse. Underwater video recordings of rudds and roaches in the Meuse River (Belgium) showed that the fishes actively avoided high-speed boats (Boussard, 1981). The flight reactions started at distances of approximately 5 m. Based on playback experiments in the lab, the author concluded that this reaction is exclusively induced by the acoustic stimuli when an amplitude of 125 dB is reached. Comparing this to data on herring and cod reactions, it seems likely that several fish species can hear a noise source over wide distances. Stress and avoidance reactions during the race must be assumed because noise levels causing stress reactions and elevation in cortisol levels in otophysines (white noise: Smith *et al.*, 2004; ship noise: Wysocki *et al.* 2004) were reached in vicinity of the powerboats. Our assumption is supported by reports of local fishermen who observed an increase in catch rates 1 to 2 days after the powerboat race (J. Wanzenböck, pers. comm.). Higher catch rates are probably due to a higher swimming activity of fishes following startling by the powerboat noise in Lake Traunsee.

V. CONCLUSIONS

Powerboats generated noise levels of ~180 dB re 1 μPa at 1 m distance during the first Class 1 powerboat race at an Alpine lake. Cyprinids, which along with whitefish and charr constitute most of the Lake Traunsee fish community, can probably detect the boat noise at distances of nearly 400 meters. Depending on species, this distance drops in hearing generalists down to 200 (perch) and 30 m (whitefish) due to their poorer hearing abilities. Although direct observations are lacking and echo-sounder measurements were only undertaken three days after the powerboat race (and are not available to the authors), we conclude—based on the fact that main sound energies are within the best hearing range of numerous fish species and on observations reported by various authors in earlier studies—that fishes near the powerboats were disturbed by high noise levels emanating from these vessels.

ACKNOWLEDGMENTS

We would like to thank the organizers of the powerboat race for providing the motor sailer from which recordings were performed, and Peter Burian and his crew from the local yacht club for navigating the boat. We are grateful to Anton Noll from the Institute of Sound Research for his help with sound analysis and noise spectra calculations, and to Michael Stachowitsch for professional scientific English proofreading. Albert Jagsch of the fish hatchery Scharfling-Kreuzstein, Heinz Tunner and Josef Wanzenböck provided fish. This study was supported by the Austrian Science Fund (FWF Grant No. 15873 to F.L.).

Amoser, S., and Ladich, F. (**2003**). ''Diversity in noise-induced temporary hearing loss in otophysine fishes,'' J. Acoust. Soc. Am. **113**(4), 2170–2179.

Arveson, P. T., and Vendittis, D. J. (**2000**). ''Radiated noise characteristics of a modern cargo ship,'' J. Acoust. Soc. Am. **107**, 118–129.

Boussard, A. (**1981**). ''The reactions of roach (*Rutilus rutilus*) and rudd (*Scardinius erythrophthalmus*) to noises produced by high speed boating,'' Proc. 2nd Brit. Freshw. Fish. Conf., 188–200.

Chapman, C. J., and Hawkins, A. D. (**1973**). ''A field study of hearing in the cod, *Gadus morhua* L.,'' J. Comp. Physiol. **85**, 147–167.

Enger, P. S. (**1981**). ''Frequency Discrimination in Teleosts—Central or Peripheral?'' in *Hearing and Sound Communication in Fishes*, edited by W. N. Tavolga, A. N. Popper, and R. R. Fay (Springer, New York), pp. 269–318.

Fay, R. R. (**1974**). ''Masking of tones by noise for the goldfish (*Carassius auratus*),'' J. Comp. Physiol. **87**, 708–716.

Fay, R. R. (**1988**). *Hearing in Vertebrates: A Psychophysics Databook* (Hill-Fay Associates, Winnetka, Illinois).

Fay, R. R., and Popper, A. N. (**1974**). ''Acoustic stimulation of the ear of the goldfish (*Carassius auratus*),'' J. Exp. Biol. **61**, 243–60.

Fernandes, P. G., Brierley, A. S., Simmonds, E. J., Millard, N. W., McPhail, S. D., Stevenson, P., and Squires, M. (**2000**). ''Fish do not avoid survey vessels,'' Nature (London) **604**, 35–36.

Handegard, N. O., Michalsen, K., and Tjostheim, D. (**2003**). ''Avoidance behavior in cod (*Gadus morhua*) to a bottom-trawling vessel,'' Aquat. Living Resour. **16**, 265–270.

Hawkins, A. D., and Chapman, C. J. (**1975**). ''Masked auditory thresholds in the cod, *Gadus morhua* L.,'' J. Comp. Physiol. **103**, 209–226.

Hawkins, A. D., and Johnstone, A. D. F. (**1978**). ''The hearing of the Atlantic salmon, *Salmo salar*,'' J. Fish Biol. **13**, 655–673.

Hawkins, A. D., and Myrberg, A. A., Jr. (**1983**). ''Hearing and sound communication under water,'' in *Bioacoustics, a comparative approach*, edited by B. Lewis (Academic, London), pp. 373–387.
Kenyon, T. N., Ladich, F., and Yan, H. Y. (**1998**). ''A comparative study of hearing ability in fishes: the auditory brainstem response approach,'' J. Comp. Physiol., A **182**, 307–318.
Ladich, F., and Popper, A. N. (**2004**). ''Parallel evolution in fish hearing organs,'' in *Evolution of the Vertebrate Auditory System*, edited by G. Manley, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 95–127.
Ladich, F., and Wysocki, L. E. (**2003**). ''How does tripus extirpation affect auditory sensitivity in goldfish?'' Hear. Res. **182**, 119–129.
Ladich, F. (**1999**). ''Did auditory sensitivity and vocalization evolve independently in otophysan fishes?'' Brain Behav. Evol. **53**, 288–304.
Lesage, V., Barette, C., Kingsley, M. C. S., and Sjare, B. (**1999**). ''The effect of vessel noise on the vocal behavior of belugas in the St. Lawrence river estuary, Canada,'' Mar. Mam. Sci. **15**, 65–84.
Malinowski, S. J., and Gloza, I. (**2002**). ''Underwater noise characteristics of small ships,'' Acta Acustica United Acustica **88**, 718–721.
Mann, D. A., Higgs, D. M., Tavolga, W. N., Souza, M. J., and Popper, A. N. (**2001**). ''Ultrasound detection by clupeiform fishes,'' J. Acoust. Soc. Am. **109**, 3048–3054.
McCauley, R. D., Fewtrell, J., and Popper, A. N. (**2003**). ''High intensity anthropogenic sound damages fish ears,'' J. Acoust. Soc. Am. **113**, 638–642.
Mitson, R. B. (**1993**). ''Underwater noise radiated by research vessels,'' ICES Mar. Sci. Symp. **196**, 147–152.
Mitson, R. B. (**1995**). ''Underwater noise of research vessels: review and recommendations,'' ICES Coop. Res. Rep. 209.
Mitson, R. B., and Knudsen, H. P. (**2003**). ''Causes and effects of underwater noise on fish abundance estimation,'' Aquat. Living Resour. **16**, 255–263.
Morgenroth, A. (**2002**). ''Verträglichkeitsprüfung zum Powerbootrennen 2002 in Lübeck–Travemünde,'' Bericht an die Class 1 Betriebsgesellschaft mbH (''Environmental impact assessment of the Powerboat race 2002 in Lübeck-Travemünde. Report to the Class 1 operating company'').
Morgenroth, A. (**2003**). ''Class 1 Championship—Powerboatrennen 2003 in Lübeck—Travemünde. Verträglichkeitsprüfung. Bericht an die Class 1 Betreiber GmbH'' (''Class 1 championship powerboat race 2003 in Lübeck—Travemünde. Environmental impact assessment. Report to the Class 1 operating company'').
Myrberg, Jr., A. A. (**1990**). ''The effects of man-made noise on the behavior of marine mammals,'' Environ. Int. **16**, 575–586.
Popper, A. N. (**1972**). ''Pure-tone auditory thresholds for the carp, *Cyprinus carpio*,'' J. Acoust. Soc. Am. **52**, 1714–1717.
Popper, A. N. (**2003**). ''Effects of anthropogenic sounds on fishes,'' Fish. Res. **28**, 24–31.
Popper, A. N., and Fay, R. R. (**1993**). ''Sound detection and processing by fish: critical review and major research questions,'' Brain Behav. Evol. **41**, 14–38.
Richardson, W. J., Greene, C. R., Malme, C. J., and Thomson, D. H. (**1995**). *Marine Mammals and Noise* (Academic, London).
Santulli, A., Modica, A., Messina, C., Ceffa, L., Curatolo, A., Rivas, G., Fabi, G., and D'Amelio, V. (**1999**). ''Biochemical responses of European Sea Bass (*Dicentrarchus labrax* L.) to the stress induced by off shore experimental seismic prospecting,'' Mar. Pollut. Bull. **38**, 1105–1114.
Scholik, A. R., and Yan, H. Y. (**2002**). ''Effects of boat engine noise on the auditory sensitivity of the fathead minnow, *Pimephales promelas*,'' Environ. Biol. Fishes **63**, 203–209.
Schummer, M. L., and Eddleman, W. R. (**2003**). ''Effects of disturbance on activity and energy budgets of migrating waterbirds in south-central Oklahoma,'' J. Wildl. Manag. **67**, 789–795.
Schwarz, A. L., and Greer, G. L. (**1984**). ''Responses of pacific herring, *Clupea harengus pallasi*, to some underwater sounds,'' Can. J. Fish. Aquat. Sci. **41**, 1183–1192.
Smith, M. E., Kane, A. S., and Popper, A. N. (**2004**). ''Noise-induced stress response and hearing loss in goldfish (*Carassius auratus*),'' J. Exp. Biol. **207**, 427–435.
Southall, B. L., Schusterman, R. J., and Kastak, D. (**2000**). ''Masking in three pinnipeds: underwater, low-frequency critical ratios,'' J. Acoust. Soc. Am. **108**, 1322–1326.
Sverdrup, A., Kjellsby, E., Krüger, P. G., Floysand, R., Knudsen, F. R., Enger, P. S., Serck-Hanssen, G., and Helle, K. B. (**1994**). ''Effects of experimental seismic shock on vasoactivity of arteries, integrity of the vascular endothelium and on primary stress hormones of the Atlantic salmon,'' J. Fish Biol. **45**, 973–995.
Vabø, R., Olsen, K., and Huse, I. (**2002**). ''The effect of vessel avoidance of wintering Norwegian spring spawning herring,'' Fish. Res. **58**, 59–77.
Wanzenböck, J., Gassner, H., Lahnsteiner, B., Hassan, Y., Hauseder, G., Doblander, C., and Köck, G. (**2002a**). ''Ecological integrity assessment of lakes using fish communities: an example from Traunsee exposed to intensive fishing and to effluents from the soda industry,'' Water, Air, Soil Pollut. **2**, 227–248.
Wanzenböck, J., Gassner, H., Hassan, Y., Lahnsteiner, B., and Hauseder, G. (**2002b**). ''Ecology of European white fish, *Coregonus lavaretus*, in two Austrian lakes in relation to fisheries management and lake productivity,'' in *Management and Ecology of Lake and Reservoir Fisheries*, edited by Cowx, I. G. (Blackwell Science Ltd., Oxford), pp. 58–69.
Wolff, D. L. (**1967**). ''Das Hörvermögen des Flußbarsches (*Perca fluviatilis* L.)'' (''The hearing ability of the European perch (*Perca fluviatilis* L.)''), Biol. Zentralbl. **86**, 449–460.
Wolff, D. L. (**1968**). ''Das Hörvermögen des Kaulbarsches (*Acerina cernua* L.) und des Zanders (*Lucioperca sandra* Cuv. und Val.)'' (''The hearing ability of the ruffe (*Acerina cernua* L.) and the pike perch (*Lucioperca sandra* Cuv. und Val.)''), Z. Vergl. Physiol. **60**, 14–33.
Wysocki, L. E., and Ladich, F. (**2002**). ''Can fishes resolve temporal characteristics of sounds? New insights using auditory brainstem responses,'' Hear. Res. **169**, 36–46.
Wysocki, L. E., and Ladich, F. (**2003**). ''The representation of conspecific sounds in the auditory brainstem of teleost fishes,'' J. Exp. Biol. **206**, 2229–2240.
Wysocki, L. E., Ladich, F., and Dittami, J. (**2004**). ''Noise, stress, and cortisol secretion in teleost fishes,'' Horm. Behav. **46**, 125.
Wysocki, L. E., and Ladich, F. ''Hearing in fishes under noise conditions'' J. Assoc. Res. Otolaryngol. (in press).
Yost, W. A. (**2000**). *Fundamentals of Hearing—An Introduction* (Academic, San Diego).